

# **Optics, Light, and Lasers**

The Practical Approach to Modern Aspects of Photonics and  
Laser Physics

*Dieter Meschede*

**WILEY-VCH**

**Author****Dieter Meschede**

Institut für Angewandte Physik  
Wegeler Straße 8  
53115 Bonn

**Cover**

The front cover image was kindly provided  
by the author.

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

**Library of Congress Card No.:** applied for

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <<http://dnb.d-nb.de>>.

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA,  
Boschstr. 12, 69469 Weinheim, Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

**Print ISBN:** 978-3-527-41331-7

**ePDF ISBN:** 978-3-527-68551-6

**ePub ISBN:** 978-3-527-68549-3

**Mobi ISBN:** 978-3-527-68550-9

**eBook ISBN:** 978-3-527-68548-6

**Cover Design** Schulz Grafik-Design, Fußgönheim,  
Germany

**Typesetting** SPi Global, Chennai, India

**Printing and Binding**

Printed on acid-free paper

## Contents

### Preface *xix*

<b>1</b>	<b>Light Rays</b>	<b>1</b>
1.1	Light Rays in Human Experience	1
1.2	Ray Optics	2
1.3	Reflection	2
1.3.1	Planar Mirrors	2
1.4	Refraction	3
1.4.1	Law of Refraction	3
1.4.2	Total Internal Reflection	4
1.5	Fermat's Principle: The Optical Path Length	5
1.5.1	Inhomogeneous Refractive Index	6
1.6	Prisms	8
1.6.1	Dispersion	9
1.7	Light Rays in Wave Guides	10
1.7.1	Ray Optics in Wave Guides	11
1.7.2	Step-Index Fibers	12
1.7.2.1	Numerical Aperture of an Optical Fiber	13
1.7.2.2	Propagation Velocity	13
1.7.3	Gradient-Index Fibers	13
1.8	Lenses and Curved Mirrors	15
1.8.1	Lenses	15
1.8.2	Concave Mirrors	16
1.9	Matrix Optics	17
1.9.1	Paraxial Approximation	17
1.9.2	ABCD Matrices	18
1.9.3	Lenses in Air	19
1.9.4	Lens Systems	21
1.9.5	Periodic Lens Systems	22
1.9.6	ABCD Matrices for Wave Guides	23
1.10	Ray Optics and Particle Optics	23
	Problems	25

<b>2</b>	<b>Wave Optics</b>	<b>29</b>
2.1	Electromagnetic Radiation Fields	29
2.1.1	Static Fields	30
2.1.2	Polarizable and Magnetizable Media	30
2.1.3	Dynamic Fields	31
2.1.4	Fourier Components	32
2.1.5	Maxwell's Equations for Optics	33
2.1.6	Continuity Equation and Superposition Principle	33
2.1.7	The Wave Equation	33
2.1.8	Energy Density, Intensity, and the Poynting Vector of Electromagnetic Waves	35
2.2	Wave Types	37
2.2.1	Planar Waves	37
2.2.2	Spherical Waves	38
2.2.3	Dipole Waves	39
2.3	Gaussian Beams	40
2.3.1	The Gaussian Principal Mode or $\text{TEM}_{00}$ Mode	41
2.3.1.1	Rayleigh Zone, Confocal Parameter $b$	42
2.3.1.2	Radius of Wave Fronts $R(z)$	42
2.3.1.3	Beam Waist $2w_0$	42
2.3.1.4	Beam Radius $w(z)$	43
2.3.1.5	Divergence $\Theta_{\text{div}}$	43
2.3.1.6	Gouy Phase $\eta(z)$	43
2.3.2	The ABCD Rule for Gaussian Modes	44
2.3.3	Paraxial Wave Equation	46
2.3.4	Higher Gaussian Modes	47
2.3.5	Creation of Gaussian Modes	49
2.3.6	More Gaussian Paraxial Beams	50
2.4	Vector Light: Polarization	50
2.4.1	Jones Vectors	52
2.4.2	Stokes Parameters	52
2.4.3	Polarization State and Poincaré Sphere	53
2.4.4	Jones Matrices, Polarization Control, and Measurement	54
2.4.5	Polarization and Projection	56
2.4.6	Polarization of Light Beams with Finite Extension	57
2.5	Optomechanics: Mechanical Action of Light Beams	58
2.5.1	Radiation Pressure	58
2.5.2	Angular Momentum of Light Beams	59
2.5.3	Beth's Experiment	60
2.5.4	Optical Angular Momentum (OAM)	60
2.5.4.1	Twisted Beams	61
2.5.4.2	Laguerre–Gaussian Modes	61
2.5.4.3	Transforming Hermite–Gaussian to Laguerre–Gaussian Beams	62
2.6	Diffraction	63
2.6.1	Scalar Diffraction Theory	64
2.7	Fraunhofer Diffraction	67
2.7.1	Optical Fourier Transformation, Fourier Optics	70

2.8	Fresnel Diffraction	71
2.8.1	Babinet's Principle	74
2.8.2	Fresnel Zones and Fresnel Lenses	75
2.9	Beyond Gaussian Beams: Diffraction Integral and ABCD Formalism	77
	Problems	77
<b>3</b>	<b>Light Propagation in Matter: Interfaces, Dispersion, and Birefringence</b>	<b>83</b>
3.1	Dielectric Interfaces	83
3.1.1	Refraction and Reflection at Glass Surfaces	84
3.1.1.1	s Polarization	84
3.1.1.2	p Polarization	86
3.1.2	Total Internal Reflection (TIR)	87
3.1.3	Complex Refractive Index	88
3.2	Interfaces of Conducting Materials	89
3.2.1	Wave Propagation in Conducting Materials	90
3.2.1.1	High Frequencies: $\omega_p\tau \gg \omega\tau \gg 1$	90
3.2.1.2	Low Frequencies: $\omega\tau \ll 1 \ll \omega_p\tau$	90
3.2.2	Metallic Reflection	91
3.2.3	Polaritons and Plasmons	92
3.2.3.1	Surface Plasmon Polaritons (SPPs)	92
3.2.3.2	Properties of Surface Plasmon Polaritons (SPPs)	93
3.3	Light Pulses in Dispersive Materials	94
3.3.1	Pulse Distortion by Dispersion	98
3.3.2	Solitons	101
3.4	Anisotropic Optical Materials	103
3.4.1	Birefringence	103
3.4.2	Ordinary and Extraordinary Light Rays	106
3.4.3	Construction of Retarder Plates	107
3.4.3.1	Lyot Filter	108
3.4.4	Birefringent Polarizers	109
3.5	Optical Modulators	110
3.5.1	Pockels Cell and Electro-optical Modulators	110
3.5.2	Liquid Crystal Modulators	112
3.5.3	Spatial Light Modulators	113
3.5.4	Acousto-Optical Modulators	114
3.5.5	Faraday Rotators	117
3.5.6	Optical Isolators and Diodes	118
	Problems	119
<b>4</b>	<b>Light Propagation in Structured Matter</b>	<b>121</b>
4.1	Optical Wave Guides and Fibers	122
4.1.1	Step-Index Fibers	123
4.1.1.1	Weakly Guiding Step Fibers	125
4.1.1.2	$\ell = 0$ : TE and TM Modes	127
4.1.1.3	$\ell \geq 1$ : HE and EH Modes	128

4.1.1.4	$\ell \geq 1$ : LP Modes	128
4.1.2	Graded-Index Fiber	129
4.1.3	Fiber Absorption	130
4.1.4	Functional Types and Applications of Optical Fibers	130
4.1.4.1	Multimode Fibers	130
4.1.4.2	Single-Mode Fibers	131
4.1.4.3	Polarization-Maintaining (PM) Fibers	131
4.1.4.4	Photonic Crystal Fibers (PCF)	132
4.2	Dielectric Photonic Materials	132
4.2.1	Photonic Crystals	132
4.2.1.1	Light Propagation in 1D Periodically Structured Dielectrics	133
4.2.2	Bloch Waves	134
4.2.3	Photonic Bandgap in 1D	135
4.2.4	Bandgaps in 2D and 3D	137
4.2.4.1	2D Photonic Crystals	137
4.2.4.2	3D Photonic Crystals	139
4.2.5	Defects and Defect Modes	139
4.2.6	Photonic Crystal Fibers (PCFs)	141
4.3	Metamaterials	143
4.3.1	Dielectric (Plasmonic) Metamaterials	143
4.3.2	Magnetic Metamaterials and negative index of refraction	144
4.3.3	Constructing Magnetic Metamaterials	145
4.3.4	Applications of Metamaterials: The Perfect Lens	146
	Problems	147
<b>5</b>	<b>Optical Images</b>	<b>149</b>
5.1	Simple Lenses	149
5.2	The Human Eye	151
5.3	Magnifying Glass and Eyepiece	152
5.4	Microscopes	154
5.4.1	Resolving Power of Microscopes	155
5.4.1.1	Rayleigh Criterion and Numerical Aperture	155
5.4.1.2	Abbe's Theory of Resolution	156
5.4.1.3	Exploiting the Abbe–Rayleigh Resolution Limit	157
5.4.2	Analyzing and Improving Contrast	159
5.4.2.1	The Modulation Transfer Function (MTF)	159
5.4.2.2	Enhancing Contrast	160
5.5	Scanning Microscopy Methods	161
5.5.1	Depth of Focus and Confocal Microscopy	161
5.5.2	Scanning Near-Field Optical Microscopy (SNOM)	162
5.5.3	Overcoming the Rayleigh–Abbe Resolution Limits with Light	163
5.5.3.1	Single-Molecule Detection	164
5.5.3.2	PALM Microscopy	165
5.5.3.3	STED Microscopy	165
5.6	Telescopes	166
5.6.1	Theoretical Resolving Power of a Telescope	166
5.6.2	Magnification of a Telescope	167

5.6.3	Image Distortions of Telescopes	168
5.6.3.1	Lens Telescopes and Reflector Telescopes	168
5.6.3.2	Atmospheric Turbulence	169
5.7	Lenses: Designs and Aberrations	169
5.7.1	Types of Lenses	170
5.7.1.1	Planar Convex Lenses	170
5.7.1.2	Biconvex Lenses and Doublets	171
5.7.1.3	Meniscus Lenses	171
5.7.2	Aberrations: Seidel Aberrations	172
5.7.2.1	Ray Propagation in First Order	172
5.7.2.2	Ray Propagation in Third Order	172
5.7.2.3	Aperture Aberration or Spherical Aberration	173
5.7.2.4	Astigmatism	174
5.7.2.5	Coma and Distortion	175
5.7.3	Chromatic Aberration	176
	Problems	177
<b>6</b>	<b>Coherence and Interferometry</b>	<b>181</b>
6.1	Young's Double Slit	181
6.2	Coherence and Correlation	182
6.2.1	Correlation Functions	183
6.2.2	Beam Splitter	184
6.3	The Double-Slit Experiment	185
6.3.1	Transverse Coherence	186
6.3.2	Optical or Diffraction Gratings	188
6.3.3	Monochromators	190
6.4	Michelson interferometer: longitudinal coherence	191
6.4.1	Longitudinal or Temporal Coherence	192
6.4.2	Mach-Zehnder and Sagnac Interferometers	195
6.4.2.1	Mach-Zehnder Interferometer	195
6.4.2.2	Sagnac Interferometer	196
6.5	Fabry-Pérot Interferometer	197
6.5.1	Free Spectral Range, Finesse, and Resolution	200
6.6	Optical Cavities	202
6.6.1	Damping of Optical Cavities	202
6.6.2	Modes and Mode Matching	203
6.6.3	Resonance Frequencies of Optical Cavities	204
6.6.4	Symmetric Optical Cavities	205
6.6.5	Optical Cavities: Important Special Cases	205
6.6.5.1	Plane Parallel Cavity: $\ell/R = 0$	205
6.6.5.2	Confocal Cavity: $\ell/R = 1$	206
6.6.5.3	Concentric Cavity: $\ell/R = 2$	207
6.7	Thin Optical Films	208
6.7.1	Single-Layer Films	208
6.7.1.1	Minimal Reflection: AR Coating, AR Layer, and $\lambda/4$ Film	209
6.7.1.2	Reflection: Highly Reflective Films	209
6.7.2	Multilayer Films	209

6.8	Holography	210
6.8.1	Holographic Recording	211
6.8.2	Holographic Reconstruction	212
6.8.2.1	Zeroth Order	213
6.8.2.2	Halo	213
6.8.2.3	Reconstructed Signal Wave	213
6.8.2.4	Conjugated Wave	213
6.8.3	Properties	214
6.8.3.1	Three-Dimensional Reconstruction	214
6.8.3.2	Partial Reconstruction	214
6.8.3.3	Magnification	214
6.9	Laser Speckle (Laser Granulation)	214
6.9.1	Real and Virtual Speckle Patterns	215
6.9.2	Speckle Grain Sizes	215
	Problems	216
<b>7</b>	<b>Light and Matter</b>	<b>219</b>
7.1	Classical Radiation Interaction	220
7.1.1	Lorentz Oscillators	220
7.1.2	Macroscopic Polarization	224
7.1.2.1	Linear Polarization and Macroscopic Refractive Index	225
7.1.2.2	Absorption and Dispersion in Optically Thin Media	226
7.1.2.3	Dense Dielectric Media and Near Fields	227
7.2	Two-Level Atoms	229
7.2.1	Are There Any Atoms with Only Two Levels?	229
7.2.2	Dipole Interaction	230
7.2.3	Optical Bloch Equations	232
7.2.4	Pseudo-spin, Precession, and Rabi Nutation	234
7.2.5	Microscopic Dipoles and Ensembles	235
7.2.6	Optical Bloch Equations with Damping	235
7.2.7	Steady-State Inversion and Polarization	236
7.2.7.1	Steady-State Inversion and Saturation Intensity	236
7.2.7.2	Steady-State Polarization	238
7.3	Stimulated and Spontaneous Radiation Processes	239
7.3.1	Stimulated Emission and Absorption	241
7.3.2	Spontaneous Emission	242
7.4	Inversion and Amplification	242
7.4.1	Four-, Three-, and Two-Level Laser Systems	243
7.4.2	Generation of Inversion	243
7.4.3	Optical Gain	244
7.4.4	The Historical Path to the Laser	245
	Problems	246
<b>8</b>	<b>The Laser</b>	<b>249</b>
8.1	The Classic System: The He–Ne Laser	251
8.1.1	Construction	251
8.1.1.1	Amplifier	251

8.1.1.2	Operating Conditions	252
8.1.1.3	The Laser Resonator	253
8.1.2	Mode Selection in the He–Ne Laser	254
8.1.2.1	Laser Line Selection	254
8.1.3	Gain Profile, Laser Frequency, and Spectral Holes	255
8.1.4	The Single-Frequency Laser	256
8.1.5	Laser Power	257
8.1.6	Spectral Properties of the He–Ne Laser	258
8.1.6.1	Laser Linewidth	258
8.1.7	Optical Spectral Analysis	259
8.1.7.1	The Fabry–Pérot Spectrum Analyzer	259
8.1.7.2	The Heterodyne Method	259
8.1.8	Applications of the He–Ne Laser	261
8.2	Other Gas Lasers	261
8.2.1	The Argon Laser	261
8.2.1.1	The Amplifier	262
8.2.1.2	Operating Conditions	262
8.2.1.3	Features and Applications	263
8.2.2	Metal-Vapor Lasers	263
8.2.3	Molecular Gas Lasers	264
8.2.3.1	The CO <sub>2</sub> Laser	265
8.2.3.2	Gain	265
8.2.3.3	Operating Conditions	267
8.2.3.4	The Excimer Laser	267
8.3	The Workhorses: Solid-State Lasers	268
8.3.1	Optical Properties of Laser Crystals	268
8.3.2	Rare-Earth Ions	269
8.4	Selected Solid-State Lasers	271
8.4.1	The Neodymium Laser	271
8.4.1.1	The Neodymium Amplifier	271
8.4.1.2	Configuration and Operation	272
8.4.2	Applications of Neodymium Lasers	273
8.4.2.1	Frequency-Doubled Neodymium Lasers	273
8.4.2.2	The Monolithically Integrated Laser (Miser)	274
8.4.3	Erbium Lasers, Erbium-Doped Fiber Amplifiers (EDFAs)	275
8.4.4	Fiber Lasers	276
8.4.4.1	Cladding Pumping	276
8.4.4.2	Fiber Bragg Gratings	277
8.4.5	Ytterbium Lasers: Higher Power with Thin-Disc and Fiber Lasers	278
8.5	Tunable Lasers with Vibronic States	279
8.5.1	Transition-Metal Ions	279
8.5.2	Color Centers	280
8.5.3	Dyes	281
8.6	Tunable Ring Lasers	281
	Problems	283

<b>9</b>	<b>Laser Dynamics</b>	285
9.1	Basic Laser Theory	285
9.1.1	The Resonator Field	285
9.1.2	Damping of the Resonator Field	286
9.1.3	Steady-State Laser Operation	288
9.1.3.1	Saturated Gain	289
9.1.3.2	Mode Pulling	289
9.1.3.3	Field Strength and Number of Photons in the Resonator	290
9.1.3.4	Laser Threshold	290
9.1.3.5	Laser Power and Outcoupling	291
9.2	Laser Rate Equations	291
9.2.1	Laser Spiking and Relaxation Oscillations	292
9.3	Threshold-Less Lasers and Micro-lasers	295
9.4	Laser Noise	298
9.4.1	Amplitude and Phase Noise	298
9.4.1.1	Amplitude Fluctuations	298
9.4.1.2	Phase Fluctuations	299
9.4.2	The Microscopic Origin of Laser Noise	301
9.4.3	Laser Intensity Noise	302
9.4.3.1	Quantum Limit of the Laser Amplitude	302
9.4.3.2	Relative Intensity Noise (RIN)	303
9.4.4	Schawlow–Townes Linewidth	304
9.5	Pulsed Lasers	305
9.5.1	“Q-Switch”	305
9.5.1.1	Technical Q-Switches	306
9.5.1.2	Cavity Dumping	306
9.5.2	Mode Locking	306
9.5.3	Methods of Mode Locking	309
9.5.4	Measurement of Short Pulses	312
9.5.5	Tera- and Petawatt Lasers	312
9.5.6	Coherent White Light	313
9.5.7	Frequency Combs	315
	Problems	316
<b>10</b>	<b>Semiconductor Lasers</b>	319
10.1	Semiconductors	319
10.1.1	Electrons and Holes	319
10.1.2	Doped Semiconductors	320
10.1.3	pn Junctions	321
10.2	Optical Properties of Semiconductors	322
10.2.1	Semiconductors for Optoelectronics	322
10.2.2	Absorption and Emission of Light	323
10.2.3	Inversion in the Laser Diode	325
10.2.4	Small Signal Gain	327
10.2.5	Homo- and Heterostructures	329
10.3	The Heterostructure Laser	330
10.3.1	Construction and Operation	330

10.3.1.1	Laser Crystal	330
10.3.1.2	Laser Operation	331
10.3.2	Spectral Properties	332
10.3.2.1	Emission Wavelength and Mode Profile	332
10.3.2.2	Electronic Wavelength Control	333
10.3.3	Quantum Films, Quantum Wires, and Quantum Dots	334
10.3.3.1	Inversion in the Quantum Film	334
10.3.3.2	Multiple Quantum Well (MQW) Lasers	336
10.3.3.3	Quantum Wires and Quantum Dots	337
10.3.4	Quantum Cascade Lasers	338
10.4	Dynamic Properties of Semiconductor Lasers	339
10.4.1	Modulation Properties	340
10.4.1.1	Amplitude Modulation	340
10.4.1.2	Phase Modulation	341
10.4.2	Linewidth of the Semiconductor Laser	341
10.4.3	Injection Locking	342
10.5	Laser Diodes, Diode Lasers, and Laser Systems	345
10.5.1	Tunable Diode Lasers (Grating Tuned Lasers)	345
10.5.2	DFB and DBR Lasers and VCSEL	346
10.6	High-Power Laser Diodes	348
	Problems	350
<b>11</b>	<b>Sensors for Light</b>	<b>353</b>
11.1	Characteristics of Optical Detectors	354
11.1.1	Sensitivity	354
11.1.2	Quantum Efficiency	354
11.1.3	Signal-to-Noise Ratio	355
11.1.4	Noise Equivalent Power (NEP)	356
11.1.5	Detectivity “D-Star”	356
11.1.6	Rise Time	356
11.1.7	Linearity and Dynamic Range	357
11.2	Fluctuating Optoelectronic Quantities	357
11.2.1	Dark Current Noise	357
11.2.2	Intrinsic Amplifier Noise	358
11.2.3	Measuring Amplifier Noise	358
11.3	Photon Noise and Detectivity Limits	359
11.3.1	Photon Statistics of Coherent Light Fields	360
11.3.2	Photon Statistics in Thermal Light Fields	361
11.3.3	Shot Noise Limit and “Square-Law” Detectors	363
11.4	Thermal Detectors	364
11.4.1	Thermopiles	365
11.4.2	Bolometers	366
11.4.3	Pyroelectric Detectors	366
11.4.4	The Golay Cell	366
11.5	Quantum Sensors I: Photomultiplier Tubes	366
11.5.1	The Photoelectric Effect	366
11.5.2	Photocathodes	367

11.5.2.1	Amplification	368
11.5.2.2	Counting Mode and Current Mode	368
11.5.2.3	Noise Properties of PMTs	369
11.5.2.4	Microchannel Plates and Channeltrons	370
11.6	Quantum Sensors II: Semiconductor Sensors	370
11.6.1	Photoconductors	370
11.6.1.1	Sensitivity	371
11.6.1.2	Noise Properties	372
11.6.2	Photodiodes or Photovoltaic Detectors	372
11.6.2.1	pn and Pin Diodes	373
11.6.2.2	Operating Modes	373
11.6.3	Avalanche Photodiodes	374
11.7	Position and Image Sensors	374
11.7.1	Photo-Capacitors	375
11.7.2	CCD Sensors	375
11.7.3	Image Intensifiers	377
	Problems	377
<b>12</b>	<b>Laser Spectroscopy and Laser Cooling</b>	<b>379</b>
12.1	Laser-Induced Fluorescence (LIF)	379
12.2	Absorption and Dispersion	380
12.2.1	Saturated Absorption	381
12.3	The Width of Spectral Lines	382
12.3.1	Natural Width and Homogeneous Linewidth	383
12.3.2	Doppler Broadening and Inhomogeneous Linewidth	383
12.3.3	Pressure Broadening	385
12.3.4	Time-of-Flight (TOF) Broadening	386
12.4	Doppler-Free Spectroscopy	388
12.4.1	Spectroscopy with Molecular Beams	388
12.4.2	Saturation Spectroscopy	388
12.4.3	Two-Photon Spectroscopy	391
12.5	Light Forces	394
12.5.1	Radiation Pressure in a Propagating Wave	395
12.5.2	Damping Forces	397
12.5.3	Heating Forces, Doppler Limit	399
12.5.4	Dipole Forces in a Standing Wave	401
12.5.5	Generalization	403
12.5.6	Optical Tweezers	403
	Problems	404
<b>13</b>	<b>Coherent Light–Matter Interaction</b>	<b>407</b>
13.1	Weak Coupling and Strong Coupling	407
13.1.1	AC Stark Effect and Dressed-Atom Model	408
13.2	Transient Phenomena	410
13.2.1	$\pi$ Pulses	411
13.2.2	Free Induction Decay	411
13.2.3	Photon Echo	413

13.2.4	Quantum Beats	414
13.2.5	Wave Packets	415
<b>14</b>	<b>Photons: An Introduction to Quantum Optics</b>	417
14.1	Does Light Exhibit Quantum Character?	417
14.2	Quantization of the Electromagnetic Field	418
14.3	Spontaneous Emission	421
14.3.1	Vacuum Fluctuations Perturb Excited Atoms	422
14.3.2	Weisskopf and Wigner Theory of Spontaneous Emission	423
14.3.3	Suppression of Spontaneous Emission	425
14.3.4	Interpretation of Spontaneous Emission	426
14.3.5	Open Quantum Systems and Reservoirs	426
14.4	Resonance Fluorescence	427
14.4.1	The Spectrum of Resonance Fluorescence	427
14.4.2	Spectra and Correlation Functions	428
14.4.3	Spectra and Quantum Fluctuations	431
14.4.4	Coherent and Incoherent Contributions of Resonance Fluorescence	432
14.4.4.1	The Mollow Triplet	433
14.5	Light Fields in Quantum Optics	435
14.5.1	Fluctuating Light Fields	435
14.5.1.1	First-Order Coherence	435
14.5.1.2	Second-Order Coherence	436
14.5.1.3	Hanbury Brown and Twiss Experiment	437
14.5.2	Quantum Properties of Important Light Fields	438
14.5.2.1	Fock States or Number States	439
14.5.2.2	Coherent Light Fields and Laser Light	439
14.5.2.3	Thermal Light Fields	441
14.5.3	Photon Number Distribution	441
14.5.4	Bunching and Anti-bunching	443
14.5.4.1	Bunching	443
14.5.4.2	Anti-bunching	443
14.6	Two-Photon Optics	444
14.6.1	Spontaneous Parametric Fluorescence, SPDC Sources	445
14.6.2	Hong–Ou–Mandel Interferometer	446
14.7	Entangled Photons	448
14.7.1	Entangled States According to Einstein–Podolsky–Rosen	448
14.7.1.1	The Einstein–Podolsky–Rosen (EPR) Paradox	448
14.7.2	Bell's Inequality	450
14.7.3	Bell's Inequality and Quantum Optics	451
14.7.4	Polarization-Entangled Photon Pairs	452
14.7.5	A Simple Bell Experiment	453
	Problems	455
<b>15</b>	<b>Nonlinear Optics I: Optical Mixing Processes</b>	457
15.1	Charged Anharmonic Oscillators	457
15.2	Second-Order Nonlinear Susceptibility	459

15.2.1	Mixing Optical Fields: Three-Wave Mixing	459
15.2.2	Symmetry Properties of Susceptibility	461
15.2.2.1	Intrinsic Permutation Symmetry	461
15.2.2.2	Real Electromagnetic Fields	461
15.2.2.3	Loss-Free Media	461
15.2.3	Two-Wave Polarization	462
15.2.3.1	Contracted Notation	462
15.2.3.2	Kleinman Symmetry	462
15.2.4	Crystal Symmetry	463
15.2.5	Effective Value of the Nonlinear $d$ Coefficient	463
15.3	Wave Propagation in Nonlinear Media	464
15.3.1	Coupled Amplitude Equations	464
15.3.2	Coupled Amplitudes for Three-Wave Mixing	465
15.3.3	Energy Conservation	466
15.4	Frequency Doubling	466
15.4.1	Weak Conversion	467
15.4.2	Strong Conversion	468
15.4.3	Phase Matching in Nonlinear and Birefringent Crystals	469
15.4.3.1	Angle or Critical Phase Matching	471
15.4.3.2	Noncritical or 90° Phase Matching	471
15.4.4	Frequency Doubling with Gaussian Beams	472
15.4.5	Resonant Frequency Doubling	474
15.4.5.1	Passive Resonators	474
15.4.6	Quasi-phase Matching	476
15.5	Sum and Difference Frequency	477
15.5.1	Sum Frequency	477
15.5.2	Difference Frequency and Parametric Gain	478
15.6	Optical Parametric Oscillators	479
	Problems	482
<b>16</b>	<b>Nonlinear Optics II: Four-Wave Mixing</b>	<b>485</b>
16.1	Frequency Tripling in Gases	485
16.2	Nonlinear Refraction Coefficient (Optical Kerr Effect)	487
16.2.1	Self-Focusing	488
16.2.1.1	Kerr Lens Mode Locking	489
16.2.1.2	Spatial Solitons	490
16.2.1.3	Nonlinear Optical Devices	491
16.2.2	Phase Conjugation	491
16.3	Self-Phase Modulation	494
	Problems	495
<b>A</b>	<b>Mathematics for Optics</b>	<b>497</b>
A.1	Spectral Analysis of Fluctuating Measurable Quantities	497
A.1.1	Correlations	500
A.1.2	Schottky Formula	501
A.2	Time Averaging Formula	502

<b>B</b>	<b>Supplements in Quantum Mechanics</b>	503
B.1	Temporal Evolution of a Two-State System	503
B.1.1	Two-Level Atom	503
B.1.2	Temporal Development of Pure States	503
B.2	Density Matrix Formalism	504
B.3	Density of States	505
<b>Bibliography</b>		507
<b>Index</b>		519

## Preface

The field of optics – in modern times perhaps equivalent to photonics – has become an important branch of the physical sciences and technologies. The saying that the twenty-first century will be the century of the photon, following the era of the electron, is gaining more and more credibility.

This textbook has attempted – and continues to do so – to link the central topics of optics that were established 200 years ago to the most recent research topics. Since the very first German language edition published already in 1999, an impressive evolution of entire new directions has emerged.

Optical and photonics research is no longer conceivable without a wide range of structured materials shaping the propagation of light fields. Optical fibers, photonic materials, and metamaterials offer an unprecedented degree of control over propagating optical fields. Inspiring concepts such as the “perfect lens” and “optical angular momentum” are posing new conceptual and experimental challenges and promise new and wide applications. Optical imaging has overcome the classic and seemingly unsurmountable Rayleigh–Abbe limit of resolution. Femtosecond laser frequency combs allow us today to perfectly count every cycle of an optical frequency, that is, with ultimate precision. The next development is looming: optical sciences are going to marry with information technology – perhaps a natural consequence of the photon following the footsteps of the electron.

While a single book can impossibly be big enough to discuss all these new topics at depth, this new edition picks them up selectively and links them to the fundamentals of optical science. It can serve as a reference for teaching the foundations of modern optics: classical optics, laser physics, laser spectroscopy, concepts of quantum optics, and nonlinear optics – those remain the building blocks of all experimental efforts and applications, but are frequently taken as “given” and no longer scrutinized for a deeper understanding.

As for the former edition, graphical educational material is available through the following website: <http://tiny.iap.uni-bonn.de/oll/OLLWS/index.html>

## 1

## Light Rays

### 1.1 Light Rays in Human Experience

The formation of an image is one of our most fascinating emotional experiences (Figure 1.1). Even in ancient times it was realized that our “vision” is the result of rectilinearly propagating light rays, because everybody was aware of the sharp shadows of illuminated objects. Indeed, rectilinear propagation may be influenced by certain optical instruments, for example, by mirrors or lenses. Following the successes of Tycho Brahe (1546–1601), knowledge about *geometrical optics* led to the consequential design and construction of magnifiers, microscopes, and telescopes. All these instruments serve as aids to vision. Through their assistance, “insights” have been gained that added to our world picture of natural science, because they enabled observations of the world of both micro- and macro-cosmos.



**Figure 1.1** Light rays.

Thus it is not surprising that the terms and concepts of optics had tremendous impact on many areas of natural science. Even such a giant instrument as the Large Hadron Collider (LHC) particle accelerator in Geneva is basically nothing other than an admittedly very elaborate microscope, with which we are able to observe the world of elementary particles on a subnuclear length scale. Perhaps as important for the humanities is the wave theoretical description of optics, which spun off the development of quantum mechanics.

In our human experience, rectilinear propagation of light rays – in a homogeneous medium – stands in the foreground. But it is a rather newer understanding that our ability to see pictures is caused by an optical image in the eye. Nevertheless, we can understand the formation of an image with the fundamentals of ray optics. That is why this textbook starts with a chapter on ray optics.

## 1.2 Ray Optics

When light rays spread spherically into all regions of a homogeneous medium, in general we think of an idealized, point-like, and isotropic luminous source at their origin. Usually light sources do not fulfill any of these criteria. Not until we reach a large distance from the observer, may we cut out a nearly parallel beam of rays with an aperture. Therefore, with an ordinary light source, we have to make a compromise between intensity and parallelism to achieve a beam with small divergence. Nowadays optical demonstration experiments are nearly always performed with laser light sources, which offer a nearly perfectly parallel, intense optical beam to the experimenter.

When the rays of a beam are confined within only a small angle with a common optical axis, then the mathematical treatment of the propagation of the beam of rays may be greatly simplified by linearization within the so-called paraxial approximation. This situation is met so often in optics that properties such as those of a thin lens, which go beyond that situation, are called “aberrations.”

The direction of propagation of light rays is changed by refraction and reflection. These are caused by metallic and dielectric interfaces. Ray optics describes their effect through simple phenomenological rules.

## 1.3 Reflection

We observe reflection of, or mirroring of, light rays not only on smooth metallic surfaces but also on glass plates and other dielectric interfaces. Modern mirrors may have many designs. In everyday life they mostly consist of a glass plate coated with a thin layer of evaporated aluminum. But if the application involves laser light, more often dielectric multilayer mirrors are used; we will discuss these in more detail in the chapter on interferometry (Chapter 6). For ray optics, the type of design does not play any role.

### 1.3.1 Planar Mirrors

We know intuitively that at a planar mirror like in Figure 1.2, the *angle of incidence*  $\theta_1$  is identical with the *angle of reflection*  $\theta_2$  of the reflected beam,

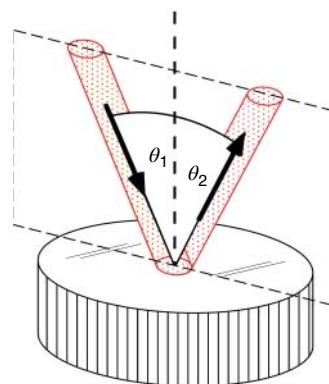
$$\theta_1 = \theta_2, \quad (1.1)$$

and that incident and reflected beams lie within a plane together with the surface normal. Wave optics finally gives us a more rigid reason for the laws of reflection. Therefore also details, for example, the intensity ratios for dielectric reflection (Figure 1.3), are explained, which cannot be derived by means of ray optics.

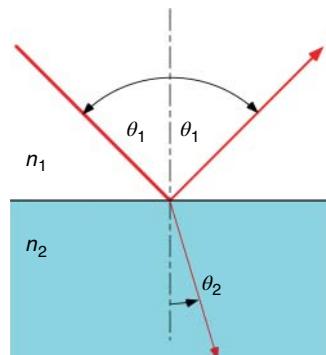
## 1.4 Refraction

At a planar dielectric surface, for example, a glass plate, reflection and transmission occur concurrently. Therefore the transmitted part of the incident beam is “refracted.” Its change of direction can be described by a single physical quantity, the “index of refraction” (also refractive index). It is higher in an optically “dense” medium than in a “thinner” one.

In ray optics a general description in terms of these quantities is sufficient to understand the action of important optical components. But the refractive index plays a key role in the context of the macroscopic physical properties of dielectric matter and their influence on the propagation of macroscopic optical waves as well. This interaction is discussed in more detail in Chapter 7.



**Figure 1.2** Reflection at a planar mirror.



**Figure 1.3** Refraction and reflection at a dielectric surface.

### 1.4.1 Law of Refraction

At the interface between an optical medium “1” with refractive index  $n_1$  and a medium “2” with index  $n_2$  (Figure 1.3), Snell’s law of refraction (Willebrord Snell, 1580–1626) is valid,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2, \quad (1.2)$$

where  $\theta_1$  and  $\theta_2$  are called the angle of incidence and angle of emergence at the interface. It is a bit artificial to define two absolute, material-specific refractive indices, because according to Eq. (1.2) only their ratio  $n_{12} = n_1/n_2$  is determined at first. But considering the transition from medium “1” into a third material “3” with  $n_{13}$ , we realize that, since  $n_{23} = n_{21}n_{13}$ , we also know the properties of refraction at the transition from “2” to “3.” We can prove this relation, for example, by inserting a thin sheet of material “3” between “1” and “2.” Finally, fixing the refractive index of vacuum to  $n_{\text{vac}} = 1$  – which is argued within the context of wave optics – the specific and absolute values for all dielectric media are determined.

**Table 1.1** Optical properties of selected glasses.

Name	Boron crown	Heavy flint glass	Barium crown	Flint glass
Abbreviation	BK7	SF11	LaSF N9	BaK 1
Abbe number A	64.17	25.76	32.17	57.55
<i>Refractive index n for selected wavelengths</i>				
$\lambda = 486.1 \text{ nm}$	1.5224	1.8065	1.8690	1.5794
$\lambda = 587.6 \text{ nm}$	1.5168	1.7847	1.8503	1.5725
$\lambda = 656.3 \text{ nm}$	1.5143	1.7760	1.8426	1.5695
<i>Dispersion constants of refractive index (see Eq. (1.6))</i>				
$B_1$	1.0396	1.7385	1.9789	1.1237
$B_2$	0.2379	0.3112	0.3204	0.3093
$B_3$	1.0105	1.1749	1.9290	0.8815
$C_1$	0.0060	0.0136	0.0119	0.0064
$C_2$	0.0200	0.0616	0.0528	0.0222
$C_3$	103.56	121.92	166.26	107.30
Density $\rho (\text{g cm}^{-3})$	2.51	4.74	4.44	3.19
Expansion coefficient $\Delta\ell/\ell (-30 \text{ to } +70^\circ\text{C}) \times 10^6$	7.1	6.1	7.4	7.6
				8.2

Strain birefringence: typically  $10 \text{ nm cm}^{-1}$ . Homogeneity of the refractive index from melt to melt:  $\delta n/n = \pm 1 \times 10^{-4}$ .

In Table 1.1 we collect some physical properties of selected glasses. The refractive index of most glasses is close to  $n_{\text{glass}} = 1.5$ . Under usual atmospheric conditions, the refractive index in air varies between 1.00002 and 1.00005. Therefore, using  $n_{\text{air}} = 1$ , the refraction properties of the most important optical interface, that is, the glass–air interface, may be described adequately in terms of ray optics. Nevertheless, small deviations and variations of the refractive index may play an important role in everyday optical phenomena in the atmosphere (e.g., a mirage; p. 7).

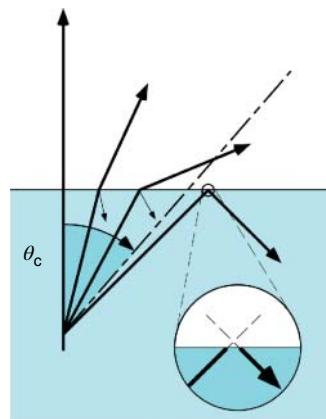
#### 1.4.2 Total Internal Reflection

According to Snell's law, at the interface between a dense medium "1" and a thinner medium "2" ( $n_1 > n_2$ ), the condition (1.2) can only be fulfilled for angles smaller than the critical angle  $\theta_c$ :

$$\theta < \theta_c = \sin^{-1}(n_2/n_1). \quad (1.3)$$

For  $\theta > \theta_c$  the incident intensity is totally reflected at the interface. We will see in the chapter on wave optics that light penetrates into the thinner medium for a distance of about one wavelength with the so-called "evanescent" wave and that the point of reflection does not lie exactly at the interface (Figure 1.4). The existence of the evanescent wave enables the application of the so-called "frustrated" total internal reflection, for example, for the design of polarizers (Section 3.4.4).

**Figure 1.4** Total internal reflection at a dielectric surface occurs for angles  $\theta > \theta_c$ . The point of reflection of the rays does not lie exactly within the interface, but slightly beyond (the Goos–Hänchen effect [1, 2]).



## 1.5 Fermat's Principle: The Optical Path Length

As long as light rays propagate in a homogeneous medium, they seem to follow the shortest geometric path from the source to a point, making their way in the shortest possible time. If refraction occurs along this route, then the light ray obviously no longer moves on the geometrically shortest path.

The French mathematician Pierre de Fermat (1601–1665) postulated in 1658 that in this case the light ray should obey a *minimum principle*, moving from the source to another point along the path that is *shortest in time*.

For an explanation of this principle, one cannot imagine a better one than that given by the American physicist Richard P. Feynman (1918–1988), who visualized Fermat's principle with a human example: One may imagine Romeo discovering his great love Juliet at some distance from the shore of a shallow, leisurely flowing river, struggling for her life in the water. Without thinking, he runs straight toward his goal – although he might have saved valuable time if he had taken the longer route, running the greater part of the distance on dry land, where he would have achieved a much higher speed than in the water.

Considering this more formally, we determine the time required from the point of observation to the point of the drowning maiden as a function of the geometric path length. Therefore we find that the shortest time is achieved exactly when a path is chosen that is refracted at the water–land boundary. It fulfills the refraction law (1.2) exactly if we substitute the indices of refraction  $n_1$  and  $n_2$  by the inverse velocities in water and on land, that is,

$$\frac{n_1}{n_2} = \frac{v_2}{v_1}.$$

According to Fermat's minimum principle, we have to demand the following. The propagation velocity of light in a dielectric  $c_n$  is reduced in comparison with the velocity in vacuum  $c$  by the refractive index  $n$ :

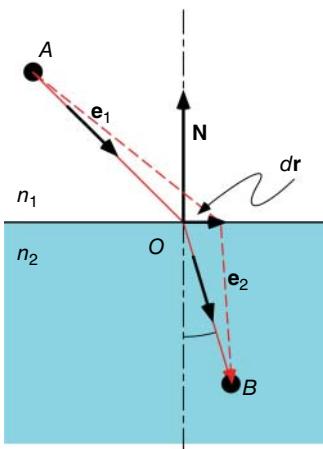
$$c_n = c/n.$$

Now the *optical path length* along a trajectory  $C$ , where the refractive index  $n$  depends on the position  $\mathbf{r}$ , can be defined in general as

$$\mathcal{L}_{\text{opt}} = c \int_C \frac{ds}{c/n(\mathbf{r})} = \int_C n(\mathbf{r}) ds. \quad (1.4)$$

With the tangential unit vector  $\mathbf{e}_t$ , the path element  $ds = \mathbf{e}_t \cdot d\mathbf{r}$  along the path can be calculated.

### Example: Fermat's principle and refraction



**Figure 1.5** Fermat's principle and refraction at a dielectric surface.

As an example of the use of the integral principle, we will again consider refraction at a dielectric surface and this time vary the length of the optical path between the points  $A$  and  $B$  in Figure 1.5 ( $\mathbf{r}_{AO}$  = vector from  $A$  to  $O$ , etc.,  $\mathbf{e}_{1,2}$  = unit vectors). Since the path must be minimal, it cannot change with small modifications  $\delta\mathbf{r}' = \mathbf{r}'_{OB} - \mathbf{r}_{OB} = \mathbf{r}_{OA} - \mathbf{r}'_{OA}$ . Thus

$$\begin{aligned}\mathcal{L}_{\text{opt}} &= n_1 \mathbf{e}_1 \cdot \mathbf{r}_{AO} + n_2 \mathbf{e}_2 \cdot \mathbf{r}_{OB}, \\ \delta\mathcal{L}_{\text{opt}} &= (n_1 \mathbf{e}_1 - n_2 \mathbf{e}_2) \cdot \delta\mathbf{r}'.\end{aligned}$$

In the homogeneous regions, light has to follow a line; thus variations can only occur along the surface with the normal  $\mathbf{N}$ , that is,  $\delta\mathbf{r}' = \mathbf{N} \times \delta\mathbf{r}$ . We use the commutativity of the triple product,

$$\begin{aligned}(n_1 \mathbf{e}_1 - n_2 \mathbf{e}_2) \cdot \delta\mathbf{r}' &= (n_1 \mathbf{e}_1 - n_2 \mathbf{e}_2) \cdot (\mathbf{N} \times \delta\mathbf{r}) \\ &= ((n_1 \mathbf{e}_1 - n_2 \mathbf{e}_2) \times \mathbf{N}) \cdot \delta\mathbf{r},\end{aligned}$$

and find minimal variation for

$$(n_1 \mathbf{e}_1 - n_2 \mathbf{e}_2) \times \mathbf{N} = 0.$$

This relation is a vectorial formulation of Snell's law (1.2), reproducing it immediately.

#### 1.5.1 Inhomogeneous Refractive Index

In general, the index of refraction of a body is not spatially homogeneous, but has underlying, continuous, even though small, fluctuations like the material itself, which affect the propagation of light rays:  $n = n(\mathbf{r})$ . We observe such fluctuations in, for example, the flickering of hot air above a flame. From the phenomenon of mirages, we know that efficient reflection may arise like in the case of grazing incidence at a glass plate, even though the refractive index decreases only a little bit toward the hot bottom.

Again using the idea of the integral principle, this case of propagation of a light ray may also be treated by applying Fermat's principle. The contribution of a path element  $ds$  to the optical path length is  $d\mathcal{L}_{\text{opt}} = n ds = n \mathbf{e}_t \cdot d\mathbf{r}$ , where  $\mathbf{e}_t = d\mathbf{r}/ds$  is the tangential unit vector of the trajectory. On the other hand,  $d\mathcal{L}_{\text{opt}} = \nabla \mathcal{L}_{\text{opt}} \cdot$

$d\mathbf{r}$  is valid in accordance with Eq. (1.4), which yields the relation

$$n\mathbf{e}_t = n \frac{d\mathbf{r}}{ds} = \nabla \mathcal{L}_{\text{opt}} \quad \text{and} \quad n^2 = (\nabla \mathcal{L}_{\text{opt}})^2,$$

which is known as the *eikonal equation* in optics. We get the important *ray equation* of optics by differentiating the eikonal equation after the path<sup>1</sup>:

$$\frac{d}{ds} \left( n \frac{d\mathbf{r}}{ds} \right) = \nabla n. \quad (1.5)$$

A linear equation may be reproduced for homogeneous materials ( $\nabla n = 0$ ) from (1.5) without difficulty.

### Example: Mirage

As a short example, we will treat reflection at a hot film of air near the ground, which induces a decrease in air density and thereby a reduction of the refractive index. (Another example is the propagation of light rays in a gradient wave guide; Section 1.7.3.) We may assume in good approximation that for calm air the index of refraction increases with distance  $y$  from the bottom, for example,  $n(y) = n_0(1 - \varepsilon e^{-\alpha y})$ . Since the effect is small,  $\varepsilon \ll 1$  is valid in general, while the scale length  $\alpha$  is of the order  $\alpha = 1 \text{ m}^{-1}$ . We look at Eq. (1.5) for  $\mathbf{r} = (y(x), x)$  for all individual components and find the  $x$  coordinate with constant  $C$ :

$$n \frac{dx}{ds} = C.$$

We may use this result as a partial parametric solution for the  $y$  coordinate:

$$\frac{d}{ds} \left( n \frac{dy}{ds} \right) = \frac{d}{dx} \left( n \frac{dy}{dx} \frac{dx}{ds} \right) \frac{dx}{ds} = \frac{d}{dx} \left( C \frac{dy}{dx} \right) \frac{C}{n} = \frac{\partial n(y)}{\partial y}.$$

The constant may be chosen to be  $C = 1$ , because it is only scaling the  $x$  coordinate. Since  $2n \partial n / \partial y = \partial n^2 / \partial y$  and  $n^2 \simeq n_0^2(1 - 2\varepsilon e^{-\alpha y})$ , we get for  $\varepsilon \ll 1$

$$\frac{d^2y(x)}{dx^2} = \frac{1}{2} \frac{\partial}{\partial y} n^2(y) = n_0^2 \varepsilon \alpha e^{-\alpha y}.$$

This equation can be solved by fundamental methods, and it is convenient to write the solution in the form

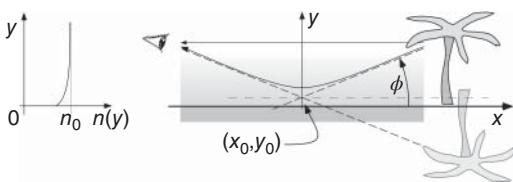
$$y = y_0 + \frac{1}{\alpha} \ln [\cosh^2(\kappa(x - x_0))] \xrightarrow{\kappa(x-x_0) \gg 1} y_0 + \frac{2\kappa}{\alpha}(x - x_0),$$

where the new parameter  $\kappa$  has to be determined from boundary conditions. For large distances from the point of reflection at  $x = x_0$ , we find straight propagation as expected. The maximum angle  $\phi = \arctan(2\kappa/\alpha)$ , where reflection is still possible, is defined by  $\kappa \leq n_0 \alpha (\varepsilon/2)^{1/2}$ . As in Figure 1.6, the observer registers two images – one of them is upside down and corresponds to a mirror image. The curvature of the light rays declines quickly with increasing distance from the bottom and therefore may be neglected for the “upper” line of sight. At  $(x_0, y_0)$  a “virtual” point of reflection may be defined.

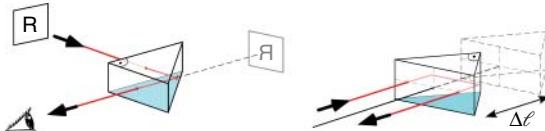
---

<sup>1</sup> Therefore we apply  $d/ds = \mathbf{e}_t \cdot \nabla$  and

$$\frac{d}{ds} \nabla \mathcal{L} = (\mathbf{e}_t \cdot \nabla) \nabla \mathcal{L} = \frac{1}{n} (\nabla \mathcal{L} \cdot \nabla) \nabla \mathcal{L} = \frac{1}{2n} \nabla (\nabla \mathcal{L})^2 = \frac{1}{2n} \nabla n^2.$$



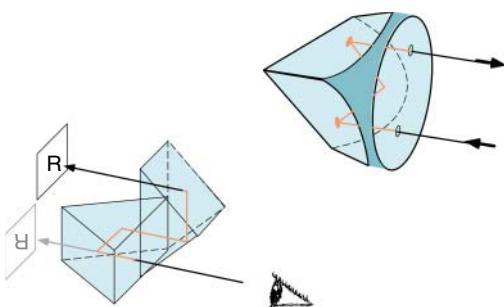
**Figure 1.6** Profile of the refractive index and optical path for a mirage.



**Figure 1.7** Reflection in  $90^\circ$  prism. This prism is used for rectangular beam deflection. It may also be used for the design of a retroreflector, whereby an optical delay  $\Delta t = 2\Delta\ell/c$  is realized by simple adjustment.

## 1.6 Prisms

The technically important rectangular reflection is achieved with an angle of incidence of  $\theta_i = 45^\circ$ . For ordinary glasses ( $n \approx 1.5$ ), this is above the angle of total internal reflection  $\theta_c = \sin^{-1} 1.5 = 42^\circ$ . Glass prisms are therefore often used as simple optical elements, which are applied for beam deflection (Figure 1.7). More complicated prisms are realized in many designs for multiple reflections, where they have advantages over the corresponding mirror combinations due to their minor losses and more compact and robust designs.



**Figure 1.8** The Porro prism (left) is combined out of two rectangular prisms, which rotate the image plane of an object such that in combination with lenses, one gets an upright image. The cat's-eye retroreflector (right) throws back every light ray independently of its angle of incidence, causing a parallel shift.

Often used designs are the Porro prism and the “cat’s-eye” retroreflector (Figure 1.8) – other names for the latter are “corner cube reflector,” “cat’s eye,” or “triple mirror.” The Porro prism and its variants are applied, for example, in telescopes to create upright images. The retroreflector not only plays an important role in optical distance measurement techniques and interferometry but also enables functioning of safety reflectors – cast in plastics – in vehicles.

We may also regard cylindrical glass rods as a variant of prisms where total internal reflection plays an important role. In such a rod (see Figure 1.11), a light ray is reflected back from the surface to the interior again and again, without changing its path angle relative to the rod axis. Such fiber

prisms where total internal reflection plays an important role. In such a rod (see Figure 1.11), a light ray is reflected back from the surface to the interior again and again, without changing its path angle relative to the rod axis. Such fiber

rods are used, for example, to guide light from a source toward a photodetector. In miniaturized form they are applied as *wave guides* in optical telecommunications. Their properties will be discussed in the section on beam propagation in wave guides (Section 1.7) and later on in the chapter on wave optics (Section 4.1) in more detail.

### 1.6.1 Dispersion

Prisms played a historical role in the spectral decomposition of white light into its constituents. The refractive index and thus also the angle of deflection  $\delta$  in Figure 1.9 actually depend on the wavelength,  $n = n(\lambda)$ , and therefore light rays of different colors are deflected with different angles. Under *normal dispersion*, blue wavelengths are refracted more strongly than red ones,  $n(\lambda_{\text{blue}}) > n(\lambda_{\text{red}})$ .

Refractive index and dispersion are very important technical quantities for the application of optical materials. The refractive index is tabulated in manufacturers' data sheets for various wavelengths, and (numerous different) empirical formulae are used for the wavelength dependence. The constants from Table 1.1 are valid for this formula, which is also called the *Sellmeier equation*:

$$n^2 = 1 + \frac{B_1 \lambda^2}{\lambda^2 - C_1} + \frac{B_2 \lambda^2}{\lambda^2 - C_2} + \frac{B_3 \lambda^2}{\lambda^2 - C_3} \quad (\lambda \text{ in } \mu\text{m}). \quad (1.6)$$

By geometrical considerations we find that the angle of deflection  $\delta$  in Figure 1.9 depends not only on the angle of incidence  $\theta$  but also on the aperture angle  $\alpha$  of the symmetrical prism and of course on the index of refraction,  $n$ :

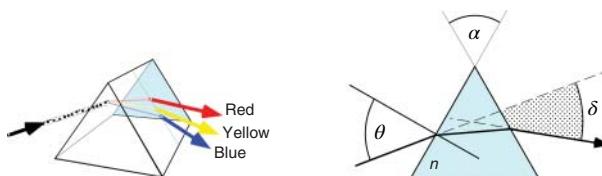
$$\delta = \theta - \alpha + \arcsin \left[ \sin(\alpha \sqrt{n^2 - \sin^2 \theta}) - \cos \alpha \sin \theta \right],$$

$$\delta_{\min} = 2\theta_{\text{symm}} - \alpha.$$

The minimum deflection angle  $\delta_{\min}$  is achieved for symmetrical transit through the prism ( $\theta = \theta_{\text{symm}}$ ) and enables a precise determination of the refractive index. The final result is expressed straightforwardly by the quantities  $\alpha$  and  $\delta_{\min}$ :

$$n = \frac{\sin[(\alpha + \delta_{\min})/2]}{\sin(\alpha/2)}.$$

For quantitative estimation of the dispersive power  $K$  of glasses, the Abbe number  $A$  may be used. This relates the refractive index at a yellow wavelength (at  $\lambda = 587.6$  nm, the D line of helium) to the change of the refractive index, estimated from the difference of the refractive indices at a blue wavelength ( $\lambda = 486.1$  nm,



**Figure 1.9** Refraction and dispersion at a symmetrical prism. The index of refraction  $n$  can be calculated from the minimum angle of deflection  $\delta = \delta_{\min}$  in a simple manner.

Fraunhofer line F of hydrogen) and a red wavelength ( $\lambda = 656.3 \text{ nm}$ , Fraunhofer line C of hydrogen):

$$A = K^{-1} = \frac{n_D - 1}{n_F - n_C}.$$

According to the aforementioned, a large Abbe number means weak dispersion, and a small Abbe number means strong dispersion. The Abbe number is also important when correcting chromatic aberrations (see Section 5.7.3).

The index of refraction describes the interaction of light with matter, and we will come to realize that it is a complex quantity, which describes not only the properties of dispersion but also those of absorption. Furthermore, it is the task of a microscopic description of matter to determine the dynamic polarizability and thus to establish the connection to a macroscopic description.

## 1.7 Light Rays in Wave Guides



**Figure 1.10** Historic station No. 51 of the Berlin–Cologne–Coblenz optic-mechanical “sight” transmission line on the tower of the St Pantaleon Church, Cologne. Picture by Weiger (1840).

The transmission of messages via light signals is a very convenient method that has a very long history of application. For example, in the nineteenth century, mechanical pointers were mounted onto high towers and were observed with telescopes to realize transmission lines of many hundreds of miles. An example of a historic relay station from the 400 mile Berlin–Cologne–Coblenz transmission line is shown in Figure 1.10. Basically, in-air transmission is also performed nowadays, but with laser light. But it is always affected by its scattering properties even at small distances, because turbulence, dust, and rain can easily inhibit the propagation of a free laser beam.

Ideas for guiding optical waves have been in existence for a very long time. In analogy to microwave techniques, for example, at first, hollow tubes made of copper were applied, but their attenuation is too large for transmission over long

distances. Later on periodical lens systems have been used for the same purpose, but due to high losses and small mechanical flexibility, they also failed.

The striking breakthrough happened to “optical telecommunication” through the development of low-loss *wave guides*, which are nothing other than elements for guiding light rays. In 2009 the Nobel prize was awarded to Charles K. Kao (1933–) “for groundbreaking achievements concerning the transmission of light in fibers for optical communication,” which in fact created the new field of *fiber optics* and is today the backbone of worldwide communication including the Internet.

Optical fibers can be distributed like electrical cables, provided that adequate transmitters and receivers are available. With overseas cables, significantly shorter signal transit times and thus higher comfort for phone calls can be achieved than via geostationary satellites, where there is always a short but unpleasant and unnatural break between question and answer.

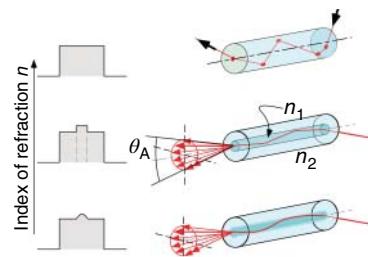
Therefore, propagation of light rays in dielectric wave guides is an important chapter in modern optics. Some basics may yet be understood by the methods of ray optics.

### 1.7.1 Ray Optics in Wave Guides

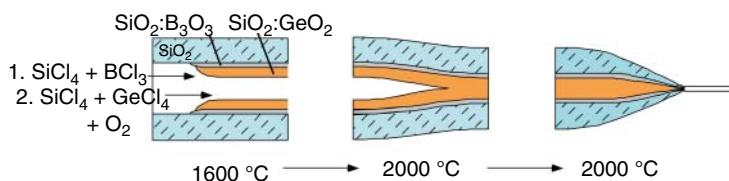
Total internal reflection in an optically thick medium provides the fundamental physical phenomenon for guiding light rays within a dielectric medium. Owing to this effect, for example, in cylindrical homogeneous glass fibers, rays whose angle with the cylinder axis stays smaller than the angle of total internal reflection  $\theta_c$  are guided from one end to the other. Guiding of light rays in a homogeneous glass cylinder is affected by any distortion of the surface, and a protective cladding could even suppress total internal reflection.

Therefore, various concepts have been developed, where the optical waves are guided in the center of a wave guide through variation of the index of refraction. These wave guides may be surrounded by cladding and entrenched like electrical cables.

We will present the two most important types. Step-index fibers consist of two homogeneous cylinders with different refractive indices (Figure 1.11). To achieve beam guiding, the higher index of refraction must be in the core, the lower one in the cladding. Gradient-index (GRIN) fibers with continuously changing (in good approximation, parabolic) refractive index are more sophisticated to manufacture (Figure 1.12), but they have technical advantages, for example, a smaller group velocity dispersion.



**Figure 1.11** Profiles of the refractive index and ray path in optical wave guides. Upper: Wave guide with homogeneous refractive index. Center: Wave guide with stepped profile of refractive index (step-index fiber). Lower: Wave guide with continuous profile of refractive index (gradient-index fiber).



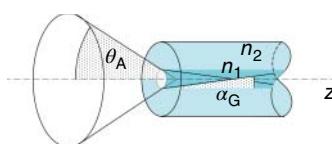
**Figure 1.12** Manufacturing of wave guides. The preform is manufactured with appropriate materials with distinct indices of refraction, which are deposited on the inner walls of a quartz tube by a chemical reaction.

### Excuse: Manufacturing optical fibers

The starting material is an ordinary tube made of quartz glass. It rotates on a lathe and is blown through on the inside by a gas mixture (chlorides such as highly purified  $\text{SiCl}_4$  and  $\text{GeCl}_4$ ). An oxyhydrogen burner heats a small zone of only a few centimeters up to about  $1600^\circ\text{C}$ , in which the desired materials are deposited as oxides on the inner walls (*chemical vapor deposition (CVD)*). Thus, by multiple repetition, a refractive index profile is established before the tube is melted at about  $2000^\circ\text{C}$  to a massive glass rod of about 10 mm diameter, a so-called preform. In the last step a fiber pulling machine extracts the glass fiber out of a crucible with viscous material. Typical cross sections are 50 and 125  $\mu\text{m}$ , which are coated with a cladding for protection.

The creation of optical fibers is a consequence of the advantageous properties of one of the oldest materials used in our culture: glass. In more recent years, more complex structured materials were developed to control the properties of light propagation in wave guides, for example, the so-called photonic crystal fibers (see Chapter 4.2.6).

### 1.7.2 Step-Index Fibers



**Figure 1.13** Critical angle in a step-index fiber.

The principle of total internal reflection is applied in *step-index fibers* (Figure 1.13), which consist of a *core* with refractive index  $n_1$  and a *cladding* with  $n_2 < n_1$ . The relative difference in the index of refraction

$$\Delta = \frac{n_1 - n_2}{n_1} \quad (1.7)$$

is not more than 1–2%, and the light rays are only guided if the angle  $\alpha$  toward the fiber axis is shallow enough to fulfill the condition for total internal reflection.

For example, for quartz glass ( $n_2 = 1.45$  at  $\lambda = 1.55 \mu\text{m}$ ), whose core index of refraction has been enhanced by  $\text{GeO}_2$  doping up to  $n_1 = n_2 + 0.015$ , according to  $\theta_c = \sin^{-1}(n_2/n_1)$ , one finds the critical angle  $\theta_c = 81.8^\circ$ . The complementary beam angle relative to the fiber axis,  $\alpha_G = 90^\circ - \theta_c$ , can be approximated by

$$\alpha_G \simeq \sin \alpha_G \simeq \sqrt{2\Delta}, \quad (1.8)$$

since  $n_2/n_1 = 1 - \Delta$ , and thus is set in relation to  $\Delta$ , which yields  $\alpha \leq 8.2^\circ$  for this case.

When light rays cross the axis of a fiber, propagation takes place in a cut plane, which is called the *meridional plane*. *Skewed rays* do not pass the axis and are guided on a polygon around the circle. It can be shown that the rays must confine an angle  $\alpha < \alpha_G$  with the  $z$  axis to be guided by total internal reflection.

### 1.7.2.1 Numerical Aperture of an Optical Fiber

To guide a light ray in an optical fiber, the angle of incidence at the coupler must be chosen small enough. The maximum aperture angle  $\theta_a$  of the acceptance cone can be calculated according to the refraction law,  $\sin \theta_a = n_1 \sin \alpha_G = n_1 \cos \theta_c$ . The sine of the aperture angle is called the *numerical aperture* (NA). According to Eq. (1.8) and  $\cos \theta_c \approx \sqrt{2\Delta}$ , it can be related with the physical parameters of the optical fiber:

$$\text{NA} = \sin \theta_a \approx n_1 \sqrt{2\Delta}. \quad (1.9)$$

This yields, for example, NA = 0.21 for the aforementioned quartz glass fiber, which is a useful and typical value for standard wave guides.

### 1.7.2.2 Propagation Velocity

Light within the core of the wave guide propagates along the trajectory with a velocity  $v(r(z)) = c/n(r(z))$ . Along the  $z$  axis the beam propagates with a reduced velocity,  $\langle v_z \rangle = v \cos \alpha$ , which can be calculated for small angles  $\alpha$  to the  $z$  axis according to

$$\langle v_z \rangle \approx \frac{c}{n_1} \left( 1 - \frac{1}{2} \alpha^2 \right). \quad (1.10)$$

In Section 4.1 on the wave theory of light, we will see that the propagation velocity is related to the phase velocity.

### 1.7.3 Gradient-Index Fibers

Beam guiding can also be performed by means of a GRIN fiber, where the quadratic variation of the index of refraction is important. To determine the curvature of a light ray induced by the refractive index, we apply the ray equation (1.5). This is greatly simplified in the paraxial approximation ( $ds \approx dz$ ) and for a cylindrically symmetric fiber:

$$\frac{d^2r}{dz^2} = \frac{1}{n} \frac{dn}{dr}.$$

A parabolic profile of the refractive index with a difference of the refractive index of  $\Delta = (n_1 - n_2)/n_1$ ,

$$n(r \leq a) = n_1 \left[ 1 - \Delta \left( \frac{r}{a} \right)^2 \right] \quad \text{and} \quad n(r > a) = n_2, \quad (1.11)$$

decreases from the maximum value  $n_1$  at  $r = 0$  to  $n_2$  at  $r = a$ . One ends up with the equation of motion of a harmonic oscillator,

$$\frac{d^2r}{dz^2} + \frac{2\Delta}{a^2}r = 0,$$

and realizes immediately that the light ray performs oscillatory motion about the  $z$  axis. The period is

$$\Lambda = 2\pi a / \sqrt{2\Delta}, \quad (1.12)$$

and a light ray is described with a wave number  $K = 2\pi/\Lambda$  according to

$$r(z) = r_0 \sin(2\pi z/\Lambda).$$

The maximum elongation allowed is  $r_0 = a$ , because otherwise the beam loses its guiding. Therefore also the maximum angle  $\alpha_G = \sqrt{2\Delta}$  for crossing the axis occurs. It is identical with the critical angle for total internal reflection in a step-index fiber and yields also the same relation to the NA (Eq. (1.9)).

### Example: Propagation velocity in gradient-index fibers

Along the fiber the velocity of the light ray slightly oscillates with elongation from the axis. We calculate the travel time for a period  $\Lambda$ :

$$T = \int_0^\Lambda dt = \int_0^\Lambda dz/v_z(z).$$

The speed depends on the refractive index profile (Eq. (1.11)),  $v(z) = c/n(r(z))$ . The  $z$ -component is calculated from  $v_z(z) = v(z)/\sqrt{1 + r'(z)^2}$ , yielding

$$T = \int_0^\Lambda \frac{n_1}{c} \left( 1 - \Delta(r_0/a)^2 \sin^2 Kz \right) \left( 1 + 2\Delta(r_0/a)^2 \cos^2 Kz \right)^{1/2} dz.$$

Due to  $\Delta(r_0/a)^2 \ll 1$ , the contribution of the oscillatory term is very small, and we can expand

$$\begin{aligned} T &= \int_0^\Lambda \frac{n_1}{c} \left( 1 - \Delta \frac{r_0^2}{a^2} \sin^2 Kz \right) \left( 1 + \Delta \frac{r_0^2}{a^2} \cos^2 Kz + \dots \right) dz \\ &= \frac{n_1 \Lambda}{c} \int_0^{2\pi} \left( 1 - \Delta \frac{r_0^2}{a^2} \sin^2 x + \Delta \frac{r_0^2}{a^2} \cos^2 x + \mathcal{O}(\Delta^2(r_0/a)^4) \right) dx. \end{aligned}$$

Elementary integration then delivers the remarkable result:

$$T \sim n_1 \Lambda / c.$$

In contrast with step-index fibers, the propagation velocity within a GRIN fiber does not depend on the initial propagation angle, at least to the order of  $(\Delta(r_0/a)^2)^2$ .

An interesting situation occurs for bundles of parallel rays: They are focused after a length  $\Lambda/4$ . Short fiber sections with this length are used as so-called GRIN lenses. As outlined in Chapter 4.1, the reduced dependence on the propagation angle plays an important role for signal transmission in optical fibers.

## 1.8 Lenses and Curved Mirrors

The formation of an image plays a major role in optics, and *lenses* and *curved mirrors* are essential parts in optical devices. First, we will discuss the effect of these components on the propagation of rays; owing to its great importance, we have dedicated an extra chapter (Chapter 5) to the formation of images.

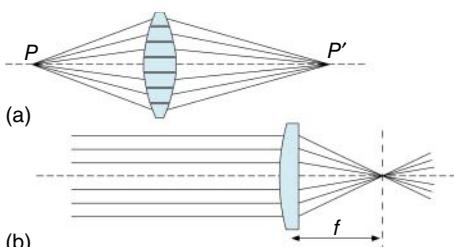
### 1.8.1 Lenses

We define an *ideal lens* as an optical element that merges all rays of a point-like source into one point again. An image where all possible object points are transferred into image points is called a *stigmatic* image (from the Greek word *stigma*, meaning point). The source may even be far away and illuminate the lens with a parallel bundle of rays. In this case the point of merger is called the *focal point* or *focus*. In Figure 1.14, we consider a beam of parallel rays that passes through the lens and is merged at the focal point. According to Fermat's principle, the optical path length must be equal for all possible paths, which means that they are independent of the distance of a partial beam from the axis. Then the propagation of light must be delayed most on the symmetry axis of the lens and less and less in the outer areas!

For a simplified analysis, we neglect the thickness of the lens body, consider the geometrical increase of the path length from the lens to the focal point at a distance  $f$ , and expand the term as a function of distance  $r$  from the axis:

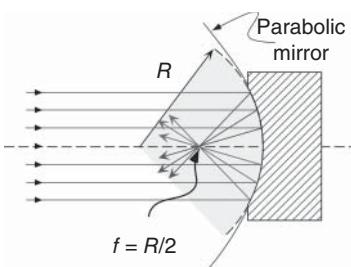
$$\ell(r) = \sqrt{f^2 + r^2} \approx f \left( 1 + \frac{r^2}{2f^2} \right).$$

To compensate for the quadratic increase of the optical path length  $\ell(r)$ , the delay by the path within the lens glass – that is, the thickness – must also vary quadratically. This is actually the condition for spherical surfaces, which have been shown to be extremely successful for convergent lenses! The result is the same with much more mathematical effort, if one explores the properties of refraction at a lens surface assuming that a lens is constructed of many thin prisms (Figure 1.14). In the chapter on lens aberrations, we will deal with the question of which criteria should be important for the choice of a planar convex or biconvex lens.



**Figure 1.14** (a) Stigmatic lens imaging. All rays starting at object point  $P$  are merged again at image point  $P'$ . The light rays are delayed more near the axis of the lens body than in the outermost areas, so that all rays make the same optical path length to the image point. A lens may be figured as a combination of several prisms. (b) A parallel beam of rays originating from a source at infinite distance is focused at the focal point at focal distance  $f$ .

### 1.8.2 Concave Mirrors



**Figure 1.15** Path of rays for a concave mirror. For near-axis incident light, spherical mirrors are used.

and the effect on a parallel beam Figure 1.15.

The reflected partial rays meet at the focal point or focus on the mirror axis, as they do in the case of a lens. It is known from geometry that the reflection points must then lie on a parabola. Near the axis, parabolic mirrors may in good approximation be substituted by spherical mirrors, which are much easier to manufacture. In Figure 1.16b the geometrical elements are shown, from which the dependence of the focal length (defined here by the intersection point with the optical axis) on the axis distance  $y_0$  of a parallel incident beam may be calculated:

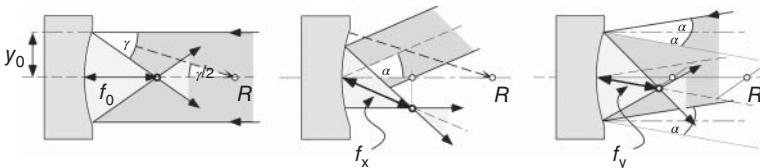
$$f = R - \frac{R}{2 \cos \alpha} \approx \frac{R}{2} \left[ 1 - \frac{1}{2} \left( \frac{y_0}{R} \right)^2 + \dots \right].$$

In general we neglect the quadratic correction, which causes an aperture error and is investigated in more detail in Section 5.7.2.3.

In laser resonators a situation often occurs in which spherical mirrors are simultaneously used as deflection mirrors, for example, in the “bowtie resonator” shown in Figure 8.33. Then the focal width of the rays within the ray plane ( $f_x$ ) and within the plane perpendicular to that one ( $f_y$ ) will differ from  $f_0 = R/2$ :

$$f_x = \frac{R}{2 \cos \alpha} = \frac{f_0}{\cos \alpha} \quad \text{and} \quad f_y = \frac{R \cos \alpha}{2} = f_0 \cos \alpha.$$

The geometrical situations in the top view (Figure 1.16, center) are easy to see. In the side view one looks at the projection onto a plane perpendicular to the direction of emergence. The projections of the radius and focal length are reduced to  $R \cos \alpha$  and  $f \cos \alpha$ , respectively. The difference between the two planes occurring here is called *astigmatic aberration* and sometimes can be compensated by simple means (see, e.g., p. 175).



**Figure 1.16** Focusing an incident beam that is parallel to (left) and oblique to (centre: top view; right: side view) the optical axis.

Among curved mirrors, concave spherical or parabolic mirrors play the most important role. They are very well known from huge astronomical telescopes (see Chapter 5), because we entered the fascinating world of the cosmos with their aid. But they are used much more often in laser resonators (Section 6.6). Taking into account the tangential plane at the intercept of the surface normal at the lens surface, we can transfer the conditions of planar reflection to curved mirror surfaces. Concave mirrors mostly have axial symmetry,

## 1.9 Matrix Optics

As a result of its rectilinear propagation, a free light ray may be treated like a straight line. In optics, systems with axial symmetry are especially important, and an individual light ray may be described sufficiently well by the distance from and angle to the axis (Figure 1.17). If the system is not rotationally symmetric, for example, after passing through a cylindrical lens, then we can deal with two independent contributions in the  $x$  and  $y$  directions with the same method.

The modification of the beam direction by optical components – mirrors, lenses, and dielectric surfaces – is described by a trigonometric and therefore not always simple relation. For near-axis rays, these functions can often be linearized, and thus the mathematical treatment is simplified enormously. This becomes obvious, for example, for a linearized form of the law of refraction (1.2):

$$n_1 \theta_1 = n_2 \theta_2. \quad (1.13)$$

Here we have made use of this approximation already with the application of Fermat's principle for ideal lenses. Near-axis rays allow the application of spherical surfaces for lenses, which are much easier to manufacture than mathematical ideal surfaces. Furthermore, ideal systems are only “ideal” for selected ray systems; otherwise they suffer from image aberrations like other systems.

When treating the modification of a light ray by optical elements in this approximation by linear transformation, matrices are a convenient mathematical tool for calculating the fundamental properties of optical systems. The development of this method led to the denomination *matrix optics*. The introduction of transformation matrices for ray optics may be visualized very easily, but they achieved striking importance because they do not change their form when treating near-axis rays according to wave optics (see Section 2.3.2). Furthermore this formalism is also applicable for other types of optics such as “electron optics” or the even more general “particle optics.”

### 1.9.1 Paraxial Approximation

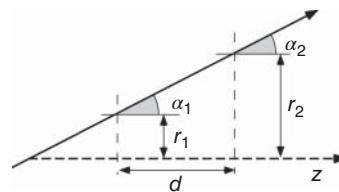
Let us consider the propagation of a light ray at a small angle  $\alpha$  to the  $z$  axis. The beam is fully determined by the distance  $r$  from the  $z$  axis and the slope  $r' = \tan \alpha$ . Within the so-called paraxial approximation, we now linearize the tangent of the angle and substitute it by its argument,  $r' \approx \alpha$ , and then merge  $r$  with  $r'$  to end up with a vector  $\mathbf{r} = (r, \alpha)$ . At the start, a light ray may have a distance to the axis and a slope of  $\mathbf{r}_1 = (r_1, \alpha_1)$ . Having passed a distance  $d$  along the  $z$  axis, then

$$r_2 = r_1 + \alpha_1 d,$$

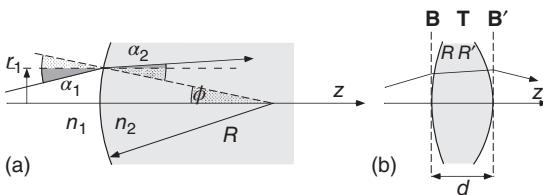
$$\alpha_2 = \alpha_1,$$

hold. One may use  $2 \times 2$  matrices to write the translation clearly:

$$\mathbf{r}_2 = \mathbf{T} \mathbf{r}_1 = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \mathbf{r}_1. \quad (1.14)$$



**Figure 1.17** Key variables of an optical ray for simple translation.



**Figure 1.18** Modification of a light ray at curved refracting surfaces.

A bit more complicated is the modification by a refracting optical surface. For that purpose we look at the situation shown in Figure 1.18, where two optical media with refractive indices  $n_1$  and  $n_2$  are separated by a spherical interface with radius  $R$ . If the radius vector subtends an angle  $\phi$  with the  $z$  axis, then the light ray is obviously incident on the surface at an angle  $\theta_1 = \alpha_1 + \phi$  and is related to the angle of emergence by the law of refraction. In paraxial approximation according to Eq. (1.2),  $n_1 \theta_1 \approx n_2 \theta_2$  and  $\phi \approx r_1/R$  is valid, and one finds

$$n_1 \left( \alpha_1 + \frac{r_1}{R} \right) = n_2 \left( \alpha_2 + \frac{r_2}{R} \right).$$

The linearized relations may be described easily by the refraction matrix  $\mathbf{B}$ :

$$\begin{pmatrix} r_2 \\ \alpha_2 \end{pmatrix} = \mathbf{B} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ (n_1 - n_2)/n_2 R & n_1/n_2 \end{pmatrix} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix}. \quad (1.15)$$

### 1.9.2 ABCD Matrices

The most important optical elements may be specified by their transformations, also called *ABCD matrices*  $\mathbf{M}_{\text{ABCD}}$ ,

$$\begin{pmatrix} r_2 \\ \alpha_2 \end{pmatrix} = \mathbf{M}_{\text{ABCD}} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix}, \quad (1.16)$$

which we collect in Table 1.2 for look-up purposes and will be presented in the following in more detail.

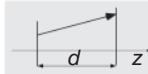
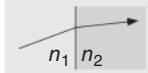
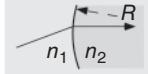
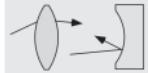
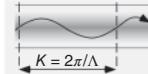
According to Figure 1.18, the effect of a lens on a light ray is characterized by a refraction  $\mathbf{B}$  at the entrance, a translation  $\mathbf{T}$  in the glass, and one further refraction  $\mathbf{B}'$  at the exit. Now the matrix method shows its strength, because the linearity allows to calculate the properties of an entire system from the product of the component matrices:

$$\begin{pmatrix} r_2 \\ \alpha_2 \end{pmatrix} = \mathbf{M}_{\text{sys}} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix} = \cdots \mathbf{M}_3 \cdot \mathbf{M}_2 \cdot \mathbf{M}_1 \cdot \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix}. \quad (1.17)$$

Before we discuss the lens and some more examples in detail, we have to fix some conventions, which in general are used in matrix optics:

- 1) The ray direction goes from left to right in the positive direction of the  $z$  axis.
- 2) The radius of a convex surface is positive,  $R > 0$ , and that of a concave surface is negative,  $R < 0$ .
- 3) The slope is positive when the beam moves away from the axis and negative when it moves toward the axis.

**Table 1.2** Important ABCD matrices.

Operation	ABCD matrix
Translation	 $\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}$
Refraction (planar surface)	 $\begin{pmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{pmatrix}$
Refraction (curved surface)	 $\begin{pmatrix} 1 & 0 \\ \frac{n_1-n_2}{n_2 R} & \frac{n_1}{n_2} \end{pmatrix}$
Lenses and curved mirrors (focal length $f$ )	 $\begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}$
Optical fiber GRIN (length $\ell$ )	 $\begin{pmatrix} \cos K\ell & K^{-1} \sin K\ell \\ -K \sin K\ell & \cos K\ell \end{pmatrix}$

- 4) An object distance or image distance is positive (negative) when lying in front of (behind) the optical element.
- 5) Object distances are defined to be positive (negative) above (below) the  $z$  axis.
- 6) Reflective optics is treated by flipping the ray path after every element.

A useful property of the ABCD matrices is their determinant<sup>2</sup>:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = n_1/n_2, \quad (1.18)$$

which is easily seen from Table 1.2 and can be understood that for all local operations without translation, we must have  $B = 0$  while translations change the slope only in infinitesimal ways ( $C = 0$ ). Since the determinant of the product of matrices is equal to the product of their determinants, Eq. (1.18) holds also for an entire system with  $|\mathbf{M}_{\text{sys}}| = 1$  for rays entering from and exiting in air.

### 1.9.3 Lenses in Air

The effect of a lens can easily be expressed as a product  $\mathbf{L} = \mathbf{B}'\mathbf{T}\mathbf{B}$  of three operations:

$$\begin{pmatrix} r_2 \\ \alpha_2 \end{pmatrix} = \mathbf{L} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix} = \mathbf{B}'\mathbf{T}\mathbf{B} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix}. \quad (1.19)$$

<sup>2</sup> Siegman's convention  $\tilde{\alpha}_i = n_i \alpha_i$  yields in general  $|\mathbf{M}| = 1$  [3]. For actual calculations we have to decide for one convention.

Explicit calculation of  $\mathbf{L}$  according to Eq. (1.16) and taking into account the index of refraction  $n_{\text{air}} = 1$  in Eqs. (1.14) and (1.15) yields

$$\mathbf{L} = \begin{pmatrix} 1 - \frac{n-1}{n} \frac{d}{R} & \frac{d}{n} \\ (n-1) \left[ \frac{1}{R'} - \frac{1}{R} - \frac{d(n-1)^2}{RR'n} \right] & 1 + \frac{n-1}{n} \frac{d}{R'} \end{pmatrix}.$$

It makes a complicated and not very convenient expression at first sight. Though it may allow the treatment of very thick lenses, by far the most important are the predominantly used “thin” lenses, whose thickness  $d$  is small compared to the radii of curvature  $R, R'$  of the surfaces. With  $d/R, d/R' \ll 1$  or by direct multiplication  $\mathbf{B}'\mathbf{B}$ , we find the much simpler form

$$\mathbf{L} \approx \begin{pmatrix} 1 & 0 \\ (n-1) \left( \frac{1}{R'} - \frac{1}{R} \right) & 1 \end{pmatrix}$$

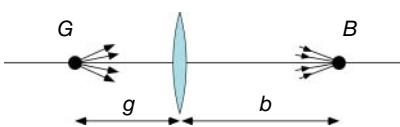
and introduce the symbol  $D$  for the *refractive power* in the lens maker’s equation:

$$D = -(n-1) \left( \frac{1}{R'} - \frac{1}{R} \right) = \frac{1}{f}. \quad (1.20)$$

Thus the ABCD matrix for thin lenses becomes very simple,

$$\mathbf{L} = \begin{pmatrix} 1 & 0 \\ -D & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}, \quad (1.21)$$

where the sign is chosen such that convergent lenses have a positive refractive power. The refractive power is identical with the inverse focal length,  $D = 1/f$ . The refractive power  $D$  is measured in units of dioptres ( $1 \text{ dpt} = 1 \text{ m}^{-1}$ ).



**Figure 1.19** Point image formation with a lens.

To support the interpretation of Eq. (1.21), we consider a bundle of rays that originates from a point source  $G$  on the  $z$  axis (Figure 1.19). Such a bundle of rays can be described at a distance  $g$  from the source according to

$$\begin{pmatrix} r \\ \alpha \end{pmatrix} = \alpha \begin{pmatrix} g \\ 1 \end{pmatrix}. \quad (1.22)$$

We calculate the effect of the lens in the form

$$\mathbf{L} \begin{pmatrix} r \\ \alpha \end{pmatrix} = \alpha \begin{pmatrix} g \\ 1 - g/f \end{pmatrix} = \alpha' \begin{pmatrix} -b \\ 1 \end{pmatrix}. \quad (1.23)$$

The lens transforms the incident bundle of rays into a new bundle, which again has the form (1.22). It converges for  $\alpha' < 0$  to the axis, crosses it at a distance  $b > 0$  (convention 4) behind the lens, and creates there an image of the point source. If  $b < 0$ , then the virtual image of the point source lies in front of the lens and the lens has the properties of a dispersive lens.

By comparison of coefficients, we obtain the relation between object distance  $g$  and image distance  $b$  from Eq. (1.23) for lens imaging:

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b}. \quad (1.24)$$

This equation is the known basis for optical imaging. We refer to this topic again in Chapter 4 in more detail.

### Example: ABCD matrix of an imaging system

For imaging by an arbitrary ABCD system, we must claim that a bundle of rays  $(r_1, \alpha_1)$  is again merged at a point at a certain distance  $d = d_1 + d_2$ :

$$\begin{pmatrix} r_2 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 1 & d_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix}.$$

For stigmatic imaging,  $r_2$  must be independent of  $\alpha_1$ , and by calculation one finds the condition  $d_1 D + d_2 A + d_1 d_2 C + B = 0$ , which for  $B = 0$  can be fulfilled by suitable choice of  $d_1$  and  $d_2$ , even if  $C < 0$ . Thus, the ABCD matrix takes exactly the form that we know already from lenses and lens systems.

#### 1.9.4 Lens Systems

The matrix method enables us to explore the effect of a system consisting of two lenses with focal lengths  $f_1$  and  $f_2$  at a distance  $d$ . We multiply the ABCD matrices according to Eqs. (1.21) and (1.14) and get the matrix  $\mathbf{M}$  of the system:

$$\begin{aligned} \mathbf{M} &= \mathbf{L}_2 \mathbf{T} \mathbf{L}_1 = \begin{pmatrix} 1 & 0 \\ -1/f_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \frac{d}{f_1} & d \\ -\left(\frac{1}{f_2} + \frac{1}{f_1} - \frac{d}{f_1 f_2}\right) & 1 - \frac{d}{f_2} \end{pmatrix}. \end{aligned} \quad (1.25)$$

The system of two lenses substitutes a single lens with focal length given by

$$\frac{1}{f} = \frac{1}{f_2} + \frac{1}{f_1} - \frac{d}{f_1 f_2}. \quad (1.26)$$

We consider the following three interesting special cases:

- 1)  $d \ll f_{1,2}$ : Two lenses that are mounted directly next to each other, with no space between them, add their refractive powers,  $\mathbf{M} \simeq \mathbf{L}_2 \mathbf{L}_1$ , with  $D = D_1 + D_2$ . This circumstance is used, for example, when adjusting eyeglasses, when refractive powers are combined until the required correction is found. Obviously a biconvex lens can be constructed out of two planar convex lenses, expecting that the focal length of the system is divided by 2.
- 2)  $d = f_1 + f_2$ : If the focal points coincide, a telescope is realized. A parallel bundle of rays with radius  $r_1$  is widened or collimated into a new bundle of parallel

rays with a new diameter  $(f_2/f_1)r_1$ . The refractive power of the system vanishes according to Eq. (1.26):  $D = 0$ . Such systems are called *afocal*.

- 3)  $d \gg f_1 f_2$ : In microscopes and other applications, effective short focal length lenses are constructed by combining two or more lenses. Application of Eq. 1.26 yields  $f \simeq -f_1 f_2/d$  where  $d$  could be the tubus length (see Chapter 5).

A thin lens is one of the oldest optical instruments, and thin lenses may have many different designs due to their various applications. But since lens aberrations are of major interest, we will dedicate a specific section to the various designs (Section 5.7.1).

### 1.9.5 Periodic Lens Systems

Periodic lens systems had already been analyzed in early times to realize optical light transmission lines. For such an application, it is important that a light ray does not leave the system even after long distances. We consider a periodic variant of the lens system with focal lengths  $f_1$  and  $f_2$  at a distance  $d$ . For that purpose we add one more identical translation to the transformation matrix from Eq. (1.25), which yields a system equivalent to a system of two concave mirrors (Figure 1.20):

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1/f_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & d \\ -1/f_2 & 1 - d/f_2 \end{pmatrix} \begin{pmatrix} 1 & d \\ -1/f_1 & 1 - d/f_1 \end{pmatrix}.$$

Now for  $n$ -fold application, the individual element will cause total transformation

$$\begin{pmatrix} A_n & B_n \\ C_n & D_n \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^n.$$

Introducing

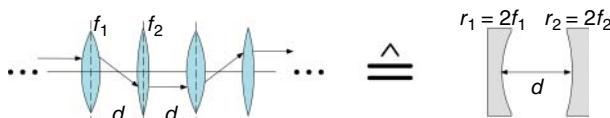
$$\cos \Theta = \frac{1}{2}(A + D) = 2 \left( 1 - \frac{d}{2f_1} \right) \left( 1 - \frac{d}{2f_2} \right) - 1, \quad (1.27)$$

this matrix can be evaluated algebraically. Thus one calculates

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^n = \frac{1}{\sin \Theta} \begin{pmatrix} A \sin n\Theta - \sin(n-1)\Theta & B \sin n\Theta \\ C \sin n\Theta & D \sin n\Theta - \sin(n-1)\Theta \end{pmatrix}.$$

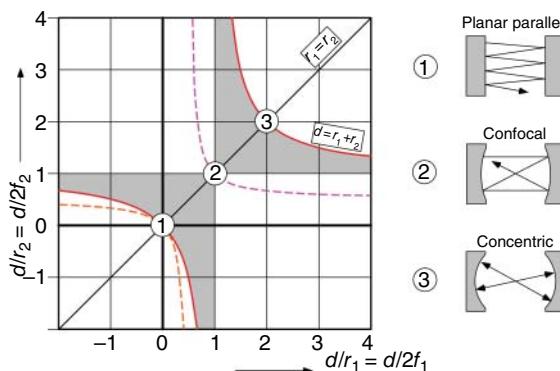
The angle  $\Theta$  must remain real to avoid the matrix coefficients increasing to infinity. Otherwise the light ray would actually leave the lens system. Thus from the properties of the cosine function,

$$-1 \leq \cos \Theta \leq 1,$$



**Figure 1.20** Periodic lens system and equivalence to a two-mirror resonator.

**Figure 1.21** Stability diagram for lenses and optical resonators according to the condition (1.28). Stable resonator configurations are within the hatched area. The dashed lines indicate the positions of confocal resonators,  $d = (r_1 + r_2)/2$ . Symmetric planar parallel, confocal, and concentric resonators are at the positions circled 1, 2, and 3, respectively.



and in combination with Eq. (1.27), we get

$$0 \leq \left(1 - \frac{d}{2f_1}\right) \left(1 - \frac{d}{2f_2}\right) \leq 1. \quad (1.28)$$

This result defines a stability criterion for the application of a wave guide consisting of lens systems, and the corresponding important stability diagram is shown in Figure 1.21. We will deal with this in more detail later on, because multiple reflection between concave mirrors of an optical resonator can be described in this way as well (Section 6.6).

### 1.9.6 ABCD Matrices for Wave Guides

According to Section 1.7 and with the aid of the wave number constant  $K = 2\pi/\lambda$  (Eq. (1.12)), a simple ABCD matrix for the transformation of a ray by a graded-index fiber of length  $\ell$  can be specified:

$$\mathbf{G} = \begin{pmatrix} \cos K\ell & K^{-1} \sin K\ell \\ -K \sin K\ell & \cos K\ell \end{pmatrix}. \quad (1.29)$$

With short pieces of fiber ( $K\ell < \pi/4$ ), also thin lenses can be realized, and it can be shown that the focal point lies at  $f = K^{-1} \cot K\ell$ . These components are called *GRIN lenses*.

## 1.10 Ray Optics and Particle Optics

Traditional optics, which deals with light rays and is the topic of this textbook, was conceptually in every respect a role model for “particle optics,” which started around the year 1900 with the exploration of electron beams and radioactive rays. Since ray optics describes the propagation of light rays, it is convenient to look for analogies in the trajectories of particles. We will see in the chapter on coherence and interferometry (Chapter 6) that the wave aspects of particle beams are widely described in terms of the ideas of optics as well.

To reestablish the analogy explicitly, we refer to considerations about Fermat’s principle (p. 5), because there a relation between the velocity of light and the index

of refraction is described. This relation is particularly simple if a particle moves in a conservative potential (potential energy  $E_{\text{pot}}(\mathbf{r})$ ), for example, an electron in an electric field. As a result of energy conservation,

$$E_{\text{kin}}(\mathbf{r}) + E_{\text{pot}}(\mathbf{r}) = E_{\text{tot}},$$

we can immediately infer from  $E_{\text{kin}} = mv^2/2$  that

$$v(\mathbf{r}) = \sqrt{\frac{2}{m}[E_{\text{tot}} - E_{\text{pot}}(\mathbf{r})]}$$

if the particles do not move too fast, and we can adopt classical Newtonian mechanics (in a particle accelerator, the special theory of relativity has to be applied).

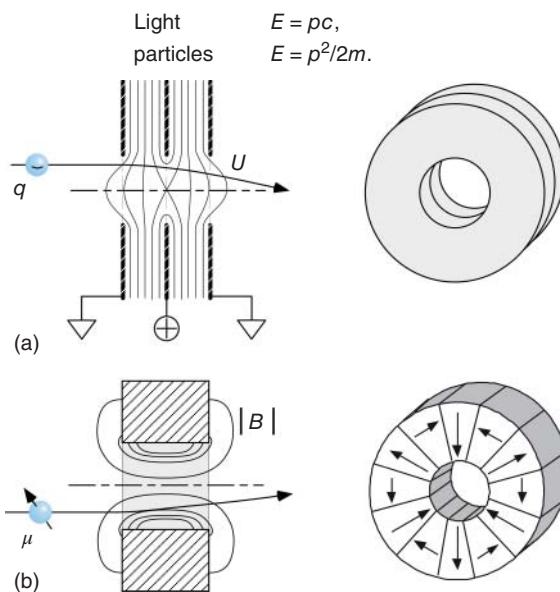
We can define an effective relative index of refraction by

$$\frac{v(\mathbf{r}_1)}{v(\mathbf{r}_2)} = \frac{n_{\text{eff}}(\mathbf{r}_2)}{n_{\text{eff}}(\mathbf{r}_1)} = \frac{\sqrt{[E_{\text{tot}} - E_{\text{pot}}(\mathbf{r}_2)]}}{\sqrt{[E_{\text{tot}} - E_{\text{pot}}(\mathbf{r}_1)]}}.$$

As in the case of light, it must satisfy an additional condition to be defined absolutely. For example, we may claim that  $n_{\text{eff}} = 1$  for  $E_{\text{pot}} = 0$ . But then it is obvious that  $n_{\text{eff}}$  depends extremely on the velocity outside of the potential – particle optics has properties that are very much chromatic! The fundamental reason for this difference is the different relation between kinetic energy  $E$  and momentum  $p$  for light and for particles having mass, which is also called the *dispersion relation*:

light	$E = pc,$
particles	$E = p^2/2m.$

In charged particle beams a narrow velocity distribution can be created by acceleration, which makes the difference not very pronounced. But the broadness

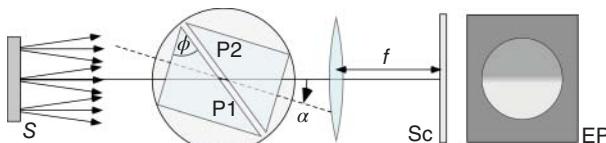


**Figure 1.22** Lenses for particle optics. (a) The so-called single lens for electron optics with equipotential surfaces  $qU$  [4]. The potential is created by symmetric positioning of three conducting electrodes, the two outer ones lying on the same potential. (b) Magnetic lens for atom optics with equipotential surfaces  $|\mu \cdot B|$  [5, 6]. An axial magnetic hexapole is formed out of circle segments, which are manufactured from a homogeneously magnetized permanent magnet (e.g., NdFeB or SmCo). The strength of the magnetic field rises as a square function of the radial distance.

of the velocity distribution in thermal beams of neutral atoms induces significant problems. Indeed, this velocity distribution can be manipulated by the so-called supersonic jets or by laser cooling (see Section 12.5) in such a way that even “atom optics” can be established [5, 6]. We present some important devices of electron and atom optics in Figure 1.22.

## Problems

- 1.1 **Sun spots** In the shade of a densely leaved tree (in the light transmitted through the gaps in window blinds), one observes with clear sky conditions numerous round light spots. What is their origin? How do they depend on the shape of the gaps in the leaves?
- 1.2 **Mirror images** Why does a mirror exchange left and right but not up and down?
- 1.3 **Parabolic mirror with skewed rays** What is the image of a bundle of parallel rays entering a parabolic mirror with a nonvanishing angle relative to its axis?
- 1.4 **Rainbow** Explain the origin of the rainbow. Estimate the dispersion of water  $dn/d\lambda$  from the appearance of colors. The index of refraction of water is about  $n = 1.33$ . (The precise value depends on the specific color.) To a good approximation and over a wide range of parameters, we can consider raindrops to be spherical. The wavelengths limiting the visible spectrum are  $\lambda = 700 \text{ nm}$  on the red side and  $\lambda = 400 \text{ nm}$  on the violet side. (Reminder:  $d/dx(\arcsin(x)) = 1/(1-x^2)^{1/2}$ .)
- 1.5 **Refractometer** Abbe’s refractometer is an instrument used for the determination of the refractive index of liquids. In this instrument, a droplet of the liquid is dabbed onto a glass prism, and then a second glass prism is folded onto it. The double prism is now rotated until a sharp border between light and darkness appears on the screen or in the eyepiece of a telescope (Figure 1.23). The refractive index of the liquid can then be determined from the rotation angle. (Sometimes there is a second ruler indicating immediately the index of refraction.) Can you imagine how this device works? If the index of refraction of the glass prisms is  $n$ , what is the range of refractive indices that can be measured with this method?



**Figure 1.23** Abbe’s refractometer, illuminated with a diffusive source  $S$ . The liquid is filled in the thin gap between parallel prisms  $P_1$  and  $P_2$ . The double prism is rotated by an angle  $\alpha$  until a half-bright/half-dark image  $EP$  appears on the screen  $Sc$ . The lens has focal length  $f$ .

- 1.6 Halo** The most frequent halo phenomenon is a circumferential ring for the sun or the moon with  $22^\circ$  opening angle and with a weak red color on the inner side. It is caused by tiny ice crystals in the atmosphere with refractive index 1.31. They are formed in the high troposphere and have predominantly prism shape with equilateral triangular cross section. How does the  $22^\circ$  halo arise?
- 1.7 Fermat's principle** In a simplified way Fermat's principle states that a light ray propagates from one point to another on a trajectory that finds the shortest travel time. Derive the laws of reflection and of refraction using this principle.
- 1.8 Curvature of light rays** Derive the measure for the curvature of a light ray (the second derivative of the trajectory) in a medium with inhomogeneous though steady refractive index by purely geometric optical means. Avoid the case where the ray travels orthogonally to the gradient of  $n$ .
- 1.9 Deflection by a prism (I)** Deflection is minimal for symmetric transit of an equilateral prism. Show that this property is a direct consequence of the reversibility of the path of the light ray.
- 1.10 Deflection by a prism (II)** Show that the refractive index can be extracted from the minimal deflection angle of a light ray  $\delta_{\min}$  by a symmetric prism from  $n = \sin[(\delta_{\min} + \alpha)/2]/\sin(\alpha/2)$ . What is the best choice for  $\alpha$  to obtain maximum precision?
- 1.11 Optical fibers (I)** Consider an optical fiber formed by a core with refractive index  $n_1 = 1.465$  and the cladding with  $n_2 = 1.4500$ . Calculate the largest angle aperture (half the opening angle of the light cone hitting the fiber) for which the light ray is still guided by the straight fiber. Suppose the core diameter is  $50 \mu\text{m}$ . How large is the smallest radius of curvature by which the fiber may be bent before strong losses occur?
- 1.12 Modal dispersion** Consider an optical light pulse of length  $T$ . If on insertion of the pulse into a fiber the light field is uniformly distributed across all angles above the critical angle for total internal reflection, the partial rays will disperse in time and propagate with different velocities along the fiber axis. How short a duration is acceptable for the pulse if the pulse length must not grow by more than 50% while propagating a distance  $\ell$ ?
- 1.13 Cylindrical lens, astigmatism** What are the imaging properties of a cylindrical lens (the refracting surface is cut out from a cylinder)? Is it possible to realize point-like imaging with two cylindrical lenses? Are two cylindrical lenses equivalent to a single spherical lens? Explain why the optometrist calls astigmatism also cylindrical aberration.
- 1.14 Determination of focal length (I)** Find out how the focal length of a lens can be rapidly estimated and how to determine it precisely. If you are wearing glasses, try it out. How many dioptres do your glasses have?

- 1.15 Bessel's method for determination of focal length (II)** See Problem 5.3.
- 1.16 Newton's equation** Show by calculation and by geometrical construction that the equation for image formation with a lens (Eq. (1.24)) is equivalent to  $(g - f)(b - f) = f^2$ . (See also Eq. (4.1).)
- 1.17 Depth of focus** How large is the depth of focus for image formation with a spherical mirror? How do you meaningfully define the depth of focus for observation with the naked eye and for photographers? How can one enhance the depth of focus?
- 1.18 Lens and glass plate** Use the ABCD method to investigate the influence of a plain parallel glass plate with thickness  $d$  on the focusing action of a lens with  $f > d$ . Assume that the glass plate is oriented normally to the beam axis and located within the focal length of the lens. Use this system to determine the refractive index of the glass plate. Estimate the accuracy of the method.
- 1.19 Optical fibers (II)** A small glass sphere (radius  $R$ , refractive index  $n$ ), which is placed immediately at the entrance facet of an optical fiber, can be used to couple light into the fiber. Calculate the ABCD matrix for a glass sphere and the transformation of a collimated bundle of light rays transiting the glass sphere. Discuss the optimal parameters ( $R, n$ ) for the sphere in order to couple light most efficiently into the fiber. For a realistic example, use the optical fiber parameters of Problem 1.11.
- 1.20 The determinant of ABCD matrices** The determinants of both the translation matrix  $\mathbf{T}$  (Eq. (1.14)) and the refraction matrix  $\mathbf{B}$  (Eq. (1.15)) are obviously identical to unity. Why does this imply for the lens matrix  $|\mathbf{L}| = 1$  as well? Show that furthermore Newton's equation for thin lenses,  $(f - g)(f - b) = f^2$ , is a consequence of this condition.
- 1.21 Thick lenses and principal planes** For image formation the result of the example from p. 21,  $bD + gA + bgC + B = 0$ , must also hold for thick lenses. Here  $\{b, g\}$  designate the separation of object and image from the intersection of the lens with the  $z$  axis. We can then identify  $C = -1/f$  in the ABCD matrix with the focal length. Show first that the relation  $(fA - g)(fD - b) = f^2$  holds. What is the position of the focal points of the lens? Rewrite the equation in the form  $[f - (g - g_p)][f - (b - b_p)] = f^2$ . Points  $\{b_p, g_p\}$  define the conjugated planes, or principal points. Interpret the result and give the corresponding Newton's equation.
- 1.22 Gardener's fantasies?** Sometimes gardeners advise not to water flowers if the sun is shining, since the focusing action of the water droplets could destroy leaves. Do you accept this advice?

## 2

### Wave Optics

At the beginning of the nineteenth century, a few phenomena were known that could not be reconciled with simple rectilinear, ray-like propagation of light and made wave theory necessary. The beginning is marked by Huygens' principle (after the Dutch mathematician and physicist C. Huygens, 1629–1695), an explanation of wave propagation often used up to now and very intuitive. About 100 years later, T. Young (1773–1829) from England and A. P. Fresnel (1788–1827) from France developed a very successful wave theory, which could explain all the phenomena of interference known at that time. After G. Kirchhoff (1824–1887) had given a mathematical formulation of Huygens' principle, the final breakthrough occurred with the famous Maxwell's equations, which will serve also here as a systematic basis for the wave theory of light.

The development of a common theoretical description of electric and magnetic fields by the Scottish physicist J. C. Maxwell (1831–1879) had a crucial influence not only on physics but also on the science and technology of the twentieth as well as the twenty-first century. Maxwell's equations, which had at first been found through empirical knowledge and aesthetic considerations, caused, for example, Heinrich Hertz in 1887 to excite radio waves for the first time, thereby laying the foundation for modern telecommunications techniques.

#### 2.1 Electromagnetic Radiation Fields

Electromagnetic fields are defined by two vector fields:

$$\begin{aligned}\mathbf{E}(\mathbf{r}, t), & \text{ electric field,} \\ \mathbf{H}(\mathbf{r}, t), & \text{ (magnetic) } H\text{-field.}\end{aligned}$$

They are caused by electric charges and currents. The fundamental magnetic field is

$$\mathbf{B}(\mathbf{r}, t) = \mu_0 \mu_r \mathbf{H}(\mathbf{r}, t),$$

where  $H$ -field and  $B$ -field differ only in the presence of magnetic materials ( $\mu_r \neq 1$ ) (see Eq. (2.6)). In optics it is convenient to use the  $\mathbf{H}$ -field.

### 2.1.1 Static Fields

Charges are the *sources* of electric fields. The formal relation between field strength and charge density  $\rho$  and total charge  $Q$ , respectively, in a volume with surface  $S$  is described by Gauss's law in differential or integral form:

$$\nabla \cdot \mathbf{E} = \rho / \epsilon_0 \quad \text{or} \quad \oint_S \mathbf{E} \cdot d\mathbf{f} = Q / \epsilon_0. \quad (2.1)$$

Furthermore, an electrostatic field is irrotational (curl-free), which means that  $\nabla \times \mathbf{E} = 0$ , and it may be described as the gradient of a scalar electrostatic potential  $\Phi(\mathbf{r})$ :

$$\mathbf{E}(\mathbf{r}) = -\nabla \Phi(\mathbf{r}).$$

The sources of the magnetic field are not charges, because it is known that

$$\nabla \cdot \mathbf{B} = \mu_0 \nabla \cdot \mathbf{H} = 0, \quad (2.2)$$

but instead *curls*, which are caused by currents (current density  $\mathbf{j}$ , total current  $I$  crossing a surface with contour  $C$ ). According to Stokes' law, we have

$$\nabla \times \mathbf{H} = \mathbf{j} \quad \text{or} \quad \oint_C \mathbf{H} \cdot d\mathbf{l} = I. \quad (2.3)$$

### 2.1.2 Polarizable and Magnetizable Media

The considerations of the preceding section are only valid for free charges and currents. But usually these are bound to materials that we can roughly divide into two classes, *conductors* and *insulators*. In conducting materials charges can move freely; in insulators they are bound to a center, but an external field causes a macroscopic dielectric polarization  $\mathbf{P}(\mathbf{r}, t)$  through displacement of charge.<sup>1</sup> For example, polar molecules may be oriented in a water bath, or a charge asymmetry may be induced in initially symmetric molecules (Figure 2.1). In a homogeneous sample, negative and positive charges compensate, and there is left only an effective charge density at the border of the polarized volume. If the polarization varies continuously, then the compensation is incomplete, and one gets an effective charge density

$$\rho_{\text{pol}} = -\nabla \cdot \mathbf{P}(\mathbf{r}, t).$$

Of course, polarization charges contribute to the electric field as well as the density of free charges, and therefore in dielectric matter, it holds that

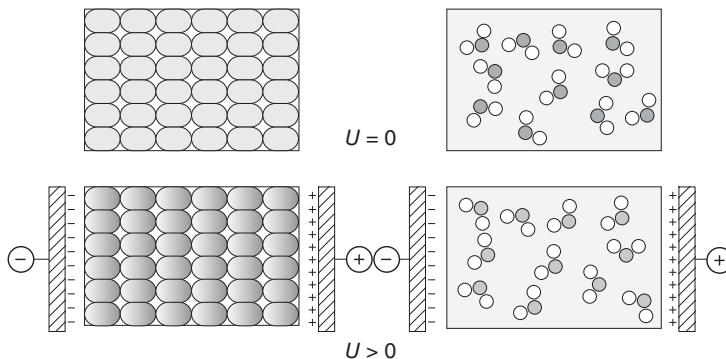
$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} (\rho_{\text{free}} + \rho_{\text{pol}}).$$

In many important optical materials, the polarization charge is proportional to the external field strength, and the coefficient is called the *dielectric susceptibility*  $\chi$ :

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}.$$

---

<sup>1</sup> More precisely it is a polarization density.



**Figure 2.1** In a solid (a), charges are separated. In a glass with polar molecules (b), existing dipoles are oriented.

Here we introduce the *dielectric displacement* phenomenologically with relative dielectric constant (permittivity)  $\epsilon_r = 1 + \chi$ ,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = \epsilon_0 \epsilon_r \mathbf{E}, \quad (2.4)$$

and thus we can write much more simply

$$\nabla \cdot \mathbf{D} = \rho_{\text{free}}. \quad (2.5)$$

It is the task of a microscopic theory or model to calculate the dynamical dielectric function  $\epsilon(\omega) = \epsilon_0 \epsilon_r(\omega) = \epsilon_0 [1 + \chi(\omega)]$  to be used in Eq. (2.4). For simple cases we will discuss this question in the chapter on the interaction of light with matter (Chapter 7). For this we need to take the response of charged, polarizable matter into account, which is described by the Lorentz force.

The analogue to the dielectric polarization for magnetic fields is the magnetic polarization  $\mathbf{M}(\mathbf{r}, t) = \chi_m \mathbf{H}(\mathbf{r}, t)$ , or *magnetization*. In a homogeneous material,  $\mathbf{B}$  and  $\mathbf{H}$  are related through

$$\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M} = \mu_0 (1 + \chi_m) \mathbf{H} = \mu_0 \mu_m \mathbf{H}. \quad (2.6)$$

For most natural materials, the relative magnetic susceptibility  $\mu_m$  is small at optical frequencies; thus  $\mu_m = 1 + \chi_m \approx 1$  is valid. Although of less importance than dielectric phenomena, some magneto-optical effects (e.g., the Faraday effect) play a significant role in optical applications (see Section 3.5.5).

Furthermore, the advent of nanostructured artificial optical materials in the early twenty-first century has paved the route for applications with large (and even negative) magnetic susceptibilities at optical frequencies, giving rise to, for example, negative indices of refraction (see Section 4.3).

### 2.1.3 Dynamic Fields

It is known that a change of the magnetic field within a circuit loop  $C$  induces a voltage. Thus we formulate the law of induction as the third Maxwell equation, where  $S$  is an arbitrary surface bounded by the closed loop  $C$ :

$$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t} \mathbf{B} = -\mu_0 \frac{\partial}{\partial t} \mu_m \mathbf{H} \quad \text{or} \quad \oint_C \mathbf{E} \cdot d\mathbf{l} = -\mu_0 \frac{\partial}{\partial t} \oint_S \mu_m \mathbf{H} \cdot d\mathbf{f}. \quad (2.7)$$

**Table 2.1** Summary: Maxwell–Lorentz equations.

In vacuum	In matter	In $(\omega, \mathbf{k})$ space
Charges are sources of electric field: $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$	$\nabla \cdot \mathbf{D} = \rho$	$i\mathbf{k} \cdot \mathcal{E} = \rho$
No magnetic charges exist: $\nabla \cdot \mathbf{B} = 0$		$i\mathbf{k} \cdot \mathcal{B} = 0$
Law of induction: $\nabla \times \mathbf{E} = -\frac{\partial}{\partial t} \mathbf{B}$		$i\mathbf{k} \times \mathcal{E} = i\omega \mathcal{B}$
Currents are curls of the magnetic field: $c^2 \nabla \times \mathbf{B} = \frac{1}{\epsilon_0} \mathbf{j} + \frac{\partial}{\partial t} \mathbf{E}$	$\nabla \times \mathbf{H} = \mathbf{j} + \frac{\partial}{\partial t} \mathbf{D}$	$i\mathbf{k} \times \mathcal{H} = \mathbf{j} - i\omega \mathcal{D}$
Coulomb–Lorentz force: $m \frac{d^2}{dt^2} \mathbf{r} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$		–

In analogy, a changing dielectric field strength causes a displacement current,  $\mathbf{j}_{\text{dis}} = \epsilon_0(\partial/\partial t)\mathbf{E}$ , and a time-dependent polarization causes a polarization current,  $\mathbf{j}_{\text{pol}} = (\partial/\partial t)\mathbf{P}$ . This yields the complete fourth Maxwell equation for time-varying fields if we consider these contributions in Eq. (2.3) (with  $(\partial/\partial t)\mathbf{D} = \epsilon_0(\partial/\partial t)\epsilon_r \mathbf{E} = \mathbf{j}_{\text{dis}} + \mathbf{j}_{\text{pol}}$ ):

$$\nabla \times \mathbf{H} = \mathbf{j} + \epsilon_0 \frac{\partial}{\partial t} \epsilon_r \mathbf{E}. \quad (2.8)$$

Accounting for the Lorentz force as well, we have now five equations also called the *Maxwell–Lorentz equations*. They are summarized in Table 2.1 in differential and integral form.

Electric and magnetic  $E$ - and  $H$ -fields are frequently and elegantly derived from a common vector potential  $\mathbf{A}(\mathbf{r})$ :

$$\mathbf{E} = -\frac{\partial}{\partial t} \mathbf{A} \quad \text{and} \quad \mathbf{H} = \frac{1}{\mu_0} \nabla \times \mathbf{A}. \quad (2.9)$$

For a complete definition of the potential  $\mathbf{A}$ , an additional condition to ensure so-called gauge invariance is necessary. For this purpose the so-called Coulomb gauge ( $\nabla \cdot \mathbf{A} = 0$ ) is a suitable choice, but in other situations alternatives like the Lorentz gauge for relativistic problems might offer advantages.

#### 2.1.4 Fourier Components

Electric and magnetic fields with harmonic time development are central to optical wave theory. When talking of *Fourier components* of an electromagnetic field, we mean the Fourier amplitudes  $\mathcal{E}$  and  $\mathcal{H}$ :<sup>2</sup>

$$\begin{aligned} E(\mathbf{r}, t) &= \Re \{ \mathcal{E}(\omega, \mathbf{k}) e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})} \}, \\ H(\mathbf{r}, t) &= \Re \{ \mathcal{H}(\omega, \mathbf{k}) e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})} \}. \end{aligned}$$

<sup>2</sup> We will write dynamic electromagnetic fields mainly in complex notation. Therefore the physical fields should always be considered as the real parts, even when this is not expressed explicitly like here.

In general the relation for an amplitude in position and time space,  $\mathcal{A}(\mathbf{r}, t)$ , and the corresponding Fourier or  $(\omega, \mathbf{k})$  dimension can be stated as

$$\begin{aligned}\mathcal{A}(\mathbf{r}, \omega) &= \frac{1}{(2\pi)^{1/2}} \int \mathbf{A}(\mathbf{r}, t) e^{-i\omega t} dt, \\ \mathcal{A}(\mathbf{k}, t) &= \frac{1}{(2\pi)^{3/2}} \int \mathbf{A}(\mathbf{r}, t) e^{i\mathbf{k}\mathbf{r}} d^3r.\end{aligned}$$

Of course, time and space variables may be Fourier-transformed simultaneously. It is particularly convenient to describe monochromatic fields, which have a fixed harmonic frequency  $\omega = 2\pi\nu$ , by Fourier components. Applying Maxwell's equations to this, the differential equations result in vector equations. We collect an overview of all variants in Table 2.1 and add the Coulomb–Lorentz force, which acts on a charge  $q$  at the point  $\mathbf{r}$  and with velocity  $\mathbf{v} = d\mathbf{r}/dt$ .

### 2.1.5 Maxwell's Equations for Optics

For most applications in optics, we can assume that no free charges and currents exist. Concentrating on  $\mathbf{E}$ - and  $\mathbf{H}$ -fields, we find a set of Maxwell's equations convenient for optics with high symmetry:

$$\begin{aligned}\nabla \cdot \epsilon_r \mathbf{E} &= 0, & \nabla \times \mathbf{E} &= -\mu_0 \frac{\partial}{\partial t} \mu_r \mathbf{H}, \\ \nabla \cdot \mathbf{H} &= 0, & \nabla \times \mathbf{H} &= \epsilon_0 \frac{\partial}{\partial t} \epsilon_r \mathbf{E}.\end{aligned}\tag{2.10}$$

### 2.1.6 Continuity Equation and Superposition Principle

We can draw two important conclusions from Maxwell's equations:

- 1) Charges are conserved, as can be found easily by applying the divergence to Eq. (2.3) and applying Eq. (2.1) for the *continuity equation*:

$$\nabla \cdot \mathbf{j} = -\frac{\partial}{\partial t} \rho.$$

- 2) The *superposition principle* is a consequence of the linearity of Maxwell's equations. Two independent electromagnetic fields,  $\mathbf{E}_1$  and  $\mathbf{E}_2$ , are superimposed linearly to yield a superposition field,  $\mathbf{E}_{\text{sup}}$ :

$$\mathbf{E}_{\text{sup}} = \mathbf{E}_1 + \mathbf{E}_2.\tag{2.11}$$

The superposition field is important as a basis for the discussion of interference.

### 2.1.7 The Wave Equation

Propagating electromagnetic fields are a direct consequence of Maxwell's equations. From Eq. (2.10) and the vector identity  $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$ , we find the wave equation in vacuum with  $\epsilon_r = \mu_r = 1$ :

$$\left( \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{E}(\mathbf{r}, t) = 0.\tag{2.12}$$

where we have introduced  $c^2 = 1/\mu_0 \epsilon_0$ . In homogeneous matter we have  $\epsilon_r \mu_r = n^2$  with  $n$  the refractive index (see Chapter 3). The simplified one-dimensional

wave equation in vacuum describes a plane wave with infinite extension into the two transverse directions and can be written in the form

$$\left( \frac{\partial}{\partial z} - \frac{1}{c} \frac{\partial}{\partial t} \right) \left( \frac{\partial}{\partial z} + \frac{1}{c} \frac{\partial}{\partial t} \right) \mathbf{E}(z, t) = 0.$$

By some straightforward algebra, one finds general solutions of the form

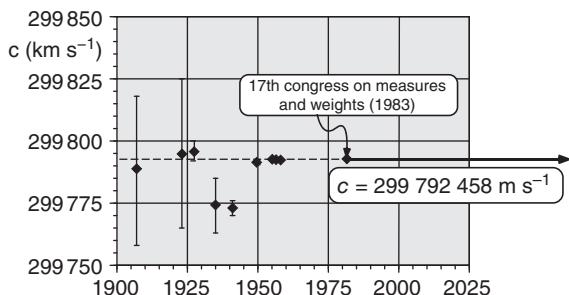
$$\mathbf{E}(z, t) = \mathbf{E}(z \pm ct).$$

The solutions propagate with the *phase velocity*  $c$ , the value of which in vacuum is called the velocity of light  $c$ . The velocity of light is one of the most important universal constants. Its numerical value was measured ever more precisely up to 1983, since when it has been set by definition once and for all to the vacuum value of ( $c$  stands for the Latin word *celeritas*, which means velocity):

$$\text{Velocity of light, } c = 299\,792\,458 \text{ m s}^{-1}.$$

### Excuse: Velocity of light and theory of relativity

According to our direct experience, light propagates "instantaneously." The Danish astronomer Olaf Roemer (1644–1710) discovered in 1676 that the phases of the innermost Jupiter moon Io get shorter when the planet approaches the Earth and longer when it moves away from Earth. From that, he concluded that the propagation of light rays does not occur on an unmeasurably short time scale, but with a finite velocity. From his data Huygens determined the speed of light to be  $225\,000 \text{ km s}^{-1}$  (see also Problem 2.1).



**Figure 2.2** Values of the velocity of light before and after the 17th Congress on Weights and Measures (1983). The diamonds indicate the measured values of various laboratories including uncertainties [7].

Since 1983 the value of the velocity of light has been fixed once and for all by international convention. At first sight it may seem surprising that one may just define a physical universal constant. But it must be considered that velocities are determined by the physical quantities *time* and *distance*, and therefore independent measurements of time and distance are always necessary. Time measurements can be performed by comparison with an atomic time standard (*atomic clock*) with extreme precision, but for distance measurements, such a measuring unit is not available. Therefore the procedure has been inverted, and now – at least in principle – any distance measurement is derived from a much more precise time measurement:

One meter is the distance that light covers in vacuum within  $1/299\,792\,458 \text{ s}$ .

The velocity of light plays a central role for the special theory of relativity, invented by Einstein [8]. In a famous interference experiment, US physicists Michelson and Morley had obtained first evidence in 1886 that in the rest frame of an observer, light *always* propagates with the same velocity  $c$ , no matter what the motional state is of the light source itself. One of the consequences of this theory is that no particle or object, nor even any *action* of physical origin, can move or propagate faster than the velocity of light  $c$ .

The theory of relativity epitomizes an outstanding intersection point of classical and modern physics. Owing to the theory, it is necessary that the equations of mechanical motion are modified for very high velocities. From the very start, Maxwell's equations, describing the propagation of light, have been consistent with the theory of relativity. This property is called "relativistic invariance".

The modification of the phase velocity in homogeneous media was already introduced in Section 1.4.1 from the ray optical point of view by means of the index of refraction  $n$ . According to Eq. (2.12), the value for the phase velocity is now modified to  $c^2 \rightarrow c^2/\epsilon_r\mu_r = c^2/n^2$ . Hence the index of refraction is

$$n = \sqrt{\epsilon_r\mu_r}. \quad (2.13)$$

Note that the sign of  $n$  is not determined by this relation, in contrast to our phenomenological assumption of a positive  $n$  in Section 1.4.1! We will further discuss this interesting observation in Section 4.3.

The wave equation is simplified if monochromatic waves with harmonic time evolution are considered. We use complex numbers, because in that way many waveforms can be discussed formally in a clear manner.

In general only the real part of the complex amplitude is considered as a physically real quantity. Inserting into Eq. (2.12)

$$\mathbf{E}(\mathbf{r}, t) = \Re\{\mathbf{E}(\mathbf{r}) e^{-i\omega t}\}$$

yields, with  $k^2 = \epsilon_r\mu_r\omega^2/c^2 = n^2\omega^2/c^2$ , the *Helmholtz equation*, which depends only on the position  $\mathbf{r}$ :

$$(\nabla^2 + k^2) \mathbf{E}(\mathbf{r}) = 0. \quad (2.14)$$

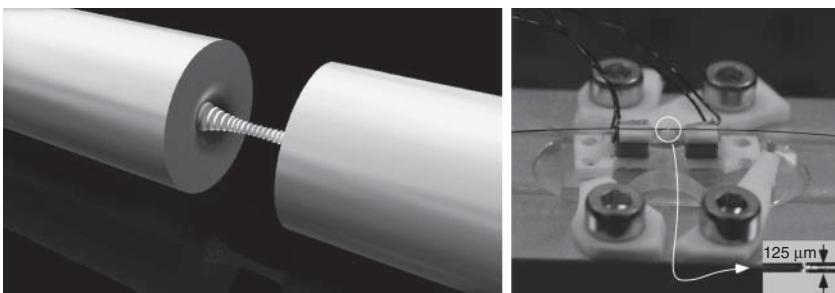
### 2.1.8 Energy Density, Intensity, and the Poynting Vector of Electromagnetic Waves

The instantaneous energy density  $U$  of an electromagnetic field in free space (in isotropic dielectric matter we have to replace  $\epsilon_0\mathbf{E}^2 \rightarrow \mathbf{D} \cdot \mathbf{E} = \epsilon_0 n^2 \mathbf{E}^2$ ) is

$$U = \frac{1}{2}(\epsilon_0\mathbf{E}^2 + \mu_0\mathbf{H}^2). \quad (2.15)$$

The contribution of the  $\mathbf{H}$ -field is equal to  $\mu_0\mathbf{H}^2/2 = \epsilon_0\mathbf{E}^2/2$  and hence  $U = \epsilon_0\mathbf{E}^2$  as well. Physical significance is associated with the total energy  $\mathcal{U}$  of an electromagnetic field. It is obtained by integration over the relevant volume  $V$ :

$$\mathcal{U} = \int_V U d^3r = \frac{1}{2} \int_V (\epsilon_0\mathbf{E}^2 + \mu_0\mathbf{H}^2) d^3r. \quad (2.16)$$



**Figure 2.3** A photon box made from two optical fibers with mirrors fabricated on their end facets (a Fiber Fabry Perot Cavity). The fibers guide the light beams to the photon box. Images by J. Gallego and L. Ratschbacher.

### Example: Average field strength of a photon in a box

Quantum physics tells us that a “photon” wave oscillating with frequency  $\omega$  has energy  $\mathcal{U} = \hbar\omega$ , evenly distributed onto the electric and magnetic degrees of freedom. The field energy of an optical photon can be stored for some 10 nano- to microseconds in the little box shown in Figure 2.3 by reflecting the wave back and forth several 10 000 times (see the discussion of Fabry–Pérot cavities in Section 6.5. For a volume  $V$  we find from Eq. (2.16) the average field amplitude

$$\langle \mathcal{E} \rangle = \sqrt{\frac{\hbar\omega}{2\epsilon_0 V}} = \sqrt{\frac{hc}{2\epsilon_0}} \sqrt{\frac{1}{\lambda V}} = 1.06 \cdot 10^3 \sqrt{\frac{[\mu\text{m}]^4}{\lambda V}} \frac{\text{V}}{\text{cm}}$$

An effective photon box (also called “mode volume”) is formed by the Gaussian mode (Section 2.3) sustained by the two curved mirrors. For simplicity consider a cylindrical tube with length 40  $\mu\text{m}$  (mirror separation) and cross section  $\pi(5 \mu\text{m})^2$  (waist of the Gaussian mode), yielding a mode volume  $V \simeq 3000 \mu\text{m}^3$ . At a typical wavelength of  $\lambda = 800 \text{ nm}$ , we find  $\mathcal{E} \sim 20 \text{ V cm}^{-1}$ . Thus a single photon can be associated with substantial field strength, leading to strong light–matter coupling at the quantum level (see Section 13.1).

The field energy density propagates at the speed of light  $c$ , and the corresponding energy current density or intensity averaged over one period  $T = 2\pi/\omega$  is most easily measured in an experiment: The total power  $P$  of a light beam traveling through an area  $A$  is  $P = \int I dA$ . The relation of the field amplitude of a harmonic electromagnetic wave  $E(t) = \mathcal{E}_0 e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})}$  and the intensity is obtained by application of the time averaging formula (A.16), yielding

$$I = \langle c\mathcal{U} \rangle = c\epsilon_0 \frac{1}{2} \Re \{ E \cdot E^* \} = \frac{1}{2} c\epsilon_0 \mathcal{E}_0^2.$$

In dielectric matter with refractive index  $n$ , we have the energy density  $\epsilon_0 n^2 \mathbf{E}^2$ , which travels at  $v = c/n$ , and hence

$$I = (c/n)\epsilon_0 n^2 \frac{1}{2} \Re \{ E \cdot E^* \} = \frac{1}{2} c\epsilon_0 n \mathcal{E}_0^2. \quad (2.17)$$

The energy current density ( $\propto U \cdot c$ ) has furthermore vectorial character, which is given by the *Poynting vector*  $\mathbf{S}$  (Figure 2.4):<sup>3</sup>

$$\mathbf{S} = c^2 \mathbf{g} = \mathbf{E} \times \mathbf{H} = c\epsilon_0 |\mathbf{E}|^2 \mathbf{e}_k. \quad (2.18)$$

In relativistic relation energy  $E$  and mechanical momentum  $P$  are connected by  $E/c = P$ . The equivalence with a mechanical momentum density associated with  $\mathbf{g} = \mathbf{S}/c^2$  was already recognized by Abraham [10]. This momentum gives rise to optomechanical phenomena (Section 2.5).

## 2.2 Wave Types

Now let us discuss several limiting cases of simple and important wave types.

### 2.2.1 Planar Waves

*Planar waves* are the characteristic solution of the Helmholtz equation (2.14) in Cartesian coordinates  $(x, y, z)$ :

$$\left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + k^2 \right) \mathbf{E}(\mathbf{r}) = 0. \quad (2.19)$$

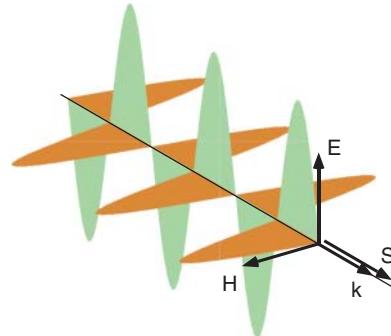
Planar waves are vector waves with constant polarization vector  $\epsilon$  and amplitude  $\mathcal{E}_0$ , and they correspond directly with a single Fourier component:

$$\mathbf{E}(\mathbf{r}, t) = \Re \{ \mathcal{E}_0 \epsilon e^{-i(wt - \mathbf{k}\mathbf{r})} \}.$$

In general they have two independent orthogonal polarization directions  $\epsilon$ , which we will discuss later in Section 2.4. Through the wave vector, we define by  $\mathbf{k} \cdot \mathbf{r} = \text{const}$  planes with identical phase  $\Phi = wt - \mathbf{k}\mathbf{r}$  (Figure 2.5).

Plane wave fields are transverse in homogeneous space, that is, they are orthogonal with respect to the wave vector  $\mathbf{k}$  (Figure 2.4),<sup>4</sup> as proven by the two left equations of the Fourier variant of Eq. (2.10):

$$\begin{aligned} \mathbf{k} \cdot \epsilon_r \mathcal{E} &= 0 \\ \mathbf{k} \cdot \mathcal{H} &= 0 \\ \mathbf{k} \times \mathcal{E} &= \omega \mu_0 \mu_r \mathcal{H} \\ \mathbf{k} \times \mathcal{H} &= -\omega \epsilon_0 \epsilon_r \mathcal{E} \end{aligned} \quad (2.20)$$



**Figure 2.4** Snapshot of a linearly polarized perfectly plane electromagnetic wave in vacuum ( $\epsilon_r = \mu_r = 1$ ).  $\mathbf{E}$ - and  $\mathbf{H}$ -field, propagation vector  $\mathbf{k}$ , and Poynting vector  $\mathbf{S}$  make a right-handed triad, respectively.

Materials with  $\{\epsilon_r, \mu_r\} > 0$  are naturally prevailing and also called *right-handed materials (rhms)*: Eq. (2.20) proves that the triad  $\{\mathcal{E}, \mathcal{H}, \mathbf{k}\}$  forms a right-handed

3 The formal proof that the Poynting vector gives the current density transporting energy out of a closed volume is called *Poynting's theorem* (see, e.g., [9]).

4 Static fields of charge distributions are called *longitudinal*.

set for  $\mu_r > 0$ . For both  $\{\epsilon_r, \mu_r\} < 0$ , the arrangement is changed to a left-handed triad. *Left-handed materials* or *metamaterials* (*lhms*) are realized with artificial structures (see Section 4.3). Metamaterials exhibit exotic properties such as negative refractive indices.

The infinite planar wave fulfills all Maxwell's equations, especially  $\nabla \cdot \mathbf{E} = 0$ , and has perfectly transverse field components. Realistic optical beams have transverse intensity variations, giving rise to small longitudinal field components, which are the origin of more subtle but relevant phenomena such as the optical angular momentum (OAM) of propagating light fields, which is discussed in more detail in Section 2.4.6.

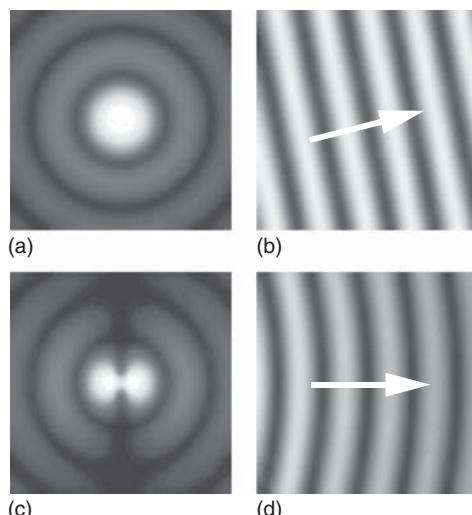
### 2.2.2 Spherical Waves

In our experience, light propagates into all directions of space, while the intensity declines. Because of this, it would be convenient to describe ray propagation by spherical waves as indicated in Figure 2.5a. In spherical coordinates  $(r, \theta, \phi)$ , the Helmholtz equation (2.14) can be written as

$$\left( \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} + k^2 \right) \mathbf{E}(\mathbf{r}) = 0. \quad (2.21)$$

But since electromagnetic waves have vector character, we have to look for solutions for "vector" spherical waves. These are known and common but are mathematically too complex in our case. In optics often only a small solid angle in a distinct direction is of practical importance. There the polarization of the light field varies only to a small extent, and in good approximation we can apply the simplified scalar solution of this wave equation. The amplitude of an isotropic spherical wave ( $\mathbf{k} = k \mathbf{e}_r$ ) has the form

$$E(\mathbf{r}, t) = \Re e \left\{ \mathcal{E}_0 \frac{e^{-i(wt - kr)}}{kr} \right\}. \quad (2.22)$$



**Figure 2.5** Snapshot of important wave types: (a) the isotropic (scalar) spherical wave has a simple structure, but it cannot describe electromagnetic waves correctly, which are always vector fields; (b) a planar wave with wave vector; (c) the dipole wave corresponds to a spherical wave with anisotropic intensity distribution; (d) yet at a distance of only a few wavelengths from the source, the dipole wave is very similar to a planar wave.

It decreases inversely with the distance  $E \propto r^{-1}$ , and its intensity with the square of the inverse distance  $I \propto r^{-2}$ . With the scalar spherical wave approximation, the wave theory of diffraction can be described in good approximation according to Kirchhoff and Fresnel (see Section 2.6).

### 2.2.3 Dipole Waves

Dipole radiators are the most important sources of electromagnetic radiation. This is true for radio waves at wavelengths in the range of meters or kilometers, which are radiated by macroscopic antennas, and for optical wavelengths as well, where induced dipoles of atoms or solids take over the role of microscopic antennas. A positive charge and a negative charge  $\pm q$  at a separation  $\mathbf{x}$  have a dipole moment  $\mathbf{d}(t) = q\mathbf{x}(t)$ . Dipoles can be induced by an external field displacing the center-of-mass charge of the positive and negative charge distributions, for example, of a neutral atom. Charge oscillations  $\mathbf{x} = \mathbf{x}_0 e^{-i\omega t}$  cause an oscillating dipole moment,

$$\mathbf{d}(t) = \mathbf{d}_0 e^{-i\omega t},$$

which radiates a dipole wave and forms the simplest version of a vector spherical wave. Let us assume that the distance of observation is large compared with the wavelength  $r \gg \lambda = 2\pi c/\omega$ . Under these circumstances we are located in the *far field* of the radiation field.

Since the separation  $|\mathbf{x}|$  between the charges is small compared with the wavelength, we may describe the intensity distribution by means of the *Hertzian dipole*.<sup>5</sup> The simplest form is shown by a linear dipole along the  $z$  axis,  $\mathbf{d} = d_0 e^{-i\omega t} \mathbf{e}_z$ , and the field amplitude is stated in spherical coordinates  $(r, \theta, \phi)$ :

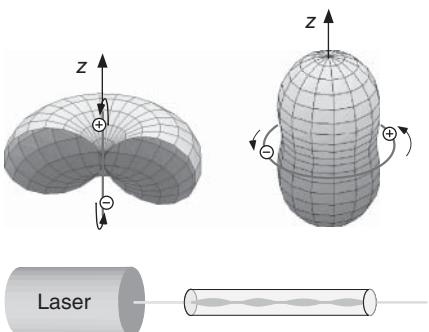
$$\mathbf{E}_{\text{lin}} = \frac{k^3 d_0}{\epsilon_0} \sin \theta \frac{e^{-i(\omega t - kr)}}{kr} \mathbf{e}_\theta.$$

The faces of constant phase are spherical faces again. Only the angle factor  $\sin \theta$ , which specifies exactly the component perpendicular to the direction of propagation, creates the antenna characteristics of a dipole. For a circular dipole,  $\mathbf{d} = d_0 e^{-i\omega t} (\mathbf{e}_x + i\mathbf{e}_y)$ , we find

$$\mathbf{E}_{\text{circ}} = \frac{k^3 d_0}{\epsilon_0} \cos \theta \frac{e^{-i(\omega t - kr)}}{kr} (\cos \theta \mathbf{e}_\theta + i\mathbf{e}_\varphi).$$

In Figure 2.6 the intensity distribution of oscillating dipoles is shown. In contrast to a circular dipole, for a linear one, directions occur into which no energy is radiated. The dipole character can be observed very nicely with the *Tyndall effect* by relatively simple means. One needs only a linearly polarized laser beam and a Plexiglas rod (Figure 2.7). The double refraction of the Plexiglas rod causes a modulation of the polarization plane, and the observer, standing at the side, sees a periodic increase and decrease of the scattered light in the Plexiglas rod.

<sup>5</sup> The Hertzian dipole has vanishing spatial extent ( $\mathbf{x} \rightarrow 0$ ) but a nonzero dipole moment  $\mathbf{d}$ .



**Figure 2.6** Angular distribution of the intensity ( $\propto |E|^2$ ) of a linearly and a circularly oscillating dipole.

**Figure 2.7** Tyndall effect in a Plexiglas rod. By birefringence the plane of polarization gets modulated. Therefore an observer at the side sees a periodic increase and decrease of the scattered light intensity.

### 2.3 Gaussian Beams

Now we want to establish the connection between ray optics and wave optics, that is, we want to describe in particular laser beam propagation through the methods of wave optics. Observation of a laser beam yields characteristic properties that we will use to construct the so-called Gaussian principal mode of laser beam propagation: laser beams are extremely well bundled, that is, they do not seem to change over distances of many meters, and they are axially symmetric. They truly form rays without any effort. Along the propagation direction  $z$ , a light ray behaves very similarly to a planar wave with constant amplitude  $\tilde{A}_0$ , which is a known solution of the wave equations (2.12) and (2.19):

$$E(z, t) = \tilde{A}_0 e^{-i(\omega t - kz)}.$$

On the other hand, we know that, at large distances from a source, also laser light behaves more like another known solution of Eqs. (2.12) and (2.22), which is divergent like the spherical or dipole wave discussed before with the amplitude inversely decreasing with distance from the source:

$$E(z, t) = A_0 \frac{e^{-i(\omega t - kz)}}{kz}. \quad (2.23)$$

A wave that combines the properties of plane waves and spherically diverging waves could have the structure

$$E(z, t) = A_0 \frac{e^{-i(\omega t - kz)}}{k(z - iz_0)}, \quad (2.24)$$

where  $z = z_0$  separates the region where the wave will behave more like a plane wave ( $|z| < z_0$ ) and more like a spherically divergent wave ( $|z| > z_0$ ). Here  $z_0$  is a real number, while the origin of the imaginary term will become transparent later. We will use this *ad hoc* approach to “construct” the fundamental mode of coherent beam propagation. The 3D extension of the wave will be introduced by replacing  $kz \rightarrow \mathbf{k}r$  and expanding  $\mathbf{k}r$  in the vicinity of the  $z$  axis.

Note that for now we have completely neglected the vectorial properties of the electric field. This is justified since the local  $\mathbf{E}$ - and  $\mathbf{H}$ -fields will always have

transverse polarization with respect to the local propagation vector  $\mathbf{k} \simeq k\mathbf{e}_z$ , and hence the modification of their orientation can be neglected in most cases. Once we become interested in the detailed evolution of the propagation vector or the Poynting vector (2.18), however, we will find that this approach is not sufficient (see Section 2.4.6).

### 2.3.1 The Gaussian Principal Mode or $\text{TEM}_{00}$ Mode

We now consider a cutout of a spherical wave close to the  $z$  axis (“paraxial”) and separate longitudinal ( $z$  coordinate) and transverse contributions. Rays with axial symmetry depend only on the transverse coordinate  $\rho$ , and when substituting  $\mathbf{kr} = kr$ , we may furthermore use the approximation  $r = \sqrt{z^2 + \rho^2} \simeq z + \rho^2/2z$  within the so-called Fresnel approximation for  $\rho \ll z, r$ :

$$E(\mathbf{r}) = \frac{\mathcal{A}(\mathbf{r})}{|\mathbf{kr}|} e^{ikr} \simeq \frac{\mathcal{A}(z, \rho)}{kz} \exp\left(i \frac{k\rho^2}{2z}\right) e^{ikz}. \quad (2.25)$$

This form of course much resembles Eq. (2.23) where the spatial phase is transversely modulated, respectively curved, with the Fresnel factor  $\exp(ik\rho^2/2z)$ .

The linear substitution  $z \rightarrow z - iz_0$  is similar to a coordinate transformation and simply realizes one more solution that also introduces a phase shift for small  $z$  due to the imaginary term  $iz_0$ . With this substitution we already arrive at the *Gaussian principal mode*<sup>6</sup> if we use a constant amplitude  $\mathcal{A}_0$ :

$$E(z, \rho) \simeq \frac{\mathcal{A}_0}{k(z - iz_0)} \exp\left(i \frac{k\rho^2}{2(z - iz_0)}\right) e^{ikz}. \quad (2.26)$$

Gaussian modes propagate in free and in isotropic space, in contrast to, for example, waves in a dielectric wave guide, which depend on the inhomogeneous optical properties of the material. In isotropic space the electric and magnetic fields, as well, are transverse to the direction of propagation, and the waveforms are called *transverse electric and magnetic* ( $\text{TEM}_{m,n}$ ) *modes* with indices  $(m, n)$ . The basic solution is called the  $\text{TEM}_{00}$  mode. It is by far the most important form of all used wave types and therefore will be analyzed in more detail before we consider the higher modes in Section 2.3.4.

The presentation of the field distribution in Eq. (2.26) is not yet very transparent. Therefore we introduce the new quantities  $q(z)$ ,  $R(z)$ , and  $w(z)$  through

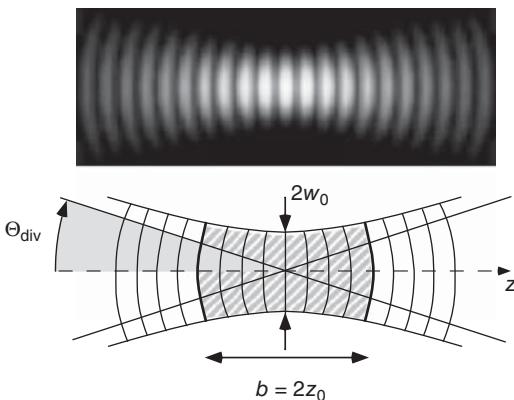
$$\frac{1}{q(z)} = \frac{1}{z - iz_0} = \frac{z + iz_0}{z^2 + z_0^2} = \frac{1}{R(z)} + i \frac{2}{kw^2(z)}. \quad (2.27)$$

The decomposition of the Fresnel factor into real and imaginary parts creates two factors, a complex phase factor that describes the curvature of the phase front and a real factor that specifies the envelope of the beam profile:

$$\exp\left(i \frac{k\rho^2}{2(z - iz_0)}\right) \rightarrow \exp\left(i \frac{k\rho^2}{2R(z)}\right) \exp\left(-\left(\frac{\rho}{w(z)}\right)^2\right).$$

---

<sup>6</sup> The notion of “mode,” which appears here for the first time, is derived from the Latin word *modus*, meaning measure or melody.



**Figure 2.8** A Gaussian principal mode close to the beam waist. In the center nearly planar wave fronts are achieved, while outside the waves quickly adopt a spherical form. In the lower part the Rayleigh zone is hatched.

The form of the Gaussian principal mode in Figure 2.8 is fully characterized by the parameter pair  $(w_0, z_0)$ . The following definitions and notations have been established to lend physical meaning to important parameters.

### 2.3.1.1 Rayleigh Zone, Confocal Parameter $b$

$$b = 2z_0 \quad (2.28)$$

The Gaussian wave exhibits its largest variation for  $-z_0 \leq z \leq z_0$ , within the so-called Rayleigh length  $z_0$  from Eq. (2.27). This region is called the Rayleigh zone and is also characterized with the confocal parameter  $b = 2z_0$ . The Rayleigh zone marks positions in the *near field* of the smallest beam cross section or focal point ("focus"). At  $z \ll z_0$  a nearly planar wave propagates, and the wave front changes only marginally. The shorter the Rayleigh zone is, the more the beam is focused. In the context of images, we also use the notion *depth of focus* (see Section 5.5.1). In the *far field* ( $z \gg z_0$ ), the propagation is again similar to a spherical or dipole wave.

### 2.3.1.2 Radius of Wave Fronts $R(z)$

$$R(z) = z[1 + (z_0/z)^2]$$

Within the Rayleigh zone,  $R(z) \rightarrow \infty$  holds at  $z \ll z_0$ , whereas in the far field,  $R(z) \simeq z$ . The largest curvature or the smallest radius occurs at the border of the Rayleigh zone with  $R(z_0) = 2z_0$ .

### 2.3.1.3 Beam Waist $2w_0$

$$w_0^2 = \lambda z_0 / \pi \quad (2.29)$$

The beam waist  $2w_0$ , or beam waist radius  $w_0$ , specifies the smallest beam cross section at  $z = 0$ . If the wave propagates within a medium of refractive index  $n$ , then  $\lambda$  must be substituted by  $\lambda/n$ . The diameter of the beam waist is then  $w_0^2 = \lambda z_0 / \pi n$ .

### 2.3.1.4 Beam Radius $w(z)$

$$w^2(z) = w_0^2 \left[ 1 + \left( \frac{z}{z_0} \right)^2 \right]$$

Within the Rayleigh zone, the beam radius  $w(z)$  stays approximately constant. But in the far field, it increases linearly according to  $w(z) \simeq w_0 z / z_0$ .

### 2.3.1.5 Divergence $\Theta_{\text{div}}$

$$\Theta_{\text{div}} = \frac{w_0}{z_0} = \sqrt{\frac{\lambda}{\pi z_0 n}}$$

In the far field ( $z \gg b$ ), the divergence can be determined from the relation  $\Theta(z) = w(z)/z$ ,  $z \rightarrow \infty$ .

### 2.3.1.6 Gouy Phase $\eta(z)$

$$\eta(z) = \tan^{-1}(z/z_0) \quad (2.30)$$

Passing through the focus, the Gaussian wave receives a bit more curvature, that is, shorter wavelength than a planar wave. For illustration, alternatively to (2.27), we can make the substitution

$$\frac{i}{z - iz_0} = -\frac{1}{z_0} \frac{w_0}{w(z)} e^{-i \tan^{-1}(z/z_0)}.$$

The imaginary factor establishes the common convention to find a real amplitude or vanishing phase at  $z = 0$ . By this function the small deviation from the linear phase evolution of the planar wave can be described,  $-\pi/2 \leq \eta(z) \leq \pi/2$ . This extra phase is known by the name *Gouy phase*; half of it is collected within the Rayleigh zone. In traveling through the focus, the phase is effectively inverted, which is reminiscent of two partial rays exchanging relative positions when crossing at a focal point.

With these notations the total result of the Gaussian principal mode or  $\text{TEM}_{00}$  mode is the following:

$$E(\rho, z) = \mathcal{A}_0 \frac{w_0}{w(z)} e^{-[\rho/w(z)]^2} e^{ik\rho^2/2R(z)} e^{i[kz - \eta(z)]}. \quad (2.31)$$

The first factor describes the transverse amplitude distribution, the second (Fresnel) factor the spherical curvature of the wave fronts, and the last one the phase evolution along the  $z$  axis. In the majority of applications in physics and optical technology, a Gaussian principal mode or  $\text{TEM}_{00}$  mode is used.

### Example: Intensity of a $\text{TEM}_{00}$ mode

The intensity distribution within a plane perpendicular to the propagation direction corresponds to the known Gaussian distribution,

$$I(\rho, z) = \frac{c\epsilon_0}{2} EE^* = \frac{c\epsilon_0}{2} |\mathcal{A}_0|^2 \left( \frac{w_0}{w(z)} \right)^2 e^{-2[\rho/w(z)]^2},$$

with the axial peak value

$$I(0, z) = \frac{c\epsilon_0}{2} |\mathcal{A}_0|^2 \left( \frac{w_0}{w(z)} \right)^2.$$

In general the “cross section” of a Gaussian beam is specified as the width  $2w(z)$ , where the intensity has dropped to  $1/e^2$  or 13% of the peak value. Some 87% of the total power is concentrated within this radius.

Along the  $z$  axis the intensity follows a Lorentzian profile  $1/[1 + (z/z_0)^2]$ . It declines from its peak value  $I(0, 0) = (c\epsilon_0/2)|\mathcal{A}_0|^2$  (see Figure 2.8) and reaches half of this value at  $z = z_0$ . The confocal parameter  $b$  then is also a measure for the longitudinal half-width of the focal zone.

The total power traveling in the Gaussian beam is  $P = 2\pi \int I(\rho, z) \rho d\rho$ . Owing to energy conservation, it cannot change, as can be verified by explicit integration,

$$\frac{2P}{c\epsilon_0} = 2\pi \mathcal{A}_0^2 w_0^2 \int_0^\infty \frac{\rho d\rho}{w^2(z)} e^{-2[\rho/w(z)]^2} = \pi w_0^2 \mathcal{A}_0^2.$$

### 2.3.2 The ABCD Rule for Gaussian Modes

The usefulness of Gaussian modes for analysis of an optical beam path is supported particularly by the simple extension of the ABCD rule (Section 1.9.2), known from ray optics. At every position  $z$  on the beam axis, a Gaussian beam may be characterized either by the pair of parameters  $(w_0, z_0)$  or alternatively by the real and imaginary parts of  $q(z)^{-1} = (z - iz_0)^{-1}$  according to Eq. (2.27). We know that both parameters of a light ray are transformed linearly according to Eq. (1.16) and that for every optical element, a distinct type of matrix  $\mathbf{T}$  with elements  $A, B, C, D$  exists. The parameters of the Gaussian beam are transformed by linear operations with coefficients that are identical to the ones from ray optics:

$$q_1 = \hat{\mathbf{T}} \otimes q_0 = \frac{Aq_0 + B}{Cq_0 + D}. \quad (2.32)$$

Note that this relation is equivalent to

$$\frac{1}{q_1} = \frac{Cq_0 + D}{Aq_0 + B} = \frac{C + D/q_0}{A + B/q_0},$$

which for a thin lens yields the transformation

$$\frac{1}{q_1} = -\frac{1}{f} + \frac{1}{q_0}$$

and with  $\Re\{1/q\} = 1/R$  according to (2.27)

$$\frac{1}{R_1} = -\frac{1}{f} + \frac{1}{R_0}. \quad (2.33)$$

A thin lens therefore changes the radius of curvature but not the width of a Gaussian light beam, as expected.

Now it is not very difficult to show that these operations may be applied multiple times and that the total effect  $\hat{\mathbf{T}}$  corresponds to the matrix product  $\hat{\mathbf{T}}_2 \hat{\mathbf{T}}_1$ :

$$\begin{aligned} q_2 &= \hat{\mathbf{T}}_2 \otimes (\hat{\mathbf{T}}_1 \otimes q_0) = \frac{A_2(A_1 q_0 + B_1/C_1 q_0 + D_1) + B_2}{C_2(A_1 q_0 + B_1/C_1 q_0 + D_1) + D_2} \\ &= \frac{(A_2 A_1 + B_2 C_1) q_0 + \dots}{\dots} \end{aligned}$$

Thus we obtain the transformation of a Gaussian beam for any composition of elements known from the matrices of Table 1.2. through Eq. (2.32) and its derivatives.

The basic principles for propagating coherent beams with arbitrary shape are based on diffraction theory (Section 2.6). The results can be combined with the ABCD concept to describe the propagation of paraxial light beams with complex shape through a system of ABCD components (see Section 2.9).

### Example: Focusing with a thin lens

As an important and instructive example, we now choose the effect of a thin lens of focal length  $f$ , with which a Gaussian beam with  $\text{TEM}_{00}$  mode has to be focused, and make a comparison with the predictions from ray optics. Let us look at the parameters of the wave in planes 1 (directly in front of the lens (Figure 2.9)), 2 (directly after the lens), and 3 (in the focus).

**Plane E1.** A Gaussian beam with large beam waist  $2w_{01}$  and infinitely large radius of curvature  $R(z = 0) = \infty$  is very close to our expectations of a planar wave. Then, the Rayleigh length is also very large, since  $z_{01} = \pi w_{01}^2 / \lambda$ ; for example, it is 124 m long for a beam diameter of only 1 cm and a wavelength of 632 nm! Let us assume that the beam waist of the incident beam is at  $z = 0$ , and due to that  $q(z)$  is purely imaginary:

$$q_1 = -iz_{01} = -i \frac{\pi w_{01}^2}{\lambda}.$$

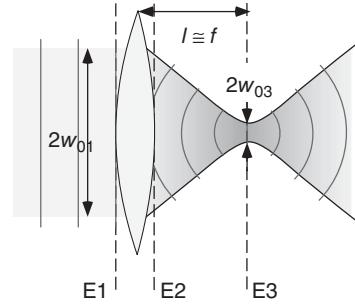


Figure 2.9 Focusing a Gaussian beam with a thin lens of focal length  $f$ .

**Plane E2.** The beam radius is not changed by the thin lens at once ( $w_{02} = w_{01}$ ), but the radius of curvature is changed and is now given by  $1/R_2 = -1/f$ :

$$\frac{1}{q_2(z = 0)} = -\frac{1}{f} + i \frac{1}{z_{01}}.$$

Formal application of the lens transformation from Table 1.2 and with Eq. (2.32) would have yielded the same result.

**Plane E3.** For the translation from the lens to the new focus, we get

$$q_3(\ell) = q_2(0) + \ell,$$

but the  $\ell$  position of plane 3 is initially unknown and must be determined from the condition that at the focus  $q_3^{-1} = i\lambda/\pi w_{03}^2$  is purely imaginary. For that purpose we determine the real and imaginary parts of  $q_2$ :

$$q_2 = -\frac{f}{1 + (f/z_{01})^2} \left( 1 + i\frac{f}{z_{01}} \right).$$

Obviously the real part of  $q_3$  is compensated exactly at

$$\ell = \frac{f}{1 + (f/z_{01})^2} = \frac{f}{1 + (\lambda f/\pi w_{01}^2)^2},$$

which means that we again find planar waves there. According to ray optics, we would have expected the focus to be located exactly at  $\ell = f$ . But if the focal length is short compared with the Rayleigh length of the incident beam,  $f \ll z_{01}$ , or equivalently, which is usually the case,  $\lambda f/w_{01}^2 \ll 1$ , then the position of the focal point will differ only marginally from that.

More interesting is the question of how large the diameter of the beam is in the focus. We know that ray optics does not answer that, and we have to take into account diffraction at the aperture of the lens (compare also example p. 69). At first we calculate the Rayleigh parameter

$$\frac{1}{z_{03}} = \frac{1}{f} \frac{1 + (f/z_{01})^2}{f/z_{01}}$$

and then determine the ratio of the beam diameter at the lens and in the focus:

$$\frac{w_{03}}{w_{01}} = \left( \frac{z_{03}}{z_{01}} \right)^{1/2} = \frac{f/z_{01}}{\sqrt{1 + (f/z_{01})^2}}. \quad (2.34)$$

Replacing  $1/z_{01}$  with  $\lambda/\pi w_{01}^2$  results in

$$w_{03} = \frac{\lambda f}{\pi w_{01}} \frac{1}{\sqrt{1 + (\lambda f/\pi w_{01}^2)^2}} \approx \frac{\lambda f}{\pi w_{01}},$$

and the first factor yields the Rayleigh criterion for the resolving power of a lens, known also from diffraction theory, which will be treated once more in the section on diffraction (Section 2.7, Eq. (2.72)).

### 2.3.3 Paraxial Wave Equation

For a more formal treatment of the Gaussian modes, we now also decompose the Helmholtz equation (2.14) into transverse and longitudinal contributions,

$$\nabla^2 + k^2 = \frac{\partial^2}{\partial z^2} + \nabla_T^2 + k^2 \quad \text{and} \quad \nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},$$

and apply it to an electric field  $E = Ae^{ikz}$ . We are interested in solutions, where the amplitude  $A(x, y, z)$  is varying only very slowly on a wavelength scale, that is, in comparison with  $e^{ikz}$ . Using

$$\frac{\partial}{\partial z} A = A' \ll kA \quad \text{and} \quad \frac{\partial^2}{\partial z^2} A = A'' \ll kA'$$

when inserting the electric field into the original Helmholtz equation (2.14), we find the approximation that is called *paraxial Helmholtz equation*:

$$\left( \nabla_T^2 + 2ik \frac{\partial}{\partial z} \right) A(x, y, z) = 0. \quad (2.35)$$

The most fundamental solution of the paraxial Helmholtz equation has already been found by intuition and construction (Eq. (2.26)), but the Gaussian principal mode is only one particular, although important, solution. With a bit of algebra, one can show that the principal mode, from Eq. (2.26), with constant amplitude  $A_0$  and  $q(z) = z - iz_0$ ,

$$A(x, y, z) = \frac{A_0}{q(z)} \exp \left( i \frac{k(x^2 + y^2)}{2q(z)} \right) e^{ikz},$$

fulfills the paraxial equation (2.35) indeed. This is not surprising, because by this we have just verified that, close to the  $z$  axis, the applied spherical wave fulfills the paraxial Helmholtz equation.

One shortcoming of the paraxial wave equation is that  $\nabla \cdot \mathbf{E} = 0$  is fulfilled to first order only. Some consequences are discussed in Section 2.4.6.

### 2.3.4 Higher Gaussian Modes

We look now for the higher-order solutions by allowing the amplitude to depend (slowly) on  $(x, y, z)$ :

$$A(x, y, z) = \frac{A(x, y, z)}{q(z)} \exp \left( i \frac{k(x^2 + y^2)}{2q(z)} \right).$$

Initially we want to use Cartesian coordinates, which deliver the best-known solutions, called *Hermite–Gaussian modes*. But there are also other solutions, for example, the *Laguerre–Gaussian (LG) modes* (see Problem 2.5), which are found when applying cylindrical coordinates. From the paraxial Helmholtz equation (2.35), we find a new equation for the amplitude distributions  $A(x, y, z)$ :

$$\left( \frac{\partial^2}{\partial x^2} + \frac{2ikx}{q(z)} \frac{\partial}{\partial x} + \frac{\partial^2}{\partial y^2} + \frac{2iky}{q(z)} \frac{\partial}{\partial y} + 2ik \frac{\partial}{\partial z} \right) A(x, y, z) = 0. \quad (2.36)$$

Obviously this is valid for  $A = \text{const}$ , which reproduces the Gaussian principal mode. Let us investigate amplitudes that depend symmetrically on  $x$  and  $y$  and along the longitudinal direction cause only a small correction of the phase evolution:

$$A(x, y, z) = \mathcal{F}(x)\mathcal{G}(y) \exp[-i\mathcal{H}(z)].$$

We substitute this form in Eq. (2.36) and take into account that  $1/(z - iz_0) = 2(1 - iz/z_0)/ikw^2(z)$ . By claiming exclusively real solutions for  $\mathcal{F}$ ,  $\mathcal{G}$  and  $\mathcal{H}$ , imaginary contributions cancel, and we get

$$\begin{aligned} & \frac{1}{\mathcal{F}(x)} \left[ \frac{\partial^2}{\partial x^2} \mathcal{F}(x) - \frac{4x}{w^2(z)} \frac{\partial}{\partial x} \mathcal{F}(x) \right] \\ & + \frac{1}{\mathcal{G}(y)} \left[ \frac{\partial^2}{\partial y^2} \mathcal{G}(y) - \frac{4y}{w^2(z)} \frac{\partial}{\partial y} \mathcal{G}(y) \right] + 2k \frac{\partial}{\partial z} \mathcal{H}(z) = 0. \end{aligned}$$

Expecting that the distribution of transverse amplitudes does not change along the  $z$  axis, we execute the variable transformation

$$u = \sqrt{2} x/w(z) \quad \text{and} \quad v = \sqrt{2} y/w(z)$$

(the factor  $\sqrt{2}$  is necessary to normalize the new equations):

$$\frac{1}{\mathcal{F}(u)} [\mathcal{F}''(u) - 2u\mathcal{F}'(u)] + \frac{1}{\mathcal{G}(v)} [\mathcal{G}''(v) - 2v\mathcal{G}'(v)] + kw^2(z)\mathcal{H}'(z) = 0.$$

By this transformation we have achieved a separation of the coordinates, and the equation can be solved via eigenvalue problems:

$$\begin{aligned} \mathcal{F}''(u) - 2u\mathcal{F}'(u) + 2m\mathcal{F}(u) &= 0, \\ \mathcal{G}''(v) - 2v\mathcal{G}'(v) + 2n\mathcal{G}(v) &= 0, \\ kw^2(z)\mathcal{H}'(z) - 2(m+n) &= 0. \end{aligned} \quad (2.37)$$

The equation for the  $(u, v)$  coordinates is known as Hermite's differential equation. Its solutions are called *Hermitian polynomials*  $H_j(x)$ , which are easy to determine according to the recursion relations

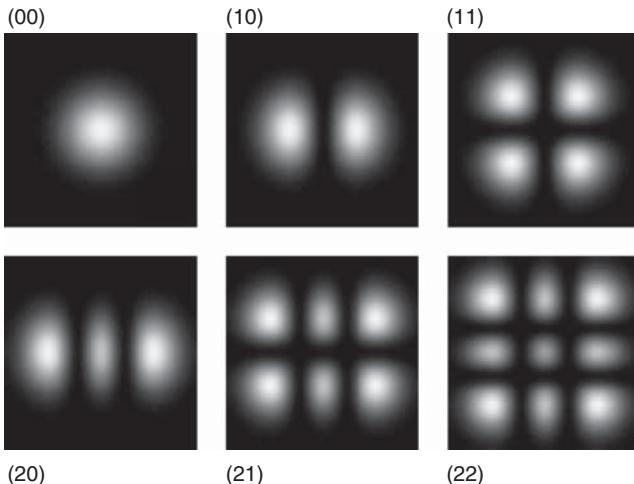
$$\begin{aligned} H_{j+1}(x) &= 2xH_j(x) - 2jH_{j-1}(x), \\ H_j(x) &= (-)^j e^{x^2} \frac{d^j}{dx^j} (e^{-x^2}). \end{aligned} \quad (2.38)$$

The Hermitian polynomials of lowest order are

$$H_0(x) = 1, H_1(x) = 2x, H_2(x) = 4x^2 - 2, H_3(x) = 8x^3 - 12x.$$

The modulus squared specifies the transverse intensity distribution and is illustrated in Figure 2.10 for the mode of lowest order. They form a system of orthonormal functions with the orthogonality condition

$$\int_{-\infty}^{\infty} H_j(x)H_{j'}(x) e^{-x^2} dx = \frac{\delta_{jj'}}{2^j j! \sqrt{\pi}}. \quad (2.39)$$



**Figure 2.10** Calculated transverse intensity distribution of Hermite–Gaussian modes of low order ( $|\mathcal{A}_{mn}(x, y)|^2 = |H_m(x)H_n(y)|^2$ ). Compare Figure 2.11.

The third equation from (2.37) is solved by

$$\mathcal{H}(z) = (n + m)\eta(z) \quad (2.40)$$

with  $\eta(z) = \tan^{-1}(z/z_0)$  (see Eq. (2.30)). It enhances the phase shift of the Gouy phase and plays an important role in the calculation of the resonance frequencies of optical resonators (see Section 6.6).

Thus the result for the modulation factor of the amplitude distribution for higher-order Gaussian or  $\text{TEM}_{mn}$  modes is

$$\mathcal{A}_{mn} = H_m \left( \sqrt{2} x/w(z) \right) H_n \left( \sqrt{2} y/w(z) \right) e^{-i(m+n)\eta(z)},$$

which with the substitution  $\rho^2 = x^2 + y^2$  and the definitions for  $w_0$ ,  $w(z)$ , and  $R(z)$  from Section 2.3.1 yields

$$E_{mn}(x, y, z) = E_0 H_m \left( \frac{\sqrt{2}x}{w(z)} \right) H_n \left( \frac{\sqrt{2}x}{w(z)} \right) \frac{w_0}{w(z)} e^{-(\rho/w(z))^2} \times e^{ik\rho^2/2R(z)} e^{i(kz-(m+n+1)\eta(z))}. \quad (2.41)$$

The result for the  $\text{TEM}_{00}$  mode is reproduced, of course. All higher modes are described by a Gaussian envelope, modulated by Hermitian polynomials. Therefore they are called *Hermite–Gaussian (HG) modes*.

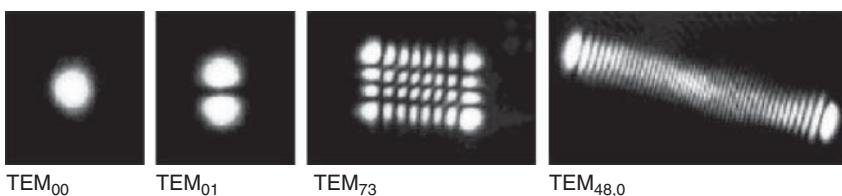
We have assumed full cylindrical symmetry up to now. This not being necessary, we may straightforwardly introduce astigmatic beams: they have different Rayleigh lengths  $z_{0x}$  and  $z_{0y}$ , and the Gouy phase is replaced by

$$e^{-i(m+n+1)\eta(z)} \rightarrow e^{-i(m+1/2)\eta_x(z)} e^{-i(n+1/2)\eta_y(z)}. \quad (2.42)$$

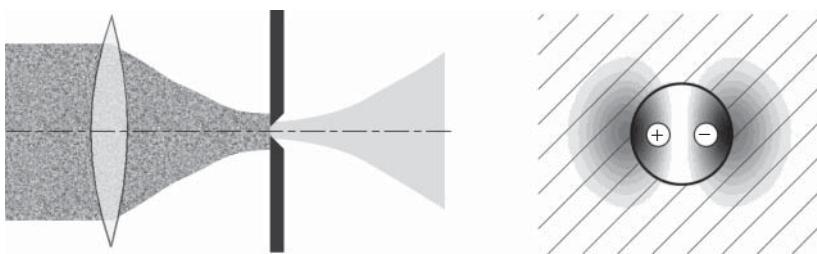
### 2.3.5 Creation of Gaussian Modes

In most experiments interest is focused on the  $\text{TEM}_{00}$  principal mode. By nature it is preferred in a laser resonator, because it has the smallest diffraction losses. According to Figure 2.10, it is obvious that the effective size of a mode increases with the orders  $(m, n)$ , so that the openings of a resonator (mirror edges, apertures) are of increasing importance. On the other hand, since the spatial amplification profile also has to match the desired mode, modes of very high order can be excited by intentionally misaligning a resonator (Figure 2.11).

Controlled shaping of light fields can also be achieved by a filter; therefore the notion of *spatial filter* is used. Such a spatial filter is shown in Figure 2.12, which in its most simple form consists of a convex lens (e.g., a microscope



**Figure 2.11** Gaussian modes of higher order from a simple titanium–sapphire laser. For the  $\text{TEM}_{48,0}$  mode, the scale has been reduced a little bit. The asymmetry of the higher-order modes is caused by technical inaccuracies of the resonator elements (mirrors, laser crystal) [11].



**Figure 2.12** Spatial filter. In front of the aperture, the beam consists of a superposition of many Gaussian modes. It is shown, for the example of a  $\text{TEM}_{01}$  mode, how higher-order modes can be suppressed by the aperture. The fields in both “ears” of the mode oscillate with opposite phase.

objective) and a so-called pinhole, with a diameter adjusted to the  $\text{TEM}_{00}$  principal mode.

Transmission of higher-order Gaussian modes is not only inhibited by the aperture, because the diameter increases rapidly with the order, but is also suppressed by the spatially alternating phase distribution. Therefore the aperture is not excited dipole-like, as is the case for the  $\text{TEM}_{00}$  principal mode, but with a higher order, which, as everybody knows, radiates with lower intensity.

At the output a “cleaned” Gaussian beam propagates, which has lost intensity, of course. Excellent suppression of higher-order modes is achieved when a single-mode optical wave guide is used instead of a pinhole (see Section 4.1).

### 2.3.6 More Gaussian Paraxial Beams

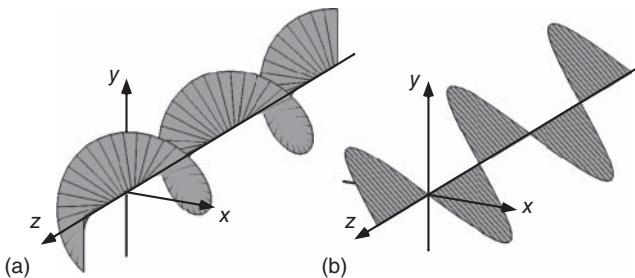
Another question might remain: Why have we chosen the Cartesian form of the paraxial Helmholtz equation, and why do cylindrical coordinates not immediately appear since they seem so natural for pencil-like laser beams? The reason is of technical nature, because at the interior of mirrors and windows, small deviations from cylindrical symmetry are always present, and thus Cartesian Gaussian modes are preferred to Laguerre modes, which are found as solutions of equations with perfect cylindrical symmetry.

While LG paraxial beams do not emanate naturally from laser oscillators, they are straightforward to make as is discussed in Section 2.5.4.3. LG beams have an interesting property not present in HG beams: they exhibit a mechanical optical angular momentum about the propagation axis (see Section 2.5.4.2).

## 2.4 Vector Light: Polarization

We have already noticed in the previous section that electromagnetic waves are vector waves; the propagating fields exhibit a direction (Figure 2.13), which can be described in terms of two orthogonal polarization vectors  $\boldsymbol{\varepsilon}$  and  $\boldsymbol{\varepsilon}'$  in free space.<sup>7</sup>

<sup>7</sup> The notion of “polarization” is also used as *dielectric polarization* elsewhere. The kind of application for which it is used is always clear from the context.



**Figure 2.13** The field of a circularly polarized wave (a) rotates around the propagation axis everywhere with the same amplitude. The linearly polarized wave (b) is a common sine wave.

We consider a transverse plane wave propagating in the  $\mathbf{e}_z$  direction. The polarization must lie within the  $xy$  plane (unit vectors  $\mathbf{e}_x$  and  $\mathbf{e}_y$ ), and we consider two components that may have different time-variant phases:

$$\mathbf{E}(z, t) = \mathcal{E}_x \mathbf{e}_x \cos(kz - wt) + \mathcal{E}_y \mathbf{e}_y \cos(kz - wt + \phi). \quad (2.43)$$

For  $\phi = 0, 2\pi, 4\pi, \dots$ , these components have equal phases, and the wave is linearly polarized:

$$\mathbf{E}(z, t) = (\mathcal{E}_x \mathbf{e}_x \pm \mathcal{E}_y \mathbf{e}_y) \cos(kz - wt).$$

For  $\phi = \pm\pi/2, \pm3\pi/2, \dots$ , they oscillate out of phase and in general yield an elliptically, or for  $\mathcal{E}_x = \mathcal{E}_y$  circularly, polarized wave:

$$\mathbf{E}(z, t) = \mathcal{E}_x \mathbf{e}_x \cos(kz - wt) \pm \mathcal{E}_y \mathbf{e}_y \sin(kz - wt).$$

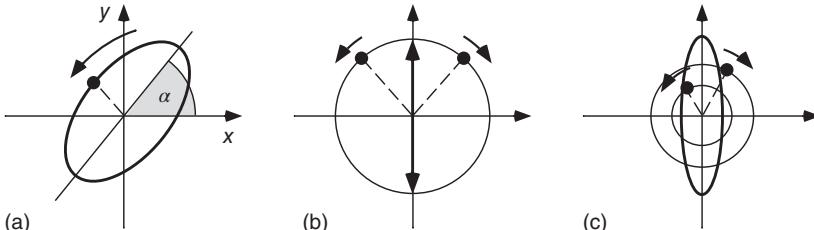
Instead of Eq. (2.43) the field amplitude may also be written in the form

$$\begin{aligned} \mathbf{E}(z, t) = & E_{\cos}(\alpha \mathbf{e}_x + b \mathbf{e}_y) \cos(kz - wt + \alpha) \\ & + E_{\sin}(-b \mathbf{e}_x + \alpha \mathbf{e}_y) \sin(kz - wt + \alpha), \end{aligned}$$

with  $a^2 + b^2 = 1$ , which corresponds to the ellipse in Figure 2.14 rotated by the angle  $\alpha$ . By comparison of the coefficients at  $(kz - wt) = 0, \pi/2$ , one may calculate the angle  $\alpha$ :

$$\tan(2\alpha) = \frac{2\mathcal{E}_x \mathcal{E}_y \cos \phi}{\mathcal{E}_x^2 - \mathcal{E}_y^2}. \quad (2.44)$$

Furthermore in Figure 2.14, the decomposition of a linear and elliptical polarization into two counter-running circular waves is illustrated.



**Figure 2.14** (a) Elliptically polarized wave. The linearly polarized wave (b) and the elliptical wave (c) can be decomposed into two counter-running circular waves.

### 2.4.1 Jones Vectors

In general any transverse polarized light wave can be decomposed into either two orthogonal and linear or two counterrotating and circularly polarized waves. For example, we find for the field of Eq. (2.43):

$$\begin{aligned} \mathbf{E}(z, t) &= \Re\{\mathcal{E}_x \mathbf{e}_x + \mathcal{E}_y e^{-i\phi} \mathbf{e}_y\} e^{-i(wt-kz)} \\ &= \frac{1}{\sqrt{2}} \Re\{(\mathcal{E}_x - i e^{-i\phi} \mathcal{E}_y) \mathbf{e}_+ + (\mathcal{E}_x + i e^{-i\phi} \mathcal{E}_y) \mathbf{e}_-\} e^{-i(wt-kz)} \\ &= \Re\{(\mathcal{E}_+ e^{-i\alpha_+} \mathbf{e}_+ + \mathcal{E}_- e^{i\alpha_-} \mathbf{e}_-) e^{-i(wt-kz)}, \end{aligned}$$

which defines two new complex field amplitudes  $\mathcal{E}_{\pm} = (\mathcal{E}_x \pm i e^{i\phi} \mathcal{E}_y)/\sqrt{2}$ . The phase angles  $\alpha_{\pm}$  are related to Eq. (2.44) through  $\alpha = (\alpha_+ + \alpha_-)/2$ . For instance, for  $\phi = \pi/2$ , we find especially  $\mathcal{E}_{\pm} = (\mathcal{E}_x \pm \mathcal{E}_y)/\sqrt{2}$ , that is, for  $\mathcal{E}_x = \mathcal{E}_y$ , a perfect right-handed circularly polarized wave. Jones suggested in 1941 [12] the orthogonal complex unit vectors

$$\left\{ \begin{array}{l} \mathbf{e}_+ = (\mathbf{e}_x + i \mathbf{e}_y)/\sqrt{2} \\ \mathbf{e}_- = (\mathbf{e}_x - i \mathbf{e}_y)/\sqrt{2} \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{l} \mathbf{e}_x = (\mathbf{e}_+ + \mathbf{e}_-)/\sqrt{2} \\ \mathbf{e}_y = -i(\mathbf{e}_+ - \mathbf{e}_-)/\sqrt{2} \end{array} \right\} \quad (2.45)$$

for the characterization of a polarized field:  $\mathbf{e}_{x,y}$  and  $\mathbf{e}_{\pm}$  for linear and circular components, respectively. Writing this for the individual components, we find

$$\mathbf{e}_x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{e}_{\pm} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \end{pmatrix}.$$

It is obvious from Eq. (2.45) that any linearly polarized wave may be decomposed into two counterrotating circularly polarized waves and vice versa, for example,

$$\frac{1}{\sqrt{2}}(\mathbf{e}_+ + \mathbf{e}_-) = \frac{1}{2} \left( \begin{pmatrix} 1 \\ i \end{pmatrix} + \begin{pmatrix} 1 \\ -i \end{pmatrix} \right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The polarization state described by the Jones vectors can be mapped onto a single quantity  $\sigma$  through

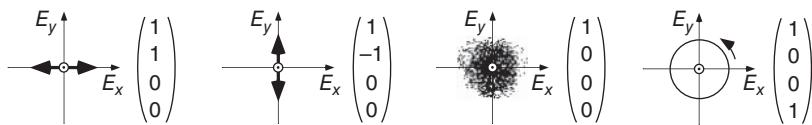
$$a\mathbf{e}_x + b\mathbf{e}_y \rightarrow \sigma = 2\Re\{iab^*\}. \quad (2.46)$$

Inspection of the Jones vectors of (2.45) reveals that we have  $\sigma = \pm 1$  and  $\sigma = 0$  for right- and left-handed circularly and linearly polarized light fields. For cylindrical coordinates we can replace  $\{\mathbf{e}_x, \mathbf{e}_y\} \rightarrow \{\mathbf{e}_\rho, \mathbf{e}_\varphi\}$ , and relation (2.46) remains valid.

### 2.4.2 Stokes Parameters

In applications we measure typically intensities, that is,  $|\mathcal{E}_j|^2$ , and not amplitudes  $\mathcal{E}_j$ . Furthermore, every detector averages over some integration time  $T$ , and the reading is proportional to  $\langle |\mathcal{E}_j|^2 \rangle_T$ . For the characterization of a polarization state of a wave by Jones vectors, however, we need the amplitudes and directions for two orthogonal components ( $\mathbf{e}_{x,y}$  or  $\mathbf{e}_{\pm}$ ) at any given time. This is the case for idealized, perfectly polarized fields only, not for waves with a (partially) fluctuating polarization.

For the complete characterization of the polarization state of a wave, we can employ the two different sets of orthogonal polarization components (2.45),



**Figure 2.15** Normalized Stokes parameters and vectors for distinct polarization states. From left to right: linearly  $x$  polarized, linearly  $y$  polarized, unpolarized, and right circularly polarized. The  $\odot$ -symbol indicates the propagation direction out of the drawing.

either  $\mathbf{e}_{x,y}$  or  $\mathbf{e}_\pm$ , and their relative phase. G.G. Stokes (1819–1903) suggested the use of the four time-averaged quantities ( $\mathcal{E}_\pm, \mathcal{E}_{x,y}, \tilde{\phi}, \tilde{\alpha}$  may fluctuate in time):

$$\begin{aligned} S_0 &= \langle \mathcal{E}_x^2 \rangle + \langle \mathcal{E}_y^2 \rangle = \langle \mathcal{E}_+^2 \rangle + \langle \mathcal{E}_-^2 \rangle \\ S_1 &= \langle \mathcal{E}_x^2 \rangle - \langle \mathcal{E}_y^2 \rangle = \langle 2\mathcal{E}_+\mathcal{E}_- \cos 2\tilde{\alpha} \rangle = S_0 \cos 2\gamma \cos 2\alpha \\ S_2 &= \langle 2\mathcal{E}_x\mathcal{E}_y \cos \tilde{\phi} \rangle = \langle 2\mathcal{E}_+\mathcal{E}_- \sin 2\tilde{\alpha} \rangle = S_0 \cos 2\gamma \sin 2\alpha \\ S_3 &= \langle 2\mathcal{E}_x\mathcal{E}_y \sin \tilde{\phi} \rangle = \langle \mathcal{E}_+^2 \rangle - \langle \mathcal{E}_-^2 \rangle = S_0 \sin 2\gamma . \end{aligned} \quad (2.47)$$

The example on p. 56 outlines how to measure four component vectors  $\mathbf{S} = (S_0, S_1, S_2, S_3)$ . Stokes parameters add linearly ( $S''_j = S_j + S'_j$ ) only for the special case of superposing two incoherent waves  $\mathbf{E}$  and  $\mathbf{E}'$  since in this case for all components  $\langle \mathcal{E}_j \mathcal{E}'_k \rangle = 0$  (see Chapter 6).

The first parameter  $S_0$  is obviously proportional to the intensity, and it is straightforward to show that  $S_0 = (S_1^2 + S_2^2 + S_3^2)^{1/2}$ . Hence there are only three independent parameters, which suggests normalizing the  $S$  parameters to  $s_i = S_i/S_0$ . Then  $s_0 = 1$  is always valid, and furthermore

$$\Pi^2 = s_1^2 + s_2^2 + s_3^2 \leq 1 \quad (= 1 \text{ for perfectly polarized light})$$

defines the degree of polarization, which varies from 0 (unpolarized) to 1 (perfectly polarized) (Figure 2.15). Here “unpolarized” means that the instantaneous polarization of the light field – which has fundamental vector character – fluctuates so rapidly that on average no definite polarization state is measured.

### 2.4.3 Polarization State and Poincaré Sphere

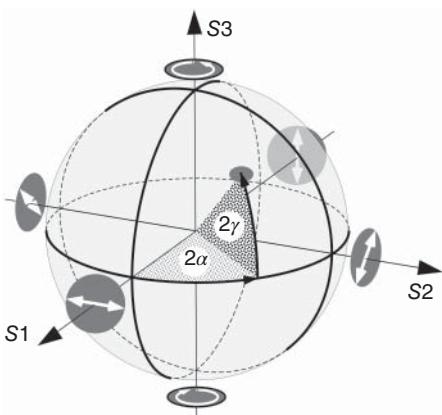
The three normalized Stokes parameters  $\{s_1, s_2, s_3\}$  completely determine the state of polarization. According to Figure 2.14, the “circularity” of a wave can be expressed by the ratio of the axes of the ellipse,

$$\tan \gamma = \frac{\mathcal{E}_+ + \mathcal{E}_-}{\mathcal{E}_+ - \mathcal{E}_-},$$

which varies from 1 to  $-1$  for right-handed to left-handed circular polarization. H. Poincaré (1854–1912) introduced the angles  $2\alpha$  (already known from Eq. (2.44)) and  $2\gamma$  with  $\tan 2\gamma = (\mathcal{E}_+^2 - \mathcal{E}_-^2)/(2\mathcal{E}_+\mathcal{E}_-)$  for characterizing the polarization state. The third parameter set in Eq. (2.47) is generated by this definition and corresponds to the spherical coordinates of a 3D vector given by  $\{1, 2\alpha, 2\gamma\}$ .

The polarization state of a perfectly polarized wave therefore unambiguously corresponds to a point on the unit sphere (see Figure 2.16).<sup>8</sup>

<sup>8</sup> We will find in the chapter on light and matter (Chapter 7) that this structure appears again in the Bloch vectors where atomic two-state systems replace the two state vectors of polarization.



**Figure 2.16** Representation of polarization states on the Poincaré sphere. Linear polarizations are found around the equator, and circular ones at the poles.

#### 2.4.4 Jones Matrices, Polarization Control, and Measurement

Many optical components, most notably retardation plates, transform the state of polarization. The action on the electric field expanded in terms of Jones vectors (2.45) is transparently given by  $2 \times 2$  Jones matrices,  $\mathbf{M}$

$$\mathbf{E}' = \mathbf{M} \odot \mathbf{E} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \mathbf{E}, \quad (2.48)$$

where the  $m_{ij}$  are complex coefficients reflecting the phase evolution of the optical waves. The linearity of this formalism allows to write the action of a series of components as  $\mathbf{M} = \dots \mathbf{M}_3 \odot \mathbf{M}_2 \odot \mathbf{M}_1$ .

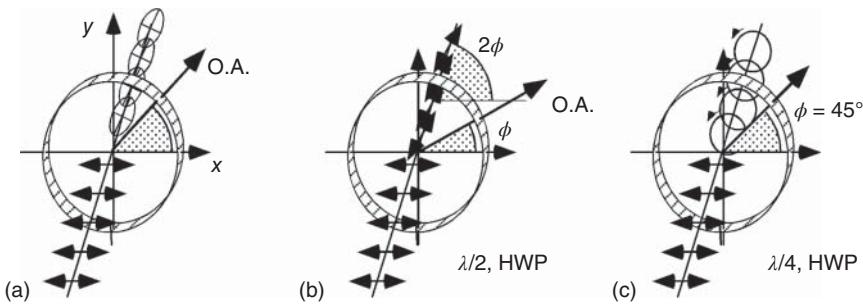
We note that Jones vectors and the Jones formalism are well suited for describing the polarization properties of coherent, that is, typically laser beams, since from intensities we can infer electric field amplitudes. For incoherent and partially polarized light fields, where phase information is lost, the Stokes parameters can still be used. Transformation of the Stokes parameters by optical components can be treated with  $4 \times 4$  Müller matrices. This formalism is omitted here but can be found in, for example, [13, 14].

Let us here concentrate on the functionality of the most relevant components for polarization control, that is, retarder plates. Their construction is discussed in Section 3.4.3. Retarder plates have an optical axis that we assume to have an angle  $\phi$  with respect to the  $x$  axis (Figure 2.17). Their Jones matrix is given by

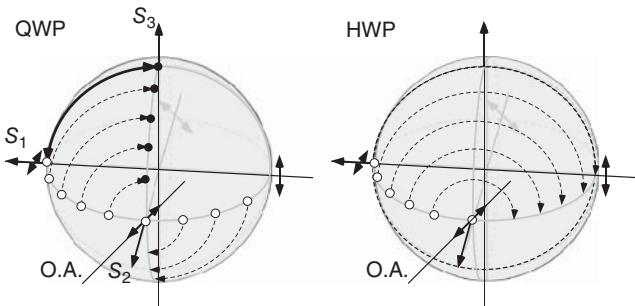
$$\mathbf{E}' = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} e^{i\alpha_o} & 0 \\ 0 & e^{i\alpha_e} \end{pmatrix} \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \mathbf{E}, \quad (2.49)$$

where the electric field is rotated into the frame of the retarder plate and back again. The retarder plate imposes a phase shift  $e^{i\alpha_o}$  ( $e^{i\alpha_e}$ ) for the component parallel (transverse) to the optical axis with  $\alpha_o \neq \alpha_e$  in general. Widely used special cases are quarter-wave plate (QWP) and half-wave plate (HWP).

QWPs ( $\lambda/4$  plate) have  $e^{i\alpha_o} = \pm ie^{i\alpha_e}$ , which corresponds to a phase difference of  $\pi/2$  or a path difference of  $\lambda/4$ , respectively. Here we restrict ourselves to the so-called fast-axis case (+ sign; the – sign for the slow-axis case does not



**Figure 2.17** Retarder plates transform an incoming state of polarization into a new state, depending the angle  $\phi$  of orientation of the optical axis O.A. (a) Important special cases are half-wave plates (HWP, b) and quarter-wave plates (QWP, c).



**Figure 2.18** Polarization transformation of linear polarization states by quarter-wave plates (QWP) and half-wave plates (HWP) visualized on the Poincaré sphere. The optical axis is parallel to the linear polarization at  $S_2$  on the equator.

introduce any new insight; see Section 3.4.3). Insertion and explicit calculation from (2.49) yields

$$\mathbf{M}_{\lambda/4} = \frac{e^{i\alpha_o}}{2} \begin{pmatrix} (1+i) + (1-i) \cos 2\phi & (1-i) \sin 2\phi \\ (1-i) \sin 2\phi & (1+i) - (1-i) \cos 2\phi \end{pmatrix}.$$

For the important orientations  $\phi = 0^\circ, 45^\circ, 90^\circ$ , we find

$$e^{i\alpha_o} \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} ; \quad e^{i\alpha_o} \frac{1+i}{2} \begin{pmatrix} 1 & -i \\ -i & i \end{pmatrix} ; \quad ie^{i\alpha_o} \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix},$$

respectively. Figure 2.18 shows the action of a QWP at  $\phi = 45^\circ$ , that is, parallel to the  $S_2$ -component of the Stokes vector, onto incoming linear polarizations. Horizontal and vertical polarizations are transformed into right(left)-handed circular polarizations.

HWP ( $\lambda/4$  plate) have  $e^{i\alpha_o} = -e^{i\alpha_e}$ , corresponding to phase and path differences of  $\pi$  or  $\lambda/2$ . We find

$$\mathbf{M}_{\lambda/2} = e^{i\alpha_o} \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ -\sin 2\phi & \cos 2\phi \end{pmatrix},$$

which shows rotation of a linear polarization by  $2\phi$ . The HWP is equivalent to two serial QWPs with parallel optical axis, which is apparent from visualization on the Poincaré sphere as well.

All polarization transformations are reversible.<sup>9</sup> Hence any elliptic polarization can be converted into a linear polarization by means of a QWP. Therefore, the combination QWP1–HWP–QWP2 allows to transform any incoming polarization into any outgoing polarization: QWP1 takes any elliptic polarization to the linearly polarized states on the equator; adjustment of the angle of the HWP converts to another linear polarization, which is transformed by QWP2 to the target state. In Section 3.4.3 we also present a fiber-based retardation component useful for applications based on this concept.

### Example: Measuring Stokes parameters

One can show (Problem 2.13) that the intensity recorded by the photodetector in Figure 2.19 is given as a function of the rotation angle  $\phi(t)$ :

$$I(\phi) = \frac{1}{2}(S_0 + S_1 \cos^2(2\phi) + S_2 \cos(2\phi) \sin(2\phi) + S_3 \sin(2\phi)).$$

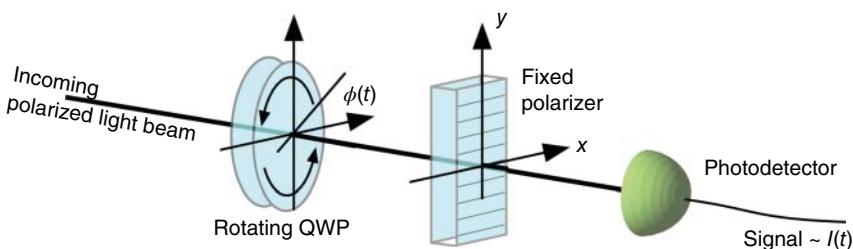
Relevant Fourier coefficients are

$$I(\phi) = \frac{1}{2}(A - B \sin(2\phi) + C \cos(4\phi) + D \sin(4\phi)),$$

which can be associated with the Stokes parameters using trigonometric identities:  $S_0 = A - C$ ,  $S_1 = 2C$ ,  $S_2 = 2D$ , and  $S_3 = B$ .

### 2.4.5 Polarization and Projection

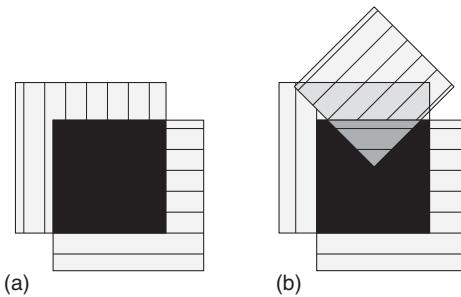
A quite astonishing property of the polarization may be demonstrated impressively with a polarizing foil. A polarizing foil generates polarized light from unpolarized light through absorption of the component that oscillates in parallel with the aligned organic molecules of the foil. More polarizing components will be treated within the chapter on wave propagation in matter (Section 3.4.4).



**Figure 2.19** Polarimeter: the Stokes parameters of the incoming light beam are extracted from the time-varying signal  $I(t)$  by Fourier transform.

<sup>9</sup> Mathematically speaking they are unitary operations. For every matrix  $\mathbf{M}$  there exists an inverse matrix  $\mathbf{M}^{-1}$  that is identical to the Hermitian conjugate matrix,  $\mathbf{M}^{-1} = \mathbf{M}^*$ .

Figure 2.20a illustrates that two crossed polarizers result in the cancellation of the transmission. But it is quite astonishing that when one more polarizer is inserted with polarization direction at  $45^\circ$  in between the two others, a quarter of the light transmitted by the first polarizer (neglecting losses) passes through the orthogonal polarizer! The polarization of the electromagnetic field is “projected” onto the transmission direction of the polarizer, the polarizer affecting the field, not the intensity. See also Problem 2.15.



**Figure 2.20** Transmission of crossed polarizers. The hatching indicates the direction of polarization. (b) In the right-hand figure, the third polarizer is inserted at  $45^\circ$  in between the other two polarizers.

#### 2.4.6 Polarization of Light Beams with Finite Extension

Perfect plane waves have, as outlined in Section 2.2.1, transverse field components only. Application of Maxwell's equation to a purely transverse wave field with finite extension, however, shows immediately that  $\nabla \cdot \mathbf{E} = 0$  does not vanish: we find, for instance,  $\nabla \cdot \mathbf{E} = E_0 e^{ikz} (\partial/\partial x) u(x, y) \neq 0$  for an  $x$ -polarized field  $\mathbf{E} = E_0 u(x, y) \mathbf{e}_x$ . If the transverse amplitude variation is slow at the wavelength scale ( $|\partial/\partial x)u(x, y)| \ll \lambda$ ), a good first-order approximation of the  $x$ -polarized vector wave field is

$$\mathbf{E} = E_0 e^{ikz} \left( u \mathbf{e}_x + \frac{i}{k} \frac{\partial}{\partial x} u \mathbf{e}_z \right), \quad (2.50)$$

where the  $z$ -component compensates the nonvanishing contribution of finite transverse fields to the divergence. The corresponding  $\mathbf{H}$ -field components are

$$\mathbf{H} = i \frac{E_0}{\mu_0 c} e^{ikz} \left( u \mathbf{e}_y + \frac{i}{k} \frac{\partial}{\partial y} u \mathbf{e}_z \right).$$

For  $y$ -polarized fields, similar arguments hold and any arbitrary polarization can be constructed then. The paraxial wave equation is still fulfilled since the new term scales as the beam width  $w(z)$ , which is in most cases much larger than the wavelength  $\lambda$  and hence remains negligible to first order, except in strongly focused regions with  $w \sim \lambda$ .

As a consequence of (2.50), small  $z$ -components of the fields cause the momentum flow and Poynting vector of optical beams to develop small azimuthal components as well. Evaluation of Eq. (2.18) shows that in addition to the linear component known already, a new contribution appears,

$$\mathbf{S} = \frac{1}{2} \epsilon_0 c |E_0|^2 \left( |u|^2 \mathbf{e}_z + \Re \left\{ \frac{i}{k} u \nabla u^* \right\} \right), \quad (2.51)$$

which we have generalized to account for all three spatial components with  $u = u(x, y, z)$ . For cylindrical beams with  $u = u_{\rho z} e^{i\ell\phi}$ , we have

$$\Re \left\{ \frac{i}{k} u \nabla u^* \right\} = -\frac{1}{2k} \left( \frac{\partial}{\partial \rho} |u_{\rho z}|^2 \mathbf{e}_\rho + \frac{\partial}{\partial z} |u_{\rho z}|^2 \mathbf{e}_z \right) + \frac{\ell}{k\rho} |u_{\rho z}|^2 \mathbf{e}_\phi. \quad (2.52)$$

The  $z$ - and  $\rho$ -components are related to the divergence of the beam; the azimuthal component is nonzero for helical wave fronts with  $\ell \neq 0$  only and gives rise to the OAM of light beams (see Section 2.5.4).

## 2.5 Optomechanics: Mechanical Action of Light Beams

Light waves do not carry mass. However, the energy density  $U$  (2.15) and the momentum density  $\mathbf{g} = \mathbf{S}/c^2$  (2.18) move at the speed of light. The momentum flux can be transferred to massive objects, for example, on reflection or refraction, giving rise to light forces such as light pressure.

### 2.5.1 Radiation Pressure

The total mechanical momentum of a propagating light field is

$$\mathbf{P} = \int_V \mathbf{g} d^3r = \frac{1}{c^2} \int_V \mathbf{E} \times \mathbf{H} d^3r,$$

and it is directed along the propagation direction of a light beam. If the direction of propagation is changed by refraction, reflection, or absorption, momentum transfer occurs, which corresponds to a force exerted by the light beam onto the object responsible for the modification.

A light beam carrying momentum density  $g$  deposits the total momentum  $\Delta P = c \cdot A \cdot g \cdot \Delta t$  per time interval  $\Delta t$  into an area  $A$ , yielding a force  $\Delta P/\Delta t = F = cgA = p_{\text{rad}} \cdot A$ . Hence we find the radiation pressure

$$p_{\text{rad}} = c \cdot g,$$

which is formally identical to the energy density (2.15),  $p_{\text{rad}} = U$ .

#### Example: Radiation pressure force of a laser pointer

A laser beam with intensity  $I$  is absorbed by a blackened beam stop. The momentum delivered to an area  $A$  is proportional to the total power deposited,  $P_L = I \cdot A = cUA = c^2gA$ , yielding a force

$$F = P_L/c.$$

The force resulting from a laser pointer (0.5 mW legal maximum power) is at  $F = 0.16 \times 10^{-11}$  N very feeble. Radiation pressure is nevertheless responsible for important phenomena including comet tails, laser cooling of atoms (see Section 12.5), and more. If the absorbing plate is finally replaced by a mirror, the momentum change is twice as large, and hence  $F = 2P_L/c$ .

### 2.5.2 Angular Momentum of Light Beams

The role of the linear momentum density  $\mathbf{g}$  for optomechanical action is to introduce a corresponding angular momentum density  $\mathbf{j}$  through  $\mathbf{j} = \mathbf{r} \times \mathbf{g} = \mathbf{r} \times \frac{1}{c^2}(\mathbf{E} \times \mathbf{H})$ . Again, physical significance can only be associated with the volume integral

$$\mathbf{j} = \frac{1}{c^2} \int_V \mathbf{r} \times (\mathbf{E} \times \mathbf{H}) dV. \quad (2.53)$$

One might expect that for circular polarization, the  $\mathbf{r} \times \mathbf{g}$  term for a plane wave propagating in the  $z$  direction, and thus  $\mathbf{g} = g\mathbf{e}_z$  gives rise to an angular momentum density. However, with  $\mathbf{r} \times \mathbf{g} \cdot \mathbf{e}_z = \mathbf{r} \cdot (g\mathbf{e}_z \times \mathbf{e}_z) = 0$ , the  $z$ -component vanishes! Now, plane waves with their infinite extensions in space and time do not exist. Replacing the magnetic field with the vector potential,  $\mathbf{H} = \epsilon_0 c^2 \nabla \times \mathbf{A}$  (2.9), performing partial integration, and letting the wave vanish at the boundaries (see problem 2.6, [9, 15]) allows to transform Eq. (2.53) to a new form,

$$\mathbf{j} = \epsilon_0 \int_V (\mathbf{E} \times \mathbf{A} + \sum_i E_i (\mathbf{r} \times \nabla) A_i) dV, \quad (2.54)$$

where the factor  $\mathbf{r} \times \nabla = \hat{\mathbf{L}}/i\hbar$  agrees with the well-known angular momentum operator of quantum mechanics. The analogy suggests a decomposition

$$\mathbf{j} = \mathbf{S}_{AM} + \mathbf{L}_{AM}. \quad (2.55)$$

The time average of the intrinsic part independent of  $\mathbf{r}$  is called *spin angular momentum* and reads with  $\mathbf{A} = \mathbf{E}/i\omega$ : Eq. (2.9)

$$\langle \mathbf{S}_{AM} \rangle = \frac{\epsilon_0}{2\omega} \int_V \mathbf{E} \times \mathbf{E}^* dV. \quad (2.56)$$

Paraxial approximation (i.e., neglecting terms  $\partial/\partial x E_i \ll kE_i$ , etc.) allows to simplify the  $\mathbf{r}$ -dependent part along the lines of Section 2.4.6. For monochromatic waves with  $\omega = ck$ , we use  $\mathbf{E} \simeq \{E_x, E_y, i(\partial/\partial x E_x + \partial/\partial y E_y)/k\}$  and  $\mu_0 \mathbf{H} \simeq \{E_y, -E_x, i(\partial/\partial x E_y + \partial/\partial y E_x)/k\}$ . Using some algebra we recover the  $z$ -component of the Poynting vector (2.18) and write

$$\langle \mathbf{L}_{AM} \rangle = \frac{\epsilon_0}{2\omega} \int_V \sum_i E_i^* (\mathbf{r} \times \nabla) A_i dV = \frac{1}{c^2} \int_V \mathbf{r} \times \mathbf{S}_{Poynt} dV. \quad (2.57)$$

The second term is associated with an OAM, which vanishes for cylindrically symmetric beams and will be discussed in Section 2.5.4.

Although the decomposition (2.55) seems very natural, it is not unique; we have selected a coordinate system (the  $z$  direction). Extensive discussions of these subtleties are found in, for example, [16, 17].

#### Example: Spin angular momentum of a circularly polarized wave

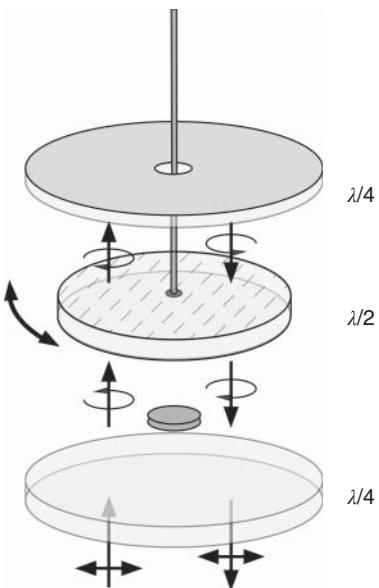
A circularly polarized electric field  $\mathbf{E}_\pm = \mathcal{E}_0 \mathbf{e}_\pm e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})}$  (2.45) propagates in the  $z$  direction. Its vector potential (Eq. (2.9)) reads  $\mathbf{A}_\pm = -\mathcal{E}_0 \mathbf{e}_\pm e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})}/i\omega$ .

To calculate the time-averaged angular momentum, we use formula (A.16),  $\mathbf{e}_\pm \times \mathbf{e}_\pm^* = \mp i\mathbf{e}_z$ , and Eq. (2.16):

$$\langle \mathbf{S}_{\text{AM}} \rangle = \frac{\epsilon_0}{2} \Re \{ \frac{1}{-i\omega} \int_V \mathbf{E}_\pm \times \mathbf{E}_\pm^* dV \} = \pm \frac{\epsilon_0}{2\omega} \int_V \mathcal{E}_0 \mathcal{E}_0^* dV \mathbf{e}_z = \pm \frac{\langle \mathcal{U} \rangle}{\omega} \mathbf{e}_z.$$

We find that the spin angular momentum density of a circularly polarized wave is proportional to the energy density. If we assume especially a total energy  $\langle \mathcal{U} \rangle = n \cdot \hbar\omega$  corresponding to  $n$  photons, we find that the spin angular momentum per photon is simply  $\langle \mathbf{S}_{\text{AM}} \rangle = \pm \hbar \mathbf{e}_z$ .

### 2.5.3 Beth's Experiment



**Figure 2.21** Beth's experiment to demonstrate the angular momentum transfer from a circularly polarized light field to a  $\lambda/2$  plate.

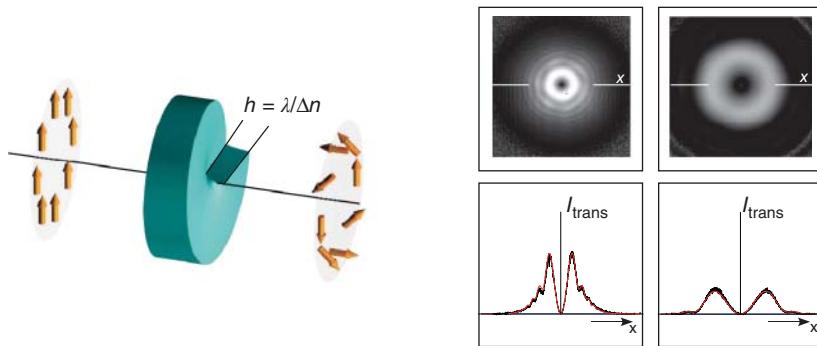
An experimental demonstration of the spin angular momentum of a light wave with circular polarization was realized already in 1936. In Beth's experiment [18] (Figure 2.21), the linear polarization of an upward-moving light field is converted into circular polarization by a fixed QWP ( $\lambda/4$ , Section 3.4.3) before it traverses a torsion pendulum made from a transparent HWP (hatched disc labeled  $\lambda/2$ ) and hung by a thin fiber. The HWP reverses the handedness of the light field, transferring angular momentum to the pendulum.

The light field is furthermore reflected back off another silvered  $\lambda/4$  plate on top. It reverses the handedness again in order to transfer the same amount of angular momentum by the upward- and downward-propagating beams. Large sensitivity is obtained by modulating (amplitude, polarization) the light field periodically at the resonance frequency of the pendulum. An opaque disc prevents light from influencing the torsion fiber.

The torque exerted onto the HWP arises from the anisotropy of the birefringent material: dielectric polarization and electric field vectors are no longer parallel (Section 3.4.1), and hence a torque volume density  $d\tau/dV = \mathbf{P} \times \mathbf{E}$  occurs.

### 2.5.4 Optical Angular Momentum (OAM)

Only in 1992 [19] it was realized that light beams can indeed have OAM with practical relevance. Light beams with helical rather than plane wave fronts have a dark hole at the center and according to Eqs. (2.51) and (2.52) must have nonvanishing azimuthal components of the field and the Poynting vector. Thus the angular momentum density (2.57) does not vanish. Such beams are called “twisted beams” or “vortex beams.”



**Figure 2.22** (a) The spiral phase plate (SPP) creates a beam with a dark spot at the center. Arrows indicate transverse phase distribution. (b) Intensity profiles within the Fresnel and the Fraunhofer regime. An analytic treatment (solid line) can hardly be distinguished from the measurement [20] (see Section 2.9).

#### 2.5.4.1 Twisted Beams

Helical wave fronts are straightforwardly generated by a vortex lens or spiral phase plate (SPP). Its action on an incoming wave front with plane transverse phase distribution is illustrated in Figure 2.22. A phase difference  $2\pi$  at the step dislocation is obtained for  $h = \lambda/\Delta n$  with  $\Delta n$  the refractive index difference of the SPP and the environment. The origin of the dark spot axis of the twisted beam is easily understood: for any light beam with cylindric amplitude profile, partial waves from opposing sides of the axis contribute with exactly opposite phases angles and hence perfectly cancel on axis.

The SPP turns an incoming Gaussian beam into a twisted beam, which varies in shape with distance from the SPP. While the transverse distribution changes with distance and can be calculated analytically in excellent agreement with observation (see Section 2.3.6), so-called LG beams propagate without losing their shape; they are the eigenmodes of light beams carrying OAM.

#### 2.5.4.2 Laguerre–Gaussian Modes

LG modes are eigenmodes of paraxial waves with cylindrical intensity distribution (like Gaussian beams) but helical wave fronts. They propagate without losing shape. The Hermitian polynomials  $H_j(x)H_k(y)$  of HG beams (Eq. (2.41)) are replaced by  $e^{i\ell\varphi}L_m^\ell(\rho)$  with Laguerre polynomials  $L_m^\ell$ . In terms of the quantities  $w_0$ ,  $w(z)$ ,  $R(z)$ , and  $\eta(z)$  defined in Section 2.3.1.1 and for cylindrical coordinates  $(\rho, z, \varphi)$ , the field distributions of LG modes are given by

$$E_{\ell m}(\rho, \varphi, z) = E_0 \left( \frac{\sqrt{2}\rho}{w(z)} \right)^\ell L_m^\ell \left( \frac{\sqrt{2}\rho}{w(z)} \right) \frac{w_0}{w(z)} e^{-(\rho/w(z))^2} \\ \times e^{i\ell\varphi} e^{ik\rho^2/2R(z)} e^{i(kz-(2m+\ell+1)\eta(z))}. \quad (2.58)$$

For  $\ell = 0$  we have  $L_0^0 = 1$  and recover once again the  $\text{TEM}_{00}$  Gaussian mode (2.31). For  $\ell \geq 1$  all modes have zero value at the center at  $\rho = 0$  associated with a hole or vortex. This annular character remains in effect even when the light beam is strongly focused since it arises from the phase singularity at the axis.

The OAM contribution of a paraxial light beam was shown in (2.57) to fulfill

$$\mathbf{L}_{\text{AM}} = \frac{1}{c^2} \int_V \mathbf{r} \times \mathbf{S} d^3 r.$$

With respect to the  $z$  axis of the light beam, only the azimuthal component of the Poynting vector will give a nonvanishing contribution with  $\mathbf{r} \times \mathbf{e}_\phi = \mathbf{p} \times \mathbf{e}_\phi$ . We find from Eq. (2.52) with  $\mathbf{S} = S_z \mathbf{e}_z + S_\phi \mathbf{e}_\phi$

$$S_\phi \mathbf{e}_\phi = \frac{1}{2} \epsilon_0 c |E_0|^2 \cdot \frac{\ell}{k\rho} |u_{\rho z}|^2 \mathbf{e}_\phi.$$

For  $\ell > 0$  the time-averaged Poynting vector  $\langle \mathbf{S} \rangle$  follows the winding of the phase fronts of LG beams around the axis of propagation, suggesting that mechanical torques can be exerted by such light fields. Evaluation of (2.57) yields

$$\langle \mathbf{L}_{\text{AM}} \rangle = \frac{\ell}{\omega} \int_V \frac{1}{2} \epsilon_0 |E_0|^2 |u_{\rho z}|^2 d^3 r \frac{\ell}{\omega} \mathbf{e}_z = \frac{\ell}{\omega} \langle U \rangle \mathbf{e}_z \quad (2.59)$$

For a “photon wave” with  $\langle U \rangle = n \cdot \hbar\omega$ , we find  $\langle L_{\text{AM}} \rangle = \hbar\ell$  per photon. The total angular momentum of a light wave is the sum of spin and orbital angular momentum (Eq. (2.55)).

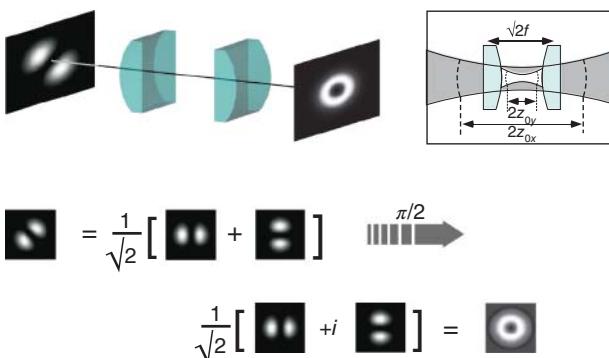
#### 2.5.4.3 Transforming Hermite–Gaussian to Laguerre–Gaussian Beams

With two cylinder lenses forming a telescope, HG beams can be transformed into LG beams [21]. The arrangement shown in Figure 2.23 introduces a  $\pi/2$ -phase shift: LG beams are constructed from a superposition of HG beams by adjusting their relative phases.

##### Example: $\pi/2$ -converter

The astigmatic lens arrangement of Figure 2.23 causes different phase shifts for the  $H_{10}$  and the  $H_{01}$  modes. Show that the cylinder lenses must be separated by  $2d = \sqrt{2}f$  in order to achieve a phase shift of  $\pi/2$ .

The astigmatic beam profile is sketched in the insert of Figure 2.23, where in between the cylinder lenses the  $x$ - and  $y$ - components have different Rayleigh



**Figure 2.23** A  $\text{TEM}_{01}$  This transformation of Hermite–Gaussian modes into Laguerre–Gaussian modes makes use of the Gouy phase.

ranges  $z_{0x}$  and  $z_{0y}$ . Using Eq. (2.42), we calculate that the phase difference for (10) and (01) modes is given by

$$(\eta_x(z) - \eta_y(z))|_{-d}^d = 2(\tan^{-1}(d/z_x) - \tan^{-1}(d/z_y)). \quad (2.60)$$

Outside the cylinder lenses the beams have equal width and curvature. Hence we have from Section 2.3.1.4, and following Eq. (2.33) and Ref. [21], the two relations

$$\frac{z_{0x}^2 + d^2}{z_{0x}} = \frac{z_{0y}^2 + d^2}{z_{0y}} \quad \text{and}$$

$$\frac{1}{f} = -\frac{1}{R_x} + \frac{1}{R_y} = -\frac{d}{z_{0x}^2 + d^2} + \frac{d}{z_{0y}^2 + d^2}.$$

We solve for  $z_{0x}$  and  $z_{0y}$  by introducing the parameter  $p^2 = (1-d)/(1+d)$  with  $p > 0$  and find

$$z_{0x} = d/p \quad \text{and} \quad z_{0y} = d \cdot p. \quad (2.61)$$

Insertion into (2.60) and application of the identity  $\tan^{-1}(1/p) + \tan^{-1}(-p) = \tan^{-1}((1-p^2)/2p) = \pi/4$  yields the final result  $p = -1 + \sqrt{2}$  and  $2d = \sqrt{2}f$ . Note that (2.61) also determines the global properties of the external mode.

The  $\pi$ -converter creates an exact mirror image of any incoming beam at separation  $2f$ . A combination of  $\pi/2$  and  $\pi$  converter allows to transform any incoming HG beam into a desired outgoing LG beam.

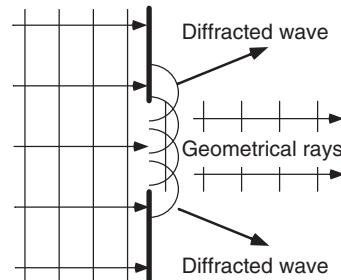
## 2.6 Diffraction

Light diffraction has played an important role in the development of the wave theory of light. Even famous physicists doubted for a long time that “light comes around a corner like sound,” but already Leonardo da Vinci (1452–1519) knew that some light falls into the shadow of an illuminated object – against the predictions of geometrical optics.

C. Huygens gave a first illustrative idea on wave theory by interpreting every point in space as an excitation source of a new wave, a concept called today *Huygens' principle* (Figure 2.24). It indeed makes many diffraction phenomena accessible to our intuition, but it assumes *ad hoc* that waves are always transmitted into the forward direction only.

The general mathematical formalism of Huygens' principle is extremely elaborate, because the electric and magnetic radiation fields are vector fields,  $\mathbf{E} = \mathbf{E}(x, y, z, t)$  and  $\mathbf{B} = \mathbf{B}(x, y, z, t)$ . Up to now there exist only a few general solved examples; the problem of planar wave propagation at an infinite thin edge solved in 1896 by A. Sommerfeld (1868–1951) counts as an exception.

An enormous simplification is achieved when substituting the vectorial field by scalar ones, whereby we have to determine the range of validity of the

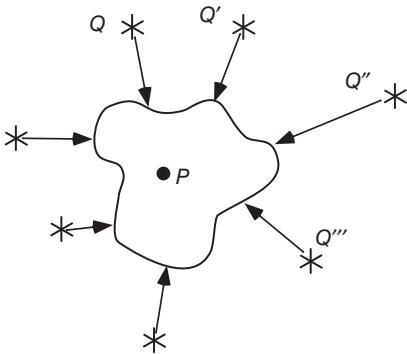


**Figure 2.24** Huygens' principle: diffraction at an aperture.

approximation. It is advantageous that light beams often propagate with only small changes of direction. Then the polarization changes only slightly, and the scalar approximation describes the behavior excellently.

### 2.6.1 Scalar Diffraction Theory

Here it is our objective to understand Huygens' principle by means of mathematics in scalar approximation by applying the superposition principle to the combined radiation field of multiple sources.<sup>10</sup>



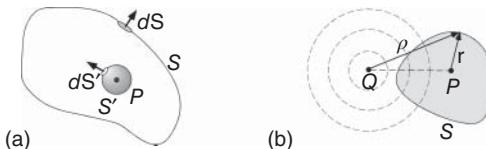
**Figure 2.25** The light field at  $P$  is fed by the sources  $Q, Q', Q'', \dots$ .

To cover all fields incident on a point  $P$ , we look at the sources on a surface  $S$  and the effect of these on a very small volume with surface element  $dS'$  around  $P$  (Figure 2.26). We can make use of Green's integral theorem, well known from mathematics, which for two solutions  $\psi$  and  $\phi$  of the Helmholtz equation (2.14) reads

$$\oint_S [\psi \nabla \phi - \phi \nabla \psi] d\mathbf{S} = \int_V [\psi \nabla^2 \phi - \phi \nabla^2 \psi] d^3 r = 0.$$

We let  $e^{ikr}/kr$  and  $\mathcal{E}(\mathbf{r}_P)$  be used for  $\psi$  and  $\phi$ , and in Figure 2.26a we cut out a sphere with very short radius  $r'$  and surface element  $dS' = r^2 d\Omega' \mathbf{e}_r$  about point  $P$  to be contracted to this point:

$$\left( \oint_S d\mathbf{S} + \oint_{S'} d\mathbf{S}' \right) \left[ \frac{e^{ikr}}{r} \nabla \mathcal{E} - \mathcal{E} \nabla \frac{e^{ikr}}{r} \right] = 0.$$



**Figure 2.26** Kirchhoff's theorem. (a) Choice of the surfaces according to Eq. (2.62). (b) The surface  $S$  is excited by a source  $Q$  and radiates toward the point  $P$ .

<sup>10</sup> This section is mathematically more tedious. The reader may skip it and simply use the results in Eqs. (2.67) and (2.68).

<sup>11</sup> This treatment requires spatial and temporal coherence of the light waves, which will be discussed in more detail in the chapter on interferometry (Chapter 6).

Within this chapter we will for simplicity exclusively discuss the propagation of monochromatic waves:<sup>11</sup>

$$E(\mathbf{r}, t) = \mathcal{E}(\mathbf{r}) e^{-i\omega t}.$$

The total field  $\mathcal{E}(P)$  at a point  $P$  (Figure 2.25) is composed of the sum of all contributions of the individual sources  $Q, Q', \dots$ . We know already that spherical waves emerging from a point-like source  $Q$  have the scalar form of Eq. (2.22):

$$\mathcal{E} = \mathcal{E}_Q e^{ikr}/kr.$$

On the surface of the small sphere around  $P$ , we have  $d\mathbf{S}' \parallel \mathbf{e}_r$  and thus  $\nabla \mathcal{E} \cdot d\mathbf{S}' = (\partial \mathcal{E} / \partial r) r^2 d\Omega'$ . We also use  $-\nabla e^{ikr}/r = (1/r^2 - ik/r) e^{ikr} \mathbf{e}_r$  and find

$$\oint_S \left[ \frac{e^{ikr}}{r} \nabla \mathcal{E} - \mathcal{E} \nabla \frac{e^{ikr}}{r} \right] d\mathbf{S} = \oint_{S'} \left[ \mathcal{E}(1 - ikr) + r \frac{\partial \mathcal{E}}{\partial r} \right] e^{ikr} d\Omega'. \quad (2.62)$$

Now we let the radius of the volume around  $P$  decrease more and more ( $r \rightarrow 0$ ), and with

$$\oint_{S'} \left( \mathcal{E} - ikr\mathcal{E} + r \frac{\partial \mathcal{E}}{\partial r} \right) e^{ikr} d\Omega' \xrightarrow{r \rightarrow 0} 4\pi\mathcal{E}|_{r=0} = 4\pi\mathcal{E}_P,$$

we can prove Kirchhoff's integral theorem:

$$\mathcal{E}(\mathbf{r}_P) = \frac{1}{4\pi} \oint_S \left[ \frac{e^{ikr}}{r} \nabla \mathcal{E} - \mathcal{E} \nabla \frac{e^{ikr}}{r} \right] d\mathbf{S}. \quad (2.63)$$

In principle we can now predict the field at point  $P$  if we know the distribution of fields on the surface  $S$ . Owing to its relatively wide generality, however, Kirchhoff's theorem does not give the impression that it might be very useful. Therefore we want to study further approximations and apply them to a point source  $Q$  illuminating the surface  $S$  (Figure 2.26b). Let us assume that a scalar spherical wave of the form

$$\mathcal{E}(\mathbf{p}, t) = \frac{\mathcal{E}_Q}{k\rho} e^{i(k\rho - wt)}$$

propagates from there. We use spherical coordinates and just insert the spherical wave into Eq. (2.63):

$$\mathcal{E}(\mathbf{r}_P) = \frac{\mathcal{E}_Q}{4\pi k} \oint_S \left[ \frac{e^{ikr}}{r} \left( \frac{\partial}{\partial \rho} \left( \frac{e^{ik\rho}}{\rho} \right) \right) \mathbf{e}_\rho - \frac{e^{ik\rho}}{\rho} \left( \frac{\partial}{\partial r} \left( \frac{e^{ikr}}{r} \right) \right) \mathbf{e}_r \right] d\mathbf{S}.$$

Then we make use of the approximation

$$\frac{\partial}{\partial \rho} \frac{e^{ik\rho}}{\rho} = k^2 e^{ik\rho} \left( \frac{i}{k\rho} - \frac{1}{(k\rho)^2} \right) \simeq e^{ik\rho} \frac{ik}{\rho} \quad (2.64)$$

for  $\rho$  and  $r$ , which is excellent already for distances of only a few wavelengths, since  $k\rho \gg 1$ . Then Kirchhoff's integral (2.63) is crucially simplified,

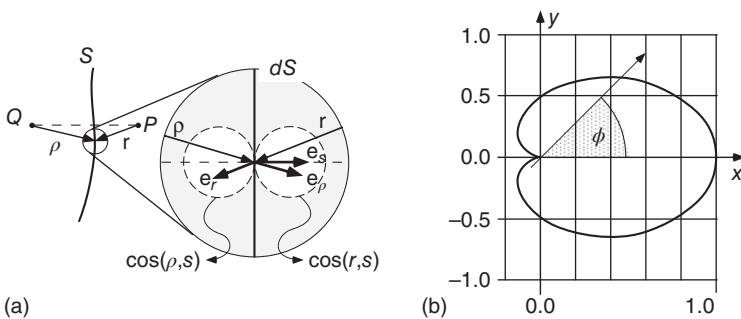
$$\mathcal{E}(\mathbf{r}_P) = -\frac{i\mathcal{E}_Q}{2\pi} \oint_S \frac{e^{ik(r+\rho)}}{r\rho} N(\mathbf{r}, \mathbf{p}) dS, \quad (2.65)$$

whereby we have introduced the Stokes factor  $N(\mathbf{r}, \mathbf{p})$ :

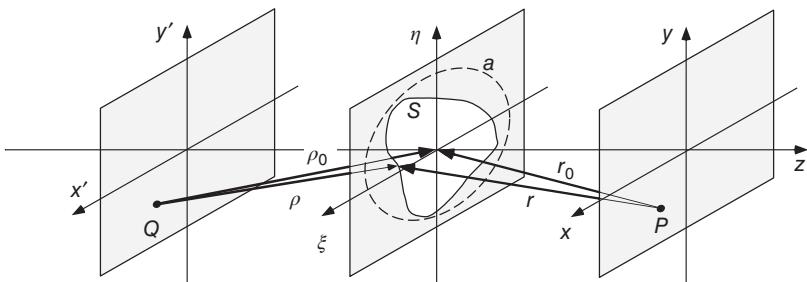
$$N(\mathbf{r}, \mathbf{p}) = -\frac{\mathbf{e}_r \mathbf{e}_s - \mathbf{e}_\rho \mathbf{e}_s}{2} = -\frac{1}{2} [\cos(\mathbf{r}, \mathbf{e}_s) - \cos(\mathbf{p}, \mathbf{e}_s)]. \quad (2.66)$$

To understand the Stokes factor (also obliquity factor) and its meaning (respectively to substitute it by the value "1" in most cases), we look at Figure 2.27. Therefore we make use of a more realistic example, in which the rays are near the axis, which means that they propagate in the vicinity of the connecting line between  $Q$  and  $P$ . We can specify the "excitation" originating from the surface element  $dS$  with

$$d\mathcal{E}_S = (\mathcal{E}_Q/k\rho) \exp(ik\rho) \cos(\mathbf{p}, \mathbf{e}_s) dS,$$



**Figure 2.27** Interpretation of the Stokes factor. (a) Geometric relations. (b) Angle dependence of the Stokes factor,  $N(\mathbf{r}, \rho \parallel \mathbf{e}_s) = [1 + \cos(\phi)]/2$ .



**Figure 2.28** Fraunhofer diffraction for a Stokes factor  $N \simeq 1$ .

the contribution at  $P$  with

$$d\mathcal{E}_P = d\mathcal{E}_S \cos(\mathbf{e}_r, \mathbf{e}_s) \exp(ikr)/r,$$

and thus find exactly the factors from Eq. (2.65).

A remarkable property of the Stokes factor is the suppression of the radiation in the backward direction, because according to Eq. (2.66),  $N \rightarrow 0$  holds for  $\mathbf{e}_\rho \rightarrow -\mathbf{e}_r$ ! In contrast to that, we find for near-axis rays in the forward direction  $N \rightarrow 1$ , and we want to restrict ourselves to this frequent case in the following. The right part of Figure 2.27 shows the total angle distribution of the Stokes factor for a planar incident wave with  $\rho = \mathbf{e}_s$ .

We finally consider the propagation of near-axis rays for  $N \simeq 1$  in the geometry and with the notations of Figure 2.28. Besides we assume that the surface  $S$  is illuminated with a planar wave. Then the field strength  $\mathcal{E}_S \simeq \mathcal{E}_Q/k\rho$  is constant, but the intensity distribution may be characterized by a transmission function  $\tau(\xi, \eta)$  (which in principle can be imaginary, if phase shifts are caused). According to Eq. (2.65), we can calculate the field strength at the point  $P$  as

$$\mathcal{E}(\mathbf{r}_P) = -\frac{i\mathcal{E}_S}{\lambda} \oint_S \tau(\xi, \eta) \frac{e^{ikr}}{r} d\xi d\eta. \quad (2.67)$$

This result is also still too difficult for a general treatment. But further approximations are made easier by the circumstance that the distance between the diffracting object and the area of observation is in general large compared with

the wavelength and the transverse dimensions, which are marked in Figure 2.28 by a circle with radius  $a$  within the plane of the diffracting object. We express the distances  $r$  and  $r_0$  by the coordinates of the respective planes:

$$r^2 = (x - \xi)^2 + (y - \eta)^2 + z^2 \quad \text{and} \quad r_0^2 = x^2 + y^2 + z^2.$$

We consider  $r$  as a function of  $r_0$ ,

$$r^2 = r_0^2 \left( 1 - \frac{2(x2.68\xi + y\eta)}{r_0^2} + \frac{\xi^2 + \eta^2}{r_0^2} \right),$$

and expand  $r$  with  $\kappa_x = -kx/r_0$  and  $\kappa_y = -ky/r_0$ ,

$$r = r_0 \sqrt{1 + \frac{2(\kappa_x \xi + \kappa_y \eta)}{kr_0} + \frac{\xi^2 + \eta^2}{r_0^2}} \simeq r_0 \left( 1 + \frac{\kappa_x \xi + \kappa_y \eta}{kr_0} + \frac{\xi^2 + \eta^2}{2r_0^2} \right).$$

Then the phase factor in Eq. (2.67) decomposes into three contributions,

$$\exp(ikr) \rightarrow \exp(ikr_0) \exp[i(\kappa_x \xi + \kappa_y \eta)] \exp\left(\frac{ik(\xi^2 + \eta^2)}{2r_0}\right).$$

The first factor just yields a general phase factor, the second depends linearly on the transverse coordinates of the diffracting plane and the plane of observation, and the last one depends only on the coordinates of the diffracting plane (we have met the last factor already when discussing Gaussian beams (“Fresnel factor” see p. 41)). In many experiments, the Fresnel factor differs only a little from 1, because  $ka^2/r_0 \ll 1$ . Therefore it delivers the distinguishing property for the two important basic diffraction types, Fraunhofer and Fresnel diffraction ( $r_0 \simeq z$ ):

- |      |                        |                                    |        |
|------|------------------------|------------------------------------|--------|
| (i)  | Fraunhofer diffraction | $a^2 \ll \lambda z$                |        |
| (ii) | Fresnel diffraction    | $a^2 \geq \lambda z$ but $a \ll z$ | (2.68) |

Since the nineteenth century, diffraction phenomena have played an important role in the development of the wave theory of light, and up to now they are closely correlated with the names of Joseph von Fraunhofer (1787–1826) and Augustin-Jean Fresnel (1788–1827). The radius  $a = \sqrt{\lambda z}$  defines the region of validity of the Fraunhofer approximation within the diffracting plane. The usual condition is that in this case the object lies completely within the first *Fresnel zone* (see also Section 2.8.2). Besides, when the distance  $z$  to the diffracting object is just chosen large enough, one always reaches the far-field limit, where Fraunhofer diffraction is valid.

## 2.7 Fraunhofer Diffraction

The Fraunhofer approximation is applied in the far field of a diffracting object (e.g., a slit with typical dimension  $a$ ), if (2.68(i)) is fulfilled. For near-axis beams we can substitute the factor  $1/r \simeq 1/r_0 \simeq 1/z$ , and we find after inserting the approximations into Eq. (2.67) the expression

$$\mathcal{E}(\mathbf{r}_P) = -\frac{i\mathcal{E}_S e^{ikr_0}}{\lambda z} \oint_S \tau(\xi, \eta) e^{i(\kappa_x \xi + \kappa_y \eta)} d\xi d\eta. \quad (2.69)$$

But we do keep the global phase factor  $e^{ikr_0}$ ,

$$\exp(ikr_0) \simeq \exp(ikz) \exp\left(\frac{ik(x^2 + y^2)}{2z}\right), \quad (2.70)$$

because here even small deviations may lead to a fast phase rotation, which then plays an important role in interference phenomena. After that the field amplitude at point  $P$  has the form of a spherical wave, which is modulated with the Fourier integral  $T(\kappa_x, \kappa_y)$  of the transmission function  $\tau(\xi, \eta)$ :

$$T(\kappa_x, \kappa_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\xi d\eta \tau(\xi, \eta) e^{i(\kappa_x \xi + \kappa_y \eta)}.$$

Finally the great impact of Fourier transformation in many areas of physics has been significantly supported by its relevance for the treatment of optical diffraction problems. Now we want to discuss some important examples.

### Example: Fraunhofer diffraction

#### 1. Fraunhofer diffraction at a long single slit

We consider a long quasi-one-dimensional slit (Figure 2.29, width  $d$ ) and assume again that the illumination may be inhomogeneous. Because we have introduced several approximations concerning ray propagation (e.g., Stokes factor  $N = 1$ ), we may not solve the one-dimensional case just by simple integration of the  $\eta$  coordinate in Eq. (2.69) from  $-\infty$  to  $\infty$ . Instead we have to work out the concept of Kirchhoff's integral theorem for a line-like (instead of a point-like) source. From a line-like source, a cylindrical wave originates, the intensity of which does not decline like  $1/z^2$  anymore, but only with  $1/z$ . It turns out that the result has a very similar structure.

The amplitude of the cylindrical wave must decline with  $1/\sqrt{z}$ , and the one-dimensional variant of Eq. (2.69) has the form

$$\mathcal{E}(\mathbf{r}_P) = -\frac{i\mathcal{E}_S e^{ikr_0}}{\lambda \sqrt{kz}} \int_{-\infty}^{\infty} \tau(\xi) e^{i\kappa_x \xi} d\xi.$$

In the case of a linear, infinitely long slit, the transmission function has the simple form  $\tau(\xi) = 1$  for  $|\xi| \leq d/2$  and else  $\tau(\xi) = 0$ . One calculates

$$\begin{aligned} \mathcal{E}(x) &= -\frac{i\mathcal{E}_S e^{ikr_0}}{\lambda \sqrt{kz}} \int_{-d/2}^{d/2} d\xi e^{i\kappa_x \xi} \\ &= \mathcal{E}_S \frac{d e^{ikr_0}}{\lambda \sqrt{kz}} \frac{\sin(kxd/2z)}{kxd/2z}. \end{aligned}$$

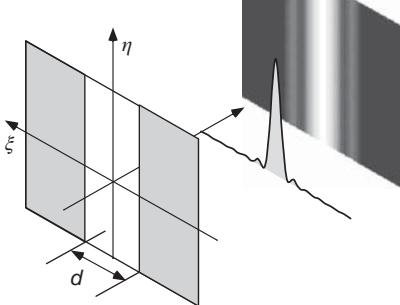


Figure 2.29 Diffraction at a long slit.

The intensity distribution  $I(x) \propto |\mathcal{E}(x)|^2$  is shown in Figure 2.29 and distorted slightly in the gray color scale for clarification.

### 2. Fraunhofer diffraction at a “gaussian transmitter”

We consider a Gaussian amplitude distribution, which one may create, for example, out of a planar wave by a filter with a Gaussian transmission profile (Figure 2.30). On the other hand, we may just use the Gaussian beam from Section 2.3 and insert an aperture only in thought – the physical result would be the same. On a screen behind the aperture, the intensity distribution must have been induced only by the diffraction at this fictitious aperture! We use the form and notation from Section 2.3 with the fictitious transmission

$$\tau(\xi) = e^{-(\xi/w_0)^2} / \sqrt{\pi}.$$

The diffraction integral

$$\mathcal{E}(x) = i\mathcal{E}_0 \frac{e^{ikr_0}}{\lambda \sqrt{kz}} \int_{-\infty}^{\infty} d\xi e^{i\kappa_x \xi} \frac{e^{-(\xi/w_0)^2}}{\sqrt{\pi}}$$

can be evaluated with

$$\int_{-\infty}^{\infty} d\xi \exp[-(\xi/w_0)^2] \exp(i\kappa_x \xi) = \sqrt{\pi} w_0 \exp[-(\kappa_x w_0/2)^2],$$

and we find, using the notion on p. 42 (beam waist  $w_0$ , the length of the Rayleigh zone  $z_0$ , etc.),

$$\mathcal{E}(x) = i\mathcal{E}_0 \frac{w_0 e^{ikr_0}}{\lambda \sqrt{kz}} e^{-(xz_0/w_0 z)^2} \simeq \mathcal{E}_0 \frac{w_0 e^{ikz}}{\lambda \sqrt{kz}} e^{ikx^2/2z} e^{-(x/w(z))^2}.$$

The last approximation is valid in the far field ( $z \gg z_0$ ), and we find after some conversions that it corresponds there exactly to the Gaussian TEM<sub>00</sub> mode from Section 2.3. However, one could have started the search for stable modes in a mirror or lens system also from the viewpoint of diffraction. The amplitude distribution must be a self-reproducing solution (or eigenfunction) of the diffraction integral, which is “diffraction limited.” However, integral equations are not very popular in teaching, which is why usually the complementary path of differential equations according to Maxwell is struck.

In our discussion we have treated the  $x$  and  $y$  coordinates completely independently from each other. That is why wave propagation according to Gaussian optics occurs independently in  $x$  and  $y$  directions, an important condition for optical systems, the axial symmetry of which is broken, for example, in ring resonators.

### 3. Fraunhofer diffraction at a circular aperture

One more element of diffraction, exceptionally important for optics, is the circular aperture, because diffraction occurs at all circle-like optical elements, among

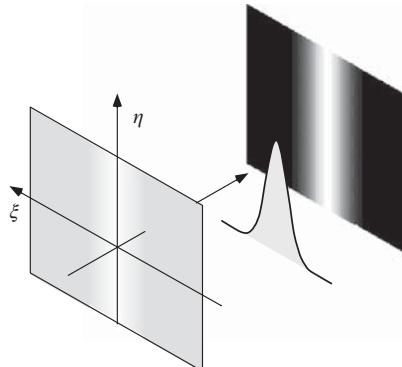
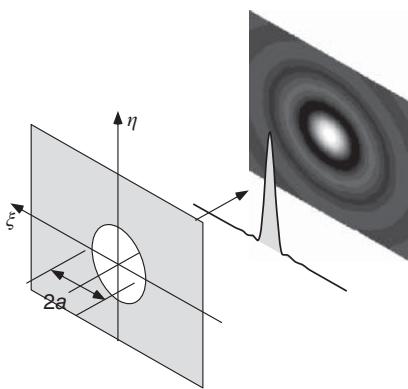


Figure 2.30 Diffraction at a “Gaussian transmitter.”

which lenses are counted, for example. We will see that the resolution of optical instruments is limited by diffraction at these apertures and that diffraction causes a fundamental limit for efficiency, the so-called diffraction limit. Only in recent years and with special methods involving nonlinear light matter interaction and computer analysis this limit has been overcome, see Section 5.5.3.

We introduce polar coordinates  $(\rho, \psi)$  within the  $(\eta, \xi)$  plane and  $(r, \phi)$  within the  $(x, y)$  plane of the screen. With these coordinates the diffraction integral from Eq. (2.69) reads as

$$\mathcal{E}(r) = -i\mathcal{E}_S \frac{e^{ikr_0}}{\lambda z} \int_0^a \rho d\rho \int_0^{2\pi} d\psi e^{-i(kr\rho/z)\cos(\phi-\psi)}.$$



**Figure 2.31** Diffraction at a circular aperture.

The intensity distribution is determined by forming the modulus:

$$I(r) = I(r=0) \left( \frac{2J_1(kar/z)}{kar/z} \right)^2. \quad (2.71)$$

The radius  $r_{\text{Airy}}$  of the Airy disc is defined by the first zero of the Bessel function  $J_1(x = 3.83) = 0$ . From  $x = kar_{\text{Airy}}/z = 3.83 = \pi \cdot 1.22$ , we find the radius

$$r_{\text{Airy}} = 1.22 \frac{z\lambda}{2a}.$$

With these specifications we may already determine the Rayleigh criterion for a lens of diameter  $2a \rightarrow D$  and with focal length  $z \rightarrow f$ ,

$$r_{\text{Airy}} = 1.22 \frac{f\lambda}{D}, \quad (2.72)$$

which matches the result of the treatment of Gaussian beams except for small constant factors (see p. 46).

### 2.7.1 Optical Fourier Transformation, Fourier Optics

According to Eq. (2.69), in the far field, a diffracting object creates an amplitude distribution that corresponds to the Fourier transform of the complex amplitude distribution in the object plane and is a function of the spatial frequencies

This can be evaluated with the mathematical relations for Bessel functions:

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp[ix \cos(\psi)] d\psi$$

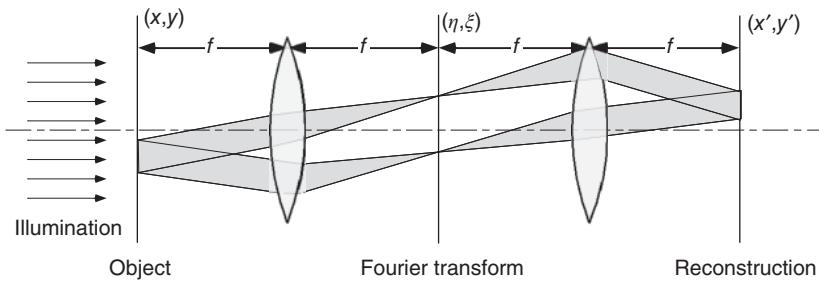
and

$$\int_0^x dx' x' J_0(x') = x J_1(x).$$

The result is

$$\mathcal{E}(r) = -i\mathcal{E}_S e^{ikr_0} \frac{ka^2}{z} \frac{J_1(kar/z)}{(kar/z)}.$$

The central diffraction maximum is also called the “Airy disc” ((Figure 2.31) do not confuse this with the Airy function!).



**Figure 2.32** A lens as an optical Fourier transformer. The image can be reconstructed by a second lens. The properties of the image can be manipulated in Fourier space, that is, the Fourier plane.

$\kappa_\eta = -k\eta/z$  and  $\kappa_\xi = -k\xi/z$ . A convex lens focuses incident beams and moves the Fourier transform of the amplitude distribution into the focal plane at the focal length  $f$  (Figure 2.32):

$$\begin{aligned}\mathcal{E}(\kappa_\eta, \kappa_\xi) &= \mathcal{A}(\eta, \xi) \int_S \tau(x, y) e^{i(\kappa_\eta x + \kappa_\xi y)} dx dy \\ &= \mathcal{A}(\eta, \xi) \mathcal{F}\{\mathcal{E}(x, y)\}.\end{aligned}$$

Therefore when observing a Fraunhofer diffraction image, one uses conveniently a lens (directly after the diffracting object) to keep the working distance short. It can be shown that the factor  $\mathcal{A}(\eta, \xi)$  is independent of  $(\eta, \xi)$  if the diffracting object is located at the front focal plane. Under these circumstances, studying the intensity distribution  $I(\eta, \xi) \propto |\mathcal{E}(\kappa_\eta, \kappa_\xi)|^2 \propto |\mathcal{F}\{\mathcal{E}(x, y)\}|^2$ , obviously one finds a power spectrum in the space frequencies of the diffracting object.

But the Fourier transformation of a diffracting object by a lens would not be that exciting if it were not for the fact that it forms the basis for the Abbe theory of imaging in a microscope (see Section 5.4.1.2) or more generally of *Fourier optics* [22]. The treatment of this goes beyond the scope of the present book, but, referring to Figure 2.32 and without going into details, we want to point out that a second lens compensates or reverses again the Fourier transformation of the first lens. Within the focal plane of the first lens, the Fourier plane, the image can now be manipulated. Just by use of simple diaphragms (amplitude modulation), certain Fourier components can be suppressed, and one obtains a smoothing of the images. On the other hand, one can also apply phase modulation, for example, by inserting glass retardation plates, which affect only selected diffraction orders. This procedure is also the basis for the phase contrast method in microscopy. Imaging can also include a magnification by application of lenses with different focal lengths.

## 2.8 Fresnel Diffraction

For Fraunhofer diffraction, not only must the screen lie in the far field, but also the size  $a$  of the radiation source must fit into the first Fresnel zone with radius

$r_0 = \sqrt{z\lambda}$ , which means that  $a \leq \sqrt{z\lambda}$  must be fulfilled. If this condition is not met, one may apply the Fresnel approximation (see Eq. (2.68(ii))), which for

$$a^2 \geq \lambda z \quad \text{but} \quad a \ll z$$

uses the full quadratic approximation in  $(x, y, \eta, \xi)$ :

$$\begin{aligned} r^2 &= (x - \eta)^2 + (y - \xi)^2 + z^2, \\ r &= z \left( 1 + \frac{(x - \eta)^2}{z^2} + \frac{(y - \xi)^2}{z^2} \right)^{1/2} \\ &= z + \frac{(x - \eta)^2}{2z} + \frac{(y - \xi)^2}{2z} + \dots . \end{aligned}$$

Then according to Eq. (2.67), the diffraction integral reads as

$$\mathcal{E}(\mathbf{r}_p) = i\mathcal{E}_0 \frac{e^{ikz}}{\lambda z} \oint_S \tau(\eta, \xi) \exp \left( \frac{ik}{2z} [(x - \eta)^2 + (y - \xi)^2] \right) d\eta d\xi. \quad (2.73)$$

Mathematically this is much more elaborate than the Fourier transformation in the Fraunhofer approximation (Eq. (2.69)) but easy to treat with numerical methods.

### Example: Fresnel diffraction

#### 1. Fresnel diffraction at a straight edge

First, we introduce the normalized variable  $u$ ,

$$\frac{k}{2z}(x - \eta)^2 := \frac{\pi}{2} u^2, \quad u_0 = u(\eta = 0) = \sqrt{\frac{k}{\pi z}} x, \quad d\eta = -\sqrt{\frac{\pi z}{k}} du,$$

into the diffraction integral and substitute ( $K$  is constant)

$$\begin{aligned} \mathcal{E}(x) &= K \int_0^\infty \exp \left[ \frac{ik}{2z}(x - \eta)^2 \right] d\eta \\ &\xrightarrow{x \rightarrow u} K \sqrt{\frac{\pi z}{k}} \int_{-\infty}^{u_0} \exp \left[ i \frac{\pi}{2} u^2 \right] du. \end{aligned}$$

At a large distance ( $x, u_0 \rightarrow \infty$ ) from the edge, we expect a homogeneous field and homogeneous intensity (Figure 2.33), which we can use for normalization:

$$I_0 = \frac{c\epsilon_0}{2} \mathcal{E}^2(x \rightarrow \infty) = \frac{c\epsilon_0}{2} \left| K \sqrt{\frac{\pi z}{k}} (1 + i) \right|^2 = \frac{c\epsilon_0}{2} K^2 z \lambda.$$

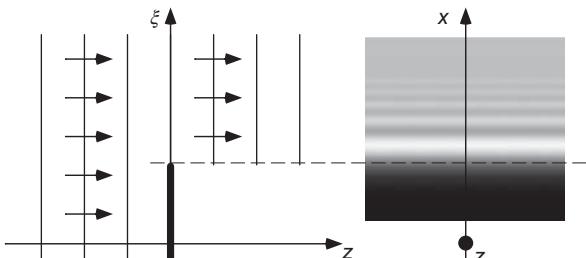
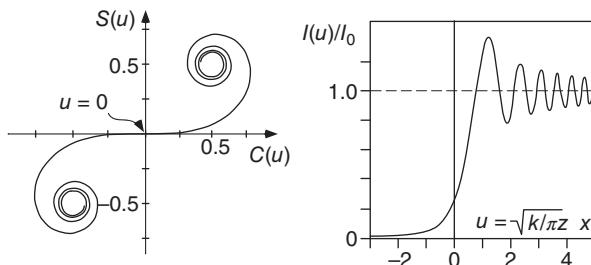


Figure 2.33 Fresnel diffraction at a straight edge.



**Figure 2.34** Cornu spiral and diffraction intensity behind a straight edge.

With that we can calculate the intensity, which can be expressed with the aid of the Fresnel integrals,

$$C(u) := \int_0^u du' \cos\left(\frac{\pi}{2}u'^2\right) \quad \text{and} \quad S(u) := \int_0^u du' \sin\left(\frac{\pi}{2}u'^2\right),$$

in a clear form:

$$\begin{aligned} I\left(x = \sqrt{\frac{\pi z}{k}} u_0\right) &= \frac{c\epsilon_0}{2} |\mathcal{E}(x)|^2 = \frac{1}{2} I_0 \left| \int_{-\infty}^{u_0} \exp\left[i\frac{\pi}{2}u^2\right] du \right|^2 \\ &= \frac{I_0}{2} \left\{ \left[ C(u_0) + \frac{1}{2} \right]^2 + \left[ S(u_0) + \frac{1}{2} \right]^2 \right\}. \end{aligned}$$

As a result we gain the Cornu spiral and the intensity distribution behind a straight edge, which are both illustrated in Figure 2.34.

## 2. Fresnel diffraction at a circular aperture

In order to evaluate the diffraction integral (2.73) for the case of near-field diffraction at a circular aperture with radius  $a$ , we use  $x = r \cos \phi'$ ,  $y = r \sin \phi'$ ,  $\eta = \rho \cos \phi$ , and  $\xi = \rho \sin \phi$ :

$$\begin{aligned} \mathcal{E}(r, \phi) &= -i\mathcal{E}_S \frac{e^{ikz} e^{ikr^2/2z}}{\lambda z} \\ &\times \int_0^a \int_0^{2\pi} e^{-ik\rho^2/2z} e^{-ir\rho \cos(\phi' - \phi)} \rho d\rho d\phi. \end{aligned} \tag{2.74}$$

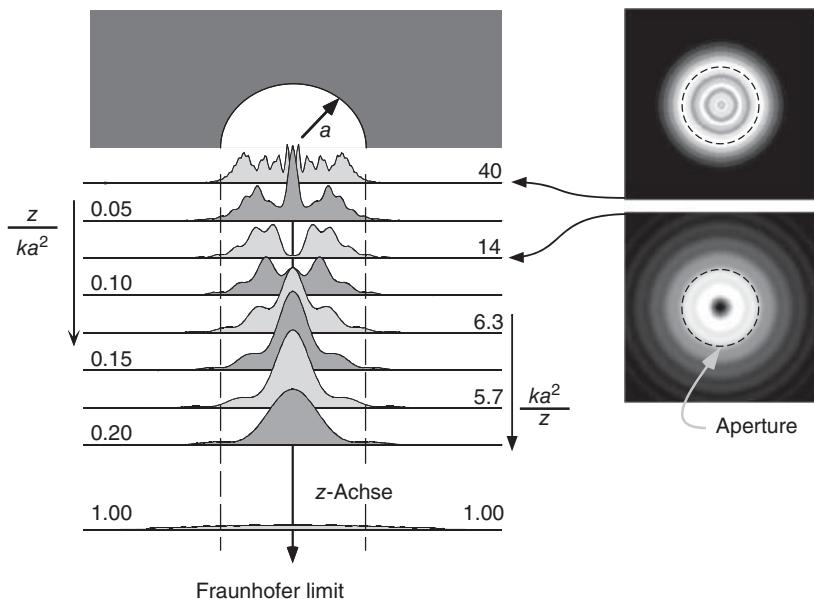
The angle integration can be carried out, and substituting  $\kappa := ka^2/z$ , it yields the expected radially symmetric result

$$\mathcal{E}(r) = -i\mathcal{E}_S e^{ikz} e^{ik(r/a)^2/2} \kappa \int_0^1 e^{-i\kappa x^2/2} J_0(\kappa xr/a) x dx.$$

Now the integral can be evaluated numerically and then yields the diffraction images from Figure 2.35. On the optical axis ( $r = 0$ ) the integral can also be solved analytically with the result

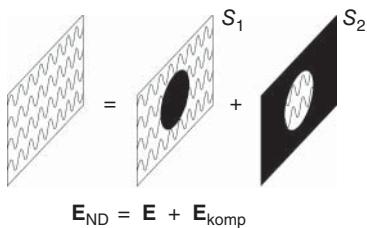
$$\begin{aligned} \mathcal{E}(r = 0) &= -i\mathcal{E}_S e^{ikz} 2 \sin(\kappa/4) e^{ik/4}, \\ I(r = 0) &= 4 \times \frac{c\epsilon_0}{2} |\mathcal{E}_S|^2 \sin^2(\kappa a^2/4z). \end{aligned} \tag{2.75}$$

Accordingly along the axis one finds up to fourfold intensity of the incident planar wave! For  $\kappa \ll 1$  the Fraunhofer approximation is reached, and there  $\sin(\kappa/4) \simeq \kappa/4 \propto 1/z$  is valid. In Section 2.8.2 we will interpret this result again with the aid of the Fresnel zones. Furthermore we will deal with the complementary problem, the circular obstacle.



**Figure 2.35** Example for Fresnel diffraction at a circular aperture from the Fresnel up to the Fraunhofer limit case. The right-hand figure shows the calculated intensity distribution at  $ka^2/z = 40$  and  $14$ .

### 2.8.1 Babinet's Principle



**Figure 2.36** Diffraction at a circular obstacle: Babinet's principle.

Babinet's principle is nothing other than an application of the superposition principle (Section 2.1.6). It often allows a clever formulation for the analysis of diffraction phenomena, because it is particularly also linear within the diffracting plane. If we consider the light field, which is created by the two geometries  $S_1$  and  $S_2$ , then the total field, which propagates without these objects, is just the sum of the two individual diffracting fields. According to

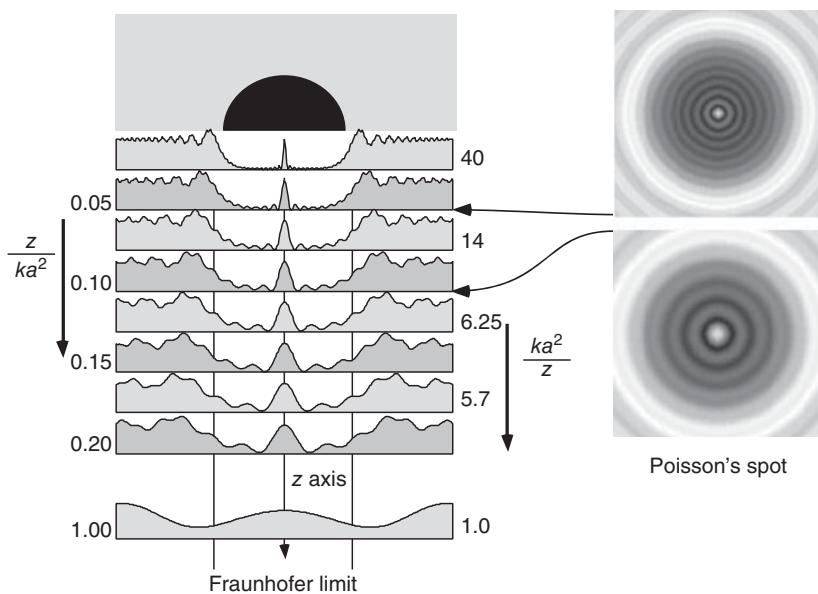
Figure 2.36, we can compose the non-diffracted field (index ND) out of the diffracted field and the corresponding complementary field:

$$\mathbf{E}_{\text{ND}}(\mathbf{r}_p) = \mathbf{E}(\mathbf{r}_p) + \mathbf{E}_{\text{comp}}(\mathbf{r}_p).$$

This statement, Babinet's principle, seems fairly banal at first sight, but it allows a clever treatment of complementary geometries.

#### Example: Circular obstacle

We can construct the light field diffracted by a disc with Babinet's principle and the result from a circular aperture. It consists of just the difference between the



**Figure 2.37** Fresnel diffraction at a circular obstacle. Right-hand side: calculated intensity distributions. In the center the “hot spot” can be recognized. Compare the complementary situation in Figure 2.35.

non-diffracted field, in the most simple case a planar wave, and the complementary field, which originates from a circular aperture:

$$\mathcal{E}(r) = \mathcal{E}_S e^{ikz} \left( 1 + i e^{i\kappa(r/a)^2/2} \kappa \int_0^1 e^{-i\kappa x^2/2} J_0(\kappa xr/a) x dx \right).$$

The diffraction image at a circular obstacle consists of the superposition of a planar wave and a diffraction wave of the circular aperture. In the center a bright spot can *always* be seen, which has become famous as the “hot spot” (Figure 2.37):

$$\mathcal{E}(r=0) = \mathcal{E}_S e^{ikz} [1 + 2i \sin(\kappa/4) e^{i\kappa/4}] \quad \text{and} \quad I(r=0) = \frac{c\epsilon_0}{2} |\mathcal{E}_S|^2.$$

According to an anecdote, Poisson opposed Fresnel diffraction theory on the grounds that the just achieved results were absurd; behind an aperture in the center of the diffraction image, a constant hot spot could not be observed, only to be disproved by experiment. Today this observation is not very difficult; a simple ball bearing (manufactured with optical precision, i.e., with only slight deviations in the micrometer range) glued onto a glass plate and held into a laser beam demonstrates Poisson’s spot nicely.

### 2.8.2 Fresnel Zones and Fresnel Lenses

In the case of Fraunhofer diffraction, we can equate the Fresnel factor from Eq. (2.74),  $\exp[-ik(x^2 + y^2)/2z]$ , with 1 according to  $ka^2/z = 2\pi(a^2/\lambda z) \ll 1$ , but this is not the case for Fresnel diffraction. This factor specifies with what kind of phase

$\Phi_F$  the partial waves of the diffracting area contribute to the interference image, for example, all with approximately  $\Phi_F = 0$  in the Fraunhofer limit case.

However, if we slowly increase the radius  $r$  of the diffracting object at a fixed distance  $z$ , then starting at  $a_1 = \sqrt{z\lambda}$  the partial waves contribute with opposite phase, since  $ka_1^2/z = \pi$ . Therefore we can refer to the criterion established by Fresnel,

$$a_N^2 = Nz\lambda, \quad (2.76)$$

to divide the diffracting plane according to the character of its phase position.

In Figure 2.38 the division with white and black zones is introduced, the outer radii of which increase according to Eq. (2.76). For clarification we look again at the diffraction at a circular aperture from the example on p. 73. According to Eq. (2.75) the brightness reaches a maximum on the axis at  $a^2/z\lambda = 1, 3, \dots$ , while at  $a^2/z\lambda = 2, 4, \dots$  a minimum appears.

In a radially symmetric aperture, every Fresnel zone contributes with the same area and intensity to the total field on the axis. Partial waves stemming from the odd Fresnel zones accumulate a path difference of  $(N - 1)\lambda/2 = 0, 2, 4, \dots, \lambda$  on the axis, which results in constructive interference. On the other hand, a contribution with opposite phase is created from the even zones ( $N\lambda/2$ ), which results in cancellation of the light field for equivalent numbers of even and odd zones.

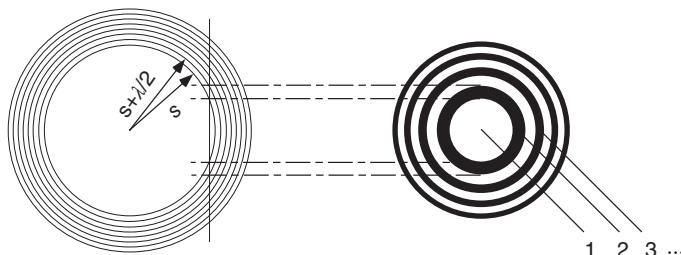


Figure 2.38 Fresnel zones and zone plate.

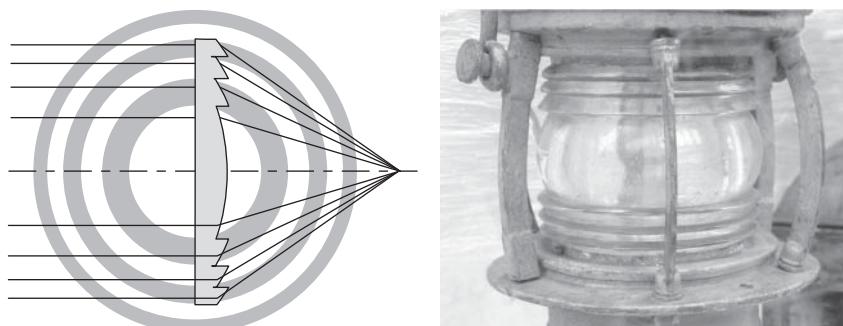


Figure 2.39 Fresnel step lens: scheme and ring-shaped application in an old boat lantern. (Anton, [https://commons.wikimedia.org/wiki/File:Fresnel\\_lampe.jpg](https://commons.wikimedia.org/wiki/File:Fresnel_lampe.jpg). Used under CC-BY-SA 3.0. <https://creativecommons.org/licenses/by-sa/3.0/deed.en>.)

The suggestion to make use of this circumstance and use a diaphragm for every second zone dates back to Fresnel. The division into zones from Figure 2.38 therefore stands exactly for the idea of a Fresnel zone plate. Alternatively one may also use a corresponding phase plate, which is better known under the name “Fresnel lens” or “Fresnel step lens” (Figure 2.39). These lenses are often used in combination with large apertures, for example, in overhead projectors.

## 2.9 Beyond Gaussian Beams: Diffraction Integral and ABCD Formalism

In applications, for example, for the creation of OAM, coherent light beams are used with increasing complexity beyond conventional Gaussian beams. In the paraxial approximation, the propagation of arbitrarily shaped beams can still be treated analytically in terms of an integral found by Collins [3, 23], which makes use of the ABCD matrix formalism. In Cartesian coordinates it is

$$\begin{aligned}\mathcal{E}(x, y, z) = & -\frac{i}{\lambda B} e^{ikz} \int \int dx' dy' \mathcal{E}_0(x', y') \\ & \times \exp \left[ \frac{ik}{2B} (A(x'^2 + y'^2) + D(x^2 + y^2) - 2(xx' + yy')) \right],\end{aligned}\quad (2.77)$$

and in cylindrical coordinates

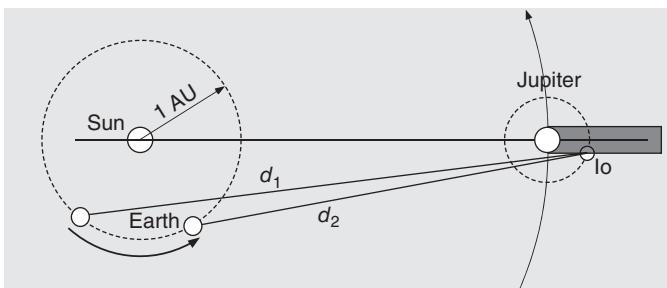
$$\begin{aligned}\mathcal{E}(x, y, z) = & -\frac{i}{\lambda B} e^{ikz} \int \int \rho' d\rho' d\phi' \mathcal{E}_0(\rho', \phi') \\ & \times \exp \left[ \frac{ik}{2B} (A\rho'^2 + D\rho^2 - 2(\rho\rho' \cos \phi - \rho')) \right].\end{aligned}\quad (2.78)$$

Note that for the simple ABCD propagation matrix (Table 1.2) with  $A = D = 1$ ,  $B = z$ , and  $C = 0$ , the Fresnel integral (2.73) is recovered. The analytic description of the twisted beams in Figure 2.22 is based on the Collins integral.

## Problems

### 2.1 Measurement of the speed of light according to Røemer and Huygens

O. Røemer observed variations in the orbital period of the innermost Jupiter moon Io ( $T_{\text{Io}} = 1.8$  days or 42.5 h) (Figure 2.40). He used these data to derive the time it takes light to traverse the orbit of the Earth around the sun (1 astronomical unit, 1 AU =  $150 \times 10^6$  km). C. Huygens calculated the first estimate of the speed of light from Rømer's observations. Examine Earth's orbit and specify which event can be observed from the Earth, entry or exit of Io into the shade of Jupiter? Where do you expect the largest deviations of the observed from the “true” orbital period? What accuracy is required for a clock such that the acceleration/deceleration can be measured within two revolutions? Rømer compared 40 revolutions before and 40 revolutions after the opposition (sun between Earth and



**Figure 2.40** O. Røemer gained the first insight about the travel time of light from the time variations he observed between the eclipses of the Jupiter moon Io.

Jupiter). He concluded that light travels across the diameter of Earth's orbit in 22 min. What value did Huygens estimate for the speed of light?

**2.2 Field distributions and Poynting vector of electromagnetic waves**

Sketch the electric and magnetic field distributions and the Poynting vector for the superposition of two plane and linearly polarized waves for (a) a plane standing wave and (b) two orthogonally crossed waves.

**2.3 Image dipoles and optical quadrupole radiation** In the vicinity of conducting planes, one can derive many properties of a radiating dipole using the superposition of its dipole field with the field of an image dipole, which is located at the mirror position of the original dipole. These arguments also hold for atomic dipoles radiating visible light. Examine the two possible orientations in Section 14.3.3 (Figure 14.5 ( $\sigma$  and  $\pi$  orientation orthogonal and parallel to the surface normal)) and assure yourself that the dipoles must have parallel and antiparallel orientation, respectively, in order to satisfy the boundary conditions of a vanishing electric field at the conductive surface. Give the far-field ( $r \gg \lambda$ ) spatial distribution of the radiation field strength as a function of  $r$ , the distance from the center of the two dipoles, and  $\theta$ , the angle with the surface normal. Assume the limiting case where the separation of the dipoles is small compared with the wavelength  $\lambda$ . How does the amplitude drop with  $r$  for the two orientations?

**2.4 Paraxial wave equation** Verify Eqs. (2.35) and (2.36) by explicitly carrying out the slowly varying envelope approximation.

**2.5 Laguerre–Gaussian modes** For cylindrically symmetric paraxial waves, an alternative description is given by Laguerre–Gaussian modes. In place of Eq. (2.41) the amplitude is in this case given with the definitions for  $w_0$ ,  $w(z)$ ,  $R(z)$ , and  $\eta(z)$  from Section 2.3.1.1 by

$$E_{\ell m}(\rho, \varphi, z) = E_0 \left( \frac{\sqrt{2}\rho}{w(z)} \right)^{\ell} L_m^{\ell} \left( \frac{\sqrt{2}\rho}{w(z)} \right) \frac{w_0}{w(z)} e^{-(\rho/w(z))^2} \\ \times e^{i\ell\varphi} e^{ik\rho^2/2R(z)} e^{i(kz - (2m+\ell+1)\eta(z))}.$$

Find out about the properties of the Laguerre polynomials (e.g., in [24]), and sketch the transverse intensity and phase angle distribution of these modes.

- 2.6 Electromagnetic fields with spin and orbital angular momentum components** Show that the volume integral

$$\mathbf{J} = \frac{1}{c^2} \int_V \mathbf{r} \times (\mathbf{E} \times \mathbf{H}) dV$$

can be decomposed into two contributions:

$$\mathbf{J} = \epsilon_0 \int_V (\mathbf{E} \times \mathbf{A}) + \sum_i E_i (\mathbf{r} \times \nabla) A_i dV.$$

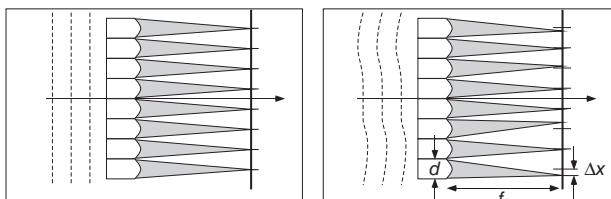
Replace  $\mathbf{H}$  by the vector potential  $\mathbf{H} = \mu_0^{-1} \nabla \times \mathbf{A}$ : show Component-wise that  $\mathbf{r} \times [\mathbf{E} \times (\nabla \times \mathbf{A}) + (\mathbf{E} \cdot \nabla) \mathbf{A}] = \sum_i E_i (\mathbf{r} \times \nabla) A_i$ . Take into account that the fields are bounded, that is, vanish at infinity.

- 2.7 Hartmann–Shack sensors: Wave front analysis without interferometry** The phase (wave front) of an optical wave cannot be directly measured. The Hartmann–Shack sensors consist of an array (rectangular or radial) of microlenses focusing portions of the incoming beam onto a camera (Figure 2.41). Explain the functioning of the sensor. How is the wave front deviation extracted from the measured deviations of the focal spots?

- 2.8 Microwave Gaussian beam** A typical satellite dish has a diameter of 50 cm and a focal length of 25 cm. Use Gaussian beam optics to estimate the diameter of the focal spot, which is created by the microwave radiation emitted by the Astra satellite at 11 GHz.

- 2.9 Intensity of microwave and optical beams** Let us send an 11 GHz signal with the antenna from Problem 2.8 to the Astra satellite. The power of the transmitter is 1 W. Use Gaussian beam optics to calculate the maximum intensity arriving at the satellite. Furthermore, estimate the intensity if we use a 1 W helium–neon laser ( $\lambda_{\text{He-Ne}} = 632 \text{ nm}$ ) instead of the microwaves.

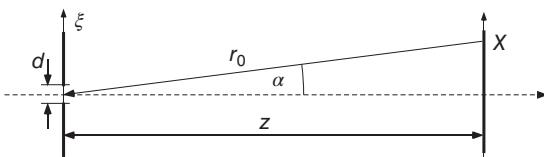
- 2.10 Gouy phase** Devise an experiment that proves the existence of the Gouy phase. (Hint: Phase relations are typically investigated with interferometric devices.) Is there an analogue of the Gouy phase in ray optics?



**Figure 2.41** Concept of the Hartmann–Shack wave front sensor.

- 2.11 Coupling into optical fibers** For an optical fiber a numerical aperture  $\text{NA} = 0.1$  is specified for  $\lambda = 850 \text{ nm}$ . How large is the diameter of the guided beam ( $2w_0$ ,  $w_0$  waist size) if we assume an approximate Gaussian profile? You want to insert a well-collimated beam with half-width 2 mm and divergence 1 mrad with a lens into this fiber. There are lenses with 10, 5, 2, and 1 cm focal lengths available. Which lens gives the best result, and at which position from the end of the fiber?
- 2.12 Axes of a polarizer** The axes of an unknown polarizer (e.g., polarizing sheets) can be determined in the following way: find a ground surface as smooth as possible in your vicinity, that is, at least with some residual reflection. Take the polarizer and observe this reflection at a distance of some 2–4 m. By rotating the polarizer, the reflection from the surface should become lighter and darker. What is the origin of the polarization, and how do you identify the axes?
- 2.13 Measuring Stokes parameters** Model the polarization elements of Figure 2.19 by means of Jones matrices, and identify the Stokes parameters.
- 2.14 Polarization and reflection** Determine the local state of polarization of the electric field from the following pairs of light beams, which are counter-propagating in the  $+z$  and  $-z$  direction, respectively, as a function of  $z$ :
- lin||lin:  $E_{+z}^{\text{lin}} \parallel E_{-z}^{\text{lin}}$ ;
  - lin $\perp$ lin:  $E_{+z}^{\text{lin}} \perp E_{-z}^{\text{lin}}$ ;
  - $\sigma^+ \sigma^+$ :  $E_{+z}^{\text{circ}+}, E_{-z}^{\text{circ}+}$ ;
  - $\sigma^+ \sigma^-$ :  $E_{+z}^{\text{circ}+}, E_{-z}^{\text{circ}-}$ .
- Which optical components are used if the counter-propagating beam is generated by retroreflection from a plane mirror?
- 2.15 Projection and rotation of the polarization** Two perfectly crossed polarizers extinguish a light beam. Insertion of another polarizer in between the crossed polarizers causes transmission of light again (see Figure 2.20). Show that the transmitted intensity is 25% for lossless polarizers. Continue the example by insertion of 2, 3, ... polarizers with equal angle settings  $30^\circ$ ,  $22.5^\circ$ , and so on.
- 2.16 Fraunhofer diffraction from simple and irregular openings** What is the diffraction image of a square aperture in the far field (Fraunhofer limit)? How is the image changed for two crossed slits? What is the influence of irregular openings, for example, punched letters?
- 2.17 Single slit: Partial waves** A plane wave (wave vector  $\mathbf{k} = 2\pi/\lambda$ ) is incident on an infinitely long slit of width  $d$ . We are interested in the intensity distribution, which we observe on a screen at the distance  $z$  from the slit

**Figure 2.42** Designations for Problems 17–20, diffraction from a slit.



(Figure 2.42). Here we assume that the Fraunhofer limit is realized, that is,  $d^2 \gg \lambda z / \pi$ . Calculate the position of the first diffraction minimum with the following method: divide the slit into two half-slits with identical width. For every point on the screen, we now calculate the difference of the average phase of the two beams from the half-slits. The partial waves annihilate each other if the phase difference equals  $\pi$ .

- 2.18 Single slit: Huygens' principle** A more realistic description than the previous problem is derived from Huygens' principle: it states that every point of the slit is the origin of a spherical wave into the half-space behind the slit (Figure 2.42). In our case we can for symmetry reasons replace the spherical wave with a cylindrical wave  $\mathcal{E}(\rho) \propto \exp(ik\rho)/\sqrt{\rho}$ , where  $\rho$  gives the distance from the origin of the cylindrical wave. The field amplitude at the screen is derived from the sum of all partial waves. Calculate with this method the intensity distribution of the diffraction image of the slit as a function of the diffraction angle  $\alpha \ll 1$ .
- 2.19 Double slit: Babinet's principle** Construct the diffraction image of the double slit using Babinet's principle using two single slits with different widths  $d_1 > d_2$ .
- 2.20 Kirchhoff's integral** Kirchhoff's diffraction integral allows the calculation of the field distribution caused by diffraction from an arbitrary object. In the far field of a diffracting slit (Figure 2.42), Kirchhoff's integral is reduced to a one-dimensional integral by taking advantage of the Fraunhofer approximation and the translational symmetry along the slit. With  $\mathcal{E}_S$  the amplitude of the incident plane wave at the slit and  $\kappa_X := -kX/r_0$  ( $\approx ka$  for  $\alpha \ll 1$ ), we have
- $$\mathcal{E}(X) = -\frac{i\mathcal{E}_S \exp(ikr_0)}{\lambda \sqrt{kz}} \int_{-\infty}^{\infty} \tau(\xi) \exp(i\kappa_X \xi) d\xi.$$
- a) Give the form of the transmission function  $\tau(\xi)$  for the diffraction slit.  
b) Calculate  $\mathcal{E}(X)$  using the formula for the case  $\alpha \ll 1$ .
- 2.21 Diffraction from a thin wire** Calculate by means of Babinet's principle the field distribution that is caused by diffraction of a plane wave from a thin wire in the far field. What is changed if the wire is replaced with a thickness of  $\ell_{\text{opt}} = (n_{\text{glass}} - 1)\ell = \lambda/2$ ?

## 3

# Light Propagation in Matter: Interfaces, Dispersion, and Birefringence

We have seen that we can describe refraction of light rays at dielectric interfaces, such as glass plates, with the help of refractive indices introduced phenomenologically. We may also consider refraction as the response of the glass plate to the incident electromagnetic light wave. The electric field shifts the charged constituents of the glass and thus causes a dynamic polarization. This in turn radiates an electromagnetic wave and acts back on the incident light wave through interference. Here we will discuss the properties of matter with macroscopic phenomenological indices of refraction. Some fundamental relations with the microscopic theory will be introduced in Chapter 7.

In the preceding chapter (Sections 1.5 and 2.1.7), we discussed wave propagation in vacuum and noticed that propagation in homogeneous matter differs only by the phase velocity  $c/n$  with  $n$ , the refractive index:

$$\left( \nabla^2 - \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{E}(\mathbf{r}, t) = 0. \quad (3.1)$$

Now we want to explore how interfaces or dielectrics with anisotropic and frequency-dependent refractive index affect the propagation of electromagnetic waves.

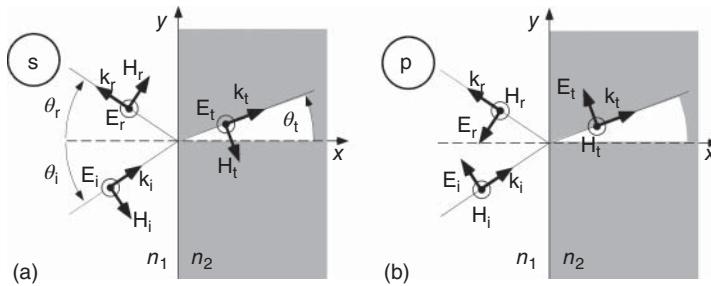
## 3.1 Dielectric Interfaces

In order to discuss dielectric interfaces, we have to know how they affect electromagnetic fields. We will only cite the relations important for optics and leave it to the reader to consult textbooks on electrodynamics for proofs of the rules (of mathematical boundary conditions) with the help of Maxwell's equations (2.10).

Suppose that an interface divides two media with refractive indices  $n_1$  and  $n_2$  and with normal unit vector  $\mathbf{e}_N$ . Then the electromagnetic radiation fields are fully characterized by

$$\begin{aligned} (\mathbf{E}_2 - \mathbf{E}_1) \times \mathbf{e}_N &= 0 & \text{and} & \quad (n_2^2 \mathbf{E}_2 - n_1^2 \mathbf{E}_1) \cdot \mathbf{e}_N = 0, \\ (\mathbf{H}_2 - \mathbf{H}_1) \times \mathbf{e}_N &= 0 & \text{and} & \quad (\mathbf{H}_2 - \mathbf{H}_1) \cdot \mathbf{e}_N = 0, \end{aligned} \quad (3.2)$$

where  $\mathbf{E}_{1,2}$  and  $\mathbf{H}_{1,2}$  are to be taken in direct proximity to, but on different sides of, the interface. We further note that in optics we may often restrict ourselves



**Figure 3.1** Electromagnetic fields at a dielectric interface for (a) s and (b) p polarization. The symbol  $\odot$  indicates the field vectors perpendicular to the plane of the drawing.

to the application of the vector products from (3.2), while the scalar products are accounted for by Snell's law (Eq. (1.2)).

### 3.1.1 Refraction and Reflection at Glass Surfaces

In the case of a transverse electromagnetic wave incident on a dielectric interface, we can distinguish two polarization configurations: the polarization may be either linearly perpendicular (s) or parallel (p) to the plane of incidence (Figure 3.1).

Waves with s (resp. p) polarization of the electric field are called s (resp. p) waves. Alternatively the notions  $\sigma$  and  $\pi$  polarizations (resp. TE and TM waves) are also used. We have to treat the two cases for each component individually. Then elliptical polarizations can be reduced to superpositions of these cases according to the superposition principle.

#### 3.1.1.1 s Polarization

We consider the  $\{\mathbf{E}, \mathbf{H}, \mathbf{k}\}_\alpha$  triads of the incident (i), reflected (r), and transmitted (t) waves and use the notation from Figure 3.1 with

$$\begin{aligned} \mathbf{E}_\alpha &= \mathcal{E}_{0\alpha} \mathbf{e}_z e^{-i(\omega_\alpha t - \mathbf{k}_\alpha \cdot \mathbf{r})}, \\ \mathbf{H}_\alpha &= \frac{\mathcal{E}_{0\alpha}}{\mu_0 c \omega_\alpha} \mathbf{k}_\alpha \times \mathbf{e}_z e^{-i(\omega_\alpha t - \mathbf{k}_\alpha \cdot \mathbf{r})}. \end{aligned} \quad (3.3)$$

The s-polarized electric field is perpendicular to the surface normal, which is why

$$\mathbf{E}_t = \mathbf{E}_i + \mathbf{E}_r \quad (3.4)$$

is valid. If this relation is fulfilled everywhere and at all times at the interface, then obviously all waves must have the same frequency, and we can consider the time  $t = 0$ . Besides, according to (3.2), at  $x = 0$ , the relation

$$\mathcal{E}_{0t} e^{ik_{yt} y} = \mathcal{E}_{0i} e^{ik_{yi} y} + \mathcal{E}_{0r} e^{ik_{yr} y}$$

must hold for arbitrary  $y$ , and thus all  $y$  components of the  $\mathbf{k}$  vectors must be equal:

$$k_{yt} = k_{yi} = k_{yr}.$$

Next we consider the components individually for the reflected and transmitted parts. Since the reflected wave propagates within the same medium as the incident wave, according to  $n_1^2 k_\alpha^2 = n_1^2 (k_{x\alpha}^2 + k_{y\alpha}^2)$ , the relations

$$k_{xr}^2 = k_{xi}^2 \quad \text{and} \quad k_{xr} = -k_{xi}$$

must be satisfied, because the positive sign creates one more incident wave, which is not physically meaningful. Thus the law of reflection (1.1) is again established. For the transmitted wave,  $k_t/n_2 = k_i/n_1$  must hold. From geometry one finds directly  $k_\alpha = k_{y\alpha}/\sin \theta_\alpha$  and thus also Snell's law (1.2) again,

$$n_1 \sin \theta_i = n_2 \sin \theta_t.$$

This is valid only for real refractive indices, but it can be generalized by the application of

$$k_{xt}^2 = k_t^2 - k_{yt}^2 = \frac{n_2^2}{n_1^2} k_i^2 - k_{yi}^2.$$

All results up to now have just confirmed the outcomes we knew already from ray optics. But by means of ray optics we could not find out about the amplitude distribution, which is now possible by means of wave optics. According to (3.3), the tangential components of the **H** field are related to the **E** components:

$$H_{y\alpha} = -\frac{\mathcal{E}_{0\alpha}}{\mu_0 c \omega} k_{x\alpha}.$$

These must be continuous due to (3.2) and therefore fulfill the equations

$$\begin{aligned} k_{xt} \mathcal{E}_{0t} &= k_{xi} \mathcal{E}_{0i} + k_{xr} \mathcal{E}_{0r} = k_{xi} (\mathcal{E}_{0i} - \mathcal{E}_{0r}), \\ \mathcal{E}_{0t} &= \mathcal{E}_{0i} + \mathcal{E}_{0r}, \end{aligned} \tag{3.5}$$

which we have extended by the condition (3.4) to complete the system of equations. It has the solutions

$$\mathcal{E}_{0r} = \frac{k_{xi} - k_{xt}}{k_{xi} + k_{xt}} E_{0i} \quad \text{and} \quad \mathcal{E}_{0t} = \frac{2k_{xi}}{k_{xi} + k_{xt}} E_{0i}.$$

With the amplitudes, the corresponding intensities can be calculated without any problems. The *reflection coefficient r* and the *transmission coefficient t* may also be described according to

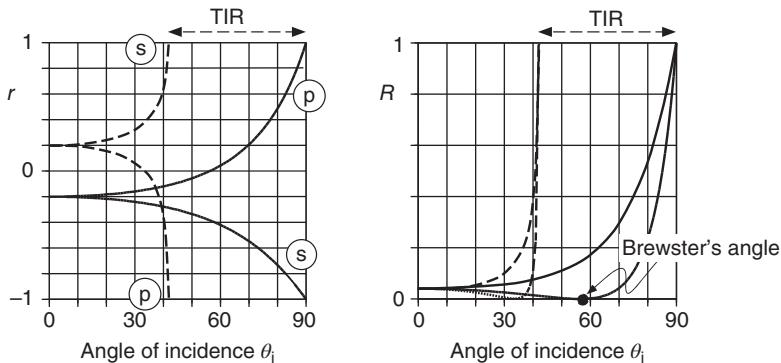
$$r = \frac{\mathcal{E}_{0r}}{\mathcal{E}_{0i}} = \frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_i + n_2 \cos \theta_t},$$

$$t = \frac{\mathcal{E}_{0t}}{\mathcal{E}_{0i}} = \frac{2n_1 \cos \theta_i}{n_1 \cos \theta_i + n_2 \cos \theta_t},$$

and by the use of  $n_1/n_2 = \sin \theta_t/\sin \theta_i$  according to Snell's law, these can be modified to yield

$$r = \frac{\mathcal{E}_{0r}}{\mathcal{E}_{0i}} = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \quad \text{and} \quad t = \frac{2 \cos(\theta_i) \sin(\theta_t)}{\sin(\theta_i + \theta_t)}.$$

The dependence of the reflection coefficient and the reflectivity on the angle of incidence  $\theta_i$  is illustrated in Figure 3.2. Among other things, the figure shows



**Figure 3.2** Reflection coefficient and reflectivity at a glass plate with refractive index  $n = 1.5$  for s and p polarization. Full line: from vacuum into glass. Dashed line: from glass into vacuum. TIR: region of total internal reflection; see Section 3.1.2.

the change of sign of the reflectivity coefficient for the reflection at a more dense medium, which at normal incidence corresponds to a phase jump by  $180^\circ$ .

A very important special case occurs when light enters perpendicularly, that is, at an angle  $\theta_i = 0^\circ$ . Then for the *reflectivity*  $R = I_r/I_i$  and *transmission*  $T$  with  $R + T = 1$ , the Fresnel formulae are valid (cf. 2.17):

$$R = \frac{|\mathbf{E}_r|^2}{|\mathbf{E}_i|^2} = \left( \frac{n_1 - n_2}{n_1 + n_2} \right)^2 \quad \text{and} \quad T = \frac{n_2 |\mathbf{E}_t|^2}{|\mathbf{E}_i|^2} = \frac{4n_1 n_2}{(n_1 + n_2)^2}. \quad (3.6)$$

It is straightforward to calculate that, at a glass-air interface ( $n_1 = 1$ ,  $n_2 = 1.5$ ), 4% of the intensity is reflected.

### 3.1.1.2 p Polarization

The discussion of a p-polarized electric field oscillating within the plane of incidence follows the procedure just outlined and can therefore be shortened, giving the conditions and the results only. Snell's law is reproduced, and for the amplitudes one the system of equations

$$\begin{aligned} k_t \mathcal{E}_{0t} &= k_i \mathcal{E}_{0i} + k_r \mathcal{E}_{0r} \\ k_i \mathcal{E}_{0t} &= k_t (\mathcal{E}_{0i} - \mathcal{E}_{0r}) \end{aligned} .$$

Solutions read

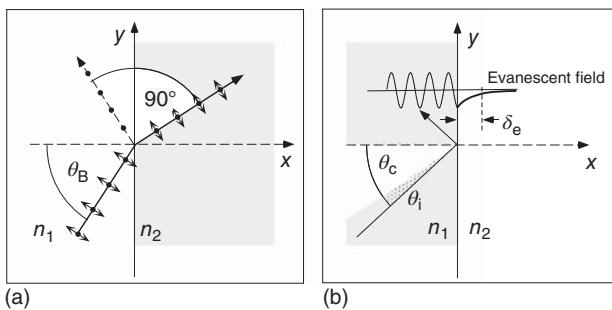
$$\mathcal{E}_{0r} = \frac{k_t^2 - k_i^2}{k_t^2 + k_i^2} \mathcal{E}_{0i} \quad \text{and} \quad \mathcal{E}_{0t} = \frac{2k_i k_t}{k_t^2 + k_i^2} \mathcal{E}_{0i} .$$

The reflection coefficient of the p wave is more transparently written as

$$r = \frac{\mathcal{E}_{0r}}{\mathcal{E}_{0i}} = -\frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)}$$

and is shown together with the reflectivity in Figure 3.2. It vanishes for

$$\theta_i - \theta_t = 0 \quad \text{and} \quad \theta_i + \theta_t = \pi/2.$$



**Figure 3.3** (a) At the Brewster angle  $\theta_B$ , only s-polarized light is reflected. (b) For total reflection at the denser medium ( $n_1 > n_2$ ), an evanescent wave field is generated in the less dense medium.

The first condition is only fulfilled trivially for  $n_1 = n_2$ . The second one leads to the *Brewster condition*

$$\frac{n_2}{n_1} = \frac{\sin \theta_B}{\sin \theta_t} = \frac{\sin \theta_B}{\sin(\pi/2 - \theta_B)} = \tan \theta_B,$$

which yields the *Brewster angle*  $\theta_B = 57^\circ$  for the glass–air transition ( $n = 1.5$ ). The Brewster condition may be interpreted physically with the angular distribution of the dipole radiation (see Section 2.2.3 and Figure 3.3): the linear dielectric polarization in the refracting medium is transverse to the refracted beam and cannot radiate into the direction of the reflected wave if the former makes a right angle with the refracted wave.

### 3.1.2 Total Internal Reflection (TIR)

As shown in Figure 3.2, total internal reflection (TIR) is a very prominent feature for light waves traveling from a denser medium into a less dense one. Let us study this situation in more detail by considering the component  $k_{xt} = k_2 \cos \theta_t$  describing the penetration of the wave into the less dense medium. We adopt the solutions for propagating waves below the critical angle  $\theta_c = \sin^{-1}(n_2/n_1)$ , now  $n_1 > n_2$ , which we know already from Eq. (1.3), by generalization of Snell's condition for  $\theta_i > \theta_c$  to imaginary values. With  $W = \sin \theta_t = \sin \theta_i / \sin \theta_c > 1$ , one may write

$$\cos \theta_t = (1 - \sin^2 \theta_t)^{1/2} = (1 - W^2)^{1/2} = iQ,$$

where  $Q$  is again a real number. Now we write the electric field in the thin medium  $\mathbf{E}_{eva}$  for angles of incidence larger than the critical angle as a propagating wave:

$$\mathbf{E}_{eva}(\mathbf{r}, t) = \mathbf{E}_{20} \exp[-i(wt - \mathbf{k}_2 \cdot \mathbf{r})].$$

With  $\mathbf{k}_2 = k_2(\cos \theta_t \mathbf{e}_x + \sin \theta_t \mathbf{e}_y)$ , one gets

$$\mathbf{E}_{eva} = \mathbf{E}_{20} \exp(-k_2 Qx) \exp[-i(wt - k_2 Wy)].$$

Thus for  $\theta_i > \theta_c$ , there is a wave propagating along the interface that penetrates into the denser medium, but is attenuated exponentially with penetration depth

$\delta_e = 1/(k_2 Q)$  (Figure 3.3). The wave within the less dense medium usually is called the *evanescent wave field*.

### Example: Penetration depth and energy transport for TIR

According to the preceding section, the penetration depth of a totally reflected wave is for  $n_2 = 1 < n_1$  and  $\theta_i > \theta_c$

$$\delta_e = \frac{1}{k_2 Q} = \frac{\lambda/2\pi}{\sqrt{n_1^2 \sin^2 \theta_i - 1}}.$$

For the case of a 90° prism from Figure 1.7 (angle of incidence 45°, refractive index  $n_1 = 1.5$ ), one calculates  $Q = 0.25$  and  $\delta_e = 0.27 \mu\text{m}$  at 600 nm.

It is instructive to consider the energy transport according to Eq. (2.18) through the interface into the evanescent wave. It turns out that the normal component of the Poynting vector is purely imaginary,

$$\begin{aligned}\langle \mathbf{S} \cdot \mathbf{e}_N \rangle &= \langle \mathbf{E} \times \mathbf{H} \cdot \mathbf{e}_N \rangle \\ &= \Re \{ c e_0 / 2 |\mathbf{E}|^2 i Q \} = 0,\end{aligned}$$

and therefore no energy transport occurs through the interface. Actually this situation changes if we position a second interface nearby, as indicated in Figure 3.4. Then the so-called frustrated total internal reflection (FTIR) occurs. The FTIR phenomenon is used not only to build optical beam splitters but also to couple light in different manners (by varying the air gap) into wave guides (see Figure 4.2) or monolithic optical resonators, or, for example, to perform spectroscopy in the immediate vicinity of a surface.

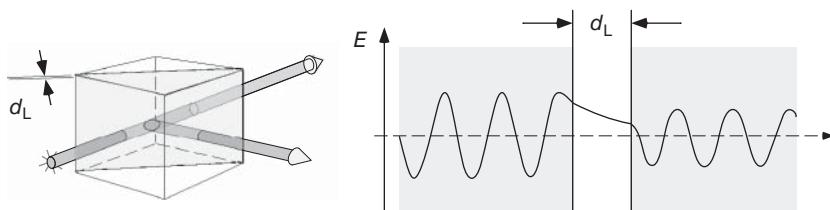
### 3.1.3 Complex Refractive Index

So far, we have considered real indices of refraction  $n$ , which are a good approximation for media with negligible losses. Absorption or other losses, for example, scattering, may be taken into account phenomenologically by the generalization of the refractive index to a complex quantity:

$$n = n' + i n''.$$

Then in a homogeneous medium wave propagation may be described according to

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{-i(wt - n' \mathbf{k} \mathbf{r})} e^{-n'' \mathbf{k} \mathbf{r}},$$



**Figure 3.4** Frustrated total internal reflection. The width of the air gap must be less than the penetration depth of the evanescent wave.

where obviously  $\alpha = 2n''k_z$  specifies the attenuation of the intensity ( $I \propto |\mathbf{E}|^2$ ), here for propagation along the  $z$  direction:

$$I(\mathbf{r}) = I(0) \exp(-\alpha z) = I(0) \exp(-2n''k_z z). \quad (3.7)$$

## 3.2 Interfaces of Conducting Materials

Nowadays dielectric multilayer mirrors are usually used for laser applications (see Section 6.7). But conventional mirrors made of evaporated metal layers also still play an important role in “everyday optical technology” due to their low price and broadband effect. Metals are characterized by enormous conductivity, which also causes their high reflectivity. We consider a classical phenomenological model for the conductivity  $\sigma$ , which goes back to Paul Drude (1863–1906). It has been shown to be extremely powerful and was later more fundamentally justified by a microscopic solid-state quantum theory. In the Drude model the motion of a free electron is attenuated by friction forces with an attenuation rate  $\tau^{-1}$ ,

$$m \left( \frac{dv}{dt} + \frac{v}{\tau} \right) = \Re e \{ q \mathcal{E}_0 e^{-i\omega t} \},$$

which takes into account all internal losses within a crystal in a lumped parameter. In equilibrium the ansatz  $v = v_0 e^{-i\omega t}$  yields an average velocity amplitude

$$v_0 = \frac{q \mathcal{E}_0}{m} \frac{1}{-i\omega + 1/\tau} = -\frac{q \mathcal{E}_0 \tau}{m} \frac{1}{1 - i\omega \tau}.$$

With charge carrier density  $\mathcal{N}$  and current density  $j = \sigma \mathcal{E} = \mathcal{N} q v$ , one may determine the frequency-dependent conductivity of a metal,

$$\sigma(\omega) = \epsilon_0 \frac{\omega_p^2 \tau}{1 - i\omega \tau} \quad \text{with} \quad \omega_p^2 = \mathcal{N} q^2 / m \epsilon_0, \quad (3.8)$$

where we have introduced the plasma frequency  $\omega_p^2$ . The plasma frequencies of typical metals with large charge carrier densities ( $\mathcal{N} = 10^{22} \text{ cm}^{-3}$ ) have values  $\omega_p \approx 10^{16} \text{ s}^{-1}$ , which is beyond the frequencies of visible light. In semiconductors the conductivity may be adjusted by doping, and this frequency can be easily shifted into the visible or infrared spectral range. Another approach is to structure the geometry of the conducting material to adjust the effective charge density (see Section 4.3).

To analyze the influence of conductivity on wave propagation, we refer to the fourth Maxwell equation (2.20), introduce the current density  $\sigma \mathbf{E}$  we have just determined, and consider a monochromatic wave at frequency  $\omega$ :

$$\nabla \times \mathbf{H} = \mu_0 \sigma \mathbf{E} - i\omega \epsilon_0 \mathbf{E}.$$

This gives rise to a modification of the Helmholtz equation (2.14):

$$\left( \nabla^2 + \frac{\omega^2}{c^2} \right) \mathbf{E}(\mathbf{r}, t) + \frac{i\omega \sigma}{\epsilon_0 c^2} \mathbf{E} = 0.$$

Using the solution  $\mathbf{E} = \mathcal{E}_0 e^{-i[\omega t - n(\omega) \mathbf{k} \cdot \mathbf{r}]}$  for the Helmholtz equation with  $k^2 = \epsilon_r(\omega)(\omega/c)^2 = n^2(\omega)(\omega/c)^2$ , we find a complex permittivity  $\epsilon_r(\omega)$  and a refractive index  $n(\omega)$ , which now depends on the conductivity of the medium:

$$\epsilon_r(\omega) = n^2(\omega) = 1 + i \frac{\sigma(\omega)}{\epsilon_0 \omega}. \quad (3.9)$$

### 3.2.1 Wave Propagation in Conducting Materials

Let us analyze the role of Eq. (3.9) for low and high frequencies.

#### 3.2.1.1 High Frequencies: $\omega_p \tau \gg \omega \tau \gg 1$

We expect this case for optical frequencies; according to (3.8), it holds that

$$\sigma \simeq i \epsilon_0 \omega_p^2 / \omega \quad \text{and} \quad n^2(\omega) \simeq 1 - (\omega_p / \omega)^2. \quad (3.10)$$

Note that for  $\omega < \omega_p$  we have  $\epsilon_r = n^2 < 0$ ! materials with negative  $\epsilon_r$  and  $\mu_r$  have gained much importance for the creation of the so-called metamaterials (see Section 4.3). In the limit of (3.10), the refractive index is purely imaginary,

$$n = i \frac{(\omega_p^2 - \omega^2)^{1/2}}{\omega} = i n'', \quad (3.11)$$

and the wave no longer propagates in this medium. Instead for  $\omega < \omega_p$ , the wave penetrates into the medium, as in the case of total internal reflection, to a depth of only

$$\delta = (n'' k)^{-1} = \frac{c}{\sqrt{\omega_p^2 - \omega^2}}.$$

For  $\tau^{-1} \ll \omega \ll \omega_p$ , we find that  $n'' \approx \omega_p / \omega$  is valid, and the penetration is called the “anomalous skin effect” with an approximately constant penetration depth  $\delta_{as}$ , which corresponds exactly to the plasma wavelength  $\lambda_p = 2\pi c / \omega_p$ :

$$\delta_{as} = c / \omega_p = \lambda_p / 2\pi. \quad (3.12)$$

For  $\omega > \omega_p$ , the index of refraction becomes essentially real again; hence propagating waves are allowed above the plasma frequency limit. The coupled system of electromagnetic and plasma wave is called volume plasmon.

#### 3.2.1.2 Low Frequencies: $\omega \tau \ll 1 \ll \omega_p \tau$

At the lower end of the frequency spectrum, the conductivity is independent of frequency to good approximation,  $\sigma(\omega) \simeq \epsilon_0 \omega_p^2 \tau$ , and in this case the imaginary part of the index of refraction is from Eq. (3.9)  $n'' \simeq \omega_p / 2\sqrt{\tau/\omega}$ . Now the refractive index determines the penetration depth, which is called the “normal skin effect” for lower frequencies:

$$\delta_{ns} = \lambda_p / (\pi \sqrt{\omega \tau}).$$

This case is relevant at radio frequencies, not in optics.

### 3.2.2 Metallic Reflection

Now we can use the results of the previous section to discuss metallic reflection. However, we confine ourselves to perpendicular incidence. Oblique incidence has many interesting properties but requires elaborate mathematical treatment, which can be found in the specialist literature [27].

For optical frequencies the limiting case of high frequencies ( $\omega\tau \gg 1$ ) from the previous section applies, and we can use the purely imaginary refractive index (3.11):  $n = in'' = i\sqrt{(\omega_p^2 - \omega^2)/\omega}$ .

We can take the boundary conditions (Figure 3.5) from Eq. (3.5) and use them directly for the air–metal interface,  $k_t/k_i = in''$ ,

$$\begin{aligned} in''\mathcal{E}_{0t} &= \mathcal{E}_{0i} - \mathcal{E}_{0r}, \\ \mathcal{E}_{0t} &= \mathcal{E}_{0i} + \mathcal{E}_{0r}. \end{aligned}$$

Without any problems, one finds

$$\mathcal{E}_{0r} = \frac{1 - in''}{1 + in''}\mathcal{E}_{0i} \quad \text{and} \quad \mathcal{E}_{0t} = \frac{2in''}{1 + in''}\mathcal{E}_{0i}.$$

An interesting result is obtained when calculating the reflectivity:

$$R = \frac{|\mathbf{E}_r|^2}{|\mathbf{E}_i|^2} = \frac{|1 - in''|^2}{|1 + in''|^2} = 1.$$

We have neglected ohmic losses (relaxation rate  $\tau^{-1}$ !), which are, of course, always present in real metals. Indeed, one finds that within the visible spectral region, important metals such as Al, Au, and Ag have reflectivities of the order of 90–98%. Normally this value is reduced by oxidized surfaces, so that metallic mirror surfaces either have to be deposited on the backside of a glass plate or are covered with a transparent and thin protective layer.

#### Example: Hagen–Rubens relation

In order to estimate role of the ohmic losses in Eq. (3.8), we use the approximation

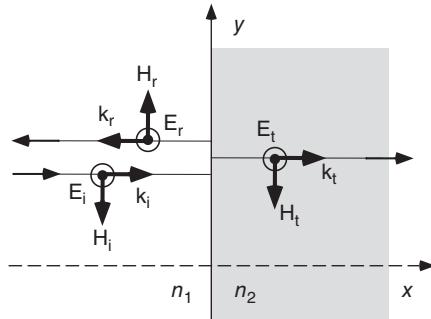
$$\sigma(\omega) = \frac{\sigma_0}{1 - i\omega\tau} \approx \frac{i\sigma_0}{\omega\tau} \left(1 - \frac{i}{\omega\tau}\right),$$

which gives for the refractive index at optical frequencies ( $\omega_p > \omega \gg \tau^{-1}$ ) the approximation  $n^2(\omega) \approx -(\omega_p/\omega)^2(1 - i/\omega\tau)$ . We take  $\omega_p/\omega \gg 1$  and  $(in'')^2 \approx (\omega_p/\omega)^2$ . Using the expansion  $1 - i/\omega\tau \approx (1 - i/2\omega\tau)^2$ , we then find

$$n^2(\omega) \approx \frac{\omega_p^2}{\omega^2} \left(i + \frac{1}{2\omega\tau}\right)^2 = n''^2 \left(i + \frac{n'}{n''}\right)^2.$$

Replacing  $in'' \rightarrow n' + in''$  in the reflectivity formula, we obtain to lowest order in  $n'/n'' = 1/2\omega\tau$  the Hagen–Rubens relation

$$R = 1 - 4n'/n''^2 \approx 1 - 2/\omega_p\tau .$$



**Figure 3.5** Electromagnetic fields reflected at perpendicular incidence.

**Table 3.1** Properties of free electron gases in selected metals [25, 26].

Metal		Ag	Au	Cu	Al
Plasma frequency	$\omega_p/2\pi$	$10^{15}$ Hz	2.2	2.2	2.1
Relaxation time	$\tau$	$10^{-15}$ s	31	9	7

The values for aluminum from Table 3.1 suggest that the reflectivity of a fresh layer is of order 92%. In reality further, surface properties such as diffuse scattering and oxidization reduce reflectivity.

### 3.2.3 Polaritons and Plasmons

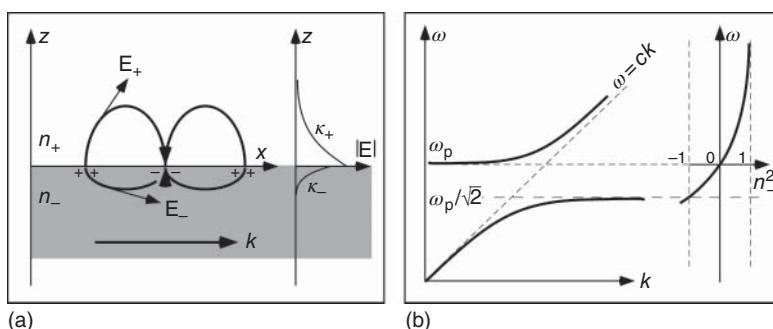
Metallic reflection is due to the very high mobility of “free” electron charge carriers, which allows to shield penetration of an incoming wave and leads to reflection at the boundary. However, there are also wave phenomena bound to and propagating along the metal–vacuum interface where the new wave is based on a joint oscillation of light and matter, for example, the metal electron gas. Mixed waves of light and polarized matter waves are called *polaritons*, and more specifically the coupling with a plasma oscillation (oscillations of the electron density in a metal) is called *plasmon polariton*. The notions “plasmon” and “polariton” relate to the concept of quantum quasiparticles in analogy with purely electromagnetic photons.

#### 3.2.3.1 Surface Plasmon Polaritons (SPPs)

Let us construct waves localized at metal–vacuum boundary with real-valued wave vector  $k$  describing propagation along the surface and real-valued  $\kappa_{\pm}$  providing confinement of the wave near the surface (Figure 3.6). A wave oscillating at frequency  $\omega$  thus reads

$$\mathbf{E}(x, z, t) = (E_{\pm}^{\parallel} \mathbf{e}_x + E_{\pm}^{\perp} \mathbf{e}_z) e^{-i\omega t} = (\mathcal{E}_{\pm}^{\parallel} \mathbf{e}_x + \mathcal{E}_{\pm}^{\perp} \mathbf{e}_z) e^{ikx} e^{-\kappa_{\pm}|z|} e^{-i\omega t}.$$

Application of the Helmholtz equation (2.14) within each of the two homogeneous half-spaces with refractive indices  $n_+^2 = 1$  in vacuum and  $n_-^2 = 1 - (\omega_p/\omega)^2$



**Figure 3.6** Electric field distribution (a) and dispersion relation  $\omega(k)$  (b) of surface waves localized at a metal–vacuum boundary.

for the loss-free metal (3.10) yields

$$-k^2 + \kappa_{\pm}^2 + n_{\pm}^2(\omega/c)^2 = 0. \quad (3.13)$$

Using boundary conditions ( $\mathcal{E}_+^{\parallel} = \mathcal{E}_-^{\parallel} = \mathcal{E}^{\parallel}$ ,  $n_+^2 \mathcal{E}_+^{\perp} = n_-^2 \mathcal{E}_-^{\perp}$ ; see (3.2)) and taking advantage of  $\nabla \cdot \mathbf{D} = \nabla \cdot n_{\pm}^2 \mathbf{E} = 0$  leading to  $ik\mathcal{E}^{\parallel} + \kappa_{\pm} \mathcal{E}_{\pm}^{\perp} = 0$ , we can eliminate  $\kappa_{\pm}$  from (3.13). We find

$$k^2 = \frac{n_+^2 n_-^2}{n_+^2 + n_-^2} \left( \frac{\omega}{c} \right)^2 = \frac{\omega^2 - \omega_p^2}{2\omega^2 - \omega_p^2} \left( \frac{\omega}{c} \right)^2. \quad (3.14)$$

The dispersion relation  $\omega(k)$  for the loss-free case is thus given by

$$\omega^2(k) = \omega_p^2/2 + (ck)^2 \pm \left( (\omega_p^2/2)^2 + (ck)^4 \right)^{1/2}. \quad (3.15)$$

We obtain physical insight by discussing the limiting cases: inspection of (3.14) shows immediately that propagating solutions (real-valued  $k$ ) exist for  $\omega > \omega_p$  and  $\omega < \omega_p/\sqrt{2}$  only. Insertion of (3.14) into (3.13) furthermore shows that for  $\omega > \omega_p$ , solutions for  $\kappa_+$  are imaginary and hence not localized ("radiative" modes). We ignore these *volume plasmon* waves in the following.

### 3.2.3.2 Properties of Surface Plasmon Polaritons (SPPs)

Based on Eqs. (3.13)–(3.15), we discuss selected properties of surface plasmon polaritons (SPPs). For a detailed account, see [28]:

- *Spatial localization.* Confinement lengths  $\kappa_{\pm}^{-1}$  for the SPPs are straightforwardly calculated from (3.13) and (3.14), yielding  $\kappa_{\pm}^2 = -n_{\pm}^4/(n_+^2 + n_-^2) \cdot (\omega/c)^2$ . With  $c/\omega = \lambda/2\pi$  and  $c/\omega_p = \lambda_p/2\pi$ , the corresponding vacuum and plasma wavelengths, we find

$$\kappa_+^{-1} = \frac{\lambda}{2\pi} \frac{(\omega_p^2 - 2\omega^2)^{1/2}}{\omega} \quad \text{and} \quad \kappa_-^{-1} = \frac{\lambda_p}{2\pi} \frac{\omega_p(\omega_p^2 - 2\omega^2)^{1/2}}{\omega_p^2 - \omega^2}.$$

For  $\omega \ll \omega_p$  we find  $\kappa_+^{-1} \rightarrow (\lambda/2\pi) \cdot (\omega_p/\omega)$  and  $\kappa_-^{-1} \rightarrow \lambda_p/2\pi$ . Metal plasma wavelength is of order 100 nm; hence confinement on the metal side is of order several 10 nm. On the vacuum side and for typical optical wavelengths, the range of the evanescent electric field remains of the order  $\lambda/2\pi$ .

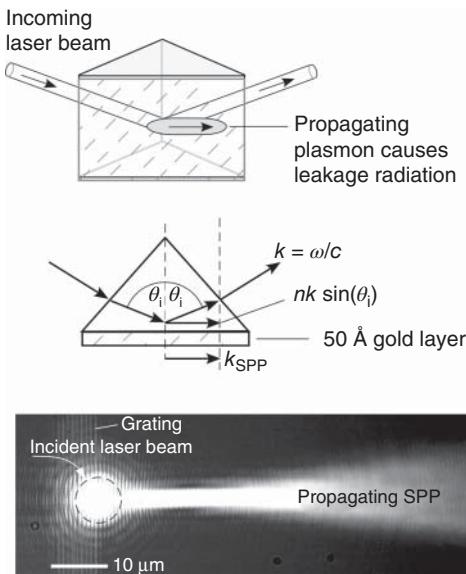
- *Coupling to free space waves.* Waves can only be coupled efficiently with each other if they propagate synchronously. From the dispersion relation in Figure 3.6, it is evident that free space waves with  $\omega = ck$  cannot be phase matched with any SPP. It is therefore necessary to control the coupling by exploiting the evanescent field emanating from TIR (see following example) or by adapting longitudinal grating structures to the plasmon wires.
- *Field enhancement.* With identical  $E^{\parallel}$  components, we find for the perpendicular components from  $\mathcal{E}_{\pm}^{\perp} = ik\mathcal{E}^{\parallel}/\kappa_{\pm}$  (compare derivation leading to (3.14))

$$\mathcal{E}_+^{\perp} = i \left( \frac{\omega_p^2}{\omega^2} - 2 \right) \mathcal{E}^{\parallel} \quad \text{and} \quad \mathcal{E}_-^{\perp} = i \left( \frac{\omega_p^2}{\omega^2} - 2 \right)^{-1} \mathcal{E}^{\parallel}.$$

The vacuum field amplitude is thus enhanced, especially in the vicinity of  $\omega^2 \rightarrow \omega_p^2$ , and of interest for applications while the metal side is reduced.

- *Damping and propagation lengths.* In our analysis we have completely neglected the ohmic losses and hence damping of propagating SPPs. Our basic understanding remains intact, however, since the oscillation is weakly damped below the %/wavelength level and hence propagation lengths extend all the way into the 10–100 μm range.

### Example:



**Figure 3.7** Coupling a free space optical wave to a surface plasmon by means of an internal prism reflection (see text) and a diffraction off a grating. (Lowest row, courtesy of S. Linden and F. Bleckmann.)

For Ag and Au films deposited on an  $n = 1.45$  glass prisms, we find angles of incidence of  $\theta_i \simeq 45^\circ$ .

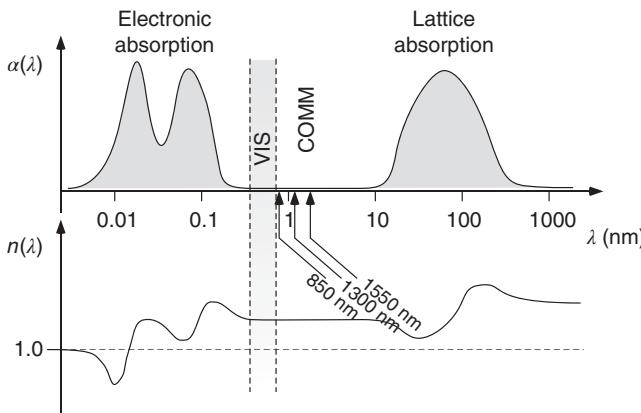
The plasmon can be observed due to radiation fields scattered off imperfections. A microscope image of the leakage radiation of a propagating SPP is shown in the lowest row of Figure 3.7 where phase matching for the excitation is achieved by a grating structure directly applied to the metallic thin film.

Surface plasmons can be excited and demonstrated with a simple arrangement (“Kretschmann configuration”): in Figure 3.7 the concept for achieving phase matching by tuning the angle of incidence  $\theta_i$  is shown. At resonance the wave along the interface of prism (refractive index  $n$ ) and the 50 nm Au layer must co-propagate with the SPP wave. We can express (3.14) in terms of  $k_{SPP} = n_{\text{eff}}(\omega/c)$ , and for Au and Ag we calculate the ratio  $\omega_p/\omega \simeq 4.6$  for the wavelength of an incoming green laser pointer at  $\lambda = 635$  nm. Hence we have  $n_{\text{eff}}^2 = 1.05$ . With the projection of the wave traveling inside the prisms and along the interface  $(n\omega/c) \sin \theta_i$ , the resonance condition is

$$\begin{aligned} k_{SPP} &= (\omega/c)n_{\text{eff}} \\ &= (\omega/c)n \sin \theta_i. \end{aligned}$$

## 3.3 Light Pulses in Dispersive Materials

Electromagnetic waves are used to transmit information. To make sure that there is enough power available at the other end of the transmission line for the message to be read by a receiver, the material (e.g., an optical fiber) in which transmission occurs must be sufficiently transparent. Of course, these conditions are valid for all kinds of electromagnetic waves used for transmission of information, for radio waves with ultrashort or long waves, and for microwave systems as well.



**Figure 3.8** Qualitative trends of the absorption coefficient and refractive index as functions of wavelength for transparent optical materials. The narrow band of visible wavelengths (VIS, 400–700 nm) and the optical windows for telecommunications (COMM, 850, 1300, 1550 nm) are indicated.

For optical wavelengths, the properties of the transparent medium are generally described by two frequency-dependent indices: absorption is described by the absorption coefficient  $\alpha(\omega)$  and dispersion by the refractive index  $n(\omega)$ .

Information is typically transmitted with digitized signals, that is, with short optical light pulses. These pulses are not only attenuated in intensity by the absorption of light energy but also deformed as a consequence of dispersion. Therefore, it is important to explore whether such a pulse is still detectable in its original shape at the end of a transmission line. We know that it is enough to describe a continuous monochromatic field by an absorption coefficient  $\alpha(\omega)$  and real index of refraction  $n(\omega)$ , the spectral properties of which are shown qualitatively in Figure 3.8. The amplitude of the field at point  $z$ , taking the propagation coefficient  $\beta(\omega) = n(\omega)\omega/c$  into account, then yields

$$\begin{aligned} \text{at the start, } z = 0 : \quad E(0, t) &= E_0 e^{-i\omega_0 t}, \\ \text{at the point } z : \quad E(z, t) &= E_0 e^{-i[\omega_0 t - \beta(\omega)z]} e^{-\alpha(\omega)z/2}. \end{aligned}$$

A light pulse can be described as a wave packet, that is, by the superposition of many partial waves. For that purpose we consider an electric field

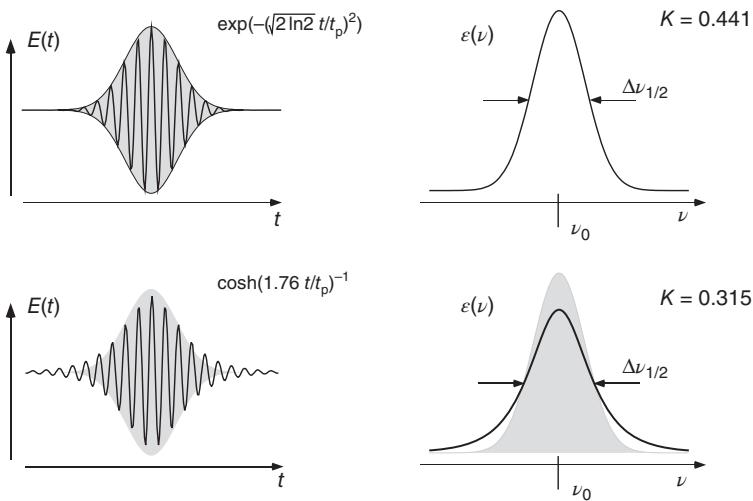
$$E(0, t) = E(0, t) e^{-i\omega_0 t},$$

with *carrier frequency*  $\nu_0 = \omega_0/2\pi$  and time-variant envelope  $E(z, t)$ , which describes the pulse shape, but in general varies slowly in comparison with the field oscillation itself:

$$\frac{\partial}{\partial t} E(t) \ll \omega_0 E(t). \quad (3.16)$$

We determine the *field spectrum*  $\mathcal{E}(z, \nu)$  of the light pulse by harmonic expansion:

$$\begin{aligned} \mathcal{E}(z, \nu) &= \int_{-\infty}^{\infty} E(z, t) e^{i2\pi\nu t} dt = \int_{-\infty}^{\infty} E(z, t) e^{i2\pi(\nu - \nu_0)t} dt, \\ E(z, t) &= \int_{-\infty}^{\infty} \mathcal{E}(z, \nu) e^{-i2\pi\nu t} d\nu = \int_{-\infty}^{\infty} \mathcal{E}(z, \omega) e^{-i\omega t} d\omega / 2\pi. \end{aligned} \quad (3.17)$$



**Figure 3.9** Two important pulse shapes in the time domain and in frequency or Fourier space. For illustration, the Gaussian pulse is superimposed on the sech or \$\cosh^{-1}\$ pulse. The amplitudes are chosen in such a way that the pulses have the same total energy (\$\int |E(t)|^2 dt\$). The \$K\$ values specify the product of half-width times pulse length from Eq. (3.19).

Usually the spectrum of the wave packet is located at \$\nu = \nu\_0\$ because of Eq. (3.16), and its width is small compared to the oscillation frequency \$\nu\_0\$. In Figure 3.9 we give two examples for important and common pulse shapes.

Characteristic quantities of pulsed laser radiation include the spectral bandwidth \$\Delta\nu\$ and the pulse length \$\Delta t\$, which are not easily defined and even more difficult to measure. We may, for instance, employ the conventional variance

$$\langle (\Delta\nu)^2 \rangle = \langle \nu^2 - \nu_0^2 \rangle = \int_{-\infty}^{\infty} (\nu - \nu_0)^2 |\mathcal{E}(\nu)|^2 d\nu / \int_{-\infty}^{\infty} |\mathcal{E}(\nu)|^2 d\nu,$$

and, accordingly, in the time domain,

$$\langle (\Delta t)^2 \rangle = \int_{-\infty}^{\infty} (t - \langle t \rangle)^2 |E(t)|^2 dt / \int_{-\infty}^{\infty} |E(t)|^2 dt,$$

and show that the general relation

$$2\pi\Delta\nu t_p \geq 1/2 \quad (3.18)$$

holds between these two quantities. The equal sign is valid only for pulses without frequency modulation, such pulses being called “Fourier limited.” From the experimental point of view, it is easier to measure half-widths \$\Delta\nu\_{1/2}\$ and \$\Delta t\_{1/2} = t\_p\$ of the intensity. Then the pulse length times bandwidth product can be written as

$$2\pi\Delta\nu_{1/2} t_p = K, \quad (3.19)$$

and this constant \$K\$ is indicated for the two examples in Figure 3.9. In general, its value is less than 0.5, because the half-width usually underestimates the variance.

In Figure 3.9 the much broader wings of the  $\cosh^{-1}$  pulse can be seen as a reason for this.

For monochromatic waves the absorption coefficient  $\alpha$  and the propagation constant  $\beta$  are often known for all partial waves of the wave packet in the frequency domain. Then Eq. (3.17 (i)) can also be described with the transfer function  $\tau(z, \nu)$ :

$$E(z, t) = E(0, t) e^{i\beta(\nu)z} e^{-\alpha(\nu)z/2} = \tau(z, \nu)E(0, t).$$

A pulse is composed of many partial waves, and the correlation between the pulse shapes at the start and the end of a transmission line is described by a linear, frequency-dependent transfer function  $\tau(z, \nu)$  in Fourier space:

$$\mathcal{E}(z, \nu) = \tau(z, \nu)\mathcal{E}(0, \nu).$$

The temporal evolution of the field amplitude at the point  $z$  can now be determined according to

$$E(z, t) = \int_{-\infty}^{\infty} \tau(z, \nu)\mathcal{E}(0, \nu) e^{-i2\pi\nu t} d\nu.$$

Incidentally, according to the convolution theorem of Fourier transformation, a nonlocal correlation in the time domain is valid,

$$E(z, t) = \int_{-\infty}^{\infty} T(z, t - t')E(0, t') dt',$$

with

$$T(z, t) = \int_{-\infty}^{\infty} \tau(z, \nu) e^{-i2\pi\nu t} d\nu.$$

The optical bandwidth of common light pulses is generally narrow compared with the spectral properties of the transparent optical materials used in optical wave guides. Therefore, the following assumptions are reasonable. The frequency dependence of the absorption coefficient plays no role in pulse propagation. In good approximation it holds that

$$\alpha(\nu) \simeq \alpha(\nu_0) = \text{const.}$$

The pulse shape is changed very sensitively by the frequency-dependent dispersion, and the propagation constant  $\beta(\nu) = 2\pi\nu n(\nu)/c$  can be described by the expansion

$$\begin{aligned} \beta(\nu) &= \beta_0 + \frac{d\beta}{d\nu}(\nu - \nu_0) + \frac{1}{2} \frac{d^2\beta}{d\nu^2}(\nu - \nu_0)^2 + \dots \\ &= \beta_0 + \beta'(\nu - \nu_0) + \frac{1}{2}\beta''(\nu - \nu_0)^2. \end{aligned} \quad (3.20)$$

Within this approximation the frequency dependence of the propagation constant  $\beta(\nu)$  is described by the material-dependent parameters  $\beta_0$ ,  $\beta'$ , and  $\beta''$ , the interpretation of which we now want to introduce. With  $\tau_0 = e^{-\alpha z/2}$  the corresponding transfer function reads as follows:

$$\tau(z, \nu) = \tau_0 e^{i\beta_0 z} e^{i\beta'(\nu - \nu_0)z} e^{i\beta''(\nu - \nu_0)^2 z/2}.$$

### 3.3.1 Pulse Distortion by Dispersion

Let us now discuss the influence of the dispersive contributions in more detail. If the dispersion is independent of frequency, then we obtain the wave equation (3.1) once more, in which the velocity of light in vacuum is substituted by the material-dependent phase velocity:

$$\beta_0 = 2\pi n(v_0)v_0/c = 2\pi v_0/v_\phi.$$

Let us first consider the case where  $\beta'' = 0$ . Indeed, this case occurs with glass, and one may realize qualitatively in Figure 3.8 that somewhere between lattice absorption and electronic absorption, the curvature of the refractive index must disappear at some point for every transparent material. For glass, this happens at a wavelength of  $\lambda = 1.3 \mu\text{m}$ , which therefore offers an important window for transmission of information by optical communication. The pulse shape after a propagation length  $z$  is obtained from

$$E(z, t) = \tau_0 e^{i\beta_0 z} \int_{-\infty}^{\infty} e^{i\beta'(v-v_0)z} \mathcal{E}(0, v) e^{-i2\pi vt} dv.$$

Substituting  $\beta'z \rightarrow 2\pi t_g$ , after some algebra, this yields the form

$$\begin{aligned} E(z, t) &= \tau_0 e^{i\beta_0 z} e^{-i2\pi v_0 t} \int_{-\infty}^{\infty} \mathcal{E}(0, v) e^{-i2\pi(v-v_0)(t-t_g)} dv \\ &= \tau_0 e^{-i(2\pi v_0 t - \beta_0 z)} \mathcal{E}(0, t - t_g). \end{aligned}$$

The only effect of dispersion is a delay of the pulse transit time by  $t_g = z/v_g$ , which we interpret as a group delay time. This can be used for the definition of a group velocity  $v_g$ , which can be associated with a “group index of refraction”  $n_g$ :

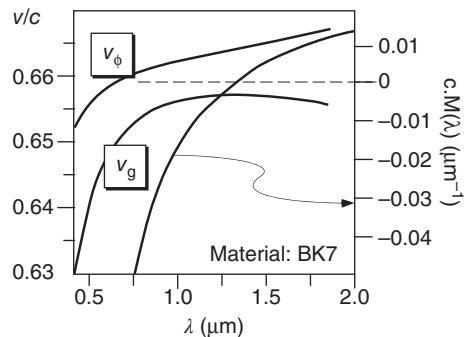
$$\frac{1}{v_g} = \frac{1}{2\pi} \frac{d}{dv} \beta = \frac{1}{c} \left( n(\omega) + \omega \frac{d}{d\omega} n(\omega) \right) = \frac{n_g(\omega)}{c}. \quad (3.21)$$

In most applications optical pulses propagate in a region of normal dispersion, that is, at  $dn/d\omega > 0$ . Then according to Eq. (3.21), it holds that  $v_g < v_\phi = c/n(\omega)$ . Red frequency (or longer wavelengths) contributions propagate faster in a medium than blue ones, but the pulse keeps its shape as long as the group velocity is constant (“dispersion-free”); this is a favorable condition for optical telecommunications, where a transmitter injects digital signals (“bit currents”) in the form of pulses into optical wave guides, which have to be decoded by the receiver at the other end. In optical fibers this situation is similar to that in BK7 glass at  $\lambda = 1.3 \mu\text{m}$ , which can be seen in Figure 3.10 for zero passage of the material parameter  $M(\lambda)$  and will be discussed in the next section.

#### Example: Phase and group velocities in glasses

We can use the specifications from Table 3.2 to determine the index of refraction and the group refractive index as a measure of the phase velocity and group velocity in important optical glasses. The wavelength 850 nm is of substantial importance for working with short laser pulses, because, on the one hand, GaAs diode lasers with high modulation bandwidth exist in this range (up to pulse durations of 10 ps and less), and, on the other, the wavelength lies in the spectral

**Figure 3.10** Example: Dispersion parameters of BK7 glass.



**Table 3.2** Indices of refraction of selected glasses at  $\lambda = 850$  nm.

Abbreviation	BK7	SF11	LaSF N9	BaK 1	F 2
Index of refraction at 850 nm					
$n$	1.5119	1.7621	1.8301	1.5642	1.6068
Group index of refraction					
$n_g$	1.5270	1.8034	1.8680	1.5810	1.6322
Material dispersion					
$cM(\lambda)$ ( $\mu\text{m}^{-1}$ )	-0.032	-0.135	-0.118	-0.042	-0.075

center of the Ti–sapphire laser, which is nowadays the most important primary oscillator for ultrashort laser pulses of 10–100 fs and below. There, with the Sellmeier formula (1.6) and the coefficients from Table 1.1, we calculate the values for Table 3.2. The values for the group refractive index are always larger than the values of the (phase) refractive index by a few percent.

For shorter and shorter pulses, the bandwidth increases according to Eq. (3.18), and the frequency dependence of the group velocity influences the pulse propagation as well. This is specified as a function of frequency or wavelength by one of two parameters: the group velocity dispersion (GVD)  $D_v(v)$  and the material dispersion parameter  $M(\lambda)$ :

$$D_v(v) = \frac{1}{(2\pi)^2} \frac{d^2}{dv^2} \beta = \frac{d}{d\omega} \left( \frac{1}{v_g} \right),$$

$$M(\lambda) = \frac{d}{d\lambda} \frac{1}{v_g} = -\frac{\omega^2}{2\pi c} D_v(v).$$

Like before, we obtain the pulse shape from

$$E(z, t) = \tau_0 e^{-i(\omega_0 t - \beta_0 z)} \times \int_{-\infty}^{\infty} \mathcal{E}(0, \nu) e^{iD_v(\omega - \omega_0)^2 z / 2} e^{-i(\omega - \omega_0)(t - t_g)} \frac{d\nu}{2\pi}. \quad (3.22)$$

The pulse is not only delayed but also distorted. We cannot specify this modification in general; instead we have to look at instructive examples.

**Example: Pulse distortion of a Gaussian pulse**

At  $z = 0$ , the optical pulse  $E(0, t) = E_0 e^{-2 \ln 2(t/t_p)^2} e^{-i\omega_0 t}$  with intensity half-width  $t_p$  has the spectrum

$$\mathcal{E}(0, \omega) = \mathcal{E}_0 e^{-[(\omega - \omega_0)t_p]^2/8 \ln 2}.$$

At the end of the propagation distance at  $z = \ell$ , the spectrum is deformed according to Eq. (3.22). For the sake of simplicity, we introduce the so-called dispersion length,

$$\ell_D = t_p^2 / 4 \ln 2 D_v, \quad (3.23)$$

and find

$$\mathcal{E}(\ell, \omega) = \mathcal{E}_0 e^{-[(\omega - \omega_0)t_p]^2/8 \ln 2} e^{i(\ell/\ell_D)[(\omega - \omega_0)t_p]^2/8 \ln 2}.$$

Inverse Fourier transformation yields the time-dependent form

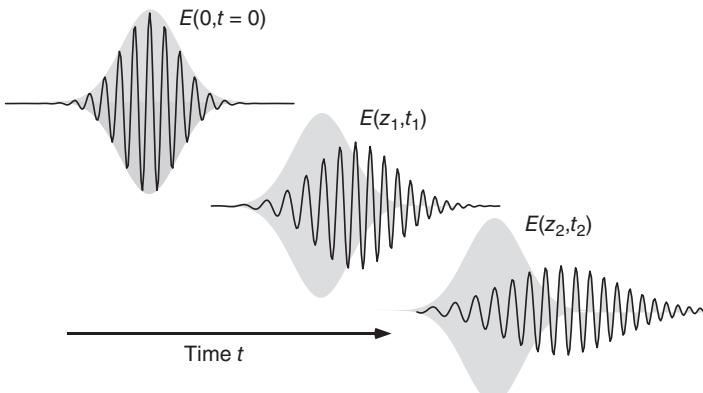
$$E(\ell, t) = \tau_0 E_0 e^{-i(2\pi v_0 t - i\beta_0 \ell)} \times \exp\left(\frac{2 \ln 2(t - t_g)^2}{t_p^2[1 + (\ell/\ell_D)^2]}\right) \exp\left(i \frac{\ell}{\ell_D} \frac{2 \ln 2(t - t_g)^2}{t_p^2[1 + (\ell/\ell_D)^2]}\right).$$

Hence not only is the pulse delayed by  $t_g$ , but it is also stretched,

$$t'_p(z = \ell) = t_p \sqrt{1 + (\ell/\ell_D)^2}, \quad (3.24)$$

and furthermore the spectrum exhibits the so-called frequency chirp, where the frequency changes during a pulse (Figure 3.11):

$$v(t) = \frac{1}{2\pi} \frac{d}{dt} \Phi(t) = v_0 + \frac{1}{\pi} \frac{\ell}{\ell_D} \frac{t - z/v_g}{t_p^2[1 + (\ell/\ell_D)^2]}.$$



**Figure 3.11** Pulse distortion manifests itself as pulse broadening and frequency chirp. The red (lower) frequency components run ahead (left-hand part of the pulse), whereas the blue (higher) ones lag behind (right-hand part). Neither distorted nor delayed pulse is also indicated for comparison.

Now we can determine how far a pulse propagates within a material without significant change of shape. For example, according to Eq. (3.24), it holds that the pulse duration has increased at  $\ell = \ell_D$  by a factor of  $\sqrt{2}$ . The dispersion length plays a similar role in the transmission of pulses as the Rayleigh zone does for the propagation of Gaussian beams (see p. 42).

For BK7 glass from Table 3.2, it holds that  $D(\lambda = 850 \text{ nm}) = 0.04 \text{ ps}^2 \text{m}^{-1}$ . Then one finds for a GaAs diode laser and a conventional Ti-sapphire laser:

$$\begin{aligned}\text{GaAs diode laser: } & t_p = 10 \text{ ps} \quad \ell_D = 200 \text{ m}, \\ \text{Ti-sapphire laser: } & t_p = 50 \text{ fs} \quad \ell_D = 5 \text{ mm}.\end{aligned}$$

It turns out that a short (50 fs) pulse is heavily distorted even by a 5 mm BK7 glass window!

### 3.3.2 Solitons

All optical materials show dispersion, resulting in pulse distortion as described earlier, and detrimental in applications. However, in some materials, one can use nonlinear properties, which will be discussed in more detail later in the chapter on nonlinear optics (Chapter 15), to compensate dynamically for the effects of dispersion. Here we are particularly interested in the optical Kerr effect, describing the intensity-dependent index of refraction:

$$n(I) = n_0 + n_2 I. \quad (3.25)$$

It is true that the values of the nonlinear index in glass are only in the range of  $n_2 \approx 10^{-15}/(\text{W cm}^{-2})$ , but since the power density in optical fibers is very high, this effect plays a role even at power levels of only a few milliwatts and enables the generation of the so-called solitons [29]. Under certain circumstances, these can propagate with a stable shape in a dispersive optical fiber for more than thousands of kilometers.

We study the influence of nonlinearity in a one-dimensional wave equation, taking the linear contribution into account by the index of refraction, that is, the propagation constant  $\beta$ , as we did before,

$$\left( \frac{\partial^2}{\partial z^2} + \beta^2(\omega) \right) E(z, t) e^{-i(\omega_0 t - \beta_0 z)} = \frac{1}{\epsilon_0 c^2} \frac{\partial^2}{\partial t^2} P_{\text{NL}}(z, t), \quad (3.26)$$

and consider a harmonic field  $E(z, t) = E(z, t) \exp[-i(\omega_0 t - \beta_0 z)]$ . In the wave equation, we separate the linear and nonlinear contributions of the polarization,

$$P = \epsilon_0(n^2 - 1)E \simeq \epsilon_0(n_0^2 - 1 + 2n_0n_2I + \dots)E = \epsilon_0(n_0^2 - 1)E + P_{\text{NL}},$$

so that

$$P_{\text{NL}}(z, t) = 2\epsilon_0 n_0 n_2 \frac{\epsilon_0 c^2}{2} |\mathbf{E}(z, t)|^2 \mathbf{E}(z, t) e^{-i(\omega_0 t - \beta_0 z)}.$$

To obtain approximate solutions, we use the so-called slowly varying envelope approximation (SVEA), where we neglect  $\partial E / \partial z \ll kE$  second derivatives:

$$\frac{\partial^2}{\partial z^2} \mathbf{E}(z, t) e^{-i(\omega_0 t - \beta_0 z)} \simeq e^{-i\omega_0 t} \left( 2i\beta_0 \frac{\partial}{\partial z} - \beta_0^2 \right) \mathbf{E}(z, t).$$

We have already used this approximation when generating the paraxial Helmholtz equation (see Eq. (2.35)).

The static dispersive properties of the materials are taken into account by  $\Delta\omega = \omega - \omega_0$  and similarly to Eq. (3.20) by

$$\beta(\omega) \approx \beta_0 + \Delta\omega/v_g + D_v(\Delta\omega)^2/2 + \dots$$

For bandwidths of the pulse that are not too large ( $\Delta\omega \ll \omega_0$ ), we can use the equivalence  $-i\Delta\omega E \simeq \partial E/\partial t$ , and so on – thereby ignoring a more stringent mathematical transformation with the aid of a Fourier transformation – and write

$$\beta^2(\omega) \approx \beta_0^2 + \frac{2i\beta_0}{v_g} \frac{\partial}{\partial t} - \beta_0 D_v \frac{\partial^2}{\partial t^2} + \dots$$

Now inserting all contributions into Eq. (3.26), we get the equation of motion of a soliton as the final result after a few algebraic steps:

$$\left[ \left( \frac{\partial}{\partial t} + \frac{1}{v_g} \frac{\partial}{\partial z} \right) + \frac{i}{2} D_v \frac{\partial^2}{\partial t^2} - i\gamma |E(z, t)|^2 \right] E(z, t) = 0. \quad (3.27)$$

Obviously, the propagation of a pulse with envelope  $E(z, t)$  is described by a non-linear coefficient

$$\gamma = \epsilon_0 c^2 n_2 \beta_0 / n_0,$$

besides the two dispersion parameters, group velocity  $v_g$  and GVD  $D_v$ .

Even with the considerable approximations that we have used so far, the solution of this equation still requires some mathematical effort. Therefore we want to restrict our discussion to the simplest solution called *solitary envelope solution*. A pulse (pulse length  $\tau_0$ ) that has the shape

$$E(0, t) = E_0 \operatorname{sech} \left( \frac{t}{\tau_0} \right)$$

at the beginning of a fiber with dispersion length  $\ell_D$  (see Eq. (3.23)) can propagate, keeping its shape

$$E(z, t) = E_0 \operatorname{sech} \left( \frac{t - z/v_g}{\tau_0} \right) e^{iz/4z_0}$$

if the conditions

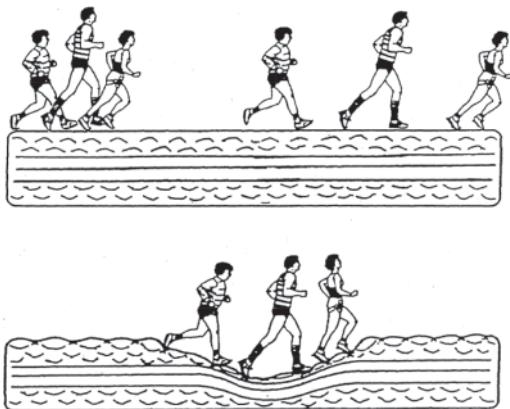
$$\gamma \propto n_2 > 0 \quad \text{and} \quad D_v < 0$$

are fulfilled and, besides, the amplitude has a value equal to [30]

$$E_0 = (|D_v|/\gamma)^{1/2} / \tau_0.$$

These conditions are found in optical fibers in the region of anomalous GVD ( $< 0$ ), typically at  $\lambda > 1.3 \mu\text{m}$ , with simultaneously moderate requirements for pulse power. Besides the fundamental solution, solitons of higher order exist, in analogy to the Gaussian modes, which are characterized by a periodic recurrence of their shape after a propagation length of  $\ell_D$ , which we do not want to discuss here.

**Figure 3.12** A soliton field of athletes. (With kind permission from Linn Mollenauer.)



Linn Mollenauer, who, together with his colleagues [31], was the first to demonstrate long-distance transmission of optical solitons in optical fibers, introduced a very instructive model to illustrate the physical properties of a soliton (Figure 3.12). He compares the differently colored wavelength contributions of a pulse with a small field of runners of different speeds, which disperses very quickly without special influences. As shown in the lower part of the figure, however, the dispersion can be compensated by a soft, nonlinear floor.

Solitons play an important role in many other physical systems as well. One more example, spatial solitons, will be given in Section 16.2.1.2. The relationship of Eq. (3.27) with the nonlinear Schrödinger equation,

$$i \frac{\partial}{\partial x} \Psi + \frac{1}{2} \frac{\partial^2}{\partial t^2} \Psi + |\Psi|^2 \Psi = 0,$$

may be demonstrated by the transformation into a moving frame of reference with  $x = z - v_g t$  and the substitutions  $\Psi = \tau_0 \sqrt{\pi \gamma / |D_v|} E$  and  $z/z_0 \rightarrow x = \pi |D_v| x / \tau_0^2$ .

## 3.4 Anisotropic Optical Materials

When discussing the propagation of light in matter, we always assume the medium to be isotropic. Because of that isotropy, the induced dielectric displacement is always parallel to the inducing field and can be described for transparent materials by just one parameter, the index of refraction,  $\mathbf{D} = \epsilon_0 n^2 \mathbf{E}$ . However, real crystals are very often anisotropic, and the refractive index depends on the relative orientation of the electric field vectors with respect to the crystal axes.

### 3.4.1 Birefringence

Birefringence in calcite (calcareous spar) has been fascinating physicists for a long time (see Figure 3.13) and is one of the most prominent optical properties of anisotropic crystals. Birefringent elements play an important role in applications as well, for example, as retarder plates (see Section 3.4.3), as birefringent filters for



**Figure 3.13** The calcareous spar crystal ( $5 \times 5 \times 15 \text{ cm}^3$ ) that Sir Michael Faraday gave to the German mathematician and physicist Julius Plücker as a present in about 1850 shows birefringence.

frequency selection (Section 3.4.3.1), or as nonlinear crystals for frequency conversion (Section 15.4). Crystal anisotropies can be induced by external influences, such as mechanical strain (strain birefringence) or electric fields (Pockels effect).

We restrict ourselves to the simplest case of uniaxial crystals, where the symmetry axis is called the optical axis (O.A.), and thus the formal problem can be reduced from three to two dimensions. Light beams that are polarized parallel to the O.A. experience a different refractive index than beams with orthogonal polarization.

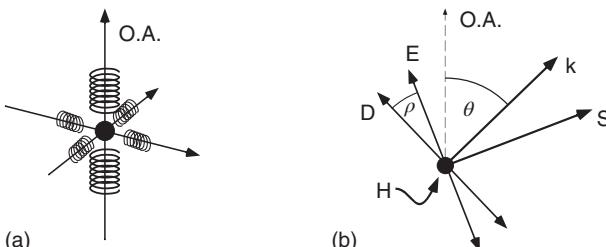
A simple microscopic model of birefringence may be illustrated by charges of the crystal bound to its axes with spring constants of different strengths (Figure 3.14). Therefore, they are displaced by different amounts for identical excitation, and the relation between dielectric displacement  $\mathbf{D}(\mathbf{r}, t)$  and the incident electric field  $\mathbf{E}(\mathbf{r}, t)$  has to be described by a tensor, which has diagonal form, if the O.A. is used as one of the coordinate axes:

$$\mathbf{D} = \epsilon_0 \begin{pmatrix} n_o^2 & 0 & 0 \\ 0 & n_o^2 & 0 \\ 0 & 0 & n_e^2 \end{pmatrix} \mathbf{E}, \quad \mathbf{E} = \begin{pmatrix} n_o^{-2} & 0 & 0 \\ 0 & n_o^{-2} & 0 \\ 0 & 0 & n_e^{-2} \end{pmatrix} \mathbf{D}/\epsilon_0.$$

In uniaxial crystals (unit vectors  $\mathbf{e}_\perp \perp$  O.A.,  $\mathbf{e}_\parallel \parallel$  O.A.), there are two identical indices (ordinary index  $n_\perp = n_o$ ) and one extraordinary refractive index ( $n_\parallel = n_e \neq n_o$ ). Selected examples are collected in Table 3.3. The difference  $\Delta n = n_o - n_e$  itself is often called *birefringence* and may have positive or negative values.

In Maxwell's equations (2.10) for optics, we also have to use the correct tensor relation instead of  $\mathbf{D} = \epsilon_0 n^2 \mathbf{E}$  and write more exactly

$$\begin{aligned} i\mathbf{k} \cdot \mathbf{D} &= 0, & i\mathbf{k} \times \mathbf{E} &= i\mu_0 \omega \mathbf{H}, \\ i\mathbf{k} \cdot \mathbf{H} &= 0, & i\mathbf{k} \times \mathbf{H} &= -i\omega \mathbf{D}. \end{aligned} \tag{3.28}$$



**Figure 3.14** (a) Microscopic model of electromagnetic crystal anisotropy. The black charge is held in different crystal directions with springs of different strengths. (b) Electromagnetic field vectors and propagation vectors in an anisotropic crystal. O.A., optical axis; S, Poynting vector; H, the  $H$  field being normal to the plane.

**Table 3.3** Birefringence of important materials at  $\lambda = 589 \text{ nm}$  and  $T = 20^\circ\text{C}$ .

Material	$n_o$	$n_e$	$\Delta n$	$\alpha_{\max}$
Quartz	1.5442	1.5533	0.0091	0.5°
Calcite	1.6584	1.4864	-0.1720	6.2°
$\text{LiNbO}_3$	2.304	2.215	-0.0890	2.3°

From that we conclude directly

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) = -\omega^2 \mathbf{D}/\epsilon_0 c^2.$$

After some algebra ( $\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) = (\mathbf{k} \cdot \mathbf{E})\mathbf{k} - k^2 \mathbf{E}$ ), we can write

$$\mathbf{D} = \epsilon_0 n^2 \left( \mathbf{E} - \frac{\mathbf{k}(\mathbf{k} \cdot \mathbf{E})}{k^2} \right),$$

introducing the index of refraction  $n^2 = (ck/\omega)^2$ , which describes the phase velocity  $v_\phi = c/n$  of the wave. Its value has to be determined including the dependence on crystal parameters.

In the next step, we decompose the propagation vector  $\mathbf{k} = k_\perp \mathbf{e}_\perp + k_\parallel \mathbf{e}_\parallel$ , and with  $D_\perp = \epsilon_0 n_\perp^2 E_\perp$ , we may write the individual components as

$$k_\perp E_\perp = \frac{n^2 k_\perp^2 (\mathbf{k} \cdot \mathbf{E})}{(n^2 - n_o^2) k^2} \quad \text{and} \quad k_\parallel E_\parallel = \frac{n^2 k_\parallel^2 (\mathbf{k} \cdot \mathbf{E})}{(n^2 - n_e^2) k^2}.$$

The sum of these two components corresponds exactly to the scalar product  $\mathbf{k} \cdot \mathbf{E}$ , and with

$$\mathbf{k} \cdot \mathbf{E} = \left( \frac{n^2 k_\perp^2}{(n^2 - n_o^2) k^2} + \frac{n^2 k_\parallel^2}{(n^2 - n_e^2) k^2} \right) (\mathbf{k} \cdot \mathbf{E}),$$

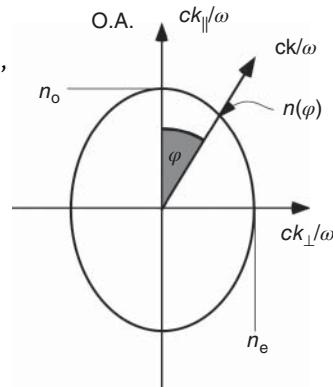
we obtain after short calculations a simplified form of the so-called Fresnel equation [27]:

$$\frac{1}{n^2} = \frac{k_\perp^2/k^2}{n^2 - n_o^2} + \frac{k_\parallel^2/k^2}{n^2 - n_e^2},$$

It yields an equation linear in  $n^2$ , because the  $n^4$  contributions cancel after multiplying ( $k^2 = k_\perp^2 + k_\parallel^2$ ). Finally, substituting the components of the propagation vector  $\mathbf{k}$  by  $k_\perp/k = \sin \phi$  and  $k_\parallel/k = \cos \phi$ , we reach the most important result for describing wave propagation in a uniaxial crystal:

$$\frac{1}{n^2(\phi)} = \frac{\cos^2 \phi}{n_o^2} + \frac{\sin^2 \phi}{n_e^2}. \quad (3.29)$$

This equation describes the so-called index ellipsoid of the refractive index in a uniaxial crystal, which we introduce in Figure 3.15.

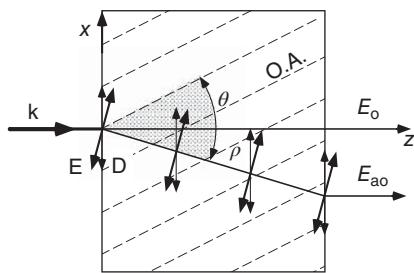


**Figure 3.15** The index ellipsoid.  
O.A., optical axis.

### 3.4.2 Ordinary and Extraordinary Light Rays

Now, we consider the incidence of a light ray onto a crystal, the crystal axis of which makes an angle  $\phi$  with the propagation direction. If the light ray is polarized perpendicular to the O.A. (Figure 3.16), then only the ordinary index of refraction plays a role. The ordinary light ray ( $E_o$ ) obeys the ordinary Snell's law (Eq. (1.2)). If the polarization lies within the plane of propagation and optical axis, then different indices of refraction affect the components of the field parallel and perpendicular to the O.A., and the light ray now propagates as an extraordinary light ray ( $E_e$ ).

Since according to the boundary conditions in Eq. (3.2) the normal ( $z$ ) component of the dielectric displacement is continuous, it must vanish for normal incidence.



**Figure 3.16** Ordinary and extraordinary rays in birefringence.

Therefore, the dielectric displacement lies parallel to the polarization of the incident electric field. According to Eq. (3.28), the vector of propagation  $\mathbf{k}$  is perpendicular to  $\mathbf{D}$  and  $\mathbf{H}$  and retains its direction in the extraordinary ray. The propagation direction of the ray continues to be determined by the Poynting vector  $\mathbf{S}$ :

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}.$$

Therefore, the direction of  $\mathbf{S}$  makes the same angle with the wave vector  $\mathbf{k}$  that occurs between  $\mathbf{E}$  and  $\mathbf{D}$ . According to Figure 3.16, it is sufficient to determine the angle

$$\tan \rho = E_z / E_x$$

from the electric field components in the crystal in order to specify the angle of deflection of the extraordinary ray.

The relation of  $\mathbf{D}$  and  $\mathbf{E}$  can be calculated without much effort if we use the system of major axes including the O.A.:

$$\begin{aligned} \begin{pmatrix} D_z \\ D_x \end{pmatrix} &= \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} n_e^2 & 0 \\ 0 & n_o^2 \end{pmatrix} \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} E_z \\ E_x \end{pmatrix} \\ &= \begin{pmatrix} n_e^2 \cos^2 \phi + n_o^2 \sin^2 \phi & (n_e^2 - n_o^2) \sin \phi \cos \phi \\ (n_e^2 - n_o^2) \sin \phi \cos \phi & n_o^2 \cos^2 \phi + n_e^2 \sin^2 \phi \end{pmatrix} \begin{pmatrix} E_z \\ E_x \end{pmatrix}. \end{aligned}$$

Because of the boundary conditions (3.2), the  $D_z$  component must vanish, and we may conclude directly that

$$\tan \rho = \frac{1}{2} \frac{(n_e^2 - n_o^2) \sin 2\phi}{n_e^2 \cos^2 \phi + n_o^2 \sin^2 \phi}.$$

The “getting out of the way” of the extraordinary beam is called *beam walk-off* and must always be considered when using birefringent components. We can find

an equivalent formulation of the *beam walk-off* angle using  $n(\phi)$  from Eq. (3.29):

$$\tan \rho = \frac{n^2(\phi)}{2} \left( \frac{1}{n_o^2} - \frac{1}{n_e^2} \right) \sin 2\phi. \quad (3.30)$$

### Example: Beam walk-off angle of quartz

We calculate the maximum deflection angle for birefringence in a quartz crystal with the common methods and find

$$\phi_{\max} = \arctan(n_e/n_o) = 45.2^\circ.$$

For  $\phi_{\max}$  we calculate the beam walk-off angle according to Eq. (3.30):

$$\rho = 0.53^\circ.$$

One could say that in general the beam walk-off angle amounts to only a few degrees; even for a material such as calcareous spar (see Table 3.3) with strong birefringence, it is only about  $6^\circ$ . In nonlinear optics, for example, in the case of the so-called angular phase matching (see Section 15.4), the efficiency of frequency conversion is limited by beam walk-off.

### 3.4.3 Construction of Retarder Plates

An important application of birefringent materials is the construction of the so-called retarder plates, the action of which we have discussed for polarization control applications and applications of the Jones formalism in Section 2.4.4 already.

With retarder plates, the state of polarization of a light beam is manipulated by adjusting the O.A. perpendicular to the direction of propagation (see Figure 2.17). Ordinary and extraordinary light rays then propagate collinearly through the crystal, and their components are given by the projection onto the O.A. The ordinary and extraordinary beams are delayed with respect to each other within a plate of thickness  $d$  with phase shifts  $\exp(i\alpha_o) = \exp(in_o\omega d/c)$ , resp.  $\exp(i\alpha_e) = \exp(in_e\omega d/c)$ . With these coefficients, the Jones matrix of Eq. (2.49) is constructed.

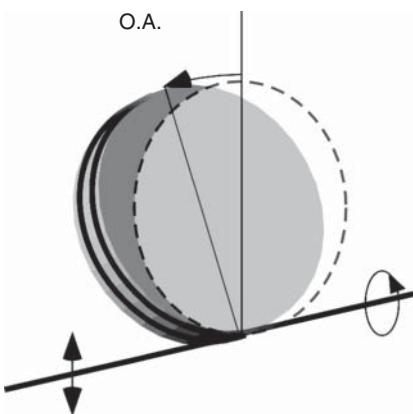
The most important special cases are the  $\lambda/4$  plate and the  $\lambda/2$  plates. The optical path lengths of the ordinary and extraordinary rays must differ by exactly a quarter and half wavelength, respectively, for the two components. However, the differences in path lengths of retarder plates are in general not exactly equal to  $\lambda/2$  and  $\lambda/4$ . Realistic materials of zero order would generally be too thin and therefore too fragile for manufacturing and applications.

One uses  $\lambda(n+1)/2$  and  $\lambda(n+1)/4$  instead, and the number of total waves  $n$  is called the *order*. They serve their purpose independently of their order, but due to the dispersion, which in addition has different temperature coefficients for  $n_o$  and  $n_e$ , retarder plates of higher order are much more sensitive to variations in frequency or temperature than retarder plates with lower order.

The so-called retarder plates of zero order consist of two plates with nearly the same thickness but unequal differences in optical path  $\lambda/2$  or  $\lambda/4$ . If two

plates with crossed O.A. are mounted on top of one another<sup>1</sup>, then the influences of higher orders are compensated, and there remains an effective plate of lower order, which is less sensitive to spectral and temperature changes.

### Example: Fiber-based retarder element



**Figure 3.17** A fiber mounted on a paddle can be used as a retarder plate.

Polarization control can be achieved by making optical fibers birefringent. It is well known that stress induces small changes in the index of refraction. Hence a fiber coiled around the circumference of a disc (Figure 3.17) will have different propagation constants for the linear polarizations parallel and perpendicular to the plane of the disc.

Thus we have an effectively birefringent medium with an optical axis in the plane of the disc. Adjustment of the angle of the disc is equivalent to rotating the optical axis of a conventional retarder plate. The total retardation, for example,  $\lambda/4$  or  $\lambda/2$ , is chosen by the length of the fiber section on the disc.

#### 3.4.3.1 Lyot Filter

A linearly polarized light field is split by a birefringent plate with O.A. oriented at  $\phi = 45^\circ$  to the direction of polarization and orthogonal to the direction of propagation evenly into ordinary and extraordinary waves. At the exit of a plate of thickness  $d$ , the relative phase shift of the two waves,  $\Delta = 2\pi(n_o - n_e)d/\lambda$ , is wavelength dependent. Combining retarder plates with polarizers, one can achieve wavelength-dependent and frequency-dependent transmission. Such applications are called *birefringent filters* or *Lyot filters*.

In Figure 3.18 a retarder plate (now of higher order) is positioned between two parallel polarizers. Only for distinct wavelengths does it serve as a  $\lambda/2$  plate, for example, and cancel transmission.

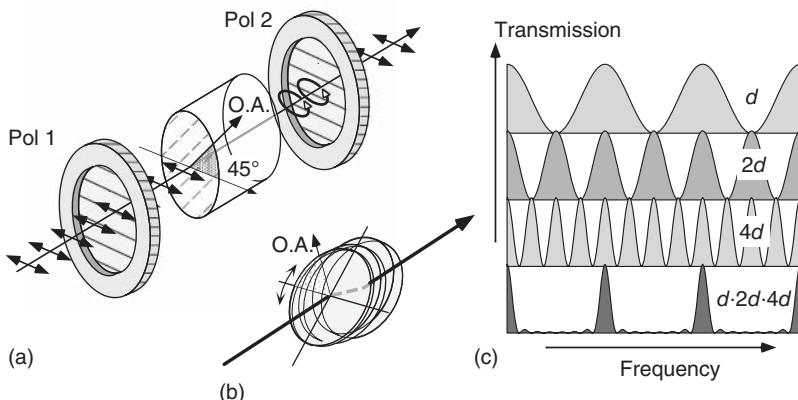
The incident light is transformed in general into elliptically polarized light depending on the orientation of the O.A. We can calculate the transmission of a light field polarized linearly in the  $x$  direction according to Eq. (2.49),

$$E'_x = \exp \left( i \frac{\alpha_o + \alpha_e}{2} \right) \left[ \cos \left( \frac{\alpha_o - \alpha_e}{2} \right) + i \sin \left( \frac{\alpha_o - \alpha_e}{2} \right) \cos 2\phi \right] E_x,$$

and with  $(\alpha_o - \alpha_e) = (n_o - n_e)2\pi v d/c$ , we find the transmitted intensity correlated with the incident intensity  $I_x$ :

$$I_T = I_x \left[ \cos^2 \left( \frac{(n_o - n_e)\pi v d}{c} \right) + \sin^2 \left( \frac{(n_o - n_e)\pi v d}{c} \right) \cos^2 2\phi \right].$$

<sup>1</sup> They are often “optically contacted,” that is, they are connected via two very well-polished surfaces (whose planarities must be much better than an optical wavelength) only by adhesive forces.



**Figure 3.18** (a) Lyot filter built from two parallel polarizers and a wavelength-dependent retarder plate oriented at  $45^\circ$ . Optimum extinction occurs for wavelengths where the retarder plate corresponds to a  $\lambda/2$  plate. (c) transmission curves of Lyot filters with single retarder plate of increasing thicknesses  $d$ ,  $2d$ , and  $4d$  and of a filter composed of the three plates. (b) three-plate Lyot filter at the Brewster angle for use within a laser resonator.

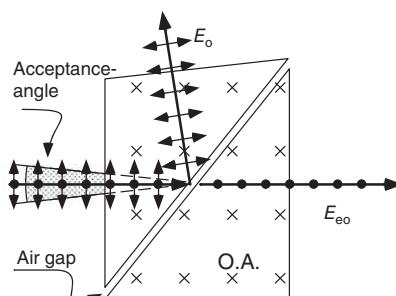
In particular, for  $\phi = 45^\circ$ , one finds a transmission modulated by 100% with the period (or the “free spectral range”)  $\Delta v = c/(n_o - n_e)d$ . Positioning several Lyot filters with thicknesses  $d_m = 2^m d$  one behind the other, the free spectral range is maintained, but the width of the transmission curve is reduced quickly.

Lyot filters, that is, birefringent filters, may be positioned in the ray path at the Brewster angle as well to reduce losses substantially (Figure 3.18). The O.A. lies within the plane of the plate, and the central wavelength of the filter with the lowest losses can be tuned by rotating the axis. Such elements are mainly used in broadband laser oscillators (e.g. Ti-sapphire lasers, dye lasers; Section 8.5.1) for rough wavelength selection.

#### 3.4.4 Birefringent Polarizers

One more important application of birefringent materials is their use as polarizers. From the many variants we introduce the Glan air polarizer. Its effect is based on the various critical angles of TIR for the ordinary beam (which is reflected for devices made of calcareous spar) and the extraordinary beam (Figure 3.19).

Applying a polarizer, both the extinction ratio and the acceptance angle are the most relevant numbers to determine the alignment sensitivity, depending on the difference of the refractive indices for ordinary (o) and extraordinary (eo) beams. With Glan polarizers, very high



**Figure 3.19** Glan air polarizer. The acceptance angle is defined by the critical angles for total internal reflection for the ordinary and the extraordinary rays.

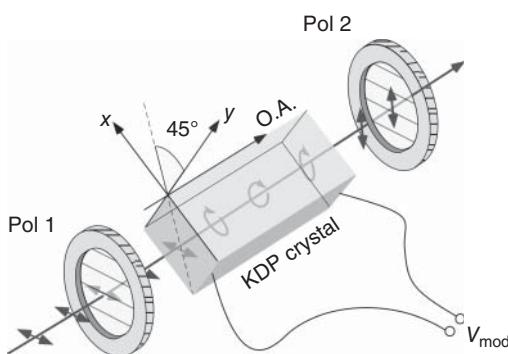
extinction ratios of  $1 : 10^6$ , and more can be achieved. One variant is the Glan–Thompson polarizer, where a glue is inserted between the two prisms with refractive index between  $n_o$  and  $n_e$ . Then TIR occurs for the partial wave with the lower index of refraction only, and the other one is always transmitted, as shown in Figure 3.19, the extraordinary wave.

## 3.5 Optical Modulators

Materials in which the index of refraction can be controlled or switched by electric or magnetic fields offer numerous possibilities to influence the polarization or phase of light fields, thereby realizing mechanically inertia-free optical modulators for amplitude, frequency, phase, or beam direction. We will pick out several important examples.

### 3.5.1 Pockels Cell and Electro-optical Modulators

The electro-optical effect addresses the linear dependence of the refractive index on the electric field strength and is also called the *Pockels effect*. If the index of refraction depends quadratically on the field strength, that is, linearly on the intensity, then we talk about the optical Kerr effect, which will be discussed in more depth in the chapter on nonlinear optics (Section 16.2). We came across self-modulation of an optical wave by the Kerr effect in Section 3.3.2 on solitons already.



**Figure 3.20** Electro-optical modulator with a KDP crystal, switched to blocking mode.

The electric field is created by electrodes attached to the faces of the crystal. The geometry of the index of refraction modulations is in general determined by crystal symmetry. Here, we discuss a simple and important example, the uniaxial KDP crystal.

The KDP crystal is mounted between two crossed polarizers, and its O.A. is adjusted parallel to the propagation direction. A longitudinal electric field is created with transparent electrodes (Figure 3.20).

In the field-free state, there is axial symmetry, which is lifted by the external field and induces an optically marginal biaxial crystal. Therefore the indices of refraction in the  $x$  and  $y$  directions, which are tilted by  $45^\circ$  against the position of the polarizer, are changed by the same modulus of the angle, but with opposite sign:

$$n_{ox} = n_o - rn_o^3 U / 2d \quad \text{and}$$

$$n_{oy} = n_o + rn_o^3 U / 2d.$$

In this arrangement, the transmission is proportional to

$$I_T = I_0 \cos^2(2\pi r n_o^3 U/d).$$

In applications of electro-optical modulators (EOMs), the half-wave voltage, where the difference in the indices of refraction creates a phase delay of the  $x$  and  $y$  components of  $\lambda/2$ , is among the most important technical criteria. The maximum modulation frequency is determined by the capacitive properties of the driver circuit. At very high frequencies (>200 MHz), transit time effects occur because the field strength across the crystal is no longer homogeneous. In this case, traveling-wave modulators can be constructed, in which the radio-frequency wave and the optical wave co-propagate.

#### Example: Half-wave voltage of KDP

The electro-optical coefficient of KDP is  $r = 11 \text{ pm V}^{-1}$  at a refractive index of  $n_o = 1.51$ . For a crystal length of  $d = 10 \text{ mm}$ , the half-wave voltage at a wavelength of  $\lambda = 633 \text{ nm}$  is calculated as ( $E = U/d$ )

$$U = 2 \times \frac{\lambda}{2} \frac{1}{r n_o^3} = 84 \text{ V cm}^{-1}.$$

In this case the half-wave voltage does not depend at all on the length of the crystal. Therefore it is more convenient to choose arrangements with transverse electro-optical coefficients.

#### Example: Phase modulation with an EOM

If one adjusts the linear polarization of a light beam parallel to the principal axis of a crystal and leaves out the polarizers as shown in Figure 3.20, then the beam experiences not an amplitude modulation (AM) but a phase modulation (PM) and frequency modulation (FM). The index of refraction depends linearly on the driving voltage and causes a phase variation at the output of the EOM,

$$\begin{aligned}\Phi(t) &= \omega t + m \sin(\Omega t), \\ E(t) &= \Re \{ E_0 \exp(-i\omega t) \exp[-im \sin(\Omega t)] \},\end{aligned}\tag{3.31}$$

where the modulation index  $m$  specifies the amplitude and is correlated with the material parameters through

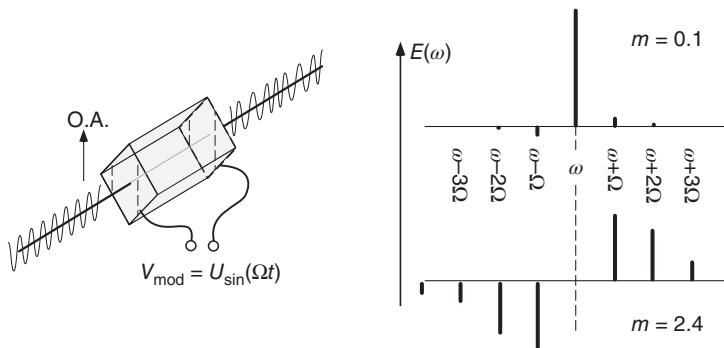
$$m = \omega r n_o^3 U / 2c.$$

The corresponding instantaneous frequency experiences a harmonic modulation as well,

$$\omega(t) = \frac{d}{dt} \Phi(t) = \omega + m\Omega \cos(\Omega t),$$

where the modulation amplitude  $M = m\Omega$  appears. Actually, we cannot strictly distinguish between PM and FM. However, the modulation index allows a rough and common categorization into distinct regions:

$$\begin{aligned}m < 1 &\quad \text{PM,} \\ m \geq 1 &\quad \text{FM.}\end{aligned}$$



**Figure 3.21** Phase modulation with an EOM. The spectra are illustrated for modulation indices  $m = 0.1$  (top) and  $m = 2.4$  (bottom). The lengths of the bars indicate the contribution of the sideband, and the direction indicates the phase position according to Eq. (3.32).

The difference gets more pronounced if we decompose the electromagnetic wave (3.31) with intensity

$$e^{-im \sin(\Omega t)} = J_0(m) + 2[J_2(m) \cos(2\Omega t) + J_4(m) \cos(4\Omega t) + \dots] - 2i[J_1(m) \sin(\Omega t) + J_3(m) \sin(3\Omega t) + \dots]$$

into its Fourier frequencies:

$$E(t) = E_0 e^{-i\omega t} [J_0(m) + J_1(m)(e^{-i\Omega t} - e^{i\Omega t}) + J_2(m)(e^{-i2\Omega t} + e^{i2\Omega t}) + J_3(m)(e^{-i3\Omega t} - e^{i3\Omega t}) + \dots]. \quad (3.32)$$

We present these spectra for the cases  $m = 0.1$  and  $m = 2.4$  in Figure 3.21. For a small modulation index (PM), the intensity at a carrier frequency  $\omega$  is barely changed, but sidebands appear offset from the carrier by the modulation frequency. The intensity of the sidebands is proportional to  $J_\ell^2(m)$ . For a large modulation amplitude (FM), the intensity is distributed to many sidebands, and in our special case the carrier is even completely suppressed due to  $J_0(2, 4) = 0$ .

In contrast to harmonic AM, where exactly two sidebands are created, many sidebands appear for PM/FM. Another important difference is that the AM variation can be shown (“demodulated”) with a simple photodetector, but PM/FM information cannot.

Liquid crystal (LC) modulators are well known from liquid crystal displays (LCDs). By “liquid crystals,” we mean a certain type of order of slab-like or disc-like organic molecules within a liquid (which appear quite often).

### 3.5.2 Liquid Crystal Modulators

In the nematic phase (there exist also smectic and cholesteric phases), all the molecular slabs point in one direction, without aligning their centers. If the molecules are exposed to a surface with a preferred direction (grooves, anisotropic plastics), then they become oriented in this direction. The enclosure

of an LC between glass plates with crossed grooves causes the rotated nematic phase shown in Figure 3.22, where the molecular axes are rotated continuously from one direction into the other.

The rotated nematic phase rotates the plane of polarization of an incident polarized light wave by  $90^\circ$ . But the molecular rods can be aligned parallel to the field lines of an electric field in the direction of propagation. Then the polarization is not changed during transmission. Thus, an electric field can be used to switch the transmitted amplitude. LCDs use the same principle but work in general in a reflection mode.

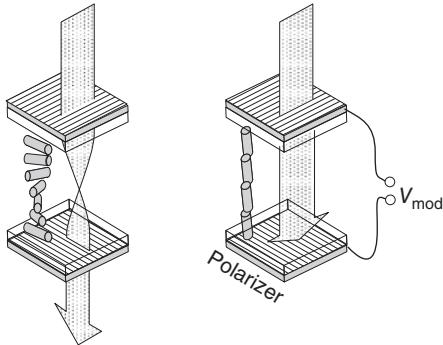


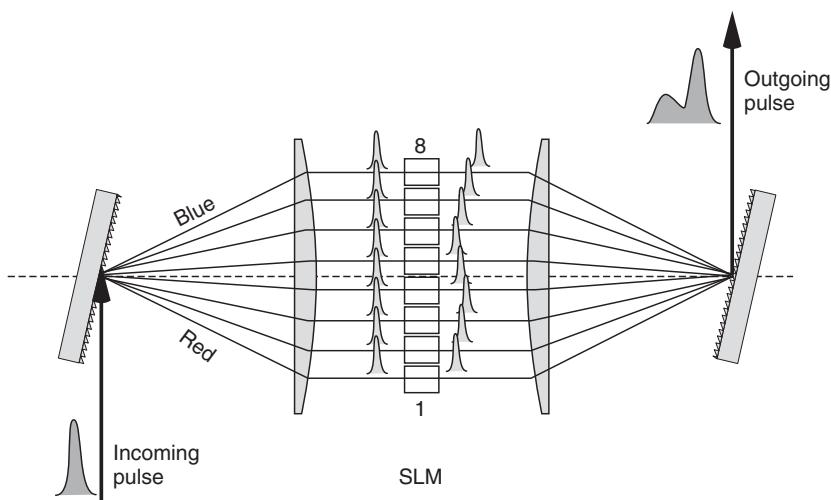
Figure 3.22 Liquid crystal modulator.

### 3.5.3 Spatial Light Modulators

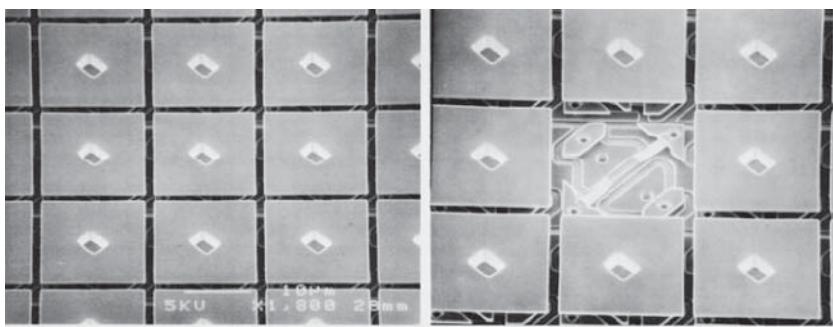
The digital revolution is more and more also entering the world of optical devices. It has led to the development of modulators, allowing spatial control of the intensity or phase of an extended light field, the so-called spatial light modulators (SLMs). Conceptually it is straightforward to use the fabrication methods of microelectronics and divide the LC described in the last section into an array of small and individually addressable pixels. LCDs have long become a ubiquitous component of electric and electronic tools. With improved optical quality, LC arrays can be used to actively control the wave front of a laser beam incident on such a device. While applications for versatile digital display technology are fairly obvious (see the following discussion), we here introduce another example where SLMs are used to control the shape of ultrashort pulses (for the generation of femtosecond pulses, see Section 16.2.1.1).

In Figure 3.23 a very short pulse is dispersed by a grating, and in combination with a lens, a parallel wave front is generated with spatially varying color. Without SLM the second grating would undo the dispersion and simply restore the original pulse. The SLM can now be configured, if necessary, by inserting additional optical elements such as polarizers to introduce attenuation or delays in each channel (typically 128 and more) individually. On recombination the pulse is now very different from the incoming pulse. This pulse shaping method is used to improve, for instance, the efficiency of chemical reactions induced by femtosecond laser pulses (“femtochemistry”) [32].

In 1987 Larry Hornbeck of Texas Instruments invented another type of SLM, the digital mirror device (DMD), which can realize more than 1.3 million hinge-mounted mirrors on a single silicon chip. Each individual mirror shown in Figure 3.24 has a square length of about  $15\text{ }\mu\text{m}$  and corresponds to a pixel of a digital image. It is separated from adjacent mirrors by  $1\text{ }\mu\text{m}$ , and it tilts up to  $12^\circ$  in less than 1 ms by microelectromechanical actuators. White and black is generated by directing each mirror in and out of the light beam from the projection lamp. Since each mirror can be switched on several thousand times



**Figure 3.23** Pulse shaping using a spatial light modulator (SLM). The incoming ultrashort pulse (typically in the femtosecond domain) is dispersed into its spectral components. The SLM modifies the intensity of individual channels (here 1–8) and generates small delays. With a second grating, the pulse is recombined.



**Figure 3.24** A sector of  $3 \times 3$  mirrors out of an array of  $1280 \times 1024$ . On the right side one mirror is removed to expose the electromechanical actuators. (With permission by Texas Instruments, from [www.dlp.com/dlp\\_technology](http://www.dlp.com/dlp_technology).)

per second, also gray scales can be realized by varying the “on” versus the “off” time of the mirror.

The DMD offers *digital light processing* (DLP<sup>TM</sup>) with excellent quality and is currently revolutionizing display technologies from large-scale cinemas to home entertainment.

### 3.5.4 Acousto-Optical Modulators

If a sound wave propagates within a crystal, it causes periodic density fluctuations, which induce a variation of the refractive index at the same frequency and wavelength. The periodic fluctuation of the refractive index has an effect like a

propagating optical grating, at which the light ray is diffracted. Diffraction may be interpreted as a Bragg scattering or Bragg refraction off this grating.

An acousto-optical modulator (AOM) consists of a crystal, at the end of which is glued a piezo element to excite ultrasonic waves (Figure 3.25), the transducer. To avoid reflection and standing waves, a sound absorber is installed at its other end.

The ultrasonic head is set vibrating mechanically with a radio frequency (typically 10–1000 MHz) and radiates sound waves through the modulator crystal. Then the light ray transits a so-called extensive sound wave field and experiences diffraction in this “Bragg domain” in one order only. If the light ray transits through a thin sound wave field, as is the case with an optical grating, then several, here undesirable, diffraction orders occur. This boundary case is called the “Raman–Nath domain.”

In order to discuss the influence of the sound wave on the propagation of the light ray in more detail, we consider the variation of the index of refraction in the  $x$  direction caused by a sound wave with frequency  $\Omega$  and wave vector  $\mathbf{q} = q\mathbf{e}_x$ :

$$n(t) = n_0 + \delta n(t) = n_0 + \delta n_0 \cos(\Omega t - qx).$$

We use the wave equation in the form of Eq. (3.1) and take into account that  $[n_0 + \delta n(t)]^2 \simeq n_0^2 + 2n_0\delta n(t) + \dots$ . Furthermore, we confine ourselves to the variations of the  $x$  components, because we do not expect any change through the sound wave in the other directions:

$$\left[ \frac{\partial^2}{\partial x^2} - k_y^2 - k_z^2 - \left( \frac{n_0^2}{c^2} - \frac{2n_0\delta n(t)}{c^2} + \dots \right) \frac{\partial^2}{\partial t^2} \right] \mathbf{E}(\mathbf{r}, t) = 0. \quad (3.33)$$

Now we shall consider how the amplitude of the incident wave evolves, which for simplification has only a linear polarization component:

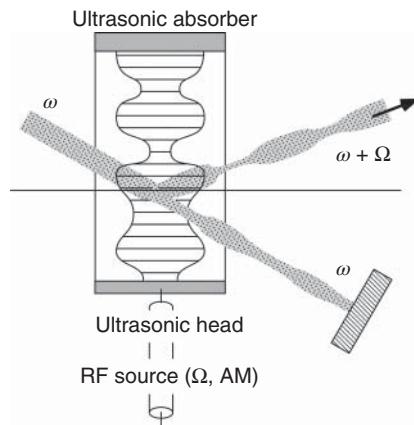
$$E_i(\mathbf{r}, t) = E_{i0}(x, t) e^{-i(\omega t - \mathbf{k}\mathbf{r})}.$$

The modulated index of refraction leads to a time-dependent variation at frequencies  $\omega \pm \Omega$ ; therefore, we can “guess” an additional field  $E_a(\mathbf{r}, t)$ , which we may interpret as a reflected field,

$$E_a(\mathbf{r}, t) = E_{a0}(x, t) e^{-i(\omega' t - \mathbf{k}'\mathbf{r})},$$

with  $\omega' = \omega + \Omega$  and  $\mathbf{k}' = \mathbf{k} + \mathbf{q}$  arising from diffraction off the sound wave. The oscillating refractive index has no influence on the propagation vector; therefore, even at this point, it must hold that (Figure 3.26)

$$\mathbf{k}'^2 = (\mathbf{k} + \mathbf{q})^2 = n_0^2(\omega + \Omega)^2/c^2 \simeq (n_0\omega/c)^2$$



**Figure 3.25** Acousto-optical modulator. The intensity variation of the optical beams reflects the intensity of the acoustic wave.

(because  $\Omega \ll \omega$ ). From that the Bragg condition immediately follows:

$$q = -2k_x.$$

Now, we study Eq. (3.33) with a total field  $E = E_i + E_a$  and again assume that the change in amplitude is negligible on the scale of a wavelength, that is,

$$\partial^2/\partial x^2 [E(x) e^{ikx}] \simeq [-k^2 + 2ikE'(x)] e^{ikx}.$$

With  $k_x^2 + k_y^2 + k_z^2 = (n_0\omega/c)^2$  and  $(k_x + K)^2 + k_y^2 + k_z^2 = [n_0(\omega + \Omega)/c]^2$ , after a short calculation, we obtain the equation

$$\begin{aligned} & \left[ 2ik_x \frac{\partial}{\partial x} + \frac{2\omega^2 n_0 \delta n_0}{c^2} \cos(\Omega t - qx) \right] E_{i0}(x) e^{-i(\omega t - k_x x)} \\ & + \left[ -2ik_x \frac{\partial}{\partial x} + \frac{2\omega^2 n_0 \delta n_0}{c^2} \cos(\Omega t - qx) \right] E_{a0}(x) e^{-i[(\omega + \Omega)t + k_x x]} = 0. \end{aligned}$$

To get a more simplified system for the two amplitudes  $E_{i0}$  and  $E_{a0}$ , we use the cos terms in their complex form, sort according to the oscillator frequencies, and ignore oscillating terms, where the incident field does not participate:

$$\begin{aligned} 2ik_x \frac{\partial}{\partial x} E_{i0} + \frac{\omega^2 n_0 \delta n_0}{c^2} E_{a0} &= 0, \\ -2ik_x \frac{\partial}{\partial x} E_{a0} + \frac{\omega^2 n_0 \delta n_0}{c^2} E_{i0} &= 0. \end{aligned}$$

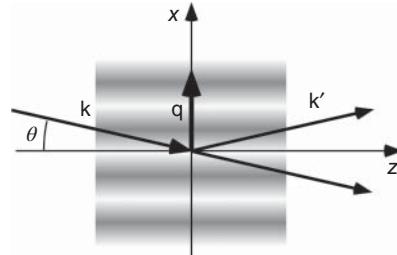


Figure 3.26 Bragg geometry.

Finally, we substitute the  $x$  dependence by the dependence along the principal propagation direction  $z$  (therefore  $E_a$  propagates in the opposite direction to  $E_i$ ). With  $k_x = k \sin \theta = (n_0\omega/c) \sin \theta$ , it holds that

$$\begin{aligned} i \frac{\partial}{\partial z} E_{i0} + \frac{k \delta n_0}{2n_0 \sin \theta} E_{a0} &= 0, \\ i \frac{\partial}{\partial z} E_{a0} + \frac{k \delta n_0}{2n_0 \sin \theta} E_{i0} &= 0. \end{aligned} \tag{3.34}$$

The solutions of this system are well-known harmonic oscillations with frequencies

$$\gamma = \frac{k \delta n_0}{n_0 \sin \theta}.$$

In general, as  $E_{a0} = 0$  is valid at the entrance of an AOM, we find the pendulum solution

$$E_i(z, t) = E_{i0} \cos(\gamma z/2) e^{-i(\omega t - \mathbf{kr})},$$

$$E_a(z, t) = E_{i0} \sin(\gamma z/2) e^{-i[(\omega + \Omega)t - (\mathbf{k} + \mathbf{q})\mathbf{r}]}.$$

So the reflected beam is actually frequency-shifted, as guessed earlier. For small  $z$  the reflected intensity is proportional to  $(\gamma z)^2$ . The modulation amplitude of the

refractive index at sound intensity  $I_S$  is

$$\delta n_0 = \sqrt{\mathcal{M} I_S / 2}.$$

The  $\mathcal{M}$  coefficient depends on the material parameters and is introduced here only phenomenologically. For small powers the reflected (in other words, diffracted) intensity is proportional to  $|E_a|^2$  and thus, according to this result, proportional to the applied sound power.

### 3.5.5 Faraday Rotators

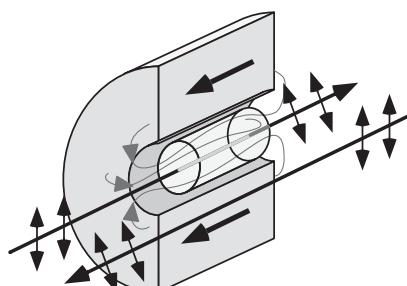
Certain materials show the Faraday effect, where the oscillation plane of linearly polarized light is rotated independently of the initial orientation proportional to a longitudinal magnetic field,

$$\mathbf{E}' = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \mathbf{E},$$

with  $\alpha = VB\ell$  where  $V$  (units  $\text{deg m}^{-1} \text{T}^{-1}$ ) is the *Verdet constant*,  $B$  is the magnetic field strength, and  $\ell$  is the crystal length. The magnetization of a Faraday crystal affects right-hand and left-hand polarized refractive indices with different indices of refraction:  $n_{\pm} = n_0 \pm VB\lambda/2\pi$ .

In contrast to the retarder plates in Section 3.4.3, the polarization transformation of an electromagnetic wave is not reversed in a Faraday rotator if the wave is returned into the same configuration (Figure 3.27). The Faraday rotator is “non-reciprocal” and therefore is suited extremely well for the design of isolators and diodes. As a result of the typically very small Verdet constants (Table 3.4), relatively high magnetic field strengths are necessary. They can be more conveniently realized with permanent magnets made of SmCo or NdFeB [33].

**Figure 3.27** Faraday rotation. Only those field lines are shown, which pass through the whole crystal.



**Table 3.4** Verdet constant of selected materials at 589 nm.

Material	Quartz	Heavy flint	TGG <sup>a)</sup>
$V$ ( $\text{deg m}^{-1} \text{T}^{-1}$ )	209	528	-145

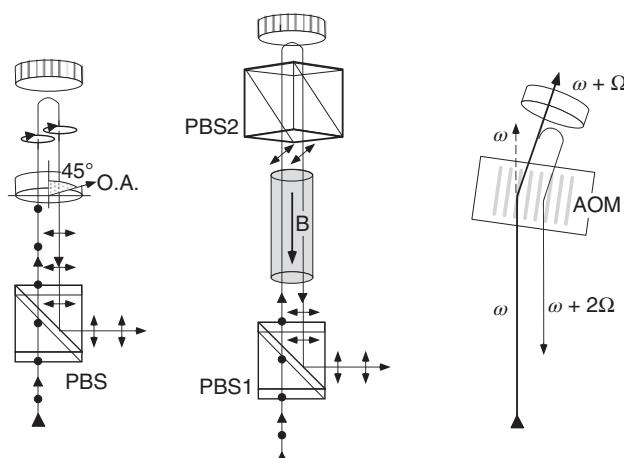
a) TGG = terbium–gallium garnet.

### 3.5.6 Optical Isolators and Diodes

In most applications laser light is sent to the device under test via various optical components. Therefore back-reflections always occur, which even for very low intensities cause undesirable amplitude and frequency fluctuations of the laser light. Optical isolators offer the possibility to decouple experiment and light source from each other. The components introduced in the preceding sections are the building blocks of such devices.

In Figure 3.28 we present three concepts that may be applied to suppress reflections from the upper reflector:

- The isolator on the left uses a  $\lambda/4$  plate, which transforms the linear polarization into circular polarization, for example, a right-handed one. After reflection, the handedness is preserved, but the wave propagates backward. After the second passage through the  $\lambda/4$  retarder plate, the action of a  $\lambda/2$  plate is realized. The polarization is thus rotated by  $90^\circ$ , and the wave is deflected at the polarizing beam splitter (see Section 3.4.4). This arrangement is only sensitive to the reflection of circularly polarized light and hence is of limited use.
- In contrast to the previous case, the Faraday isolator (center) allows the suppression of arbitrary reflections only in combination with a second polarizer between rotator and mirror. One disadvantage is the technically impractical rotation by  $45^\circ$ , which can be compensated with a  $\lambda/2$  plate or a second rotator stage [33]. A two-stage isolator also offers typically 60 dB extinction of reflections in contrast to the typical 30 dB of a single-stage unit.
- From time to time the AOM is applied for isolation purposes. Its isolation effect is based on the frequency shift of the reflected light by twice the modulation frequency, which, for example, may lie outside of the bandwidth of the laser light source to be isolated.



**Figure 3.28** Optical isolators. (i) Left:  $\lambda/4$  plate isolator. (ii) Center: Faraday isolator. (iii) Right: the effect of an AOM isolator is based on frequency shifts. PBS, polarizing beam splitter.

## Problems

- 3.1 Phase shift in total internal reflection (TIR)** Begin by showing that the coefficient of dielectric reflection for p polarization (see p. 86) is alternatively given by  $r = E_{0r}/E_{0i} = (n_1 \cos \theta_t - n_2 \cos \theta_i)/(n_1 \cos \theta_t + n_2 \cos \theta_i)$ . Consider the reflection coefficient for the case of total internal reflection ( $n_2 < n_1$ ), that is, for an angle of incidence above the critical angle  $\theta_i > \theta_c = \sin^{-1}(n_2/n_1)$ . Show that the coefficients for both s and p polarization and for  $\theta_i > \theta_c$  take on the value  $R = |r|^2 = 1$ . Use the generalized Snellius' law  $\cos(\theta_t) = (1 - \sin \theta_i / \sin \theta_c)^{1/2} = iQ$ ,  $Q$  real, for this purpose. Show furthermore that the phase shift of the reflected waves is  $\varphi_s = 2\tan^{-1}(n_1 \cos \theta_i / n_2 Q)$  and  $\varphi_p = 2\tan^{-1}(n_1 Q / n_2 \cos \theta_i)$ , respectively.

- 3.2 Phase shift at metallic surfaces** Show that on reflection from a metallic surface (refractive index  $n = n' + in'' = n(1 + ik)$ ), a phase shift

$$\tan \phi = \frac{2n\kappa}{n^2 - 1 + n^2\kappa^2}$$

occurs for normal incidence. Show that for a perfect conductor (conductivity  $\sigma \rightarrow \infty$ ), the phase shift vanishes,  $\tan \phi \rightarrow 0$ .

- 3.3 Beam splitter made from metal films** Simple beam splitters can be manufactured by depositing a thin metal film onto a glass substrate. Determine the width of the thin layer with conductivity  $\sigma$  so that the intensity is evenly distributed between transmitted and reflected beams. For the sake of simplicity, consider normal incidence.

- 3.4 Polarizers** Collect information from books and catalogues and describe the polarizing mechanism of the following polarizing components: Glan–Taylor, Glan–Thompson, Rochet, Snarmont, and Nicol prisms; thin-film polarizers; and polarization sheets (also Polaroid filter). Give specifications for transmission, extinction ratio, and acceptance angle.

- 3.5 Wollaston prisms** Two triangular rectangle prisms with even-sided cross section are cut from a uniaxial crystal (usually quartz). One crystal has the optical axis in, the other one orthonormal to, the cross section. The crystals are cemented to each other at the hypotenuses (sometimes simply with a droplet of water). What happens to an unpolarized light beam incident onto one of the four sides of the polarizer? Which angles are observed for calcite?

- 3.6 Switching laser beams** In experiments and applications, laser beams must be switched off and on. Consider several methods and discuss their limits, advantages, and disadvantages:

- mechanical shutters,
- AOMs, and
- EOMs.

- 3.7 AOM with standing acoustic wave** An AOM is excited with an acoustic standing wave of frequency  $\Omega$ . Model the index of refraction variation in terms of an optical phase grating. Study the effect on a laser beam incident at normal angles to the acoustic wave and at an angle  $\alpha$ .
- 3.8 Faraday isolator.** Using the Verdet constants from Table 3.4, calculate the field strength required for the rotator of a Faraday isolator at the wavelength  $\lambda = 589 \text{ nm}$ . How can one generate such magnetic fields? See [33].

**4**

## Light Propagation in Structured Matter

Traditional optical components, such as lenses, mirrors, and so on, are used to control the propagation of light beams along their natural paths, mostly free space trajectories. With the advent of miniaturization of almost all known materials, it has become possible to confine and control light propagation to structured materials down to the wavelength scale and even below. In this chapter we will outline (and restrict ourselves to) the conceptual and technological foundations for photonic components and materials. Photonic devices offer compact and low-energy solutions for numerous applications including sensing and data processing and complementing and extending the range of electronic technologies for the twenty-first century.

We will distinguish roughly three major classes of structured matter:

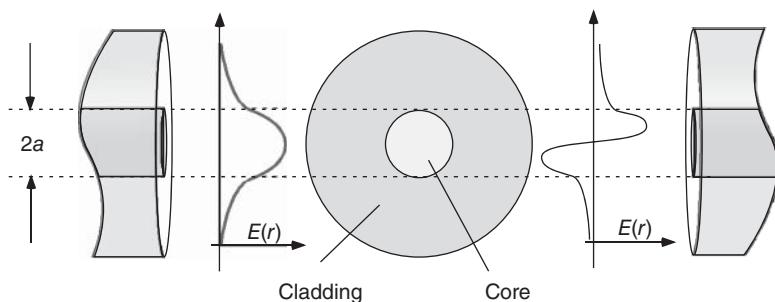
- Dielectric wave guides (optical fibers) have been used for many decades now, and today they have established the backbone of our worldwide information grid. The breakthrough occurred when Charles Kao (1933–, Nobel Prize in 2009 “for groundbreaking achievements concerning the transmission of light in fibers for optical communication”) predicted [34] that with low-loss silica wave guides, optical signals should remain strong enough to be detected even after 100 km propagation distance.
- Photonic materials have been inspired by the properties of electron motion in solids, especially in dielectric periodic structures. Engineering of artificial materials with periodic variation of the index of refraction offers a wealth of schemes for controlled light propagation and integration with other components.
- Metamaterials are structured materials designed to shape the local electric and magnetic response of materials at scales smaller than the relevant wavelengths, leading to an average effective index of refraction. With metamaterials the range of available indices of refraction is extended beyond naturally occurring materials. New functionalities are made available, such as negative indices of refraction.

## 4.1 Optical Wave Guides and Fibers

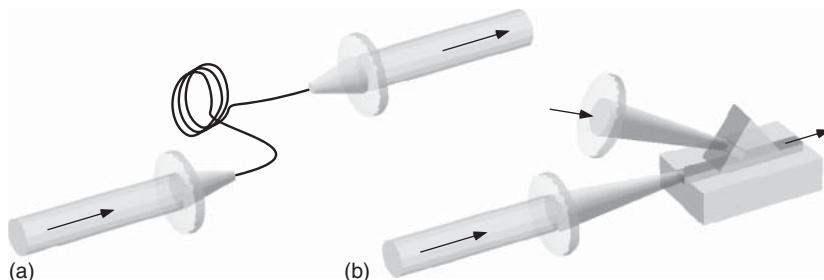
Let us extend the treatment of optical fibers in ray optics (Section 1.7) by solving the Helmholtz equation (2.14) in order to analyze the properties of wave propagation in optical fibers. We concentrate on wave guides that have a cylindrical cross section (commonly also called *optical fibers*; Figure 4.1) and, as mentioned before in Section 1.7, constitute the backbone of optical networks – from short-range interconnections for local cross-linking of devices up to overseas cables for optical telecommunications.

Wave guides are also an important basic element of integrated optics. Here, planar structures are preferred (Figure 4.2), onto which transverse structures can be fabricated by well-known techniques of semiconductor technology. In  $\text{LiNbO}_3$ , for example, the index of refraction may be varied within approximately 1% by in-diffusion of protons, creating wave guides just below the surface of planar crystals that have a nearly rectangular profile of the index of refraction.

A mathematical investigation of the waveforms of an optical fiber is rather tedious and involved. As an example, let us sketch the treatment of the cylindrical step-index fiber, the most important type for applications.



**Figure 4.1** Step-index fiber with characteristic transverse field distributions (modes). The curvature of the modes must be positive in the core and negative in the cladding. The left side shows the field distribution for the fundamental mode, and the right side for the next higher mode.



**Figure 4.2** Types of wave guides. (a) Cylindrical, mechanically very flexible fibers (light is coupled in and out by lenses) are used for transmission over long distances. (b) Wave guides with rectangular cross section just below the surface of suitable substrates (e.g.,  $\text{LiNbO}_3$ ) play an important role in integrated optics. Coupling can be performed via an edge or by frustrated total internal reflection (FTIR) with a prism on top.

### 4.1.1 Step-Index Fibers

The index of refraction in a step-index fiber (Figure 4.1) is cylindrically symmetric and homogeneous within the core and the cladding, respectively. Its value declines from  $n_1$  within the core at  $r = a$  step-like to the value  $n_2$  of the cladding. According to the geometry we look for solutions of the form  $\mathbf{E} = \mathbf{E}(r, \phi) e^{-i(\omega t - \beta z)}$  (and correspondingly for  $\mathbf{H}$ ) with

$$\mathbf{E}(r, \phi) = \mathcal{E}_r(r, \phi) \mathbf{e}_r + \mathcal{E}_\phi(r, \phi) \mathbf{e}_\phi + \mathcal{E}_z(r, \phi) \mathbf{e}_z.$$

The wave equation for cylindrical  $(r, \phi)$  components is complicated, since the  $\mathbf{e}_r$  and  $\mathbf{e}_\phi$  unit vectors are not constant. Fortunately, for the  $\mathcal{E}_z$ - and  $\mathcal{H}_z$ -components, a scalar wave equation still holds, where  $\nabla_\perp(r, \phi)$  stands for the transverse part of the nabla operator:

$$(\nabla_\perp^2 + \mathbf{k}^2 - \beta^2) \begin{Bmatrix} \mathcal{E}_z \\ \mathcal{H}_z \end{Bmatrix} = 0.$$

A complete system of solutions is obtained if one first evaluates the components  $\{\mathcal{E}_z, \mathcal{H}_z\}$  and then constructs  $\{\mathcal{E}_r, \mathcal{E}_\phi, \mathcal{H}_r, \mathcal{H}_\phi\}$  by means of Maxwell's equations,

$$\nabla \times \mathbf{H} = -i\omega \epsilon_0 n_i^2 \mathbf{E} \quad \text{and} \quad \nabla \times \mathbf{E} = i\omega \mu_0 \mathbf{H}, \quad (4.1)$$

the result of which is given in Eqs. (4.5) and (4.6).

The propagation constant  $\beta$  must still be determined, and the Helmholtz equation for  $\{\mathcal{E}_z, \mathcal{H}_z\}$  in cylindrical coordinates with  $k_{1,2} = n_{1,2}\omega/c$  is

$$\left( \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \phi^2} + (k_i^2 - \beta^2) \right) \begin{Bmatrix} E_z(r, \phi) \\ H_z(r, \phi) \end{Bmatrix} = 0.$$

With the help of the trial solutions  $\{\mathcal{E}_z, \mathcal{H}_z\} = \{e(r), h(r)\} e^{\pm i\ell\phi}$ , this is reduced to a Bessel equation for the radial distribution of the amplitudes:

$$\left( \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + k_i^2 - \beta^2 - \frac{\ell^2}{r^2} \right) \begin{Bmatrix} e(r) \\ h(r) \end{Bmatrix} = 0.$$

The curvature of the radial amplitudes  $\{e(r), h(r)\}$  depends on the sign of  $k_i^2 - \beta^2$ . Within the core we can permit positive convex curvatures corresponding to oscillating solutions, but within the cladding, the amplitude must decline rapidly and therefore must have a negative curvature – otherwise, radiation results in an unwanted loss of energy (see Figure 4.1):

$$\begin{array}{ll} \text{Within the core} & 0 < k_\perp^2 = k_1^2 - \beta^2, \\ \text{Within the cladding} & 0 > -k^2 = k_2^2 - \beta^2. \end{array} \quad (4.2)$$

In other words, the propagation constant must have a value between the wave numbers  $k_i = n_i\omega/c$  of the homogeneous core and of the cladding material,

$$n_1\omega/c \leq \beta \leq n_2\omega/c,$$

and differs only a little from  $k_{1,2}$  for small differences in the index of refraction  $\Delta = (n_1 - n_2)/n_1$  (Eq. (1.7)). Such wave guides are called *weakly guiding*.

By definition of Eq. (4.2), we have  $k_\perp^2 + \kappa^2 = k_1^2 - k_2^2$ . Since  $k_1 \approx k_2$ , the transverse wave vectors  $k_\perp$  and  $\kappa$  are small compared with the propagation constant  $\beta$ :

$$k_\perp^2 + \kappa^2 = (\omega/c)^2(n_1^2 - n_2^2) \approx 2\Delta(n_1\omega/c)^2 \ll \beta^2. \quad (4.3)$$

For  $k_1 \simeq k_2 \simeq \beta$ , the transverse wave vectors  $k_{\perp}$  and  $\kappa$  are small compared with the propagation constant  $\beta$ . The transverse solution must remain finite, thereby keeping only the Bessel functions  $J_{\ell}$  (and modified Bessel functions  $K_{\ell}$ ) of the first kind within the core and the cladding, respectively. For the sake of transparency, we introduce scaled coordinates  $X := k_{\perp}r$ ,  $Y := \kappa r$  and  $X_a := k_{\perp}a$ ,  $Y_a := \kappa a$ , respectively:

$$e(r) = \begin{cases} A \cdot \frac{J_{\ell}(X)}{J_{\ell}(X_a)} & \xrightarrow{r \rightarrow 0} (k_{\perp}r)^{\ell} \quad \text{core,} \\ A \cdot \frac{K_{\ell}(Y)}{K_{\ell}(Y_a)} & \xrightarrow{r \rightarrow \infty} e^{-\kappa r}/\sqrt{\kappa r} \quad \text{cladding,} \end{cases} \quad (4.4)$$

$$h(r) = \begin{cases} B \cdot \frac{J_{\ell}(X)}{J_{\ell}(X_a)} & \xrightarrow{r \rightarrow 0} (k_{\perp}r)^{\ell} \quad \text{core,} \\ B \cdot \frac{K_{\ell}(Y)}{K_{\ell}(Y_a)} & \xrightarrow{r \rightarrow \infty} e^{-\kappa r}/\sqrt{\kappa r} \quad \text{cladding.} \end{cases}$$

By properly defining the coefficients  $\{A, B\}$ , we have already taken care that the components  $\{\mathcal{E}_z, \mathcal{H}_z\}$  are continuous at  $r = a$ . For the  $\{\mathcal{E}_r, \mathcal{E}_{\phi}, \mathcal{H}_r, \mathcal{H}_{\phi}\}$  components, we obtain conditions from Eqs. (4.1) in cylindrical coordinates:

$$\begin{aligned} -i\omega\epsilon_0 n_i^2 \mathcal{E}_r &= \frac{i\ell h(r)}{r} e^{i\ell\phi} - i\beta \mathcal{H}_{\phi}, \\ -i\omega\epsilon_0 n_i^2 \mathcal{E}_{\phi} &= i\beta \mathcal{H}_r - \frac{\partial}{\partial r} h(r) e^{i\ell\phi}, \\ i\omega\mu_0 \mathcal{H}_r &= \frac{i\ell e(r)}{r} e^{i\ell\phi} - i\beta \mathcal{E}_{\phi}, \\ i\omega\mu_0 \mathcal{H}_{\phi} &= i\beta \mathcal{E}_r - \frac{\partial}{\partial r} e(r) e^{i\ell\phi}. \end{aligned}$$

Insertion of the solutions for  $\{e(r), h(r)\}$  and utilization of scaled coordinates  $X = k_{\perp}r$  yields radial components

$$\begin{aligned} \mathcal{E}_r(X, \phi) &= i\beta a \left( \frac{\omega\mu_0}{\beta} \frac{iB\ell}{XX_a} \frac{J_{\ell}(X)}{J_{\ell}(X_a)} + \frac{A}{X_a} \frac{J'_{\ell}(X)}{J_{\ell}(X_a)} \right) e^{i\ell\phi}, \\ \mathcal{H}_r(X, \phi) &= \beta a \left( \frac{\omega\epsilon_0 n_i^2}{\beta} \frac{A\ell}{XX_a} \frac{J_{\ell}(X)}{J_{\ell}(X_a)} + \frac{iB}{X_a} \frac{J'_{\ell}(X)}{J_{\ell}(X_a)} \right) e^{i\ell\phi}, \end{aligned} \quad (4.5)$$

and azimuthal contributions

$$\begin{aligned} \mathcal{E}_{\phi}(X, \phi) &= i\beta a \left( \frac{iA\ell}{XX_a} \frac{J_{\ell}(X)}{J_{\ell}(X_a)} - \frac{\omega\mu_0}{\beta} \frac{B}{X_a} \frac{J'_{\ell}(X)}{J_{\ell}(X_a)} \right) e^{i\ell\phi}, \\ \mathcal{H}_{\phi}(X, \phi) &= \beta a \left( \frac{B\ell}{XX_a} \frac{J_{\ell}(X)}{J_{\ell}(X_a)} + \frac{\omega\epsilon_0 n_i^2}{\beta} \frac{iA}{X_a} \frac{J'_{\ell}(X)}{J_{\ell}(X_a)} \right) e^{i\ell\phi}. \end{aligned} \quad (4.6)$$

Here solutions (4.5) and (4.6) are valid for the core. For the cladding the corresponding solution is found by replacing  $X \rightarrow Y, J_{\ell} \rightarrow K_{\ell}$ .

To determine the propagation constant  $\beta$ , we use boundary conditions (3.2) at  $r = a$  and  $X = X_a$  in Eq. (4.6). A short calculation yields a linear system of equations in  $\{A, B\}$ ,

$$\begin{aligned} B \frac{\omega \mu_0}{\beta} \left( \frac{J'_\ell(X)}{X J_\ell(X)} + \frac{K'_\ell(Y)}{Y K_\ell(Y)} \right) + -i\ell A \left( \frac{1}{X^2} + \frac{1}{Y^2} \right) &= 0, \\ A \frac{\omega \epsilon_0}{\beta} \left( \frac{n_1^2 J'_\ell(X)}{X J_\ell(X)} + \frac{n_2^2 K'_\ell(Y)}{Y K_\ell(Y)} \right) + i\ell B \left( \frac{1}{X^2} + \frac{1}{Y^2} \right) &= 0, \end{aligned}$$

and yields a characteristic eigenvalue equation

$$\left( \frac{J'_\ell(X)}{X J_\ell(X)} + \frac{K'_\ell(Y)}{Y K_\ell(Y)} \right) \left( \frac{k_1^2 J'_\ell(X)}{X J_\ell(X)} + \frac{k_2^2 K'_\ell(Y)}{Y K_\ell(Y)} \right) = \ell^2 \beta^2 \left( \frac{1}{X^2} + \frac{1}{Y^2} \right)^2. \quad (4.7)$$

An additional condition is derived by multiplying Eq. (4.3) with  $a^2$ , resulting in

$$X_{\ell m}^2 + Y_{\ell m}^2 = \left( \frac{\omega}{c} \right)^2 (n_1^2 - n_2^2) a^2 = V^2. \quad (4.8)$$

Here we have introduced the *V parameter*, which is directly related to the numerical aperture (NA) through Eq. (1.9) and corresponds to a normalized wave vector  $ka$ :

$$V = \frac{\omega a}{c} \cdot \text{NA} = ka \cdot \text{NA}. \quad (4.9)$$

For every wavelength  $\lambda$ , the *V* parameter fully accounts for the physical properties of the step-index fiber – the indices of refraction ( $n_1, n_2$ ) and the core radius  $a$ . For every  $V$  and  $\ell = 0, 1, 2, \dots$ , numerical treatment of the transcendental equation (4.7) in accord with Eq. (4.8) yields a set of solutions  $(X_{\ell m}, Y_{\ell m})$  with  $m = 1, 2, 3, \dots$  as well as the propagation constant  $\beta_{\ell m}$  according to Eq. (4.2):

$$\beta_{\ell m} = (k_1^2 - (X_{\ell m}/a)^2)^{1/2} = (k_2^2 + (Y_{\ell m}/a)^2)^{1/2}.$$

Numerical treatment of this transcendental equation, for  $\ell = 0, 1, 2, \dots$ , gives solutions  $(X_{\ell m}, Y_{\ell m})$  and a propagation constant  $\beta_{\ell m}$  for every frequency or wavelength  $\omega = 2\pi c/\lambda$ , respectively; this treatment is elaborate and is covered extensively in the literature [30]. As we did in Section 1.7 on ray optics, we restrict ourselves to the simplified case of weakly guiding waves with small differences of the indices of refraction.

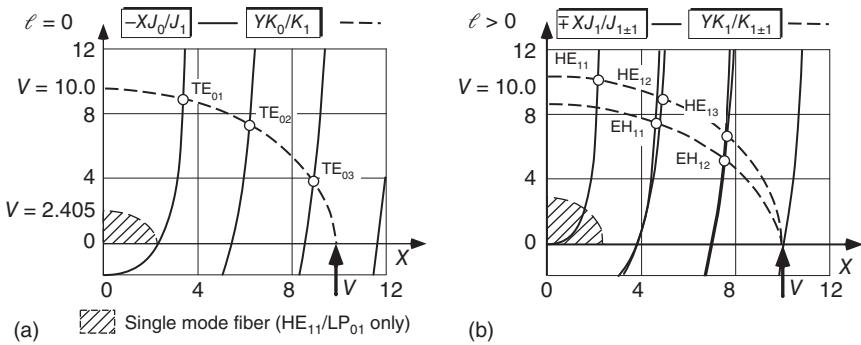
#### 4.1.1.1 Weakly Guiding Step Fibers

In weakly guiding wave guides, we have  $n_1 \simeq n_2$  and  $k_1 \simeq k_2 \simeq \beta$ . Hence Eq. (4.7) can be simplified, yielding

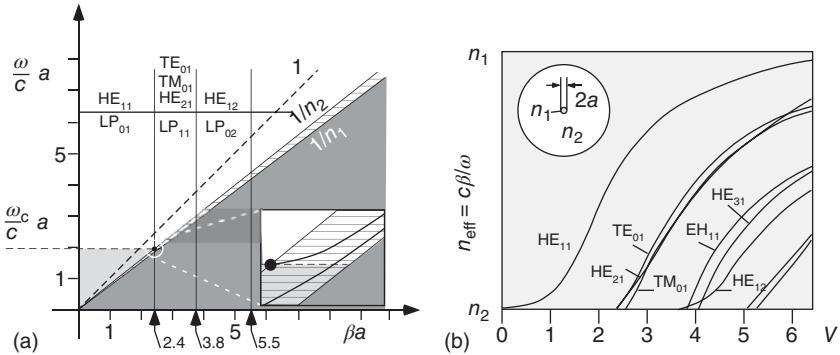
$$\left( \frac{J'_\ell(X)}{X J_\ell(X)} + \frac{K'_\ell(Y)}{Y K_\ell(Y)} \right) = \pm \ell \left( \frac{1}{X^2} + \frac{1}{Y^2} \right). \quad (4.10)$$

We furthermore substitute the derivatives with the identities

$$J'_\ell(X) = \pm J_{\ell \mp 1}(X) \mp \frac{\ell J_\ell(X)}{X}, \quad K'_\ell(Y) = -K_{\ell \mp 1}(Y) \mp \frac{\ell K_\ell(Y)}{Y}. \quad (4.11)$$



**Figure 4.3** Graphical analysis of Eq. (4.12) for a weakly guiding step-index fiber and  $V=10$ ; here we have used the inverse of Eq. (4.12), and the argument of  $YK_\ell/K_{\ell+1}$  is  $\sqrt{V^2 - X^2}$ . (a) TE and TM modes for  $\ell=0$ . The shaded area indicates the region where only one solution exists (single mode). (b) HE and EH modes for  $\ell>0$ .



**Figure 4.4** (a) The  $(\omega, \beta)$  dispersion relation for a weakly guiding step-index fiber. Within the normalized units, the air line, that is,  $\omega = c\beta$ , has unit slope. Below the cutoff frequency  $\omega_c$  (Eq. (4.15)), the wave guide supports a single mode only. (b) Effective index of refraction as a function of the  $V$  parameter (Eq. (4.9)).

For  $\ell = 0$  we find transverse electric (TE) and transverse magnetic (TM) solutions. For both the signs in  $\ell \pm 1$ , we find after some transformations the so-called hybrid modes. The conditions are as follows:

$$\begin{aligned} \ell = 0 \text{ TE}_{\ell m}, \text{ TM}_{\ell m} \quad & \frac{J_0(X_{\ell m})}{X_{\ell m} J_1(X_{\ell m})} = - \frac{K_0(Y_{\ell m})}{Y_{\ell m} K_1(Y_{\ell m})}, \\ \ell \geq 1 \text{ HE}_{\ell m} \quad & \frac{J_{\ell-1}(X_{\ell m})}{X_{\ell m} J_{\ell}(X_{\ell m})} = \frac{K_{\ell-1}(Y_{\ell m})}{Y_{\ell m} K_{\ell}(Y_{\ell m})}, \\ \ell \geq 1 \text{ EH}_{\ell m} \quad & \frac{J_{\ell+1}(X_{\ell m})}{X_{\ell m} J_{\ell}(X_{\ell m})} = - \frac{K_{\ell+1}(Y_{\ell m})}{Y_{\ell m} K_{\ell}(Y_{\ell m})}. \end{aligned} \quad (4.12)$$

In Figure 4.3 we have sketched graphical solutions for the additional condition (4.8).

From the solutions generated, for example, by the methods shown in Figure 4.3, we can construct the dispersion relation  $\omega(\beta_{\ell m})$  displayed in Figure 4.4 with normalized dimensionless units  $\omega a/c$  versus  $\beta a$ . It exhibits the following characteristic properties:

- Left: In normalized units, the dispersion relations  $\omega a/c$  versus  $\beta a$  in a homogeneous material with refractive index  $n$  have slope  $1/n$ . For a step-index fiber, the lines with slopes  $1/n_1$  and  $1/n_2$  divide the diagram into three regions. Below  $\omega a/c = 1/n_1$  in the dark shaded area, no propagation modes exist. Above the line with  $\omega a/c = 1/n_2$ , radiation modes can freely propagate in the cladding and above  $\omega a/c = 1$  (the air line) also in air. An individual mode is characterized by  $(\omega, \beta)$ , where  $\beta$  gives the component of the wave vector along the fiber axis. Thus the spectrum is continuous above  $\omega a/c = (1/n_2)\beta a$ .

In between the lines  $1/n_1$  und  $1/n_2$ , propagation is only possible for discrete  $(\omega, \beta)$  values of the guided modes. Within the light gray shaded area, the fiber has single-mode character, that is, for every frequency  $\omega$  there exists a unique propagation constant  $\beta$  (see Sections 4.1.1.2 and 4.1.1.3). The magnified cutout shows the end of the single-mode area where the next modes enter.

- Right: Sometimes it is useful to introduce an effective index of refraction instead of the propagation constant:

$$n_{\text{eff}} = \beta/k = c\beta/\omega \quad \text{with} \quad n_2 < n_{\text{eff}} < n_1. \quad (4.13)$$

The effective refractive index also allows a more transparent overview of the spectrum of modes allowed between the lines with  $1/n_1$  and  $1/n_2$ .

Equation (4.4) gives one more condition from Eq. (4.12), which fixes the ratio of the electric ( $A$ ) and magnetic amplitudes ( $B$ ). The “+” sign holds for the HE modes, and the “−” sign for the EH modes:

$$(A \pm i(\omega\mu_0/\beta)B)\ell = 0. \quad (4.14)$$

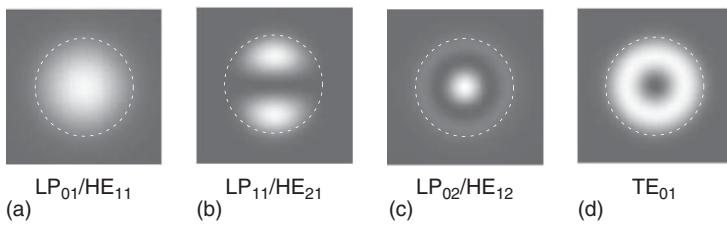
Because of the imaginary coefficient, the electric and magnetic components are temporally  $90^\circ$  out of phase. In order to obtain an overview of the geometric properties of the optical wave guide modes, let us have a look at some special cases.

#### 4.1.1.2 $\ell = 0$ : TE and TM Modes

For  $\ell = 0$  we have  $A = 0$  ( $B = 0$ ), that is, either the  $E$ - or the  $H$ -field is purely transverse. Hence for  $\ell = 0$ , the TE/TM denominations are sensibly used. We have indicated the graphical conditions for the (degenerate)  $\text{TE}_{0m}$  and  $\text{TM}_{0m}$  modes in Figure 4.3. TE/TM modes can be guided only for  $V > X_{01} = k_\perp a = 2.405$  ( $J_0(2.405) = 0$ ) or above the corresponding cutoff frequency

$$\omega_c = 2.405(n_1^2 - n_2^2)^{-1/2}(c/a), \quad (4.15)$$

which is directly obtained from Eqs. (4.2) and (4.8). Higher-order TE/TM modes appear above  $V = 5.520$  ( $J_1(5.520) = 0$ ).



**Figure 4.5** Intensity distribution ( $E$ -field) for selected step-index fiber modes. A fiber with numerical aperture ( $\text{NA}$ ) = 0.12 and core diameter  $2a = 5.1 \mu\text{m}$  was used as model. The dashed circle indicates the core. (a)  $\lambda = 850 \text{ nm}$ ,  $V = 2.26$ ; (b–d)  $\lambda = 400 \text{ nm}$  and  $V = 8.81$ .

#### 4.1.1.3 $\ell \geq 1$ : HE and EH Modes

The lowest-order mode is the  $\text{HE}_{11}$  mode (Figure 4.5), which exists down to  $X = 0$  or  $\omega = 0$ . This mode is “tacked” to the core for arbitrarily small frequencies and correspondingly small curvatures of the transverse amplitudes. With decreasing frequency the portion of the energy propagating within the cladding increases more and more. For the mathematical analysis, we assumed the cladding to have an infinite extension – with negligible error, however.

Owing to Eqs. (4.12) and (4.14), the HE and EH modes differ. This difference not only is revealed by the different propagation constants but also manifests itself in the domination of the  $z$ -components in the corresponding  $H$  (HE) and  $E$  (EH) parts.

#### 4.1.1.4 $\ell \geq 1$ : LP Modes

According to Eq. (4.14) we have  $A = \pm i(\omega\mu_0/\beta)B$  for  $\ell > 0$ . By substitution in Eqs. (4.5) and (4.6), application of the recursion formulae (4.11), and a short calculation, one can show that the modes  $\text{HE}_{\ell,m}$  have linear transverse polarization and that the transverse components  $\mathcal{E}_x = \mathcal{E}_r \cos(\phi) + \mathcal{E}_\phi \sin(\phi)$  dominate the longitudinal  $\mathcal{E}_z$  components by the factor  $\beta_{\ell,m}a/X_{\ell,m} \gg 1$ . These modes are also called (*linearly polarized*) LP modes. They are derived from the HE modes and for higher  $\ell$  values from a linear superposition of degenerate  $\{\text{HE}_{\ell+2,m}, \text{EH}_{\ell,m}\}$  modes:

$$\text{HE}_{\ell,m} \rightarrow \text{LP}_{\ell-1,m} \quad \text{and} \quad \text{HE}_{\ell+2,m}, \text{EH}_{\ell,m} \rightarrow \text{LP}_{\ell+1,m} \quad (\ell \geq 2).$$

#### Example: Core diameter of a single-mode wave guide

In manufacturers' catalogues wave guides for single-mode applications are typically presented with specifications for the numerical aperture (NA) and the cutoff wavelength, for instance:

NA	0.13	0.12	0.11
$\lambda$ cutoff (nm)	1260	800	620

From these specifications we can calculate the core diameter  $2a$  using Eq. (4.9) and  $2a = V\lambda/(\pi\text{NA})$ , yielding

2a ( $\mu\text{m}$ )	7.4	5.1	4.3
----------------------	-----	-----	-----

### 4.1.2 Graded-Index Fiber

The term “quadratic index media” covers all the common systems like gradient-index fibers with parabolic index profile (see Figure 4.6) that we have already dealt with in the section on ray optics and may be treated like the limiting case of an infinite thick lens. Realistic gradient fibers have a quadratic profile in the center only, which then continues into a step-like form again. Instead, we look at a simplified, purely quadratic system, which reflects already the properties of a graded-index fiber. The index of refraction depends on the normalized radius  $r/a$ , and making use of the difference in the index of refraction,  $\Delta = (n_1 - n_2)/n_1$  (see Section 1.7.3), we find

$$n(\rho) = n_1[1 - \Delta(r/a)^2] \quad \text{and} \quad \Delta \ll 1.$$

We seek solutions to the Helmholtz equation (2.14), whose envelope does not change along the direction of propagation, that is, of the form  $\mathcal{E}(x, y, z) = \mathcal{A}(x, y) \exp(i\beta z)$ , and get the modified equation

$$\{\nabla_{\perp}^2 + n_1^2 k^2 - 2n_1^2 k^2 \Delta[(x/a)^2 + (y/a)^2] - \beta^2\} \mathcal{A}(x, y) = 0$$

using  $\{n_1 k[1 - \Delta(r/a)^2]\}^2 \simeq (n_1 k)^2 [1 - 2\Delta(r/a) + \dots]$ . Let us assume now, as in the case of higher Gaussian modes (see Section 2.3.4), that the transverse distribution corresponds to modified Gaussian functions:

$$\mathcal{A}(x, y) = \mathcal{F}(x) e^{-(x^2/x_0^2)} \mathcal{G}(y) e^{-(y^2/y_0^2)}.$$

With this ansatz we find

$$\begin{aligned} & \left( \mathcal{F}'' - \frac{4x}{x_0^2} \mathcal{F}' - \frac{2}{x_0^2} \mathcal{F} \right) \mathcal{G} + \left( \mathcal{G}'' - \frac{4y}{y_0^2} \mathcal{G}' - \frac{2}{y_0^2} \mathcal{G} \right) \mathcal{F} + n_1^2 k^2 \mathcal{F} \mathcal{G} \\ & + \left[ \left( \frac{4}{x_0^4} - \frac{2n_1^2 k^2 \Delta}{a^2} \right) x^2 + \left( \frac{4}{y_0^4} - \frac{2n_1^2 k^2 \Delta}{a^2} \right) y^2 \right] \mathcal{F} \mathcal{G} - \beta^2 \mathcal{F} \mathcal{G} = 0, \end{aligned}$$

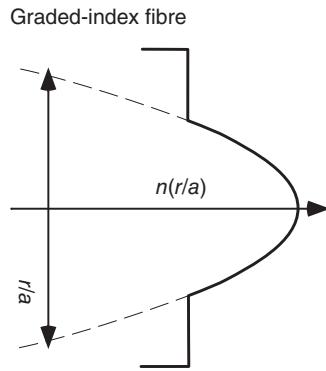
where the unpleasant quadratic term in general can be eliminated by choosing

$$kx_0 = ky_0 = (ka)^{1/2}/(2n_1^2 \Delta)^{1/4} \gg 1.$$

By substituting  $\sqrt{2} x/x_0 \rightarrow u$  and  $\sqrt{2} y/y_0 \rightarrow v$ , we transform again to Hermite's differential equation that we already know from the higher Gaussian modes. With indices  $m$  and  $n$ , we find

$$\begin{aligned} & 2(\mathcal{F}'' - 2u\mathcal{F}' + 2m\mathcal{F}) \mathcal{G} + 2(\mathcal{G}'' - 2v\mathcal{G}' + 2n\mathcal{G}) \mathcal{F} \\ & + [n_1^2 k^2 x_0^2 - \beta^2 x_0^2 - 4(m+n+1)] \mathcal{F} \mathcal{G} = 0. \end{aligned}$$

The terms of the upper row are constructed to vanish upon inserting the Hermite polynomials  $\mathcal{H}_{m,n}$  (see Eq. (2.37)). After a short calculation one gets for the



**Figure 4.6** Simplified profile of the index of refraction of a GRIN fiber.

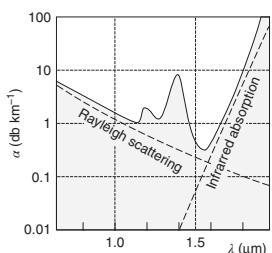
propagation constant, which is the center of interest,

$$\beta_{mn}(\omega) = n_{\text{eff}} \frac{\omega}{c} = \frac{n_1 \omega}{c} \sqrt{1 - \frac{4\sqrt{2\Delta}(m+n+1)}{n_1 k a}}.$$

The transverse distribution of the amplitudes also corresponds to the ones shown in Figure 2.10. But in contrast to the Gaussian modes, the mode diameters ( $x_0, y_0$ ) do not change. This example of a simplified GRIN fiber illustrates that multimode fibers, in addition to “material dispersion” characterized by a frequency-dependent index of refraction, show “mode dispersion”. This influences the form of pulses, because individual partial modes have different propagation velocities.

#### 4.1.3 Fiber Absorption

One could not imagine the success of optical fibers without their extraordinarily advantageous very low absorption properties (Figure 4.7). On the short-wavelength side, these are limited by Rayleigh scattering at small inhomogeneities rising like  $\propto 1/\lambda^4$ .



**Figure 4.7** Absorption properties of optical fibers made of silicate glass.

On the long-wavelength side, absorption increases again due to infrared absorption by the wings of the phonon spectrum. The wavelengths 1.3 and 1.55  $\mu\text{m}$ , very important for telecommunications, coincide with very small absorption coefficients, and simultaneously the group velocity dispersion vanishes for standard silica fibers at 1.3  $\mu\text{m}$  (see Figure 3.10). In between we find resonances that are caused, for example, by OH contamination in the glass.

#### 4.1.4 Functional Types and Applications of Optical Fibers

##### 4.1.4.1 Multimode Fibers

Inspection of Figure 4.3 shows that with growing  $V$  parameter (which corresponds to a scaled frequency),  $V = (\omega a/c)(n_1^2 - n_2^2)^{1/2}$  (Eq. (4.9)), with nearly even spacing at  $V = \pi, 2\pi, 3\pi, \dots$  a new mode occurs with initial parameters  $Y_{\ell m} = 0$  and  $\beta = n_2 \omega / c$ . All guided modes are located in Figure 4.4 between the lines with slopes  $1/n_1$  and  $1/n_2$  describing propagation in the homogeneous core and dielectric cladding. With growing frequency, ever larger  $\ell$  values are important. One can show [30] that the number of modes  $M$  increases quadratically,

$$M \approx V^2/2,$$

not linearly with the  $V$  parameter (see Problem 4.4).

When light is coupled into a multimode fiber, a superposition of multiple transverse modes is excited, which propagate with different velocities (“modal dispersion”). At the exit of the wave guide, the multimode field is transformed into a free field whose transverse and temporal profile is deformed as a result of dispersion. Such fibers are used if transverse coherence (i.e., well-controlled phase fronts) is not very important, for example, for optical pumping of high-power solid-state lasers (see Section 8.4.2).

#### 4.1.4.2 Single-Mode Fibers

The condition for single-mode operation of an optical wave guide is straightforwardly expressed by demanding that the  $V$  parameter (4.9) must not exceed the minimum value for the occurrence of the next higher (TE) mode. Thus for

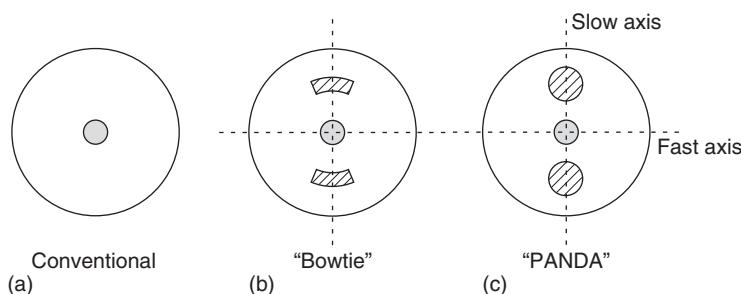
$$V < 2.405, \quad (4.16)$$

a step-index fiber supports propagation of a single mode only, the  $LP_{01}$  or  $HE_{11}$  mode, respectively. The modal properties of the step-index fiber are collected in Figure 4.4 in terms of a dispersion or propagation diagram. For guided modes, free propagation is allowed in the core but not in the cladding. It is remarkable that the narrow hatched region between the lines with slopes  $1/n_{1,2}$  spans the entire region of current optical communication technology!

The lowest-order mode of the cylindrical step-index fiber ( $HE_{11}/LP_{01}$ ) has a bell-shaped profile resembling the transverse envelope of the Gaussian  $TEM_{00}$  mode in a homogeneous medium (Figure 2.10). Therefore a free space Gaussian beam can be efficiently coupled to the ground mode of a single-mode fiber. Indeed, the single-mode fiber is frequently used as an ultimate spatial filter (see also Section 2.3.5). Of the full amplitude distribution launched into the fiber, only the ground mode component propagates along the fiber. Thus a very “clean”  $TEM_{00}$  mode emerges from the fiber output.

#### 4.1.4.3 Polarization-Maintaining (PM) Fibers

A perfect cylindrical step-index fiber is still degenerate with respect to two orthogonal polarization states. Thus the polarization state at the output of a fiber is in general not predictable since small disturbances (e.g., curvatures; see example on p. 108) make the fiber effectively birefringent. Fluctuations because of temperature variations or mechanical motion of the fiber then cause slow variations of the polarization state of the guided wave. Such problems can be avoided with polarization-maintaining (PM) single-mode fibers. They are technically realized by inserting extra elements into the cladding (Figure 4.8), which apply mechanical strain to the core of the fiber, yielding small elliptical deviations of the index of refraction from cylindrical symmetry. Thus the degeneracy of the propagation constants  $\beta$  for the principal axes is lifted, and coupling of the two



**Figure 4.8** Polarization-maintaining (PM) fibers. By insertion of additional elements into the cladding of a cylindrical fiber, (a) mechanical strain is exerted, which controls the birefringent properties of the fiber. The *bowtie* (b) and *PANDA* (c) structures are in common use.

orthogonal polarization modes is suppressed. The polarization of the coupled-in light must be parallel to the principal axis to make use of the characteristics of conserving the polarization.

#### 4.1.4.4 Photonic Crystal Fibers (PCF)

Since about 1995, optical wave guides have been manufactured with special structures widely exceeding the complexity of the index of refraction profile of step-index or GRIN fibers. Such fibers are a very active field of research since their properties (including dispersion) can be tailored with great variability. Photonic crystal fibers (PCFs) are a special case of the subject of the next section, *photonic materials*, where they are treated in a special section (Section 4.2.6).

## 4.2 Dielectric Photonic Materials

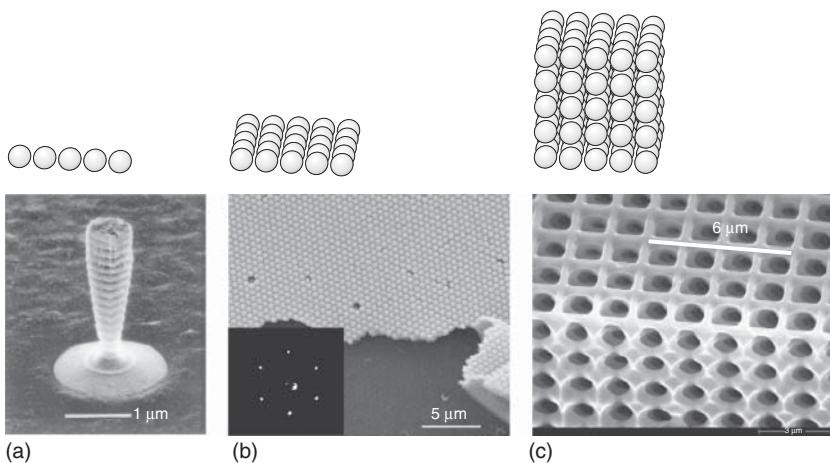
Up to now we have studied the propagation of light in more or less homogeneous materials, with the important exceptions of dielectric boundaries and slowly varying indices of refraction. Today numerous methods of microstructuring are available to tailor dielectric matter at the wavelength scale of light and below. Taking advantage of these methods leads to *photonic materials* that have properties that allow to control propagation of light in a manner not available from naturally occurring materials.

Structured dielectric materials with periodic modulation of the index of refraction in two or three dimensions have a key role for photonic materials. Since such samples resemble the periodic properties of crystals, they are also called *photonic crystals*. In contrast to natural crystals, the periods are of the order of the wavelength of light, that is, several 100 nm to  $\mu\text{m}$  for visible and near-visible wavelength ranges. Photonic crystals are usually manufactured in a laboratory. However, nature provides examples of materials whose wealth of colors is a direct consequence of periodic structures, for example, the beautiful wings of butterflies or opal gems.

### 4.2.1 Photonic Crystals

In Figure 4.9 we show examples of photonic crystals with different dimensionalities:

- Figure 4.9a: The one-dimensional (1D) pillar consists of GaAs layers with periodically varying composition [35] providing axial confinement of light. In the transverse direction the propagation of light in this *hybrid structure* is suppressed by total internal reflection (TIR) (see Section 3.1.2) such that a closed resonator is formed.
- Figure 4.9b: Two-dimensional (2D) crystals are manufactured by conventional methods of microstructuring (see p. 158) or alternatively by self-organization as for the present example [36]. 2D crystals are of interest for integrated optics since light is confined to the thin layer if the refractive index of the adjacent material is smaller. In abPcFs (Section 4.2.6), light is guided along the fiber axis by transverse photonic structures.



**Figure 4.9** Early examples of photonic crystals (see text). With permission from Y. Yamamoto ((a) [35]), M. Giersig ((b) [36]) and R. Wehrspohn ((c) [37]).

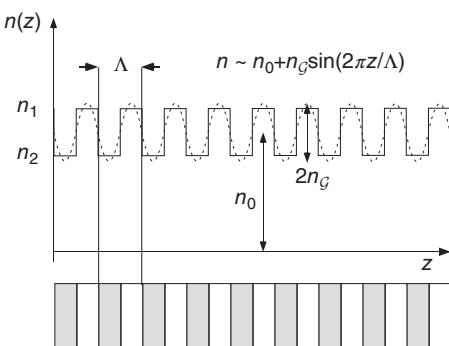
- Figure 4.9c: The photonic crystal shown here was obtained by controlled photochemical etching along the crystallographic (100) direction of a Si crystal [37]. It exhibits a so-called photonic bandgap (PBG) at the infrared wavelength of 5  $\mu\text{m}$ .

The propagation of light in photonic crystals – that is, in periodic dielectric structures – is theoretically well described with the notions from solid-state physics: point symmetry groups as well as the reciprocal lattice are important concepts. The band model of electron dynamics has inspired many proposals for applications, with a special emphasis on semiconductor materials with their bandgaps. The engineering potential of these materials has created microelectronics and suggests a similar yet unrealized potential for photonic materials. Theoretical treatment of photonic materials is even more involved than in solid-state physics since scalar electronic wave functions give satisfactory results. Here, the vector nature of Maxwell's equations needs to be accounted for.

For materials with a PBG, propagation of light is fully suppressed for certain wavelength regions. Multilayer systems with periodically varying index of refraction have been produced for a long time; they are used as dielectric mirrors or interference filters (see Section 6.7). They make an excellent 1D model to understand the origin of the bandgap. However, PBG materials are of significant interest only in 2D and 3D where they promise the realization of complex optical circuitry.

#### 4.2.1.1 Light Propagation in 1D Periodically Structured Dielectrics

For an introduction we study the propagation of light in a crystal whose index of refraction is modulated in one direction only with period  $\Lambda$ . The 1D example is closely related to the treatment of dielectric multilayer mirrors in Section 6.7. They are manufactured by deposition of the layers from evaporation sources and were introduced a long time ago.



**Figure 4.10** A periodic system of layers with alternating index of refraction is approximated by a Fourier sum (Eq. (4.18)). The dashed line shows the first-order approximation.

which is well known from solid-state physics [38], and we have for the index of refraction to first order

$$n(z) = n_0 + n_G e^{iGz} + n_{-G} e^{-iGz} + \dots$$

Since \$n(z)\$ is a real function we have \$n\_G = (n\_{-G})^\*\$. In order to find approximate solutions for Eq. (4.17), it is useful to expand \$n^{-2}(z)\$ into a Fourier series as well. For small coefficients \$n\_G \ll n\_0\$, we have

$$n^{-2}(z) \simeq n_0^{-2} - \sum_G \frac{2n_G}{n_0^3} e^{iGz}. \quad (4.18)$$

#### 4.2.2 Bloch Waves

Let us assume that a monochromatic wave traveling in the dielectric with periodic index of refraction variation can be described by a sum of plane waves with coefficients \$e\_K\$:

$$E(z) = \sum_K e_K e^{iKz}.$$

The propagating field \$E(z)\$ is not necessarily periodic in \$\Lambda\$. From the Helmholtz equation (4.17), we obtain by insertion and minor rearrangement

$$\frac{1}{n_0^2} \left[ \sum_K \sum_G \frac{2n_G}{n_0} K^2 e_K e^{i(K+G)z} - \sum_K (n_0^2 k_0^2 - K^2) e_K e^{iKz} \right] = 0.$$

The coefficients of an individual wave vector \$K\$ are obtained by multiplying this equation with \$e^{-iK'z}\$ and using \$\int e^{iK'z} e^{-iKz} dz = \delta(K - K')/\sqrt{2\pi}\$. Only for \$K' = K\$ is the integral nonzero. Hence in this case the factor

$$(n_0^2 k_0^2 - K^2) e_k + \sum_G \frac{2n_G}{n_0} (k - G)^2 e_{k-G} = 0 \quad (4.19)$$

must vanish, where we have set \$K' \rightarrow k\$. This equation connects waves whose \$k\$-vectors differ by a vector of the reciprocal lattice. Thus the solution has the

For our problem it is convenient to consider the Helmholtz equation (2.14) for a plane wave in the form

$$\left( n^{-2}(z) \frac{d^2}{dz^2} + k_0^2 \right) E(z) = 0. \quad (4.17)$$

The index of refraction \$n\$ is a real function \$n(z) = n(z + \Lambda)\$ with periodicity \$\Lambda\$ (Figure 4.10). Thus we can expand it in terms of a Fourier series with \$G = 2\pi/\Lambda\$. Fourier coefficients \$\{0, \pm G, \pm 2G, \pm 3G, \dots\}\$ constitute the *reciprocal lattice*,

form

$$E(z) = \sum_G e_{k-G} e^{i(k-G)z} = \left( \sum_G e_{k-G} e^{-iGz} \right) e^{ikz} = \mathcal{E}_k(z) e^{ikz}. \quad (4.20)$$

The Fourier series in the brackets,  $\sum_G e_{k-G} e^{-iGz}$ , is constructed from reciprocal lattice vectors only and therefore periodic in  $\Lambda$ ,  $\mathcal{E}(z) = \mathcal{E}(z + \Lambda)$ . We have thus established the *Bloch theorem*, which was initially given for electrons moving in a periodic potential. Such electrons are described by a wave function  $\psi_k(r) = u_k(r)e^{ikr}$  where  $u_k(r)$  has the periodicity of the crystal potential. Wave vectors in 1D differ by multiples of  $G = mG$ ,  $m = \pm 1, \pm 2, \pm 3, \dots$ ; hence only wave vectors from the first Brillouin zone are physically significant, that is,  $-G/2 \leq k \leq G/2$  in the description of a specific wave.

#### 4.2.3 Photonic Bandgap in 1D

Determination of the  $e_{k-G}$  coefficients in Eq. (4.20) makes solution of the infinitely large system of equations (4.19) necessary, which in general requires not only approximations but also numerical treatment. For an illustration we restrict ourselves to the special case where only coefficients  $n_0$  and  $n_{\pm 1} = n_{\pm G}$  are nonzero. The equations for the first three coefficients are then given by

$$\begin{aligned} (n_0^2 k_0^2 - k^2) e_k - \frac{2n_1}{n_0} (k - G)^2 e_{k-G} - \frac{2n_{-1}}{n_0} (k + G)^2 e_{k+G} &= 0, \\ (n_0^2 k_0^2 - k^2) e_{k-G} - \frac{2n_1}{n_0} (k - 2G)^2 e_{k-2G} - \frac{2n_{-1}}{n_0} k^2 e_k &= 0, \\ (n_0^2 k_0^2 - k^2) e_{k+G} - \frac{2n_1}{n_0} k^2 e_k - \frac{2n_{-1}}{n_0} (k + 2G)^2 e_{k+2G} &= 0. \end{aligned} \quad (4.21)$$

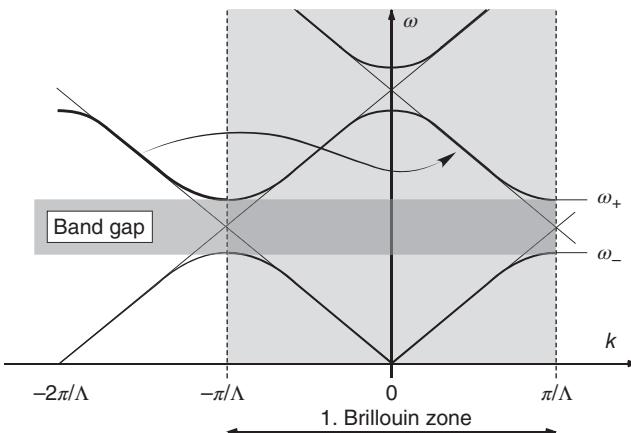
For an estimate of the coefficients, we can use Eq. (4.19):

$$e_k = \frac{\sum_{G>0} 2(n_G/n_0)(k - G)^2 e_{k-G}}{(n_0 \omega/c)^2 - k^2}.$$

In the vicinity of the origin of the Brillouin zone where  $(k, \omega/c) \ll G$ , the denominator of  $e_k$  is given by  $((n_0 \omega/c)^2 - k^2)$ , and for the other coefficients  $e_{k \pm G}$  by  $((n_0 \omega/c)^2 - (k - G)^2) \simeq G^2 \gg |(n_0 \omega/c)^2 - k^2|$ . Therefore, the  $e_k$  coefficient dominates, and to a good approximation, we find the linear relationship  $k = n_0 \omega/c$ . For long wavelengths the propagation constant is thus determined by the average index of refraction  $n_0$ , like in a homogeneous material.

A very different situation occurs at the boundary of the Brillouin zone at  $k \simeq G/2$ . Here we have  $|k - G| \simeq G$ , and at least coefficients  $e_k$  and  $e_{k-G}$  are relevant. We consider  $e_k$  and  $e_{k-G}$  and simplify Eqs. (4.21) by neglecting all other components:

$$\begin{aligned} (n_0^2 k_0^2 - k^2) e_k - \frac{2n_1}{n_0} (k - G)^2 e_{k-G} &= 0, \\ \frac{2n_{-1}}{n_0} k^2 e_k - (n_0^2 k_0^2 - k^2) e_{k-G} &= 0. \end{aligned}$$



**Figure 4.11** Dispersion relation for electromagnetic waves in a 1D photonic crystal with bandgap. Addition of a reciprocal lattice vector  $\mathcal{G}$  shifts the marked branch of the dispersion relation into the first Brillouin zone.

It is well known that this system of linear equations has a solution if the determinant vanishes:

$$((n_0\omega/c)^2 - k^2)^2 - (2|n_1|/n_0)^2(k(k - \mathcal{G}))^2 = 0.$$

The dispersion relation  $\omega$  versus  $k$  is straightforwardly derived from this equation, yielding

$$\omega_{\pm} = \frac{ck}{n_0} \left( 1 \pm \frac{2|n_1|}{n_0} \frac{k - \mathcal{G}}{k} \right)^{1/2} \quad \text{for } k \approx \mathcal{G}/2 = \pi/\Lambda, \quad (4.22)$$

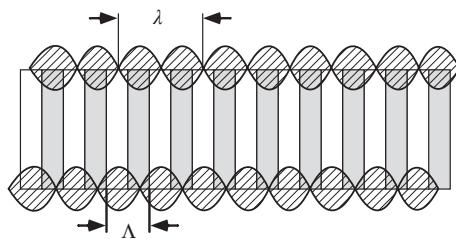
and is qualitatively displayed in Figure 4.11. No solution exists for frequencies between  $\omega_-$  and  $\omega_+$ . In this region the propagation of electromagnetic waves is suppressed in the photonic crystal. In analogy with the band model of semiconductors in solid-state physics, we speak of a *PBG*. The width of the bandgap is determined from Eq. (4.22) for  $k = \mathcal{G}/2$  and not too large  $n_1/n_0$  from

$$\Delta\omega = \omega_+ - \omega_- \simeq \omega_0 \frac{4|n_1|}{n_0},$$

where we have used  $\omega_0 = ck/m_0$ . As expected, the bandgap vanishes with vanishing modulation,  $n_1 \rightarrow 0$ . For  $k = \mathcal{G}/2$  the forward wave ( $k = \mathcal{G}/2$ ) is strongly coupled to the backward wave ( $k - \mathcal{G}/2 = -\mathcal{G}/2$ ). Thus at the bandgap, maximal mixing of waves is realized, resulting in reflection and formation of a standing wave (Figure 4.12). The situation is fully equivalent to the Bragg condition for reflection,

$$k = 2\pi/\lambda = \mathcal{G}/2 = \pi/\Lambda \quad \text{or} \quad 2\Lambda = \lambda,$$

where  $\lambda$  denotes the wavelength within the crystal. In Section 6.7 we will more closely investigate such a multilayer Bragg mirror, which is used to generate highly reflecting dielectric mirrors. The reflection coefficient shown in Figure 6.27 exhibits a large region with little wavelength dependence, which



**Figure 4.12** This drawing offers a physical explanation for the origin of the bandgap. The maxima of the upper standing wave with wavelength  $\lambda$  matching twice the photonic crystal periodicity  $\Lambda = \lambda/2$  rest on zones with higher (and for the lower standing wave with lower) index of refraction. Thus the effective index of refraction is very different and causes an increase or decrease of the frequency compared to a wave traveling in a material with the average refractive index. (See also Figure 10.28.)

is equivalent to the PBG discussed with the present model. Bragg mirrors made from alternating indices of refraction also play an important role for the construction of semiconductor lasers (see Figure 4.9 and Section 10.5.2).

#### 4.2.4 Bandgaps in 2D and 3D

In 1D periodic dielectrics a bandgap always exists, as discussed in the preceding section. It is well known from solid-state physics, however, that in two and three dimensions (2D, 3D), the occurrence of a bandgap (for electrons in semiconductors) depends on details of the crystal symmetry, a statement that holds for photonic crystals as well. PBG that suppress propagation of light in all three or at least two dimensions have received much interest for some years [39, 40]. Again in analogy with semiconductor physics, it has been suggested that dielectric defects (see Section 4.2.5) playing the role of dopant atoms can be used to generate localized (“zero-dimensional”), non-propagating states of optical electromagnetic fields.

##### 4.2.4.1 2D Photonic Crystals

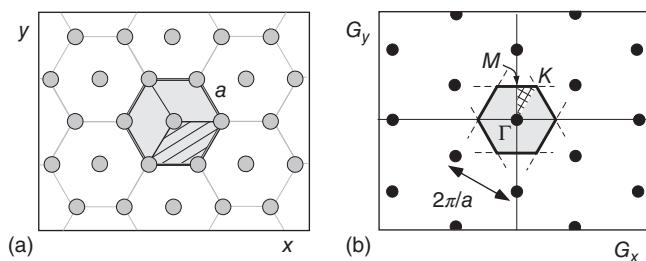
Figure 4.13 shows an important example of a (hexagonal) 2D crystal structure (a) along with its reciprocal lattice (b). Like in the 1D case, the propagation properties are characterized in terms of a dispersion diagram  $\omega(\mathbf{k})$ ,  $\mathbf{k} = (k_x, k_y)$ , which now has 2D character itself, that is, it is a surface in 3D space. An example for hexagonal 2D lattices is given in Figure 4.14. The dispersion relations are typically displayed along the boundaries of the irreducible Brillouin zone instead of giving the full 2D surface.

For 2D photonic crystals we can distinguish two types of systems:

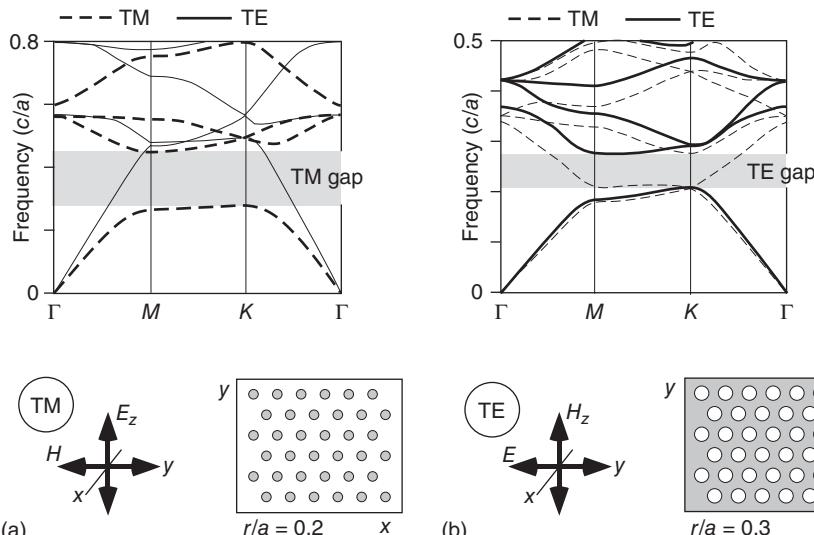
- In thin plane hybrid structures, transverse confinement of light waves to the thin layer is accomplished by index guiding, that is, by TIR from the plane boundaries (Figure 4.15). Propagation within the plane is controlled by the periodic 2D modulation of the refractive index.
- In PCFs light is guided along the axis as in conventional fibers (Section 4.1), but transverse confinement to the axis of the fiber can be realized by 2D PBG.

A special section (Section 4.2.6) is dedicated to these novel components, which have rapidly opened a new era of nonlinear optical interactions because of their very special properties.

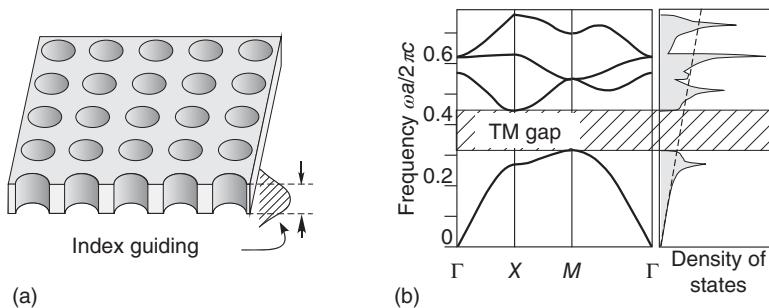
The dispersion relations of Figure 4.14 give detailed information on the propagation of a plane wave with a given  $\mathbf{k}$  vector or propagation constant. In many cases it is sufficient to analyze the so-called density of states, and a bandgap in particular shows up as a gap in the density of states as well. For homogeneous materials the evaluation of the density of states is explained in Section B.3; for periodic structures numerical treatment is in general necessary. An example is given in



**Figure 4.13** (a) Hexagonal two-dimensional crystal lattice with primitive unit cell (hatched). (b) Reciprocal lattice and first Brillouin zone. The dashed lines indicate half the separation of the next reciprocal lattice nodes from the origin. For symmetry reasons all information is already contained in the cross-hatched triangle  $\Gamma$ - $M$ - $K$ . Band structures are commonly displayed by the dispersion relations along the boundaries of this irreducible Brillouin zone (see Figure 4.14).



**Figure 4.14** Examples of bandgaps in 2D photonic crystals. The designations for the reciprocal lattice vectors are given in Figure 4.13. (a) Periodic dielectric cylinders with  $n^2 = 12$ . (b) Periodic cylindrical holes in a dielectric with  $n^2 = 12$ . With kind permission of Steven G. Johnson. Further properties in [41].



**Figure 4.15** (a) Hybrid structures. In a 2D photonic crystal the transverse field distribution of an electromagnetic wave can be confined by total internal reflection at the boundaries of the slab-shaped crystal. (b) Density of states of the TE modes of a 2D photonic crystal formed from a quadratic lattice of cylindrical holes with  $r/a = 0.2$  in a dielectric with  $n^2 = 10$  (from Ref. [42]). The dashed line indicates the expected behavior of the 2D density of states in a homogeneous material (see App. B.3) with averaged refractive index.

Figure 4.15, which shows the density of states of TE modes for the structure on the left side with  $r/a = 0.2$ .

#### 4.2.4.2 3D Photonic Crystals

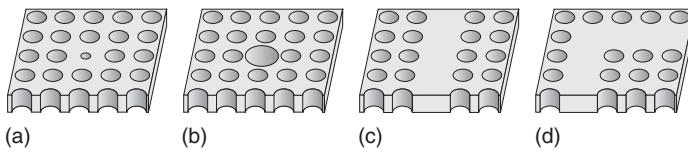
Although in three dimensions the structure of the dispersion relations is necessarily even more complex as a result of the geometry, the third dimension does not introduce novel features beyond the phenomena of 2D structures; a 3D bandgap is characterized again by a gap in the density of states.

The search for fabrication methods of suitable periodically structured dielectric materials that can suppress wave propagation in all three dimensions is a very active area of research. The prediction of the propagation properties of a given structure already requires challenging and theoretical calculation and by far exceeds the scope of this text. They are in fact more elaborate than for electron propagation in semiconductor crystals, which can be treated with scalar wave functions, while electromagnetic waves in structured dielectrics must be treated with full account of their vector character. Today even theoretically, few structures are known that offer a full 3D PBG.

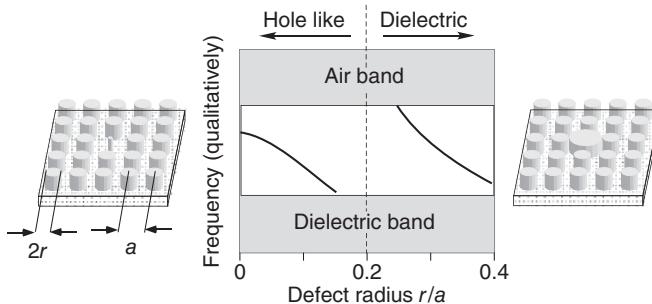
The preparation of photonic crystals exhibiting a PBG at nanometer scale poses another challenge. Novel methods of fabrication are necessary since the conventional processes of microstructuring are directed toward manipulation of thin layers at the surface of suitable substrates. An example generated with electrochemical etching processes with a bandgap at about  $5 \mu\text{m}$  wavelength is shown in Figure 4.9.

#### 4.2.5 Defects and Defect Modes

The intense interest in photonic crystals is driven by the application perspectives of the so-called defect modes as well. *Defects* are local perturbations of the perfectly periodic refractive index lattice. For instance, magnification or demagnification of a single hole in the 2D photonic crystal of Figure 4.15 leads to such defects. Examples of such defects are shown in Figure 4.16. Demagnification of a



**Figure 4.16** Defects in 2D photonic crystals: (a) dielectric defect, point-like; (b) hole-like defect, point-like; (c) dielectric defect, line-like; (d) line-like defects in 2D hybrid structures can be used to realize complex plane wave guides for light.



**Figure 4.17** A 2D photonic crystal constructed from dielectric cylinders with radius  $r/a = 0.2$  shows a bandgap for  $n^2 = 12$ ; see Figure 4.14. The position (frequency) of the defect modes depends on the defect radius. For  $r/a < 0.2$  defects are hole-like, and for  $r/a > 0.2$  dielectric.

hole leads to a *dielectric defect*, since the index of refraction is locally enhanced. Magnification of the hole reduces the index of refraction and generates a *hole-like defect*.

Defects can cause isolated and localized electromagnetic field states in the PBG. Propagation of the field is suppressed as well as spontaneous decay. The electric field of the so-called *defect mode* drops exponentially with the distance from the perturbation, much like the evanescent field of TIR (Section 3.1.2). Thus a point-like defect can store light as in an optical cavity (Section 6.6). An optical defect in a PBG thus shows phenomena resembling the properties of donor and acceptor atoms in a semiconductor, which can accept or deliver an electron from a local perturbation (Section 10.3.3).

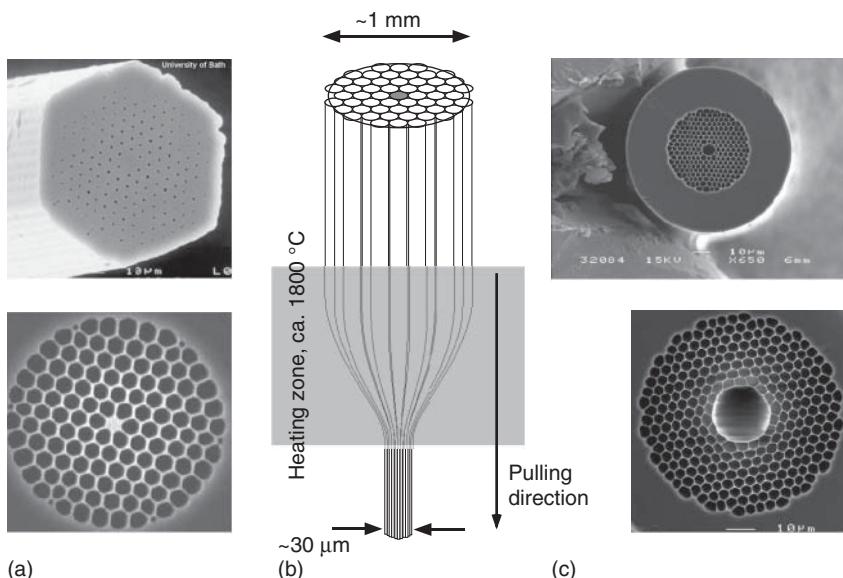
In Figure 4.17 the dielectric defects seem to grow out of the upper (air) band, the hole-like from the lower (dielectric) band. One can qualitatively understand this evolution by taking into account that frequency and propagation constant are related through  $\omega = c\beta/n_{\text{eff}}$  (Eq. (4.13)). We expect that a dielectric defect enhances  $n_{\text{eff}}$ , while a hole-like defect decreases the effective refractive index locally. In the first case the frequency of the corresponding localized state decreases; in the second it increases into the PBG. The localization or “tacking” of an electromagnetic field state to a dielectric defect can be understood in direct analogy with the guiding properties of a step-index fiber: the fiber core allows axial propagation for  $(\omega, \beta)$  values below those allowed in the homogeneous cladding material (see Figure 4.4), thus suppressing transverse propagation. For hole-like defects the “forbidden” region, the PBG, is generated by diffraction from the periodic structure.

#### 4.2.6 Photonic Crystal Fibers (PCFs)

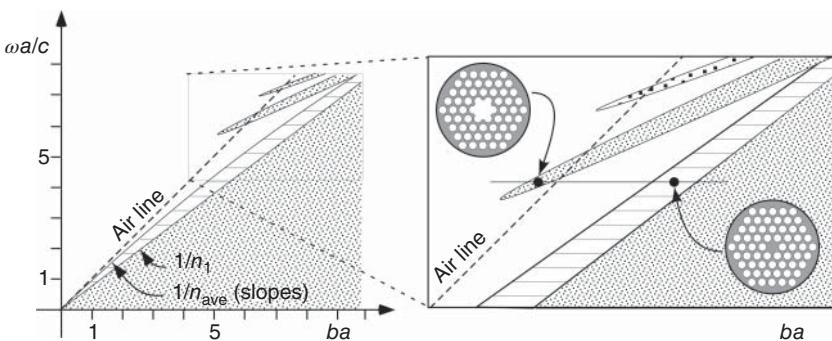
PCFs count among some of the most interesting microstructured materials in optics since their invention around 1995 [43]. They not only offer very unusual properties but also are relatively straightforward to manufacture. A preform of the desired structure is made from a bundle of hollow or solid glass capillaries. Then the preform is heated like in conventional fiber drawing processes of telecommunication fibers. Surprisingly enough, the transverse structure is essentially preserved in this process but reduced in diameter by a factor of about 30.

A sketch of the process is shown in Figure 4.18, along with two specific examples. In contrast to the 2D photonic hybrid structures of Figure 4.16 that guide light waves within a thin layer, light propagates in PCFs along the fiber core, which corresponds to the defects of Section 4.2.5. It is expected that for the characterization of PCFs, dispersion diagrams are used as for conventional fibers and in analogy with Figures 4.4 and 4.14. A conceptual diagram with characteristic properties is given in Figure 4.19

Below the line with slope  $1/n_1$  corresponding to the bulk material with refractive index  $n_1$ , propagation within the 2D photonic crystal is certainly not possible – this region is forbidden. For large wavelength or small propagation constants  $\beta$ , respectively, we can assume that the dielectric–air structure causes an average index of refraction depending on the filling fraction of the dielectric and with  $n_{\text{ave}} < n_1$ , which pushes the lower boundary for guided modes up to the line with slope  $1/n_{\text{ave}}$ .



**Figure 4.18** Photonic crystal fibers are fabricated by drawing a heated bundle of capillary tubes (b). (a) Fibers with dielectric core. (c) Fibers with hollow core (BlazePhotonics Ltd). With kind permission by P. St. Russell.



**Figure 4.19** Dispersion diagram of photonic crystal fibers. Dotted regions are forbidden for wave propagation. Below the line with slope  $1/n_1$ , propagation of light is strictly forbidden in analogy with conventional step-index fibers. In the hatched area guided modes are generated by dielectric cores. The fingerlike forbidden areas correspond to photonic bandgaps extending even beyond the air line (dashed line). Hollow cores can cause guided modes in the bandgaps.

When the wavelength of light propagating in such a fiber approaches the scale of the 2D periodicity, simple averaging is no longer possible. It turns out that light becomes more concentrated in the dielectric, and hence the boundary for guided modes gradually approaches the line of the bulk material. A dielectric core defect then generates guided modes within the hatched area of the dispersion diagram. While conventional optical fibers are weakly guiding because of their small index of refraction variation, the strong index of refraction contrast in PCFs – typically glass versus air – causes strong guiding corresponding to smaller cross sections of the guided light field. This property contributes significantly to the strong activity of nonlinear processes in PCFs.

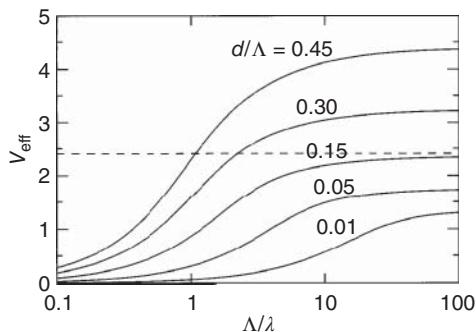
In contrast to step-index fibers, PCFs exhibit PBG as well. They generate regions in the  $(\omega, \beta)$  plane where propagation is forbidden even above the air line (see Figure 4.19). In such a region also hollow cores – hole-like defects – generate isolated guided modes.

Because of their unusual properties, strongly guiding PCFs have very versatile applications, which are treated in special publications [44]. Here we restrict ourselves to a qualitative example, which illustrates the rapidly growing interest in these microstructured optical wave guides.

#### Example: An endlessly single-mode fiber [45]

For single-mode operation of a step-index fiber, the condition  $V = (2\pi a/\lambda) (n_1^2 - n_2^2)^{1/2} < 2.405$  must be fulfilled for the  $V$  parameter (Eqs. (4.9) and (4.16)). Neglecting dispersion the  $V$  parameter is proportional to  $1/\lambda$  in conventional fibers and eventually exceeds the single-mode threshold value 2.405 with decreasing wavelength. A solid dielectric core can guide light in PCF materials made from a glass-air structure as shown in Figure 4.18 in analogy with step-index fibers. The aforementioned condition for  $V$  is modified, however, since the effective index of refraction ( $n_{\text{eff}} < n_1$ ) of the periodically structured material now depends itself on the wavelength.

In fact, with decreasing  $\lambda$ , the guided light tends to be more and more concentrated into the dielectric portion of the structure, and  $n_{\text{eff}}$  more and more approaches  $n_1$ . Hence it also modifies the  $V$  parameter and may even compensate the  $1/\lambda$  dependence. In an “endlessly single-mode” fiber, the  $V$  parameter can remain below 2.405 as shown in Figure 4.18a. The curves shown in Figure 4.20 show calculated  $V$  values for varying ratios of the diameter  $d$  of the holes and their separation  $\lambda$ .



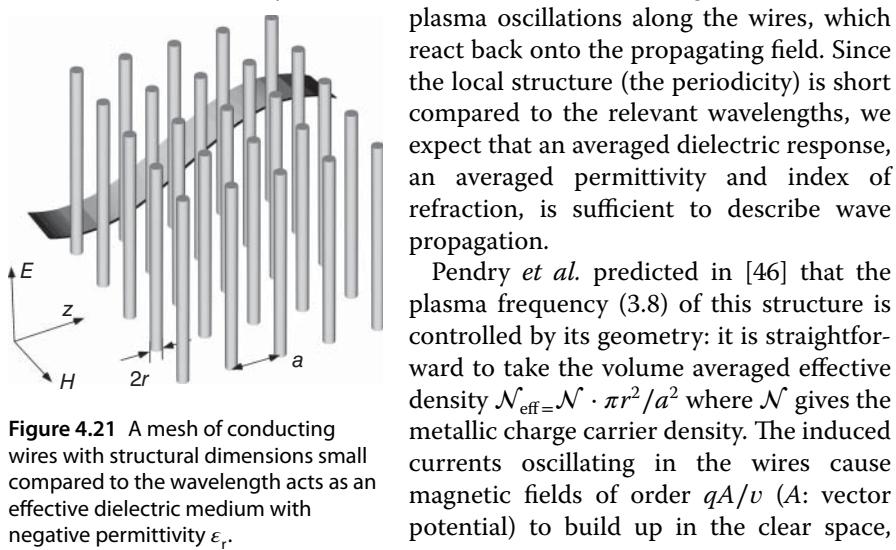
**Figure 4.20** Calculated values of the effective  $V$  parameter for photonic crystal fibers of the type shown in Figure 4.18a. With kind permission from P. St. Russell, Erlangen.

## 4.3 Metamaterials

### 4.3.1 Dielectric (Plasmonic) Metamaterials

Plasmonic materials owe their interesting properties (and applications) largely to the negative values of the dielectric susceptibility with  $\epsilon_r = 1 + \chi_{\text{diel}} = n^2 = 1 - (\omega_p/\omega)^2$  for  $\omega < \omega_p < 0$  (see Section 3.2.1.1). The negative permittivity  $\epsilon$  is the origin of the strong confinement of surface waves called surface plasmon polaritons (SPPs); for optical frequencies this occurs at sub-micrometer scales of the surface as outlined in Section 3.2.3. Localization of the fields near the surface also permits to control properties of plasmonic materials by shaping the geometry.

An example of a “dilute” plasmonic *metamaterial* is shown in Figure 4.21. It consists of a dilute array of wires. The  $E$ -field of an incoming wave will excite plasma oscillations along the wires, which react back onto the propagating field. Since the local structure (the periodicity) is short compared to the relevant wavelengths, we expect that an averaged dielectric response, an averaged permittivity and index of refraction, is sufficient to describe wave propagation.



Pendry *et al.* predicted in [46] that the plasma frequency (3.8) of this structure is controlled by its geometry: it is straightforward to take the volume averaged effective density  $\mathcal{N}_{\text{eff}} = \mathcal{N} \cdot \pi r^2/a^3$  where  $\mathcal{N}$  gives the metallic charge carrier density. The induced currents oscillating in the wires cause magnetic fields of order  $qA/v$  ( $A$ : vector potential) to build up in the clear space,

which act as an inertia to the external forces, leading to an effective charge mass  $m_{\text{eff}} = (\mu_0 r^2 q^2 \mathcal{N} / 2) \cdot \ln(a/r)$  (Probl. 4.6). We then arrive at the remarkable result

$$\omega_p^2 = \frac{\mathcal{N}_{\text{eff}} q^2}{m_{\text{eff}} \epsilon_0} = \frac{\omega_a^2}{2\pi \ln(r/a)},$$

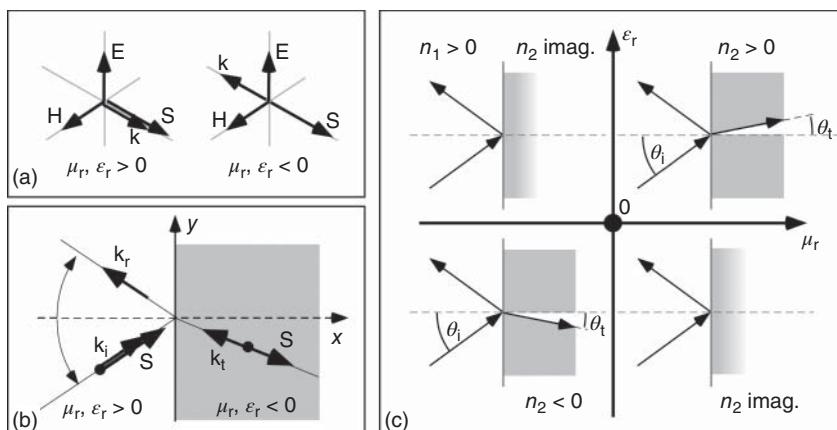
where  $\omega_a = 2\pi c/a$  depends on geometry only. Taking into account geometric modifications of the damping rate, [46, 47], we can resort to the Drude model of Section 3.2.1 to calculate basic properties of this metamaterial. Of course, extensions of the concept to 3D are needed, and manufacturing of suitable artificial materials remains a challenge, but contemporary nanostructuring methods are offering an ever-increasing catalogue of solutions.

#### 4.3.2 Magnetic Metamaterials and negative index of refraction

Let us – by analogy to the dielectric case of the preceding section – assume we have a material with  $\mu_{\text{mag}} = 1 + \chi_{\text{mag}} < 0$  in addition. What are the consequences?

In our discussion on the vectorial properties of planar waves in Section 2.2.1, we have already realized that the triad  $\{\mathcal{E}, \mathcal{H}, \mathbf{k}\}$  is left-handed in this case. We do not find such left-handed materials (lhms) in nature. Hence, in spite of a 1968 argument by Veselago [48], negligible attention was paid until the 1990s [49]. This was changed when J. Pendry pointed out that with artificial materials and contemporary nanostructuring methods, the route for construction of these so-called metamaterials could be explored for both the electric and magnetic cases.

Inspired by the discussion of Section 3.1.1 for conventional (right-handed materials (rhms)), we take a shortcut for understanding Snell's law with respect to refraction from rhm to lhm systems. The incoming wave can be characterized by the rhm triad  $\{\mathcal{E}, \mathcal{H}, \mathbf{k}_i\}$  and the Poynting vector  $\mathbf{S}$  (Figure 4.22), the transmitted wave by the lhm triad  $\{\mathcal{E}, \mathcal{H}, \mathbf{k}_t\}$  and  $\mathbf{S}$ . Then, at the interface, the normal



**Figure 4.22**  $\{\mathcal{E}, \mathcal{H}, \mathbf{k}\}$  triads for right- and left-handed materials (a). Propagation and Poynting vectors for refraction from rhm to lhm (b). Overview of (loss-free) negative materials and their influence on wave propagation (c).

component of  $\mathbf{S}$  must be continuous (otherwise energy would accumulate), and hence  $\mathbf{k}_t$  must have a negative  $k_{tx}$ -component due to the left-handedness. In addition, the projections of wave propagation  $k_{yi}$  and  $k_{yt}$  along the  $y$ -interface must be phase synchronous, as is the case for rhm $\rightarrow$ rhm refraction with  $n > 0$  as well. This suggests to accept the validity of Snell's law (1.2),

$$n_1 \sin \theta_i = n_2 \sin \theta_t,$$

where the index of refraction  $n_2$  is now negative as well as the angle of the transmitted wave:  $\theta_t$  (see Figure 4.22).

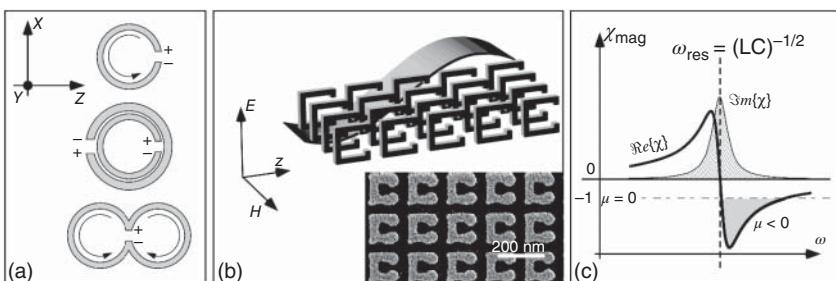
### 4.3.3 Constructing Magnetic Metamaterials

It is known that for fundamental relativistic reasons, the magnetic response of natural materials compared with the electric response is typically scaled down by  $\alpha = 1/137$ , the fine structure constant. This is the origin of our argument in Section 2.1.2 that in most cases relevant for optics, the magnetic permeability is almost always simply  $\mu_r = 1$ .

Two aspects may be used to overcome this blockade: ring geometries force currents to show a magnetic response, and resonance structures allow to significantly enhance the response to an external excitation. We thus arrive at the split-ring resonator (SRR) suggested by Pendry [49], conceptually shown in Figure 4.23a.

In an SRR, electric and magnetic fields oscillating along the  $x$  and  $y$  direction, respectively, will induce an electric dipole at the capacitor gap (capacity  $C$ ) and a magnetic dipole within the current loop (inductivity  $L$ ). Combination of SRRs (double SRRs) furthermore allows to suppress the electric or magnetic response for even better control. The response of the resonant structure follows the well-known Lorentz oscillator behavior (Figure 4.23b). In Section 7.1.2 we discuss as an example the connection of the microscopic dielectric response with the macroscopic dielectric polarization and the index of refraction for natural resonant materials. Conceptually the calculation of  $\mu_r$  would follow the same line and is thus omitted.

At least in principle a strong magnetic response can now be engineered by selecting the resonance frequency  $\omega_{\text{res}} = (LC)^{-1/2}$  for a suitable frequency range.

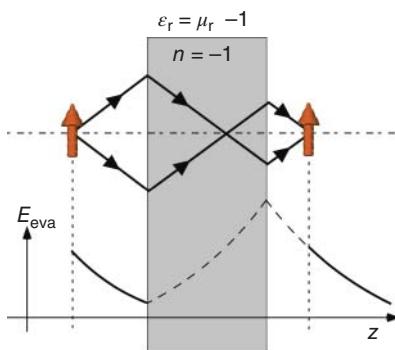


**Figure 4.23** For details, see text. (a) Resonant negative materials are constructed from split-ring resonators (SRRs). (b) SRR array offering a negative magnetic material. (c) Resonant behavior of the magnetic susceptibility  $\chi_{\text{mag}}$  showing a shaded region where  $\text{Re}\{\mu\} = 1 + \text{Re}\{\chi_{\text{mag}}\} < 0$  and absorption  $\text{Im}\{\mu\} = \text{Im}\{\chi_{\text{mag}}\}$  remains small.

Apart from more complex influences of the material, this amounts to choosing the proper dimensions at micro- and nanometer scales similar to the wire mesh case of the preceding chapter. The frequency interval useful for applications is then roughly the right half of the shaded area in Figure 4.23c. There the imaginary (absorptive) part of the magnetic response has already dwindled, while the real (dispersive) part remains significant.

#### 4.3.4 Applications of Metamaterials: The Perfect Lens

Many intriguing applications of doubly negative (both  $\epsilon_r$  and  $\mu_r$ ) materials have been suggested, including the perfect lens, optical cloaking, and more. We restrict ourselves to the example of the perfect lens and refer the reader to the specialist literature [47].



**Figure 4.24** A slab of negative index material acting as a perfect lens. Upper half: rays for image construction. Lower half: amplitude of evanescent field components.

While the ray optical picture for the plasmonic slab lens is straightforward, the central question for imaging applications – what is the resolution or information content about spatial structures – remains. As we will discuss in Section 5.4.1, the information carried by propagating waves with some wavelength  $\lambda$  is limited to structures of order  $\lambda/2\pi = \omega/c = k$  only.

The plasmonic slab lens, however, transports information based on the (non-propagating) evanescent fields emanating from the object structure as well. The dispersion relation for surface plasmons (Figure 3.6) shows clearly that for  $\omega \rightarrow \omega_p/\sqrt{2}$ , literally all  $k$  values are available and hence no  $k$ -limit to resolution is present. In addition to the spatial information, also the correct amplitude must be transmitted. This is shown qualitatively in the lower half of Figure 4.24: the evanescent fields of the object coupling to the slab lens are exciting the plasmon resonance near  $\omega \rightarrow \omega_p/\sqrt{2}$ , which according to Section 3.2.3.2 leads to strong field enhancement. Thus energy conservation is fulfilled, and the energy stored in this quasi-resonance is drawn from the fields scattered off the object.

The “perfect” lens offers – in principle – imaging without loss of information, an attractive promise for applications. The fundamental functionality of the perfect lens concept has indeed been demonstrated [50]. The real world imposes limitations caused by plasmon damping and even small deviations of the refractive index from  $n = -1$ .

The concept of the perfect lens is shown in Figure 4.24. At  $n = -1$ , according to Snell's law, the angle of the rays is simply flipped. Note that the condition  $\{\epsilon, \mu_r\} \rightarrow -1$  requires  $\omega \rightarrow \omega_p/\sqrt{2}$  (see Figure 3.6). All fields with  $\omega \rightarrow \omega_p/\sqrt{2}$  scattered off and excited from arbitrary points of the object are properly images. Thus we expect formation of a 1:1 image at an object–image separation of twice the thickness of the slab.

While the ray optical picture for the plasmonic slab lens is straightforward, the central question for imaging applications – what is the resolution or information content about spatial structures –

## Problems

- 4.1 Fiber couplers, directional couplers** (a) For applications in, for example, communication networks, optical fibers must frequently be coupled with each other face to face. Study qualitatively the influence of small transverse mismatch, axial tilts, and gaps. For simplicity assume a box-shaped mode profile (which width do you choose?).  
 (b) The signal propagating in an optical fiber must frequently be directed into multiple output ports. Directional couplers are used for this purpose, which can be manufactured by parallel fusion of two wave guides. The fusion causes the field distributions of the optical modes to overlap, and hence the electric fields in the two fibers are coupled with each other. Consider a simplified model in which for the fusion length  $\ell$  the field amplitudes are locally coupled with a constant coefficient. How does the power exiting into the two ports depend on the coupling strength and length  $\ell$ ?
- 4.2 Linear polarization of the LP/HE modes** Construct explicitly the linearly polarized transverse components  $\mathcal{E}_x, \mathcal{E}_y$  of the  $LP_{01}$  mode from the  $HE_{11}$  mode and Eqs. (4.5) and (4.6) using, for example, the route sketched in Section 4.1.1.4. Justify the designation LP mode.
- 4.3 Single-mode fiber** For a typical step-index fiber ( $NA = 0.11$ ,  $\lambda_{\text{cutoff}} = 620 \text{ nm}$ ) operated in the single-mode regime, estimate the portion of the total power of the  $LP_{01}$  mode, which travels in the core and cladding, respectively, as a function of the wavelength. How do you focus a free space Gaussian  $TEM_{00}$  mode onto the fiber end for optimal coupling efficiency?
- 4.4 Number of wave guide modes** With increasing  $V$  parameter (4.9), not only the number of allowed  $\ell$  values for  $J_\ell(X)$  increases, but also the number of roots  $X_{\ell m} < V$  continues to rise. In order to estimate the number of modes for a given  $V$  parameter, begin by assuring yourself that  $J_\ell(x) \approx (2/\pi x)^{1/2} \cos(x - (\ell + 1/2)\pi/2)$  is a good approximation for large  $x$ . Use this approximation to calculate explicitly the number  $M$  of allowed modes and show that it rises with  $M \approx 4V^2/\pi^2$ . Calculate the mode number for a step-index fiber with  $NA = 0.2$  and core diameter  $2a = 50 \mu\text{m}$ . How does the mode number change if the cladding is replaced with air, that is, if a pure solid  $50 \mu\text{m}$  diameter glass fiber is used?
- 4.5 Brillouin zone in a two-dimensional rectangular lattice** Construct the reciprocal lattice for a crystal square lattice (length  $a$ ) and for a parallelogram (side length  $a$  and  $b$ , angle  $45^\circ$ ). Construct the first Brillouin zone as well.
- 4.6 Pendry's wire mesh** Verify the expression for the effective charge density and mass for the geometry of Figure 4.21.

# 5

## Optical Images

This chapter is dedicated to the basic principles of optical image formation. Imaging by lenses and lens systems is the basis for optical instruments, which have substantially influenced the development of optics and have made possible – literally – our insights into the macro- and micro-cosmos.

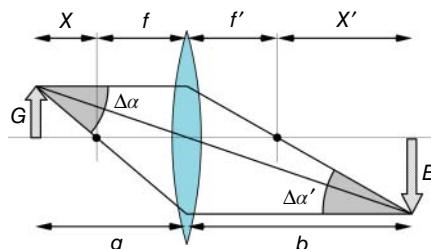
The basic element in optical imaging is the convex lens, which for stigmatic imaging merges into one point again all the rays that originated from a single point. Systems of lenses let us build microscopes, telescopes, and numerous other instruments to peek into the worlds not directly accessible with our eyes. Imaging occurs as well with the lens of our eye; thus we need to account the properties of our “own vision instrument” to understand image formation.

In addition to the basic principles, we raise the question of the resolution capability of such instruments. What objects at very large distances can we make visible? What are the smallest objects that we can observe with a microscope including their structure or at least their size?

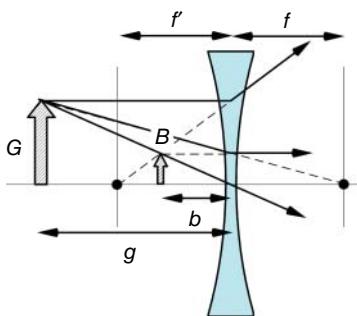
### 5.1 Simple Lenses

With the help of geometry, we can understand the most important properties of a (real and virtual) optical image (Figure 5.1):

- A beam parallel to the axis is sent through the focal point by a convex lens.
- A beam that reaches the lens via the focal point leaves the lens parallel to the axis.
- A beam passing through the center of the lens is not diffracted.



**Figure 5.1** Conventional construction of a lens image with the common notation.  $G, B$ : object and image size;  $f, f'$ : focal length with identical lengths.



**Figure 5.2** Virtual image formation by a diverging lens.

Let us summarize selected relations and leave the justification to the reader.

**Real and virtual images.** In Figure 5.1 formation of a real image is shown: all rays emanating from an object point are reconvened at a single image point (stigmatic imaging). All rays have identical optical path lengths (Section 1.8.1). The image can be recorded by, for example, a camera sensor at the image position. Note that for acceptable imaging, this condition needs to be fulfilled ideally for all image points and for all rays, a requirement that can only be achieved approximately.

In contrast, in Figure 5.2, formation of a virtual image is illustrated. The focal lengths  $\{f, f'\}$  of the diverging lens are counted negative with respect to the object–lens axis. Otherwise the construction principles of p. 149 remain. The virtual image seems to emerge from  $B$  and cannot be recorded directly.

**Newton's equation.** We connect the distances  $g$  and  $b$  with sizes  $G$  and  $B$ , respectively, and the focal length  $f$  and find with little algebra ( $f' = -f$ )

$$\frac{x'}{f} = \frac{f}{x}. \quad (5.1)$$

**Imaging equation.** We have already come across the relation of  $\{f, g, b\}$  through Eq. (1.24), when discussing matrix optics,

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b}. \quad (5.2)$$

It evolves from Newton's equation when one uses  $g = f + x$  and  $b = f + x'$ .

**Magnification.** The lateral magnification  $\mathcal{M}_\ell = B/G$  is

$$\mathcal{M}_\ell = B/G = -b/g = -f/x,$$

where the minus sign accounts for the inversion of the real image.

**Angular magnification.** For small angles we can use the paraxial approximation  $\tan \alpha \simeq \sin \alpha \simeq 1$  and find for the opening angles  $\Delta\alpha, \Delta\alpha'$  in Figure 5.1

$$\mathcal{M}_\alpha = \frac{\Delta\alpha'}{\Delta\alpha} = \frac{g}{b} = \frac{1}{\mathcal{M}_\ell}. \quad (5.3)$$

Thus large magnification can only be obtained at the expense of a strong reduction in angle spread and hence image luminosity,  $\mathcal{M}_\ell \mathcal{M}_\alpha = 1$ .

**Thin lenses and thick lenses.** The widely used thin lens approximation assumes that all refractive processes of the optical imaging system occur in a single plane. For realistic thick lenses and lens systems, also, one can split this single plane into two separate *principal planes*. Image formation can then be treated in the thin lens approach by ignoring details in between the two principal planes. The *ABCD* formalism introduced in Section 1.9 allows to treat thin as

well as thick lenses and can also be used to find out about the position of the principal planes (Problem 1.21).

*Ray tracing.* This is today widely used for the design and optimization of optical imaging systems. The concept is simple: all rays emanating from an object point  $P_{\text{obj}}$  shall be collected in a single point  $P_{\text{ima}}$  in the image plane again, which requires all individual rays  $P_{\text{obj}} \rightarrow P_{\text{ima}}$  to have identical optical path lengths. We furthermore expect that all points  $P_{\text{ima}}$  form a faithful image in a single plane, without distortions.

We have seen in Section 1.9 that linearized and hence analytic solutions can be obtained for the paraxial case. Ray tracing beyond this case is immediately affected by nonlinear terms (see Section 5.7). Numerous commercial software packages are available to address these problems numerically with powerful computers.

## 5.2 The Human Eye

Unfortunately, it is not possible here to go into the physiological origin of the vision process, and for that we refer the reader to the relevant literature [51].

For our purposes it is sufficient to construct a “reduced artificial standard eye” (Figure 5.3). The eye body in general has a diameter of 25 mm, and several important optical properties are collected in Table 5.1. The refractive power of the total eye is achieved predominantly by the curvature of the cornea (typical radius 5.6 mm, difference in refractive index with respect to air  $\Delta n \sim 0.37$ ), while the variable crystalline lens guarantees “focusing” by contraction. Laser ablation with femtosecond lasers – used as scalpels – is routinely used for reshaping the cornea, and thus a patient’s ability to see can be improved.

By adaptation of the focal length of our eyes – unfortunately and largely lost with age, we are generally able to recognize objects at a distance of 150 mm or more. As a standardized distance for optical instruments, often the conventional *least distance of distinct vision* of  $S_0 = 250$  mm is chosen, where the best results are achieved with vision aids.

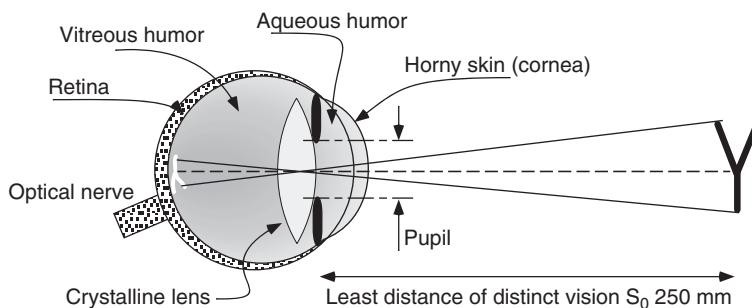


Figure 5.3 Human eye, reduced to the most important optical “components.”

**Table 5.1** Optical properties of the human eye.

Vitreous humor, aqueous humor	$n = 1.336 (\sim 4/3)$
Cornea	$n = 1.368$
Crystal lens	$n = 1.37\text{--}1.42$
Focal length, front	$f = 14\text{--}17 \text{ mm}$
Focal length, back	$f = 19\text{--}23 \text{ mm}$
Clear vision distance	150 mm to $\infty$ , $S_0 = 250 \text{ mm}$
Pupil (diameter)	$d = 1\text{--}8 \text{ mm}$
Pupil (shutter time)	$\tau = 1 \text{ s}$
Resolving power at 250 mm	$\Delta x = 10 \text{ } \mu\text{m}$
Sensitivity (retina)	$1.5 \times 10^{-17} \text{ W/vision cell} \sim 30 \text{ photon s}^{-1}$

### 5.3 Magnifying Glass and Eyepiece

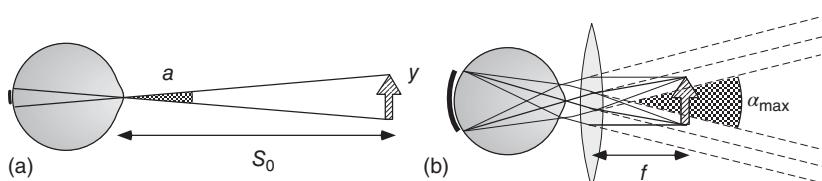
The simplest, and since ancient times very popular, optical instrument is the convex lens used as a magnifier. The effect can most quickly be understood by considering the angle  $\alpha$  at which an object of height  $y$  is seen, since this angle determines our physiological impression of its size – a mountain 1000 m high at a distance of 10 km seems to have the same size as a matchbox at a distance of 25 cm. Only our knowledge of their distance identifies objects according to their real sizes.

Without technical aid, we view an object of size  $y$  with the eye (Figure 5.4) at an angle  $\alpha = \tan(y/S_0) \approx y/S_0$ , which is determined by the least distance of distinct vision  $S_0$ . Now, we hold the magnifier directly in front of the eye: the magnifier widens the angle at which we see the object. If we bring the object close to the focal length,  $x \sim f$ , then parallel rays reach the eye, so that the object appears to be removed to infinity. From geometrical relations, we can determine that

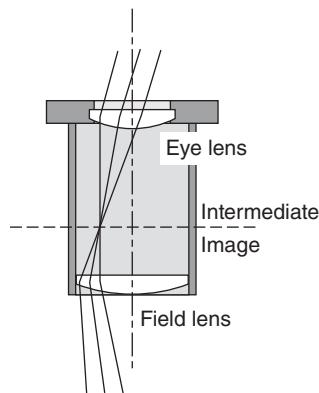
$$\alpha_{\max} = \frac{y}{f}.$$

From that we can directly deduce the maximum magnification  $M$  of the magnifying glass:

$$M = \frac{\alpha_{\max}}{\alpha} = \frac{S_0}{f}.$$



**Figure 5.4** Vision without (a) and with (b) a magnifying glass. The magnification is caused by widening of the vision angle. The object lies within the focal length of the magnifier, and the position of the virtual image was chosen to be at infinity in this example.

**Figure 5.5** Huygens eyepiece with path of rays.

Thus, the smaller the focal length of a magnifying lens, the stronger is its magnification. However, since the least distance of distinct vision is defined as  $S_0 \sim 250$  mm, and because thicker and thicker, more curved lenses are necessary at smaller focal lengths, the practicable magnification of magnifying lenses is limited to  $\mathcal{M} \leq 25$ .

In contrast to the *real* image, discussed in the first section, the magnifying glass generates a *virtual* image. If the magnifying glass is not held directly in front of the eye like in Figure 5.4, then the magnification is a little bit less, as one may straightforwardly find out from geometrical considerations. However, the difference is in general marginal, and, anyway, an individual user looks for a suitable working distance by manual variation of the distances of the magnifier, eye, and object.

In optical devices, such as microscopes and telescopes, real intermediate images are generated, which are then observed with a so-called eyepiece. The eyepiece in general consists of two lenses to correct for chromatic aberrations, which we will discuss in Section 5.7.3. In the Huygens eyepiece (Figure 5.5), a real intermediate image is generated by the field lens, which is looked at with the eyepiece. The eyepiece fulfills exactly the task of a magnifying lens with an effective focal length  $f_{\text{ocu}}$  and a magnification  $\mathcal{M}_{\text{ocu}} = S_0/f_{\text{ocu}}$  for an eye that is adapted to infinite vision distance.

### Example: Effective focal length and magnification of a Huygens eyepiece

A Huygens eyepiece consists of two lenses at a separation of

$$d = (f_1 + f_2)/2,$$

because there the minimal chromatic aberrations occur (see Section 5.7.3). We determine the effective focal length and magnification of a system consisting of two lenses with  $f_1 = 30$  mm and  $f_2 = 15$  mm, for instance, by application of the matrix formalism of Section 1.9. The system has an effective focal length of

$$\begin{aligned}\frac{1}{f_{\text{ocu}}} &= \frac{1}{f_1} + \frac{1}{f_2} - \frac{f_1 + f_2}{2f_1 f_2} \\ &= \frac{f_1 + f_2}{2f_1 f_2} = (20 \text{ mm})^{-1}\end{aligned}$$

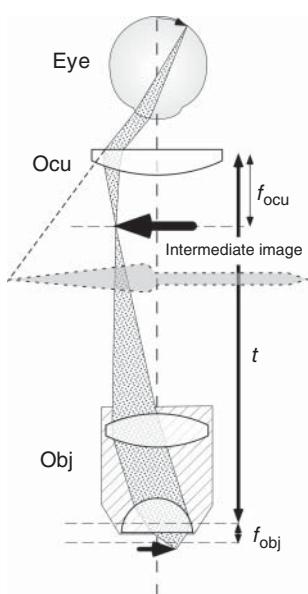
and thus a magnification of  $\mathcal{M}_{\text{ocu}} = 250/20 = 12.5\times$ .

## 5.4 Microscopes

To see small things “big” is one of the oldest dreams of mankind and continues to constitute a driving force for our scientific curiosity. The magnifying glass alone is not sufficient, as we know already, to make visible the structure of very small objects, for example, the details of a biological cell. However, by adding one or two lenses, which generate a real image at first, it has been possible since the nineteenth century to achieve up to 2000-fold magnifications – the microscope [52] “opens” our eyes.

We consider a microscope (Figure 5.6) in which an objective lens “obj” with focal length  $f_{\text{obj}}$  generates a real intermediate image. The intermediate image plane is a suitable position to install, for example, a graticule, the lines of which are seen simultaneously with the object under test. For that purpose, an eyepiece “ocu” with focal length  $f_{\text{ocu}}$  or more simply a magnification  $\mathcal{M}_{\text{ocu}} = S_0/f_{\text{ocu}} = 250 \text{ mm}/f_{\text{ocu}}$  is used, typically with enlargement factors 10 $\times$  or 20 $\times$ . In practice, objectives and eyepieces are lens combinations in order to correct for aberrations (see Section 5.7.3). The total focal length  $f_\mu$  of the composite microscope is evaluated according to Eq. (1.26) as

$$\frac{1}{f_\mu} = \frac{1}{f_{\text{obj}}} + \frac{1}{f_{\text{ocu}}} - \frac{t}{f_{\text{obj}} f_{\text{ocu}}}.$$



**Figure 5.6** Beam path within a microscope:  $t$  is the length of the tube;  $f_{\text{obj}}$  and  $f_{\text{ocu}}$  are the focal lengths of objective and eyepiece; the black arrow shows the position of the intermediate image.

In general, microscopes have tubes with well-defined lengths of  $t = 160 \text{ mm}$ , and since  $t \gg f_{\text{obj}}, f_{\text{ocu}}$ , one may approximately specify

$$f_\mu \approx -\frac{f_{\text{obj}} f_{\text{ocu}}}{t} = -\frac{f_{\text{obj}} f_{\text{ocu}}}{160 \text{ mm}}.$$

We can determine the image size in two steps: (i) The object lies approximately in the focal plane of the objective, whereas the distance between the real image and the objective differs only a little bit from the length of the tube  $t$ . According to Eq. (5.1), it then holds that  $y/f_{\text{obj}} \approx y'/t$  and the objective gives rise to a magnification  $\mathcal{M}_{\text{obj}} = y'/y \approx t/f_{\text{obj}}$ . (ii) The eyepiece further magnifies the image by the factor  $\mathcal{M}_{\text{ocu}} = y''/y' = S_0/f_{\text{ocu}}$ , as explained in Section 5.3 on the magnifying glass. The total magnification  $\mathcal{M}_\mu$  of the microscope is then

$$\mathcal{M}_\mu = \mathcal{M}_{\text{ocu}} \mathcal{M}_{\text{obj}} \approx \frac{S_0}{f_{\text{ocu}} f_{\text{obj}}} \frac{t}{f_\mu} = \frac{S_0}{f_\mu}.$$

This last result shows that the microscope in fact acts like an effective magnifying lens of extremely short focal length.

### Example: Magnification of a microscope

We construct a microscope with an eyepiece, magnification  $10\times$ , and an objective with focal length  $f_{\text{obj}} = 8 \text{ mm}$ . The magnification of the objective amounts to  $M_{\text{obj}} = 160/8 \text{ mm} = 20$ . The total magnification can be calculated according to  $M_{\mu} = 10 \times 20 = 200$ .

Standard microscopes are designed for a quick exchange of the optical elements to change the magnification easily. Both eyepiece and objective are usually specified with the magnification, for example,  $100\times$ ; the components of different manufacturers are in general interchangeable. The total magnification can be determined according to the procedure described earlier without difficulties. For precision measurements it is necessary to calibrate the magnification factor by means of a suitable length standard.

#### 5.4.1 Resolving Power of Microscopes

So far, we have looked at the microscope only from the geometrical optics point of view and assumed that a point is imaged into an ideal point, again and again. However, as a result of diffraction at the apertures of the lenses, this is not possible, so the resolving power is limited by diffraction.

##### 5.4.1.1 Rayleigh Criterion and Numerical Aperture

A first measure for the resolving power can be gained from the result for the diameter of the Airy disc and named after Lord Rayleigh: the image of a point-like object by an objective lens with diameter  $D$  and focal length  $F_{\text{obj}}$  results in an intensity distribution given by the *point spread function* (PSF). For a perfect circular aperture, this is the Airy function (2.71), and we define the resolution limit  $\Delta x_{\min}$  for two PSFs by assuming that the maxima of PSF1 and PSF2 are at least separated by the first zero of the PSFs (2.72), finding

$$\Delta x_{\min} \geq 1.22 \frac{f_{\text{obj}} \lambda}{D}. \quad (5.4)$$

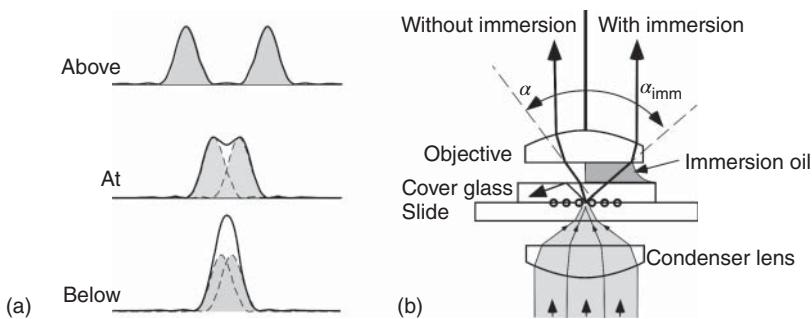
A more systematic approach is based on the numerical aperture (NA) (or Abbe's theory; see next section).

$$\text{NA} = n \sin \alpha.$$

It is defined as in the case of the (1.9) as the sine of half of the aperture angle (Figure 5.7), that is, of the extreme rays that still contribute to the image. Therefore  $n$  specifies the index of refraction in the object space. In the spirit of the Rayleigh criterion, we have to extend relation (5.4), which is valid for not too large angles  $\alpha$ ,  $f_{\text{obj}}/D = 2 \tan \alpha \sim 2 \sin \alpha \rightarrow 2\text{NA}$ , yielding

$$\Delta x_{\min} \geq 0.61 \lambda / \text{NA} \quad (5.5)$$

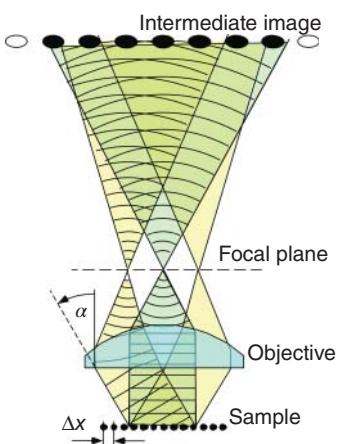
In applications used to investigate the tissue in biology and medicine, a cover slip (typical thickness 0.08–0.17 mm) protects the sample (see Figure 5.7). In order to realize the theoretical values of resolution, when designing the objective, the cover slip must be taken into account. Furthermore, total internal reflection can be a problem limiting the maximum transmission angle within the cover glass to about  $40^\circ$ .



**Figure 5.7** (a) Joint distribution of two Airy point spread functions above, at, and below the Rayleigh resolution limit. (b) Resolving power and numerical aperture with and without immersion oil. The resolution is influenced by illumination as well. Here, a condenser lens is applied, which illuminates a maximum solid angle (Köhler illumination).

Since the sample is always very close to the focal plane, the NA is a property of the objective used and is in general specified on standard components. According to (5.5), the resolving power increases with shorter wavelengths (optical microscopes use blue or even ultraviolet (UV) light for high resolution) and a large NA. For objectives with large magnification 100 $\times$ , that is, with very short focal lengths, NA values of about 0.8 are achieved in air. With blue light illumination, optical microscopes achieve resolutions of about 0.2  $\mu\text{m}$ .

#### 5.4.1.2 Abbe's Theory of Resolution



**Figure 5.8** Fourier image model of a microscope by Ernst Abbe.

To determine the resolving power of a microscope even more accurately, we follow the insight of E. Abbe and consider a periodic structure (a grating with period  $\Delta x$ ) that we observe with a microscope. Ernst Abbe (1840–1905), professor of physics and mathematics at the University of Jena, Germany, and close coworker of entrepreneur Carl Zeiss (1816–1888), provided crucial experimental and theoretical contributions to the development of modern microscopy.

The simplified situation for the grating is illustrated in Figure 5.8, and the focal plane or Fourier plane of the objective now plays a very crucial role. There, bundles of parallel rays are focused, and one observes the Fraunhofer diffraction image of the object, which is simple only for the chosen example of a

one-dimensional (1D) grating. However, the following point is crucial. Within the focal plane, the objective generates the Fourier transform of the complex amplitude distribution in the object plane, as we have already seen in Section 2.7.1. A structure with length scale  $\Delta x$  can only be reconstructed if, apart from the zero

order, at least one more diffraction order enters the objective and contributes to the image. The Abbean criterion for the resolution (or diffraction) limit can be straightforwardly taken from the diffraction properties of gratings (see Eq. (6.7)):

$$\Delta x \geq \Delta x_{\min} = \lambda/2 \sin \alpha = \lambda/2NA. \quad (5.6)$$

It is important to note that this limit can only be achieved if the extreme rays at large angles  $\alpha$  are indeed generated and contribute to the image. Since in many cases strong diffracted rays are generated for small diffraction angles, illumination schemes providing rays at all angles as, for example, shown in Figure 5.7 are very important.

In the optical microscope, the Fourier spectrum of a diffracting or scattering object is reconstructed by the eyepiece and eye or camera objective to yield a magnified image. In principle, the reconstruction can be gained by a calculation or a numerical procedure. In this sense, the scattering experiments of high-energy physics, where the far field of the diffraction of extremely short-wavelength matter waves off very small diffracting objects is measured, are nothing other than giant microscopes.

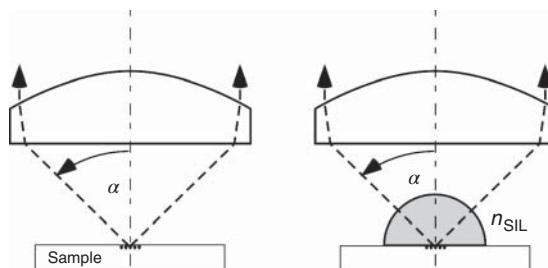
#### 5.4.1.3 Exploiting the Abbe-Rayleigh Resolution Limit

The resolution offered by the Abbe-Rayleigh limit (5.5) is optimized by choosing appropriate parameters for wavelength  $\lambda$  and NA. It can be significantly enhanced by means of immersion fluids, that is, by controlling the index of refraction  $n$  within the sample space. In Figure 5.7 a standard application for life sciences is shown where the index of refraction is adjusted to match that of the cover glass ( $n = 1.5$ ), pushing NA to 1.3.

The immersion concept makes use of the effectively shortened wavelength  $\lambda \rightarrow \lambda/n$  in high-index materials. It has been extended to so-called solid immersion lenses (SILs) [53, 54] as well. In one configuration the object is positioned on the surface and at the center of a hemisphere (Figure 5.9). The resolution

is increased by  $n_{\text{SIL}}$  where materials such as LaSFN9 and Si offer refractive indices of 1.85 (in the visible) and 3.5 (in the infrared) near  $1.8 \mu\text{m}$ .

The Abbe-Rayleigh resolution limit plays a central role for manufacturing microelectronic circuitry, and because of its enormous economic significance, intense efforts have been invested to fully exploit or even overcome these limits (see the excursion in the following text). In microscopy optical resolution beyond this limit remains very attractive since in life sciences many relevant cell structures have nanometer and not micrometer length scales. Electron microscopy, which is based on naturally much shorter wavelengths and hence much higher resolution, offers only partial solutions since it requires highly invasive



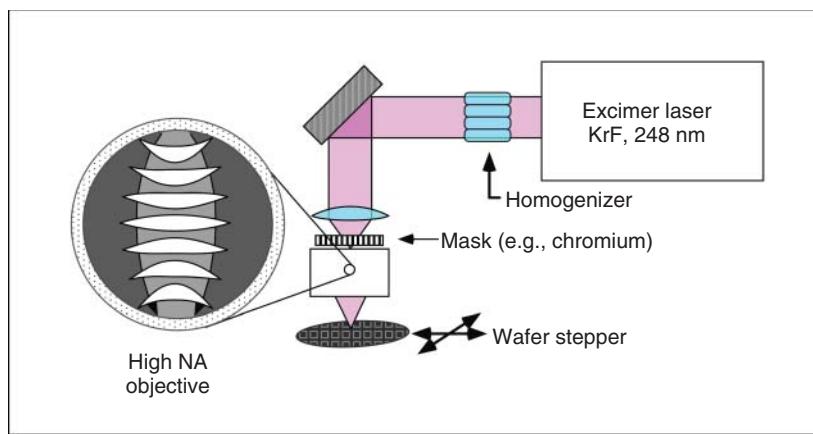
**Figure 5.9** A hemispherical solid immersion lens enhances the resolution by  $n_{\text{SIL}}$ .

sample preparation. Recent successful advances to actively transform optical microscopy into a “nanoscopy” method by actively overcoming the Rayleigh and Abbe diffraction limit have therefore received much attention (see Section 5.5.3).

### Excuse:: Optical lithography

From many points of view, optical lithography is the reverse of microscopy, because lithography, which has been for decades one of the most powerful driving forces of the world economy, is primarily concerned with the miniaturization of electronic circuits to the smallest possible dimensions. The principle is introduced schematically in Figure 5.10. A high NA objective reduces a mask (“reticule”), which contains the structure of the desired circuit, to a  $\sim \text{cm}^2$  area and illuminates wafers that have standard sizes of 12 inches (30 cm) and more. The wafer is coated with adequate film material (“resist”), which is photochemically altered such that afterward in eventually several processing stages, transistors and transmission lines can be produced. The pattern is multiple times applied to cover the entire wafer (“wafer stepper”).

Manufacturers of lithography objectives, which may consist of 60 and more individual lenses, succeeded impressively in guiding the resolution of their wafer steppers directly along the resolution limit according to Eq. (5.5). Further resolution enhancements were obtained by exploiting the nonlinear light-resist interaction on exposure. At present miniaturization of most electronic circuits is based on the wavelengths of the KrF\* laser at 248 nm and the ArF\* laser at 193 nm. In 2016 the structural sizes obtained have routinely reached the level of 45 nm, and scaling to the 10–20 nm level is expected over the next years. Still further progress concentrates on the application of 13.5 nm light. The reason is that for these extreme UV wavelengths, reflective optical components with reasonable efficiency can be manufactured [55]. Enormous costs arise, because at these short wavelengths tremendous problems arise in manufacturing and processing of suitable, that is, transparent and homogeneous, optical materials. The evolution of the state of art in microlithography is regularly published [56].



**Figure 5.10** Optical lithography. Principle of the wafer stepper and UV illumination unit. The lens systems contain numerous components for the correction of aberrations.

### 5.4.2 Analyzing and Improving Contrast

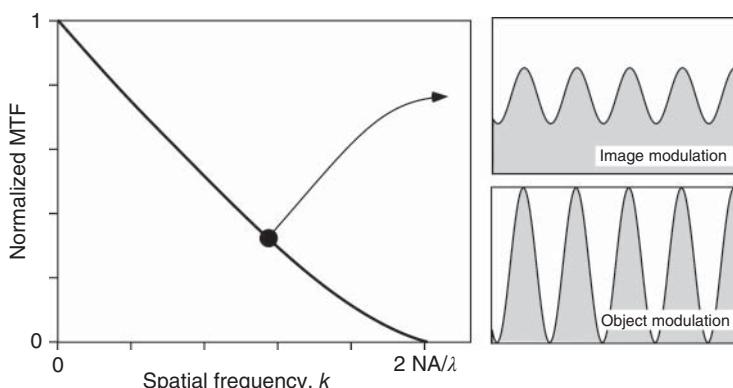
Abbe's approach to understanding the resolution limit still takes a global view: What is the minimal structural size  $\Delta x$  of an object still visible in the image? It does, for instance, not make any prediction about the contrast we can expect at this scale, an obviously relevant question for imaging applications. Analyzing and enhancing contrast in optical images is therefore an important issue in addition to resolution.

#### 5.4.2.1 The Modulation Transfer Function (MTF)

If we associate the minimal size with a maximal spatial frequency  $k = 2\pi/a$ , it becomes clear that in order to explore more deeply the practical and theoretical physical limits of optical imaging, it is useful to analyze the transfer of object structures (the transfer of information) with respect to all relevant spatial frequencies: this is the program of Fourier optics. An explicit introduction into Fourier optics is beyond the scope of our text, we have to refer the reader to the relevant literature [57]. Instead we present an illustration of selected tools useful for analyzing the quality of optical imaging instruments.

We have already introduced the *point spread function*  $\text{PSF}(x, y)$ , which describes the image amplitude distribution of a point-like object created by an optical imaging system, for example, a simple lens. As outlined in Section 2.7 on Fraunhofer diffraction, the PSF of a circular aperture is the Airy function (2.71). In Figure 5.21 the famous example of the Hubble Space Telescope (HST) is shown, which only after a repair reached its projected resolution. In general the PSF contains information about all aspects of an imaging system including aberrations.

It is known that a spatially point-like object would correspond to a flat spectral distribution in  $k$ -space. The complex spectral amplitude distribution of the PSF therefore (the Fourier transform  $\text{OTF}(k_x, k_y) = \mathcal{F}\{\text{PSF}(x, y)\}$ ) describes how this initially flat distribution is transferred to the image. It is called *optical transfer*



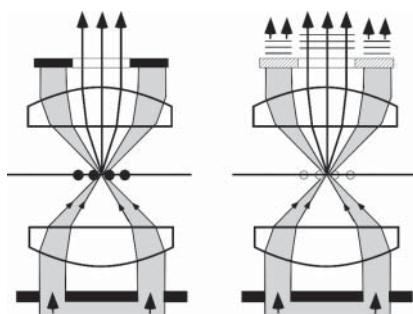
**Figure 5.11** Diffraction-limited modulation transfer function (MTF) for a circular aperture. The MTF shows the contrast  $= (I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$  (example on the right hand) as a function of spatial frequency  $k = 2\pi/\lambda$ . Disturbances and aberrations reduce the MTF of real imaging devices below this theoretical limit.

*function* (OTF) and contains spectral information about the contrast that can be expected.

For the assessment of imaging systems, the *modulation transfer function* (MTF) is in widespread use. It is defined by the normalized modulus  $MTF(k_x, k_y) = |\text{OTF}(k_x, k_y)|/\text{OTF}(0, 0)$ . In Figure 5.11 we show the MTF of a circular aperture, which defines the theoretical limit for optical imaging systems. Aberrations, other distortions, and additional optical components will reduce the MTF below this limit.

Let us finally mention a useful quantity to rate the performance of an imaging system, the *Strehl ratio*: it is defined as the ratio of the measured intensity maximum of the PSF of a point-like source versus an idealized PSF.

#### 5.4.2.2 Enhancing Contrast



**Figure 5.12** Enhancing contrast by dark-field and phase-contrast imaging.

The information content of any physical signal is ultimately limited by the signal-to-noise ratio for images: can we faithfully distinguish structures from noise? Due to the importance of microscopy, therefore numerous methods have been invented to systematically improve the visibility of structural contrast. Let us quote important examples.

*Dark-field imaging.* Every photographer knows that it is difficult to take a picture of a spider web against a bright background but straightforward against the dark and with sidewise illumination. In Figure 5.12 exactly this scheme is realized for a microscope. While the transmitted illumination light is blocked, light scattered off the object toward the axis carries the image information.

*Phase-contrast imaging.* Objects that are fully transparent for visible light but still do have structures occur frequently, for example, cell structures in biology, and they offer very little contrast. Small phase shifts caused by index of refraction differences nevertheless carry information about the structure of the object. Since we cannot see phase shifts with our eyes (or any intensity detector), we need to convert phase contrast into amplitude contrast, for example, by inserting the phase ring as shown in Figure 5.12, causing retardation of the wave fronts for phase–amplitude conversion. The phase ring also attenuates the illumination light since optimal contrast is obtained if the zeroth and the first orders have similar amplitudes (see Section 5.4.1.2). For the method of phase-contrast imaging, which has far-reaching consequences since it allows to observe, for example, living cells, in 1936, the Nobel Prize was awarded to F. Zernike (1888–1966).

*Polarization contrast imaging.* Two polarization filters (polarizer and analyzer) oriented at  $90^\circ$  and can hence efficiently suppress background light and provide a dark background. Bright structural contrast is obtained due to birefringence of the samples, which modifies incoming linear polarization. The so-called

lambda plate may rotate polarizations for certain colors back to the original and hence suppressed orientation. This method turns grayscale images into colorful pictures used, for instance, for analyzing rock composition.

*Fluorescence microscopy.* With the advent of narrowband lasers, it became possible to efficiently and selectively excite atoms and molecules. There are many molecules such as the ones used for dye lasers (see Section 8.5.3) that re-emit the absorbed energy at a redshifted wavelength with high quantum efficiency, that is, without losing it to internal degrees of freedom. They furthermore allow many absorption–re-emission cycles before being photobleached. For microscopy such fluorophores offer interesting applications since they can be used as markers for structurally relevant substances by binding them to, for example, enzymes. Popular species include fluorescein or the *green fluorescent protein* (GFP). Spectral discrimination of the fluorescing substances is then straightforward by, for example, color filters. Fluorophores can be observed down to the level of single molecules, which has contributed to breaking the Abbe and Rayleigh resolution limit (see Section 5.5.3).

## 5.5 Scanning Microscopy Methods

Instead of parallel recording of an entire image, a serial scanning method offers additional degrees of freedom through manipulation of a sample and recording its local response defined by the focus of the objective. Computers then reconstruct the image from this information point by point. Here we concentrate on purely optical scanning microscopy. The famous tunneling and force microscopy methods have strongly contributed to the success of the scanning methods by demonstrating atomic resolution but rely on nonoptical mechanisms for contrast generation.

### 5.5.1 Depth of Focus and Confocal Microscopy

Every user of a microscope knows that he or she has to adjust the objective–sample distance to render the image “sharp” and that the range of adjustments for which sharp images are generated decreases with increasing magnification. The longitudinal distance along the optical axis of the microscope between the two points where the sample can still be seen sharply is called the “depth of focus.” A quantitative measure of the depth of focus may be obtained, for example, from the geometrical considerations in Figure 5.13.

The movement of an object point by  $\delta g$  out of the “true” object plane causes a spot with diameter  $\Delta x$  in the intermediate image plane. From Eq. (5.2) we can derive that for  $\delta g/g \ll 1$  the image distance moves

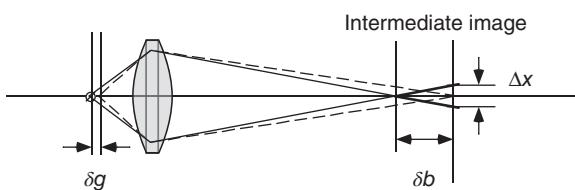


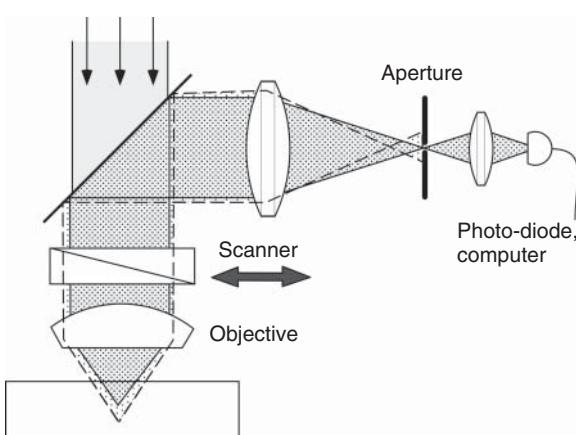
Figure 5.13 Geometry of the depth of focus.

approximately by  $\delta b/b \sim -\delta g/g$ . Geometrical considerations then yield directly the result  $\Delta x = |\delta b D/2b| = |\delta g D/2g| \simeq |\delta g D/2f|$ . If we require that this spot should stay smaller than the diffraction disc of the object point, then we find for a maximum tolerable movement  $\Delta z$ :

$$\Delta z \leq \Delta x_{\min} \frac{f}{D/2} \sim \frac{\lambda f^2}{(D/2)^2} \sim \frac{\lambda}{NA^2}.$$

Then, for larger magnifications, the depth of focus becomes very small as well; it reaches the order of a wavelength. The small depth of focus for the reverse process of microscopy, reduction in optical projection lithography, causes high demands on the mechanical tolerances of wafer steppers in optical lithography. We can draw an analogy to Gaussian ray optics (see example p. 45): the length of the Rayleigh zone of the focused coherent light beam has the same ratio with the diameter of the focal spot as the depth of focus has!

Confocal microscopy uses the short depth of focus of an image with short focal length and large NA to gain – beyond “planar” information – three-dimensional information of the device under test. In Figure 5.14 the basic principle of confocal microscopy is shown: a coherent light beam creates a narrow spot with little depth of focus within the probe. Only the light intensity reflected, or scattered, out of the spot is focused onto an aperture. Structures in other planes are projected into other planes, and therefore radiation originating from those planes is largely suppressed by the aperture.



**Figure 5.14** Principle of confocal microscopy. The dashed line shows that radiation from a deeper plane is suppressed by the aperture.

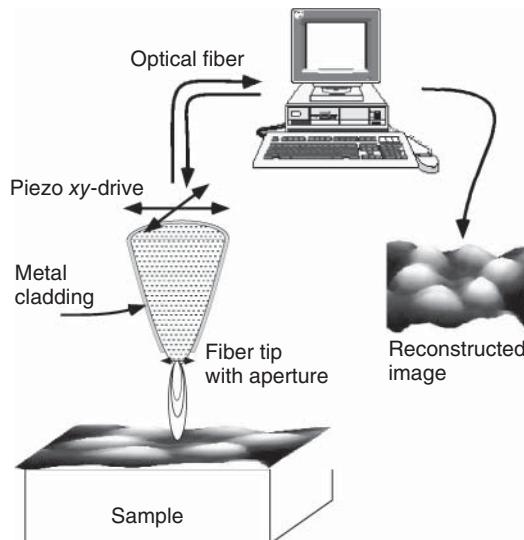
The intensity transmitted through the aperture is continuously monitored by a photodetector and recorded by a computer, which recovers an image from these data. Confocal microscopy is an example of scanning probe microscopy, since data from the sample are taken point by point and later reconstructed by computer algorithms. In Figure 5.14 this is achieved by a movable beam steerer (“scanner”).

Confocal microscopy achieves resolutions of about  $1 \mu\text{m}$ ; the advantage is access to the third dimension, which is, of course, only possible in transparent samples.

### 5.5.2 Scanning Near-Field Optical Microscopy (SNOM)

The limited resolution of a microscope is a “result” of Maxwell’s equations. In free space the curvature of the electric field cannot occur on a scale much shorter

**Figure 5.15** Scanning near-field optical microscope. An aperture at the end of an optical fiber is used as a source or detector of radiation fields with a resolution of less than optical wavelengths.



than a wavelength. An ideal point light source is imaged by an optical imaging system into a small but finite spot at best, and this limits the resolution by diffraction to a value of about half a wavelength  $\lambda/2$  of the light in use.

In the near field of a radiating system, this limit can be exceeded, since in the presence of a polarizable material, the propagation of electromagnetic waves is no longer restricted by the diffraction limit. A typical arrangement is introduced in Figure 5.15. An optical fiber is pulled by a pipette pulling device to yield a tip, the radius of curvature of which is less than 100 nm. This receives a cladding, for example, out of relatively low-loss aluminum, which leaves only a small aperture, which serves as radiation source or detector of the local light field (or both simultaneously).

The end of the fiber, which is made to oscillate by a piezo drive, experiences an attractive van der Waals force at a typical distance of micrometers and a damping force that may be used to adjust the distance as in atomic force microscopy (AFM) by some feedback control circuit. The system therefore gives information about the surface topography of the sample as well. The optical information is recorded by detecting at the end of the fiber the light picked up or reflected at the tip. With smaller and smaller apertures, the spatial resolution increases and can be pushed significantly below the wavelength in use (typically  $\lambda/20$ ); note that it depends essentially on the diameter of the aperture and not on the wavelength. On the other hand, the detection sensitivity decreases more and more, because the sensitivity decreases with a high power of the diameter of the aperture and even 1 mW of light power damages the apertures.

### 5.5.3 Overcoming the Rayleigh–Abbe Resolution Limits with Light

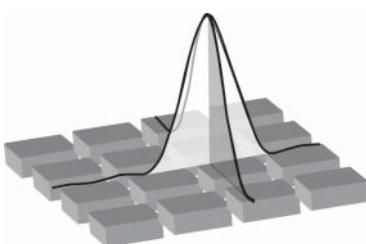
For propagating waves the Rayleigh and Abbe resolution limits do pose a fundamental limit – in the far field, we cannot concentrate the spot size of the wave field any further. Traditionally we therefore had to resort to short-wave electron

microscopes to obtain higher resolution. Attractive applications, however, which are not compatible with electron microscopy, for example, the observation and manipulation of living cells, continue to render optical methods superior for many cases.

Here we discuss two successful concepts that were awarded the Nobel Prize in Chemistry in 2014, E. Betzig (1960 –), S. Hell (1962 –), and W. Moerner (1953 –), in recognition of their potential to “see” beyond the Abbe–Rayleigh limit: Moerner took fluorescence spectroscopy to the limit where the fluorescence of single molecules is observed and their position determined below the Rayleigh limit. By optically switching the fluorescence state of molecules and fluorophores between bright and dark states, Betzig (photoactivated localization microscopy (PALM) method) and Hell (stimulated emission depletion (STED) method) were the driving persons to show that structures with length scales well below the Rayleigh limit could be uncovered with optical methods.

### 5.5.3.1 Single-Molecule Detection

Detection of single quantum radiators was not started with molecules: initially, trapped and laser-cooled single  $\text{Ba}^+$  ion set the stage in 1979 [58] for a new era of optical quantum control based on the methods presented in Chapter 14. The door to working with single large molecules was opened in 1989 with pentacene dopant molecules embedded into a host crystal [59]. The embedded molecules cause strain in the host lattice, which leads to a distribution of narrow spectral shifts of the dopants. In the wings of this distribution, it becomes possible to find spectrally isolated resonance lines of individual molecules, which can also be found by a microscope.



**Figure 5.16** Point spread function distribution of a point-like radiation source registered by a digital camera. The centroid is determined with uncertainty well below the RMS width and sub-pixel resolution.

created by adding all pixel gray values along the transverse direction. Then one can estimate the uncertainty from [60, 61]

$$(\Delta x_{\min})^2 = \frac{\text{RMS}^2}{N} \left( 1 + \frac{(\Delta_p/\text{RMS})^2}{12} + \frac{n_\perp \sigma_B^2}{N} \frac{4\sqrt{\pi}}{\Delta_p/\text{RMS}} \right). \quad (5.7)$$

The first term  $\text{RMS}^2/N$  is well known; it relates the uncertainty of the centroid of the distribution to its RMS width and the number of measurements – here

Observing the fluorescence of a point-like molecular radiator (which must survive numerous excitation–re-emission cycles to be suitable for this technique) amounts to recording the PSF with the microscope. Investing our knowledge – the PSF is “hiding” an original point-like source, we can identify the centroid of the PSF as the position with uncertainty well below its width. In modern microscopy, images are almost always recorded with digital cameras where the spatial information is organized in discrete pixels (Figure 5.16). In order to estimate the minimal uncertainty for position detection in a given direction, a 1D distribution is

the number of photons  $N$ . As with any conventional measurement, this uncertainty scales as integration time  $T^{-1/2} \sim N^{-1/2}$ . The second term accounts for the discrete structure of spatial data sampling, that is, camera pixels of size  $\Delta_p$  in the object plane with RMS value  $\Delta_p/\sqrt{12}$ . It calls for a small  $\Delta_p/\text{RMS}$ , that is, the PSF should be distributed across a small number of pixels. The third term takes the omnipresent background noise (RMS value  $\sigma_B$  per pixel; the average background signal can be subtracted) into account for  $n_{\perp}$  transverse pixels. In contrast to the second term, it scales as  $(\Delta_p/\text{RMS})^{-1}$ . Hence for optimal conditions, a compromise with values around  $\Delta_p/\text{RMS} \sim 3-10$  needs to be chosen.

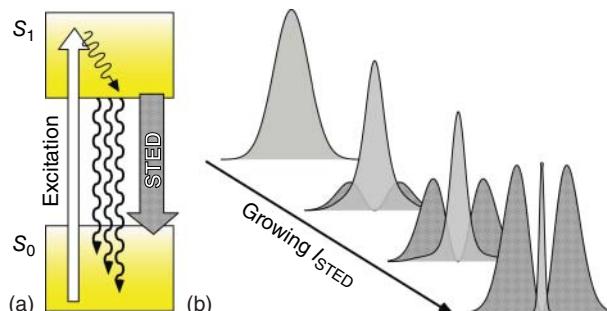
We note that position determination with the so-called super-resolution below the Rayleigh limit using the centroid of the PSF is possible for nonoverlapping single molecules only. Imaging of potentially dense structures thus requires additional measures.

### 5.5.3.2 PALM Microscopy

The PALM method [62, 63] as well as its variants like STORM relies on photo-switchable marker molecules (e.g., PA-GFP for photoactivatable green fluorescent protein), which can be switched between a fluorescent (bright) and a dark state by optical pulses. To avoid overlapping single molecule signals, the number of bright fluorescent molecules is now diluted such that position determination within the nanometric uncertainty of Eq. (5.7) is possible. The structure marked by, for example, the PA-GFPs is then reconstructed by computer from a series stochastically prepared dilute samples.

### 5.5.3.3 STED Microscopy

The STED method [64] uses an auxiliary and hollow *STED beam* to suppress fluorescence from most chromophores except for a nanometric region by **S**timulated Emission Depletion. Fluorophores follow the simple excitation scheme shown in Figure 5.17 (a). A focused Gaussian beam excites molecules in a small diffraction-limited area at some, for example, green wavelength. Non-radiative processes take them to the bottom of the upper  $S_1$  band from where they emit yellow or red color photons.



**Figure 5.17** (a) Excitation/depletion scheme of STED microscopy. (b) With increasing intensity  $l_{\text{STED}}$  of the STED beam, the island of actively fluorescing molecules is reduced to nanometric dimensions.

The hollow STED beam (a simple example for the formation of a hollow beam is discussed in Section 2.5.4.1) is superposed with the excitation beam Figure 5.17 (b) but tuned such that by stimulated emission (Section 7.3), molecules are efficiently returned to the  $S_0$  ground state; their fluorescence is thus effectively turned off. The efficiency of this process scales as  $\exp(-I_{\text{STED}}/I_0)$  where  $I_0$  is the so-called saturation intensity discussed in detail in Section 7.2.7.1. The cross section of the actively fluorescing area is reduced by the STED beam to [65]

$$\Delta x_{\text{STED}} = \frac{\lambda}{2n \sin \alpha} \frac{1}{\sqrt{1 + I_{\text{STED}}/I_0}}. \quad (5.8)$$

The application of a nonlinear optical process (stimulated emission) allows to obtain resolution at the several 10 nm scale and below the Abbe–Rayleigh limit.

## 5.6 Telescopes

Binoculars and telescopes are used to make terrestrial or astronomical objects more visible. In general, they are composed of two lenses or mirrors, the focal points of which coincide exactly, that is, their separation is  $d = f_1 + f_2$ . In the Galilean telescope in Figure 5.19, a concave (diverging) lens with negative focal length is used. Under these circumstances, the image matrix of the system reads as follows according to Eq. (1.25):

$$M_{\text{tel}} = \begin{pmatrix} -f_2/f_1 & d \\ 0 & -f_1/f_2 \end{pmatrix}. \quad (5.9)$$

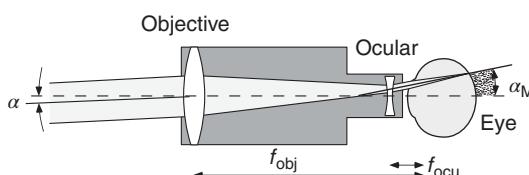
The total refractive power of this system vanishes,  $D_{\text{tel}} = 0$ . Such systems are called *afocal* [66]; their action is based on angle magnification only (Figure 5.18).

The objects are located effectively always at very large distances. From there, parallel bundles of rays originate, which are transformed into parallel bundles of rays at other, but larger, angles,  $\alpha \rightarrow \alpha_M$  as shown in Figure 5.19.

### 5.6.1 Theoretical Resolving Power of a Telescope

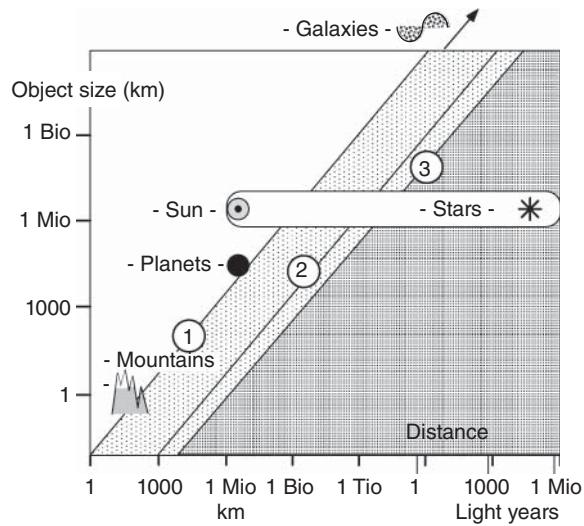
Before we determine the magnification, we want to consider the kinds of objects we might be able to recognize. Therefore, we have to recall the resolving power of a convex lens, which we have determined already in Eq. (2.68). There, we have already seen that at a fixed wavelength, the aperture of any imaging optics determines the smallest angle at which two point-like objects can still be distinguished. We reformulate this condition for telescopes:

$$\text{Minimum structural dimension} \simeq \frac{\text{wavelength} \times \text{distance}}{\text{aperture}}.$$



**Figure 5.18** Angle magnification in a Galilean telescope.

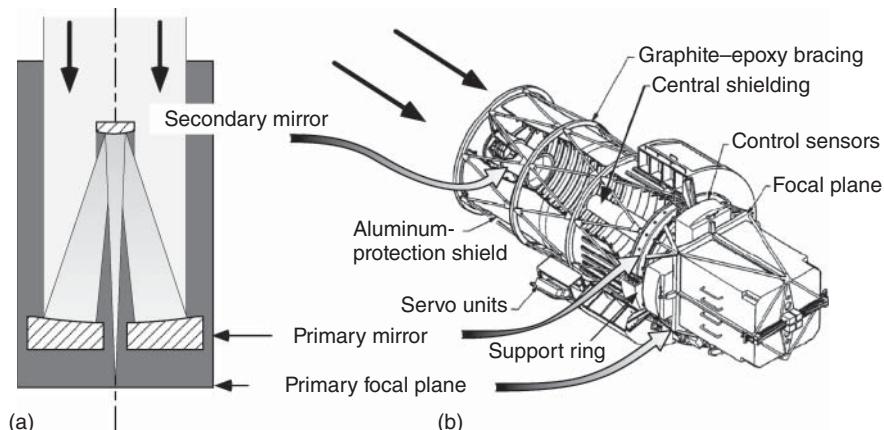
**Figure 5.19** Pattern recognition of objects far away, with the eye (1), a telescope with 10 cm mirror (2), and the Hubble Space Telescope (3) (1 light year = 9.5 trillion km).



The consequences for (1) the human eye (pupil 1 mm), (2) a telescope with 10 cm mirror, and (3) the 2.4 m mirror of the HST have been illustrated in Figure 5.20. The shapes of objects can be recognized above the limiting lines 1–3, whereas below those lines, the objects cannot be distinguished from points.

### 5.6.2 Magnification of a Telescope

In Figure 5.19 we introduced the concept invented by Galileo Galilei (1564–1642), which is composed of a convex lens with focal length  $f_{\text{obj}}$  and a concave eyepiece with a focal length  $f_{\text{ocu}}$ . Geometrical considerations, such as the calculation of



**Figure 5.20** (a) Reflector telescope of the Cassegrain type, schematic. (b) Hubble Space Telescope, in operation since 1990.

the system matrix  $\mathcal{M}_{\text{tel}}$  (5.9), show easily that the magnification of the angle by telescopes is

$$\text{Magnification} \quad \mathcal{M} = \frac{\alpha_{\mathcal{M}}}{\alpha} = -\frac{f_{\text{obj}}}{f_{\text{ocu}}}. \quad (5.10)$$

The negative sign of  $\mathcal{M}$  means that the image is inverted; therefore, the Galilean telescope of Figure 5.19 offers a non-inverted image due to the concave lens with negative  $f_{\text{ocu}}$ . Telescopes are large-volume devices, because large apertures and lengths of focus are advantageous. The minimum device length is

$$\ell_{\text{telescope}} = f_{\text{obj}} + f_{\text{ocu}}.$$

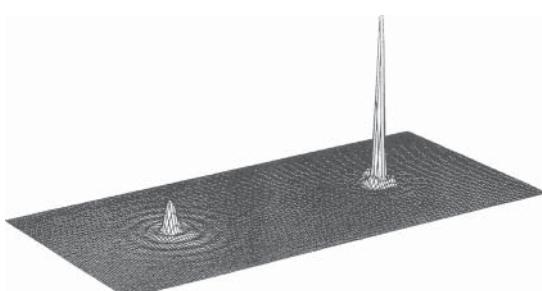
### 5.6.3 Image Distortions of Telescopes

Like all optical instruments, telescopes are affected by several aberrations (see next section). Here, we discuss two selected problems; furthermore, the Schmidt mirror, which sets an example for the correction of spherical aberrations, is described on p. 174.

#### 5.6.3.1 Lens Telescopes and Reflector Telescopes

Chromatic aberrations, which we will discuss in detail in Section 5.7.3, were identified very early as an obstacle to improving lens telescopes technically. Isaac Newton (1642–1727) was one of the first (1688) to discover that refractive lens optics, suffering from strong dispersion, should be substituted by the reflective optics of reflector telescopes, which nowadays has become the standard device layout.

In Figure 5.20 the Cassegrain concept is shown, which consists of a primary concave mirror and a secondary convex mirror. If the primary mirror has parabolic shape, then the secondary must have hyperbolic shape; however, other types (with other types of aberrations) are possible as well. One of the newest instruments of this type is the HST, which since 1990 has delivered more and more new and fascinating pictures of stars and galaxies far away, not influenced by atmospheric fluctuations [67].



**Figure 5.21** Point spread function of the Hubble Space Telescope before (left) and after (right) installation of the COSTAR (correction optics). (After [68].)

In the original HST configuration, a mistake in the calculation of the mirror properties resulted in aberrations, which inhibited realization of the total theoretically available resolution of the HST! However, there was real delight after the optics of the telescope were corrected by an additional pair of mirrors – after, so to speak, fitting “spectacles” to the HST.

To evaluate the quality of an imaging system, often the so-called PSF is used, by means of which the image of a point according to wave theory is described, taking the exact form of the imaging system into account. In Figure 5.21 the result for the calculation for the HST before and after the installation of the correction optics is shown.

#### 5.6.3.2 Atmospheric Turbulence

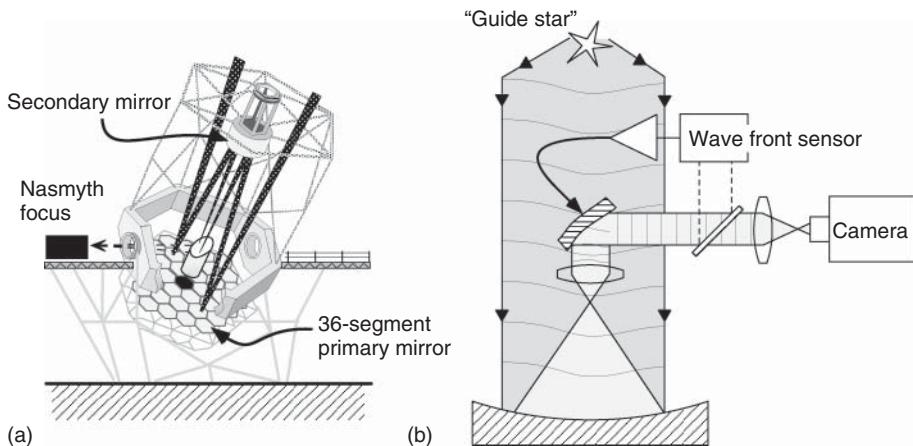
The HST, with a mirror diameter of 2.4 m, does not have an extraordinary diameter; with 10 m the telescope at the Keck Observatory in Hawaii offers much more than that. However, the resolution of the HST is much superior to that of terrestrial telescopes, because the resolving power of the latter is limited to effectively 10 cm by turbulent motion of the atmosphere (like the optical telescope in Figure 5.20)! However, owing to their collecting power, giant terrestrial telescopes offer the possibility to study faint objects with very low light power in more detail.

For the installation of huge mirror telescopes, one looks for environments with very favorable atmospheric conditions, for example, in the Andes of Chile or on the Hawaiian Islands. The 10 m telescope at the Keck Observatory is one of the most modern facilities (construction year 1992). To use the total theoretical efficiency of a mirror, the geometric shape – a sphere, hyperboloid, or whatever – must be kept to within subwavelength precision. However, with increasing size, this requirement is more and more difficult to fulfill, because these heavy mirrors are even distorted by the influence of gravity, thus causing aberrations. Therefore, the Keck mirror was manufactured with 36 segments, the positions and shapes of which can be corrected by hydraulic positioning elements in order to achieve optimum imaging results.

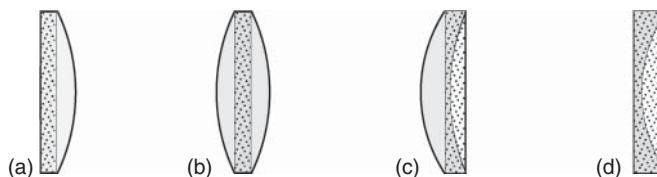
Actively tuned optical components are used more and more and are summarized under the term *adaptive optics*. With newer developments it is possible to compensate for atmospheric turbulence that changes on a time scale of about 100 ms. Typically, for that purpose, the wave front must be analyzed and used to control a deformable mirror within a feedback loop. The wave front in the upper atmosphere may be assessed, observing atmospherically, by analysis of the light from a very bright reference star or by the positioning of an “artificial star” (Figure 5.22) [69], such as a laser excited source, for example, in the upper atmosphere. For this purpose one uses a dye laser tuned to the yellow D2 line of sodium at 589 nm. At about 100 km, the laser beam hits a layer of sodium atoms created by micrometeorites, which vaporize as they enter the upper atmosphere. Resonance fluorescence of the sodium atom then acts as a glowing guide star wherever an astronomer needs it.

## 5.7 Lenses: Designs and Aberrations

The spherical biconvex lens is, so to speak, the cardinal case of a convex lens and is the lens usually illustrated in figures. All spherical lenses (Figure 5.23) cause aberrations, however, and the application of certain designs depends completely



**Figure 5.22** (a) Schematic of the 10 m telescope at the Keck Observatory. (b) Artificial or reference stars for the application of adaptive optics.



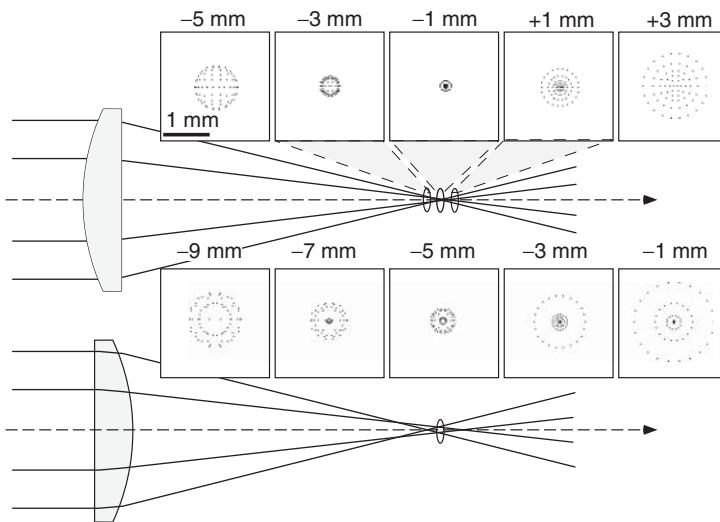
**Figure 5.23** Important lens types: (a) planar convex lens; (b) biconvex lens; (c) convergent meniscus lens; and (d) planar concave lens.

on the area of application. As a rule of thumb, we recall the paraxial approximation: the linearized form of Snell's law ( $\sin \theta \rightarrow \theta$ , Eq. (1.13)) is better fulfilled, the smaller the angles of refraction are! Therefore, it is convenient to distribute the refraction of a beam of rays, passing through a lens, as evenly as possible to the two refracting surfaces. At selected points, aberrations can be compensated by adequate choice of the surfaces. In a multi-lens system (doublet, triplet, etc.), several curved surfaces and thus degrees of freedom are available. However, the perfect lens system, correcting for several types of aberrations at the same time (see the following discussion), cannot be realized in this way, and thus all multi-lens systems ("objectives") are in general designed for specific applications. Before introducing the technical discussion of aberrations, we want to collect some intuitive arguments for dealing with one or two lenses. More complex systems must be analyzed numerically.

### 5.7.1 Types of Lenses

#### 5.7.1.1 Planar Convex Lenses

This type of lens only has one curved surface and therefore may be manufactured quite cheaply. For typical indices of refraction of technical glasses of  $n = 1.5$ , one finds according to Eq. (1.20),  $f = -1/D \simeq 2R$ . To focus a light ray, the planar convex lens may be used in two different orientations. Figure 5.24 indicates how



**Figure 5.24** Spot diagrams of a planar convex lens for two different orientations (after commercial software for the analysis of aberrations). The specifications of the distances refer to the distances to the nominal focal point (here 66 mm).

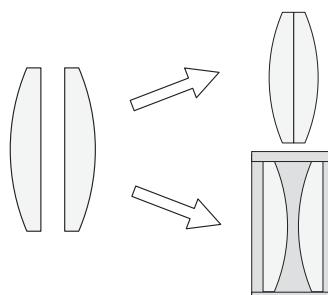
spherical aberrations primarily affect the ability to focus. The so-called spot diagrams show the evolution of the size of a spot along the optical axis. Obviously, it is convenient to distribute the refractive power to several surfaces – indeed, in the orientation of the lower part in Figure 5.24, refraction occurs only on one side of the lens, resulting in reduced quality focusing.

### 5.7.1.2 Biconvex Lenses and Doublets

We may imagine a biconvex lens as composed of two planar convex lenses back to back, as indicated in Figure 5.25. Therefore, the refractive powers add, and we find for common glasses again according to Eq. (1.20) with  $n \sim 1.5$

$$f \approx R.$$

For 1 : 1 imaging, the biconvex lens singlets have minimum spherical aberration, which is important, for example, for collimators. However, the refractive powers of planar convex lenses add in exactly the same way if they are mounted with their spherical surfaces opposing each other. Therefore, in a 1 : 1 image, the refractive power is distributed to four surfaces, and one achieves further reduction of aberrations.



**Figure 5.25** Biconvex lens and planar convex doublet.

### 5.7.1.3 Meniscus Lenses

Meniscus lenses may minimize as singlets the aberrations for a given distance between object and image. Indeed, they are first of all part of multi-lens objectives and serve, for example, to change the length of focus of other lenses,

without introducing additional aberrations or coma. Such components are called *aplanatic* [70].

### 5.7.2 Aberrations: Seidel Aberrations

Here, we briefly sketch the fundamental formal method, going back to P. Seidel (1821–1896), to classify aberrations. Since it is now necessary to deal with non-axial contributions as well, the complex numbers  $r_0 = x + iy$  are convenient for the discussion of the traces of light rays.

We use the notation from Figure 5.26 following the discussion of matrix optics from Section 1.9 and using complex numbers for convenience. The relation between the ray originating at  $r_0 = x + iy$  propagating with slope  $r'_0$  and its image point at  $r(z)$  is described by

$$\begin{aligned} r(z) &= g(z; x, x', y, y') \\ &= f(z; r_0, r_0^*, r'_0, r'^*_0). \end{aligned}$$

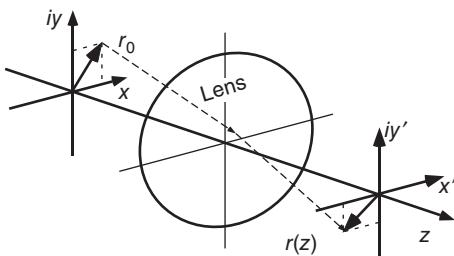


Figure 5.26 Notation of ray parameters.

One may use the Laurent expansion, known from the theory of complex numbers [24]:

$$r(z) = \sum_{\alpha\beta\gamma\delta \geq 0} C_{\alpha\beta\gamma\delta} r_0^\alpha r_0^{*\beta} r'^\gamma_0 r'^{* \delta}_0. \quad (5.11)$$

A rotation by the angle  $\Theta$  in the plane of the object,  $r_0 \rightarrow r_0 e^{i\Theta}$ , must cause a rotation by the same angle in the image plane:

$$r(z) e^{i\Theta} = \sum_{\alpha\beta\gamma\delta} C_{\alpha\beta\gamma\delta} r_0^\alpha r_0^{*\beta} r'^\gamma_0 r'^{* \delta}_0 e^{i\Theta(\alpha-\beta+\delta-\gamma)}.$$

From that one finds directly the first condition

$$\begin{aligned} \text{(i)} \quad & \alpha - \beta + \gamma - \delta = 1, \\ \text{(ii)} \quad & \alpha + \beta + \gamma + \delta = 1, 3, 5, \dots, \end{aligned} \quad (5.12)$$

while the second follows from the special case  $\Theta = \pi$ , resp.  $r(z) \rightarrow -r(z)$ , from direct reflection at the optical axis. It determines that only odd orders 1, 3, ... may occur.

#### 5.7.2.1 Ray Propagation in First Order

In first order ( $\alpha + \beta + \gamma + \delta = 1$  in Eq. (5.12)), one finds  $\beta = \delta = 0$  and

$$r(z) = C_{1000} r_0 + C_{0010} r'_0.$$

This form corresponds exactly to the linear approximation, which we already used as the basis of matrix optics and discussed in detail in Section 1.9.

#### 5.7.2.2 Ray Propagation in Third Order

In third order ( $\alpha + \beta + \gamma + \delta = 3$ ), in total six contributions arise, the prefactors of which are known as *Seidel coefficients*. We find the conditions  $\alpha + \gamma = 2$  and  $\beta + \delta = 1$ , which can be fulfilled with six different coefficients  $C_{\alpha\beta\gamma\delta}$  and are itemized in Table 5.2.

**Table 5.2** Seidel coefficients of aberrations.

Coefficient	$\alpha$	$\beta$	$\gamma$	$\delta$	$\alpha$	Aberration
$C_{0021}$	0	0	2	1	$r'^3$	Spherical aberration
$C_{1011}$	1	0	1	1	$r'^2 r$	Coma I
$C_{0120}$	0	1	2	0	$r'^2 r$	Coma II
$C_{1110}$	1	1	1	0	$r' r^2$	Astigmatism
$C_{2001}$	2	0	0	1	$r' r^2$	Curvature of the image field
$C_{2100}$	2	1	0	0	$r^3$	Distortion

From the table we will now discuss several selected aberrations and the corrections of those in more detail. The coefficients are properties of the lens or the lens system, and in the past the theoretical determination of those has been possible only for certain applications due to the enormous numerical calculation expenditure. Nowadays, these tasks are done by suitable computer software.

### 5.7.2.3 Aperture Aberration or Spherical Aberration

We have already introduced the effect of spherical aberration in Figure 5.24 with the example of a planar convex lens with spot diagrams. It depends only on the aperture angle ( $r'_0$  in Eq. (5.11)), may be reduced by limiting the aperture, and is therefore called “aperture aberration.” However, on doing this, the imaging system very quickly loses light intensity. Therefore, for practical applications, further corrections are necessary, which can be achieved by choice of a combination of convenient radii of curvature (“aplanatic systems”) or by the use of a lens system, for example. In particular, spherical aberration is often corrected at the same time as chromatic aberration (see Section 5.7.3).

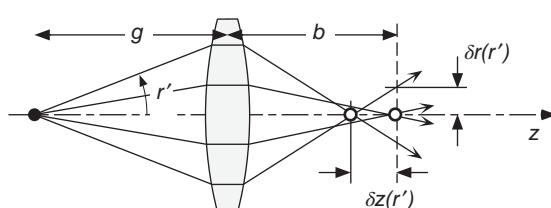
#### Example: Aperture aberration of a thin lens

Since spherical aberration is determined by the aperture angle only, we consider a point on the axis,  $r_0 = 0$ , at a distance  $g$  from the lens (Figure 5.27). As we have already discussed in more detail on p. 20, the image point must also lie at  $r(z) = 0$  and must be independent of  $r'_0$ . From the combination of the linear approximation with the Seidel approximation, one finds

$$r(z) = 0 = gz \left( \frac{1}{g} + \frac{1}{z} - \frac{1}{f} \right) r'_0 + C_{0021} r'_0{}^3.$$

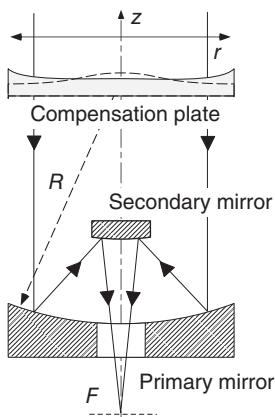
Within the paraxial approximation, Eq. (5.2) is fulfilled exactly for  $z = b$ . But here the intercept with the optical axis depends on  $r'_0$ . In linear approximation for small shifts, it holds that  $z = b + \delta z(r')$  and  $r(z) = 0$  for  $r'_0$  is valid for

$$\delta z = \frac{b}{g} C_{0021} r'_0{}^3.$$

**Figure 5.27** Spherical aberrations.

Here, we have determined the so-called longitudinal spherical aberration. In a similar way the transverse spherical aberration ( $\delta r(r')$  in Figure 5.27) may be calculated.

### Example: Schmidt mirror



**Figure 5.28** Cassegrain-Schmidt telescope.

An interesting variant of the commonly used Cassegrain concept is the so-called Schmidt telescope, which is additionally equipped with a compensator plate made of glass. It corrects not only the aperture aberration but also chromatic aberrations, coma, and astigmatism. Therefore large image fields up to  $6^\circ$  are achieved that are very suitable for a celestial survey campaign. Standard telescopes do not achieve more than about  $1.5^\circ$ .

Schmidt's idea first takes into account that a parabolic mirror may generate perfect images very close to the axis but on the other hand causes strong comatic distortions even at small distances, while a spherical mirror creates a much more regular image of a circular observation plane. In the vicinity of the axis, the location of the spherical mirror may be described according to the expansion

$$z = \frac{r^2}{4f} + \frac{r^4}{64f^3} + \dots,$$

where the first term corresponds exactly to the paraboloidal form. The compensator plate with refractive index  $n$  compensates exactly for the difference in optical path length between spherical and paraboloidal surfaces if the variation of the thickness is chosen to be

$$\Delta(r) = \frac{r^4}{(n-1)32f^3}$$

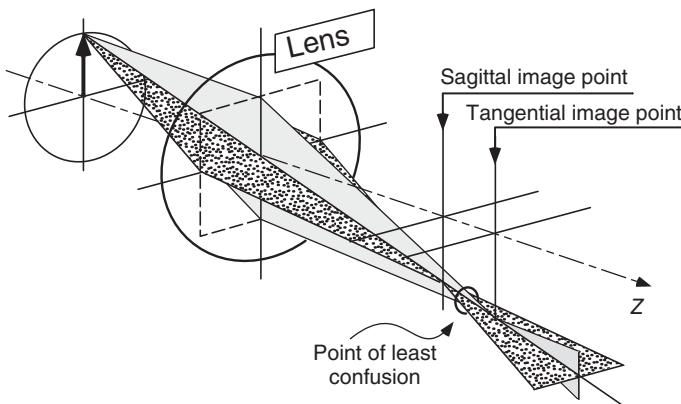
(the factor of 2 occurs due to reflection). This form – the solid line variant in Figure 5.28 – increases toward the aperture of the telescope, whereas the dashed variant in Figure 5.28 minimizes chromatic aberrations as well [27].

When the compensator plate is mounted within the plane of the center of curvature of the primary mirror (radius  $R$  in Figure 5.28), then the correction is valid also for larger angles of incidence within good approximation.

#### 5.7.2.4 Astigmatism

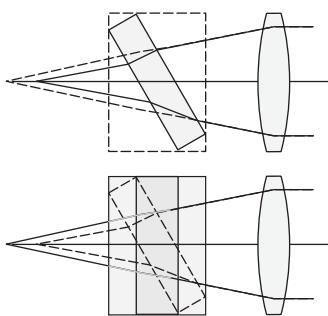
If the object points do not lie on the optical axis, then the axial symmetry is violated, and we have to discuss the “sagittal” and the “tangential” planes of beam propagation separately.<sup>1</sup> The effective length of focus of a lens depends on the angle of incidence, as can be recognized in Figure 5.29, where the light rays of

<sup>1</sup> Astigmatism of an optical lens also occurs for a component that is perfectly rotationally symmetric. It should be distinguished from astigmatism of the eye, which is caused by cylindrical asymmetry of the cornea and creates image points at different distances even for axial points.



**Figure 5.29** Astigmatism of a lens. Within the sagittal (dotted) and tangential planes (shaded), the image points lie at different distances.

the sagittal and tangential planes are concentrated into two different focal lines. Between these two lines, there exists a plane where one may identify an image point of “least confusion” as a compromise.



**Figure 5.30** Astigmatism of a tilted planar plate.

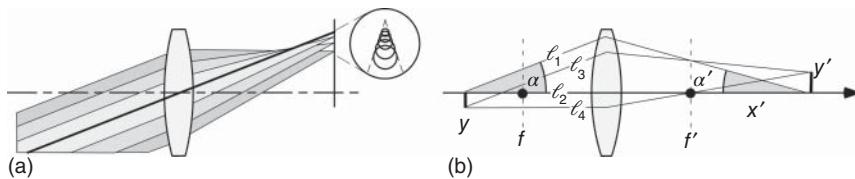
#### Example: Astigmatism of tilted planar plates

When light passes through a planar plate at an oblique angle, this leads to different effective focal lengths and thereby to astigmatism. We have illustrated this qualitatively in Figure 5.30. In turn, a planar plate may be used to compensate for the astigmatism of other components as well. For example, light beams emanating from diode lasers in edge-emitting configuration do not have axial symmetry in general (see Chapter 10). They exhibit astigmatism that can be corrected by a window at a suitable angle.

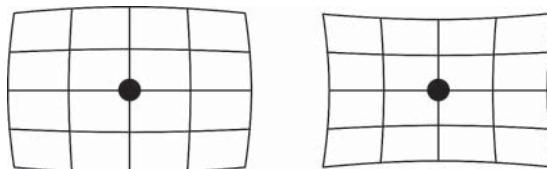
In laser resonators, optical components are often installed at the Brewster angle, which causes astigmatism. If curved concave mirrors are used, then the astigmatism of those elements (see p. 16) may be used for compensation again by suitable choice of the angle [71].

##### 5.7.2.5 Coma and Distortion

Among all image aberrations, the one called “coma” (from the Greek word for *long hair*) or asymmetry aberration is the most annoying. Coma causes a comet-like tail (i.e., where the name comes from) for non-axial object points, which we have illustrated qualitatively in Figure 5.31. In our example we take a spherically corrected lens (i.e., all rays leaving an axial point are imaged at the same distance  $x'$  while the lateral displacement at the imaging plane grows linearly with distance from the central ray).



**Figure 5.31** Comatic aberrations for a bundle of parallel rays (a). For ideal stigmatic imaging, optical path lengths  $\{\ell_1, \ell_2\}$  and  $\{\ell_3, \ell_4\}$  are equal (b).



**Figure 5.32** Distortion by field curvatures.

In order to reduce this (and other aberrations linear in  $\{y, y'\}$ ) (see Figure 5.31), we can impose a condition that the paraxial and wide angle rays of non-axial object and image points  $\ell_3$  and  $\ell_4$  have the same optical path lengths at least to first order. This is by definition the case for the axial points and a spherically corrected lens,  $\ell_1 = \ell_2$ . From  $y$  to the focal plane  $f'$ , the partial optical path lengths of parallel rays  $\ell_2$  and  $\ell_4$  are again by definition identical. Thus we have  $\ell_2 - \ell_4 = (x'^2 + y'^2)^{1/2} = x'(1 + (y'/x')^2/2 + \dots) - x' = \mathcal{O}(y'^2)$ .

To calculate the difference  $\ell_1 - \ell_3$ , we have to take into account the projections of  $y(y')$  onto  $\ell_3(\ell_4)$ : from  $y$  to the focal plane  $f'$ , the partial path of  $\ell_3$  is longer by  $y \sin \alpha$ , and from  $f'$  to the image point, it is shorter by  $y' \sin \alpha'$  than  $\ell_1$ . Thus we find  $\ell_1 - \ell_3 = -y \sin \alpha + y' \sin \alpha'$ . Straightforward algebra finally yields the condition  $\ell_3 - \ell_4 = y \sin \alpha - y' \sin \alpha' - \mathcal{O}(y'^2)$ . An optical system obeying the *Abbe sine condition*

$$y \sin \alpha = y' \sin \alpha' \quad (5.13)$$

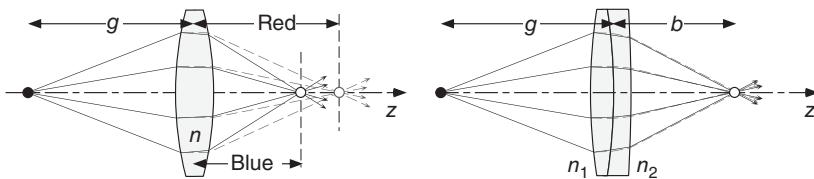
thus eliminates distortions depending linearly on axial separation and is related to the lateral magnification. Note that the sine condition generalizes the angular magnification of (5.3) for small angles.

According to Table 5.2 the image field curvature has a similar form to astigmatism, but it is axially symmetric. Distortion has two variants, pincushion and barrel distortions, which are also indicated in Figure 5.32. This contribution depends on the radius only.

### 5.7.3 Chromatic Aberration

Chromatic aberration is caused by *dispersion* of optical materials, since the index of refraction of the glasses used in lenses depends on the wavelength. The refractive power of a convex lens is in general higher for blue light than for red light. We discuss the effect of dispersion with the lens maker's equation (1.20) for a lens with refractive index  $n(\lambda)$  and radii of curvature  $R$  and  $R'$ :

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b} = -(n - 1) \left( \frac{1}{R'} - \frac{1}{R} \right).$$



**Figure 5.33** Chromatic aberration and correction with the so-called achromats.

The object distance is fixed, of course, but the image distance changes with the index of refraction,

$$\Delta \frac{1}{b} = -\Delta n \left( \frac{1}{R'} - \frac{1}{R} \right) = \frac{\Delta n}{n-1} \frac{1}{f} .$$

We know (see Section 1.9.4) that the refractive powers  $D$  of two directly neighboring lenses add, and due to  $D = 1/f$ , it holds that  $1/f_{\text{tot}} = 1/f_1 + 1/f_2$ . If the focal length of the combined system is not to change with wavelength anymore, then

$$\Delta \frac{1}{f_{\text{tot}}} = \frac{\Delta n_1}{n_1 - 1} \frac{1}{f_1} + \frac{\Delta n_2}{n_2 - 1} \frac{1}{f_2} = 0,$$

and we find the condition to correct for chromatic aberration:

$$f_2 \frac{\Delta n_1}{n_1 - 1} = -f_1 \frac{\Delta n_2}{n_2 - 1}. \quad (5.14)$$

To treat this situation more precisely, we have to use the linear expansion of the refractive index:

$$\Delta n_i = \frac{dn_i}{d\lambda} \Delta \lambda + \frac{1}{2} \frac{d^2 n_i}{d\lambda^2} (\Delta \lambda)^2 + \dots$$

However, because certain standard wavelengths for  $\Delta \lambda$  have been agreed (see Table 1.1), the aforementioned expression is sufficient. Since dispersion has the same sign for all kinds of known glasses, a lens without chromatic aberrations, which is called *achromatic*, must be composed of a convex and a concave lens (see Figure 5.33). Lenses also play an important role in particle optics; there, it is much more difficult than in light optics to construct achromatic systems, since divergent lenses cannot be constructed so easily.

Incidentally, the radii of curvature of the two lenses are not yet determined by the condition (5.14) for correction of chromatic aberration. This degree of freedom is often used to correct not only for chromatic aberration but also simultaneously for spherical aberration of a lens. Therefore, with an achromat one often gets a lens that is corrected spherically as well.

## Problems

- 5.1 Graphical image construction** Construct graphically the image of the object O, which is generated by the imaging system Figure 5.34a (e.g.,  $F_{1,2} = 3$  cm, separation of the lenses  $d = 8$  cm, distance of the object from the first focal point  $x = 2$  cm) and Figure 5.34b (e.g.,  $F_1 = -2$  cm,  $F_2 = 2$  cm,  $x = 2$  cm,  $d = 1.5$  cm).

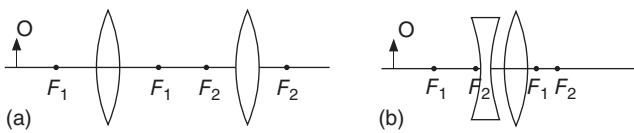


Figure 5.34 Lens systems for Problem 5.1.

- 5.2 Imaging with lenses** Show that the separation of an object and its image generated by a convex lens is at least four times the focal length. How large is the image of the sun generated by a lens with focal lens  $f$ ?
- 5.3 Bessel's method for the determination of focal lengths** The focal length of a lens can be determined with the following method. The lens is used to image some object, for example, a narrow bright slit, onto a screen. The separation  $c$  from the object  $G$  to the image  $B$  is measured. In the next step the lens is moved to the second imaging position, and the distance  $a$  between the old and the new positions of the lens is measured. Show that the focal length of the lens is given by  $f = (a^2 - c^2)/4a$ . What is the advantage of the method?
- 5.4 Projection** A projection device generates on a screen a magnified image of a small object such as a slide (Figure 5.35). For the perception of the observer, it is essential to uniformly illuminate the object. Explain the illumination scheme from the drawing, that is, the function of the filament  $F$ , reflector  $R$ , condenser lens  $CL$ , and projection objective  $PO$ . What is the requirement for the quality of the lenses (e.g., chromatic and other aberrations)?
- 5.5 The dentist's mirror** How do you construct a dentist's mirror such that for a working distance of 15 mm, an upright image with magnification 2 is obtained?
- 5.6 The near-sighted have advantages** Why is it possible for near-sighted people to see small objects better? How much can this effect account for?
- 5.7 Limits of the magnification glass** Why is it not possible to obtain with a magnification glass more than a 20- to 30-fold enlargement?
- 5.8 Cover slip and microscope imaging** Common microscope objectives are designed to achieve optimum resolution if the sample under observation

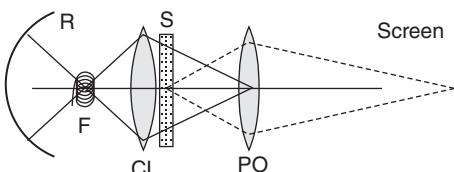


Figure 5.35 Schematic diagram of projection device, with light source. R: reflector; F: heated filament; CL: condenser lens; S: slide; PO: projection objective.

is covered with a slip of standard thickness of 0.17 mm. What is the effect on the image if the cover slip is missing, in particular with large numerical aperture?

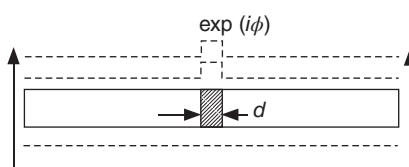
- 5.9 Contrast generation and microscope imaging** The information content of an image is in the first place an effect of contrast, for example, the distribution of gray values (or color values). The image contrast in our eye is a result of locally varying light absorption of an object, and the human eye requires – on a bright background – intensity variations of about 10–20%. Find out about the following two methods for contrast enhancement. (a) The dark-field method renders the background dark by suitable illumination of the sample. (b) With the phase-contrast method, bacteria and cell cultures can be made visible, which absorb very little light by making use of small refractive index variations between the cells and the aqueous environment.

As a model for a transparent object, consider a thin glass plate with a narrow stripe, which causes an additional phase shift of  $e^{i\phi}$  for light rays transmitted (Figure 5.36). Following Abbe we consider image formation in the microscope in analogy with Fourier optics, that is, we begin by determining the intensity distribution in the focal plane through Fourier transformation of the diffracting object. The image is then retrieved by inverse transformation of this distribution. The final image can be manipulated by introducing special filters into the focal plane, which modify the intensity and the phase distribution. Show that the diffracted field in the focal plane has the form  $\mathcal{E}(\kappa) = \delta(\kappa) + d(e^{i\phi} - 1) \sin(\kappa d/2)(\kappa d/2)$ . For the dark-field method, the axial rays (the bright field) are removed by a stop. For the phase-contrast method, they are delayed with respect to the remaining diffracted field by  $\pi/2$ . Study the effect of these operations onto the image.

- 5.10 Stars with lobes** Why do we frequently see stars with fourfold lobes on astronomical images obtained from reflector telescopes?

- 5.11 Tele lens and zoom lens** With telescopes we can observe distant objects. In order to obtain large magnification, a long focal length of the objective is essential (Eq. (5.10)). For a camera, however, telescopes are impractical because of their length. Tele lenses are constructed in such a way that a long focal length is obtained with a comparatively short overall length. Furthermore, the image is formed in the vicinity of the lenses. Begin by confirming that a large image of a distant object on a film (today more importantly on the CCD chip of a digital camera) requires a long focal length lens as well.

**Figure 5.36** Model object for the dark-field and the phase-contrast methods.



A tele lens is composed of a convex lens (focal length  $f_S$ ) and a concave lens (focal length  $-f_Z$ ). Show that for  $d > f_S - f_Z$  the lens system is equivalent to a single convex lens. Identify the range of focal lengths that can be controlled by varying the separation of the lenses. Sketch the positions at which the two lenses have to be positioned with respect to a fixed film position for different distant object positions.

- 5.12 Tricolore** Our eye is apparently well corrected for chromatic aberrations. When observing a red area and a blue surface, the accommodation muscles must increase the curvature of the eye lens more strongly for the red light, which is refracted less efficiently than blue light. What is the origin of the saying that, as painters claim, red approaches us more aggressively while blue seems to drag us into the downs? When one observes colored stained glass windows in churches, the different colors appear to be located in different planes. In the French national flag, the Tricolore, the red stripe is significantly wider (37% of the total height) than the white (33%) and the blue ones (30%). Why?

# 6

## Coherence and Interferometry

The principle of superposition from Section 2.1.6 delivers all the requirements needed to deal with the interference of wave fields. So one could treat interferometry and coherence just as part of wave optics or as an implementation of the principle of superposition. But interference in interferometry is critically determined by the phase relations of the partial waves. Therefore, we will begin by considering the somewhat unwieldy concept of coherence. We introduce quantitative measures for the role of phases, which in the real world are always subject to fluctuations.

Because of this enormous significance, we will devote this chapter to these aspects of wave optics. Nearly every field of physics dealing with wave and especially interference phenomena has taken up the concept of coherence, for example, quantum mechanics, which calls the interference of two states “coherence.” With the help of quantum mechanics, interference experiments are described and interpreted with matter waves.

The wealth of literature dealing with interferometry is not easily comprehensible, not least due to its significance, for example, for the methods of precision length measurement. In this book, we are focusing on the types of interferometers collected in Table 6.1 and underlying all variants.

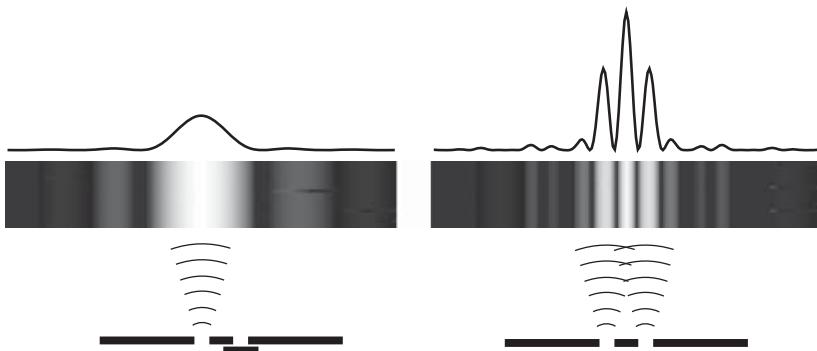
### 6.1 Young's Double Slit

The double-slit experiment first carried out by Thomas Young (1773–1829), an early advocate of the wave theory of light, is certainly among the most famous experiments of physics because it is one of the simplest arrangements to achieve interference. The concept is emulated in numerous variants in order to prove the wave properties of different phenomena, for example, of matter waves of electron beams [72] or atomic beams [73], which are discussed later in a short digression (see p. 187).

The fundamental effect of interference for light emanating from a double-slit arrangement instead of a single slit is shown in Figure 6.1. The conditions for and properties of this interference phenomenon are discussed in detail in Section 6.3.

**Table 6.1** Basic interferometer types.

Coherence type	Two-beam interferometer	Multiple-beam interferometer
Transverse	Young's double slit	Optical grating
Longitudinal	Michelson interferometer	Fabry-Pérot interferometer



**Figure 6.1** Young's double-slit experiment. On the right is shown the interference pattern from a double slit. On the left is shown the pattern of a single slit (one slit blocked) for comparison. A logarithmic gray scale has been chosen to make the interference patterns visible.

## 6.2 Coherence and Correlation

The concept of “coherence” means the “capability of interference” of wave fields. We shall see how we can also quantitatively describe it by “coherence length” and “coherence time.” These terms come from optical wave theory and state the distances or periods of time over which a fixed phase relation exists between (at least) two partial waves, so that within this interval the principle of superposition can be applied without any trouble.

When one calculates the distribution of intensity from the superposition of two *coherent* partial waves  $E_{1,2}(\mathbf{r}, t)$ , first the amplitudes have to be added, and then the square of the total field has to be taken:

$$\begin{aligned} I_{\text{coh}}(\mathbf{r}, t) &= \frac{c\epsilon_0}{2} |E_1(\mathbf{r}, t) + E_2(\mathbf{r}, t)|^2 \\ &= I_1(\mathbf{r}, t) + I_2(\mathbf{r}, t) + c\epsilon_0 \Re\{E_1(\mathbf{r}, t)E_2^*(\mathbf{r}, t)\}. \end{aligned} \quad (6.1)$$

In the *incoherent* case, however, the intensities of the partial waves are added instead:

$$I_{\text{inc}}(\mathbf{r}, t) = \frac{c\epsilon_0}{2} [|E_1(\mathbf{r}, t)|^2 + |E_2(\mathbf{r}, t)|^2] = I_1(\mathbf{r}, t) + I_2(\mathbf{r}, t).$$

We see immediately that the difference is determined by the superposition term.

This quantity  $I_{\text{coh}}$  can only be observed if there is a fixed phase correlation between  $E_1$  and  $E_2$  at least during the time of the measurement, because

every real detector carries out an average over a finite time and space interval. The times of fluctuation depend on the nature of the light source. For example, thermal light sources exhibit fluctuations on the scale of pico- and femtoseconds, which is much faster than detectors with response times at the nanosecond scale.

### 6.2.1 Correlation Functions

Quantitatively the relative time evolution of the phase of superposed fields can be understood by the concept of *correlation*. We define the general complex *correlation function*, also known as the *coherence function*, as

$$\begin{aligned}\Gamma_{12}(\mathbf{r}_1, \mathbf{r}_2, t, \tau) &= \frac{c\epsilon_0}{2} \langle E_1(\mathbf{r}_1, t + \tau) E_2^*(\mathbf{r}_2, t) \rangle \\ &= \frac{1}{T_D} \int_{t-T_D/2}^{t+T_D/2} \frac{c\epsilon_0}{2} E_1(\mathbf{r}_1, t' + \tau) E_2^*(\mathbf{r}_2, t') dt',\end{aligned}$$

which by the average (brackets  $\langle \rangle$ ) accounts for the finite integration time  $T_D$  of the detector. It is obvious that the interference term in Eq. (6.1) is a special case of this function. More exactly, this is the first-order correlation function. Fully developed theories of coherence make extensive use of correlation functions of higher orders as well. In the second order, for example, there are four field amplitudes related to each other [74]; see Section 14.5.1.

In interferometry we will consider correlations that do not change with time, so that after averaging only the dependence of the delay is left. Additionally, we will generally determine the intensity of the superposition of waves, that is, we will consider  $\Gamma_{12}$  at only *one* point  $\mathbf{r} = \mathbf{r}_1 = \mathbf{r}_2$ , so that the simplified form

$$\begin{aligned}\Gamma_{12}(\mathbf{r}, \tau) &= \frac{c\epsilon_0}{2} \langle E_1(\mathbf{r}, t + \tau) E_2^*(\mathbf{r}, t) \rangle \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \frac{c\epsilon_0}{2} E_1(\mathbf{r}, t + \tau) E_2^*(\mathbf{r}, t) dt\end{aligned}\tag{6.2}$$

is sufficient. In the case of very large delay times  $\tau$ , we expect in general the loss of phase relations between  $E_1$  and  $E_2$ . The product  $E_1(\mathbf{r}, t + \tau) E_2^*(\mathbf{r}, t)$  statistically fluctuates around 0 and vanishes on average:

$$\Gamma_{12}(\mathbf{r}, \tau \rightarrow \infty) \rightarrow 0.$$

To make the connection with Eq. (6.1), we have to take into consideration that, in a typical interferometry experiment, the partial waves are created with the help of beam splitters from the same light source. The delay  $\tau$  then reflects the different optical path lengths of the partial waves to the point of superposition. The function  $\Gamma_{12}(\mathbf{r}, \tau)$  describes their capability of forming interference stripes.

It is very convenient to define the normalized correlation function  $g_{12}(\mathbf{r}, \tau)$ , which is a quantitative measure for the interference contrast:

$$g_{12}(\mathbf{r}, \tau) = \frac{(c\epsilon_0/2) \langle E_1(\mathbf{r}, \tau) E_2^*(\mathbf{r}, 0) \rangle}{\sqrt{\langle I_1(\mathbf{r}) \rangle \langle I_2(\mathbf{r}) \rangle}}.\tag{6.3}$$

The function  $g_{12}$  is complex and takes values in the range

$$0 \leq |g_{12}(\mathbf{r}, \tau)| \leq 1.$$

An important special case of Eq. (6.3) is the autocorrelation function,

$$g_{11}(\mathbf{r}, \tau) = \frac{\langle E_1(\mathbf{r}, \tau) E_1^*(\mathbf{r}, 0) \rangle}{\langle |E_1(\mathbf{r})|^2 \rangle}, \quad (6.4)$$

which in this case relates the amplitude of an electromagnetic field to itself with delay  $\tau$ . We shall see its important role in the quantitative analysis of coherence features.

Now we can summarize the calculation of intensity for coherent and incoherent superposition by

$$\langle I(\mathbf{r}) \rangle = \langle I_1(\mathbf{r}) \rangle + \langle I_2(\mathbf{r}) \rangle + 2\sqrt{\langle I_1(\mathbf{r}) \rangle \langle I_2(\mathbf{r}) \rangle} \operatorname{Re}\{g_{12}(\mathbf{r}, \tau)\}.$$

In interferometry, the different paths of light beams coming from the same source generally cause a delay  $\tau = (s_1 - s_2)/c$ . In order also to define a quantitative measure of coherence, we introduce the *visibility*

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (6.5)$$

with  $I_{\max}$  and  $I_{\min}$  describing the maxima and minima of an interference pattern, respectively. Obviously,  $V(\tau)$  also takes values between 0 and 1. In an interferometric experiment, the degree of coherence can be measured by determination of the visibility.

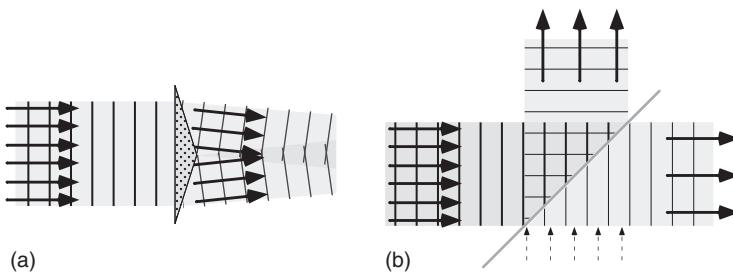
The capability of interference could not have been taken for granted and has played an important part in the development of wave theory. The reason for the great significance of interferometry for wave theory is to be found in the fact that the physical features of a wave, that is, *phase* and *amplitude*, can only be measured by superposition with another reference wave, that is, by an interferometric experiment. Whether interference can be observed is crucially dependent on the coherence properties of the waves.

### 6.2.2 Beam Splitter

The central element of an interferometric arrangement is the beam splitter. In the past, only by *separation* of an optical wave from a single light source<sup>1</sup> could one create two separated partial waves that were able to interfere. One can differentiate between two different types of BSs, as shown in Figure 6.2. The “wave front splitter” is based on the variations of the double slit as the classic form. The “amplitude splitter” is usually realized with a partially reflecting glass substrate. In the case of advanced applications, the existence of a second entrance gains importance. The second entrance can easily be seen in the right-hand interferometer shown in Figure 6.2.

---

<sup>1</sup> Today, we are able to synchronize two individual laser light sources so well that we can carry out interference experiments using them.



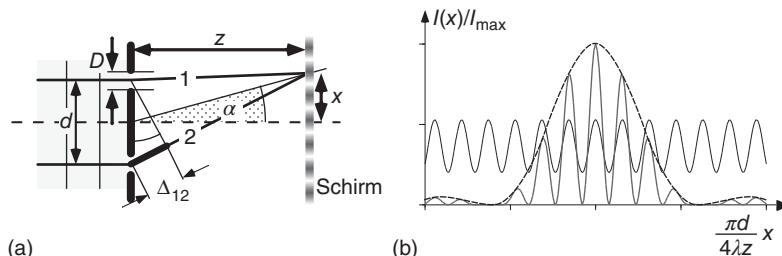
**Figure 6.2** Wave front (a) and amplitude (b) beam splitters. Beam splitters have a second entrance, which is not always as visible as for the right-hand type.

### 6.3 The Double-Slit Experiment

Let us now consider in detail the incidence of a planar wave on a double slit (Figures 6.1 and 6.3). Both slits act as new virtual and phase-synchronous (“coherent”) light sources. To understand the interference pattern on the screen, we have to determine the difference between the two optical paths “1” and “2.” If the distance  $z$  between the double slit and the screen is very much larger than the distance  $d$  between the slits themselves and the extent  $x$  of the interference pattern, that is,  $d, x \ll z$ , we can determine the path difference  $\Delta_{12}$  between paths 1 and 2 in a geometrical way according to the construction from Figure 6.1 and calculate the intensity distribution according to Eq. (6.1).

If the path difference is an integer multiple of the wavelength,  $\Delta_{12} = n\lambda$ , we expect constructive interference; in the case of half-integer multiples, we have destructive interference. The path difference  $\Delta_{12}$  and the corresponding phase difference  $\Phi_{12}$  at angle  $\alpha$  are

$$\Delta_{12} = d \sin \alpha \quad \text{and} \quad \Phi_{12} = k\Delta_{12} = \frac{2\pi d}{\lambda} \sin \alpha.$$



**Figure 6.3** Analysis of the diffraction pattern from Figure 6.1. (a) Notations and geometry of the double slit. (b) The interference pattern is understood as the product of the single-slit diffraction pattern (dashed line) and sinusoidal modulation (thin line). Here, width  $D = \text{distance}/4 = d/4$ .

For small angles  $\alpha \simeq x/z$ , we expect a periodic fringe pattern on the screen varying as

$$I(x) = \frac{I_0}{2} \left( 1 + \cos \frac{2\pi d}{\lambda} \frac{x}{z} \right),$$

with maximum intensity  $I(x = 0) = I_0$ .

During this analysis, we have assumed that the two slits are infinitesimally narrow. In a real experiment, of course, they have finite width, so we have to take into account single-slit diffraction as well. The superposition of the two phenomena can be taken into account by means of Fraunhofer diffraction at a slit according to the example on p. 68. The situation becomes very simple if we displace the slits by  $\xi = \pm d/2 = \pm \xi_0$  from the axis. Calling the box-shaped function for the slit again  $\tau(\xi) = 1$  for  $\xi \leq D/2$  and 0 otherwise, we get for the diffraction integral with scaled screen coordinate  $\kappa_x = 2\pi x/\lambda z$ :

$$\begin{aligned} \mathcal{E} &\propto \int_{-\infty}^{\infty} [\tau(\xi - \xi_0) e^{ik_x \xi} + \tau(\xi + \xi_0) e^{ik_x \xi}] d\xi \\ &= \int_{-D/2}^{D/2} e^{ik_x \xi} d\xi (e^{ik_x d/2} + e^{-ik_x d/2}). \end{aligned}$$

The intensity distribution is calculated as usual from  $I = c\epsilon_0 |\mathcal{E}|^2 / 2$  for linear polarization,

$$I = \frac{I_0}{2} \left( 1 + \cos \frac{2\pi d}{\lambda} \frac{x}{z} \right) \frac{\sin^2(\pi x D / \lambda z)}{(\pi x D / \lambda z)^2},$$

and we recognize immediately the complete interference pattern containing the product of the diffraction images of the single slit and of the double slit (Figures 6.1 and 6.3).

### 6.3.1 Transverse Coherence

Typical classical light sources<sup>2</sup> (incandescent lamps, discharge lamps, etc.) have a finite extent. We can visualize them as consisting of point-like light sources that illuminate the double slit with the same color or wavelength but with completely independent phases. In this case an additional phase difference appears that can be determined according to a similar construction as in Figure 6.1. If one of these point sources  $S$  lies at an angle  $\beta$  to the axis, the whole phase difference is

$$\Phi_{12} = k\Delta_{12} \simeq \frac{2\pi d}{\lambda}(\alpha - \beta)$$

for small angles  $\alpha$  and  $\beta$ .

According to this, displacement of the light source causes a transverse shift of the interference pattern on the screen. If all shifts between 0 and  $2\pi$  occur, the superposition of all fringe patterns causes the overall interference pattern to be washed out. Thus, in order to observe interference, the maximum phase shift  $\Delta_{\max}$  occurring between two point sources of light at a separation of

---

<sup>2</sup> The Gaussian beams of laser light sources have perfect transverse coherence!

$\Delta a = z_s(\beta - \beta')$  from each other and at a distance  $z_s$  from the double slit must not become too large:

$$k\Delta_{\max} = \frac{2\pi d \Delta a}{\lambda z_s} < 1.$$

This condition is met if the angle  $\Omega = \theta - \theta' = \Delta a/z_s$  with which both of the point sources are seen is sufficiently small, that is,

$$\Omega = \frac{\Delta a}{z_s} < \frac{1}{2\pi} \frac{\lambda}{d}. \quad (6.6)$$

According to this, for a given wavelength  $\lambda$  and a given distance  $z_s$ , the ability to interfere (“interferability”) can be achieved through a light source with a sufficiently small point-like area ( $\Delta a \leq \lambda z_s / 2\pi d$ ) or slit separation ( $d \leq \lambda z_s / 2\pi \Delta a$ ).

The coherence area of a source is to be determined by changing the slit separation  $d$ , while the source distance is fixed. The central interference fringe (which is always a maximum) with its adjacent minima is observed and evaluated according to Eq. (6.5). The distance where the value  $V = 1/2$  is obtained is defined as the *transverse coherence length*.

#### Excuse: Double-slit experiments with matter waves

We have dealt with double-slit interference as a pure wave phenomenon in the previous section, and we have also already referred to the application to other wave phenomena, in particular to matter waves. In this context, there is a very well-known phenomenon that bothers our intuition quite hard: an interference pattern is also generated by a single particle, by so-called self-interference. Although we always detect only one particle, its matter wave must have gone through both slits simultaneously! We infer this interpretation from the way in which quantum mechanics deals with it theoretically. It has been proven by experiments time and again, but stands in bizarre contradiction to our natural, that is, macroscopic, view of a “particle.”

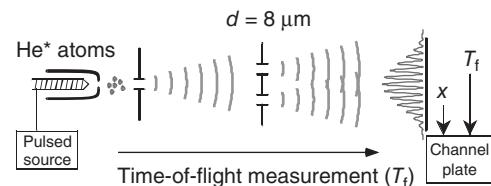
The first demonstration of the double-slit experiment with matter waves was given by Möllenstedt and Düker [72] using electron beams. For that experiment an electron beam was collimated and sent through an electric field arrangement corresponding to a Fresnel biprism. Later optics with atomic matter waves [75] has been established as yet variant. With helium atoms a double-slit experiment has been carried out in perfect analogy to Young’s experiment [73]. On the one hand, the de Broglie wavelength  $\lambda_{\text{deB}}$  of neutral atoms with mass  $m$  and velocity  $v$  within the atomic beam is very small,  $\lambda_{\text{deB}} = h/mv \simeq 20 \text{ pm}$ . That is why very tiny slit widths and separations had to be used,  $d \leq 1 \mu\text{m}$ , in order to obtain resolvable diffraction. The atomic flux was accordingly very small. On the other hand, helium atoms in the metastable  $^3\text{S}$  state can be detected nearly atom by atom by means of channel plates. This high detection sensitivity has made possible the atomic Young’s experiment with neutral atoms.

In the lower part of Figure 6.4, the result of the experiment is shown. The small atomic flux density has even an additional advantage. With a pulsed beam source, one can record

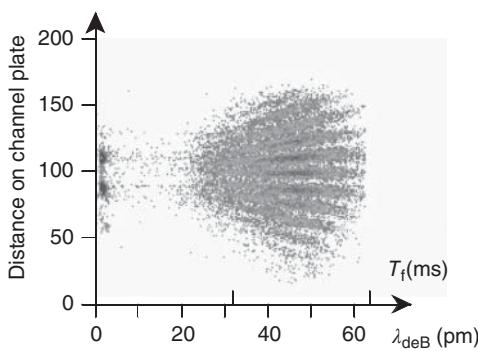
(Continued)

**Excuse: (Continued)**

the velocity of the atom by time-of-flight measurement and associate it with the change of the interference pattern. This can be directly interpreted as a consequence of the variation of the de Broglie wavelength, which can be immediately calculated from the time-of-flight measurement.



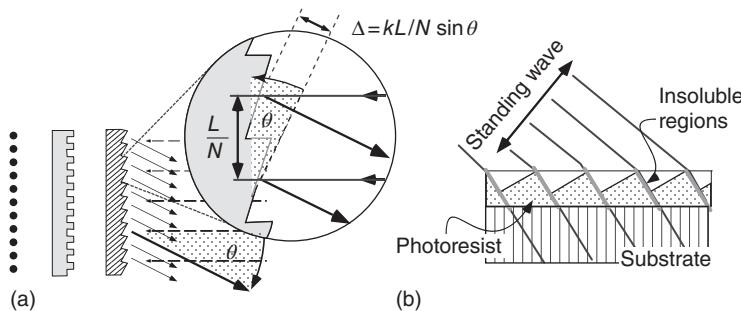
**Figure 6.4** Diffraction of matter waves at a double slit. (By courtesy of Kurtsiefer et al. [76].)



Finally we may turn to the interpretation once more and consider the light from the point of view of the particles or *photons*. For that we imagine an experiment in which the double slit is illuminated with such a weak intensity that there is only one photon at a time – the condition for self-interference is also met again. Such experiments are sometimes called *Taylor experiments* because they were conducted already in 1909 by G. Taylor. Sensitive photon-counting cameras are used to detect the interference pattern. We observe indeed a statistical pattern, which after some time generates a frequency distribution described exactly by the interference of the light waves.

### 6.3.2 Optical or Diffraction Gratings

If the number of slits is greatly increased, one obtains an optical grating, an example of multiple-beam interference. Optical gratings are used as amplitude, phase, or reflective gratings and are qualitatively introduced in Figure 6.5. They are specified according to the number of lines per millimeter, typically  $1000 \text{ lines mm}^{-1}$  or more for optical wavelengths. It is remarkable and impressive that even very fine gratings may “simply” be carved mechanically with diamonds. Optical gratings exhibit typically several *orders of diffraction*. For efficient



**Figure 6.5** (a) Amplitude, phase, and reflective grating. The blaze of a reflective grating can be chosen in such a way that the diffraction is mostly directed into a certain desired order, for instance, by shaping the grooves. (b) Manufacturing of a holographic grating with an asymmetric groove. The photoresist is illuminated by a standing-wave light field. At the nodes it is only weakly affected and thus remains undissolved.

application, grooves with special shapes are used to concentrate the intensity into a single or a few diffraction orders only; see Figure 6.5. Such gratings are called *blazed* gratings.

Mechanically manufactured gratings, though, suffer from scattering losses and additional faults with a long period (“grating ghosts”). Better optical quality is offered by components called “holographic gratings” according to the method of manufacture. They are produced by methods of optical microlithography. A film (“photoresist”) on a substrate of optical quality is exposed to a standing light wave. The solubility of the exposed film depends on the dose, and thus a remnant of film is left over at the nodes of the standing wave (see Figure 6.5). A reflection grating can be manufactured from this structure, for example, by coating with a reflecting material. One disadvantage, in the case of a holographic grating, is that it is more difficult to control the “blaze” by properly shaping the grooves.

The condition for interference is identical with that of the double slit. We consider the beams radiating from the  $N$  lines of a grating with length  $L$ . Two adjacent beams have a path difference

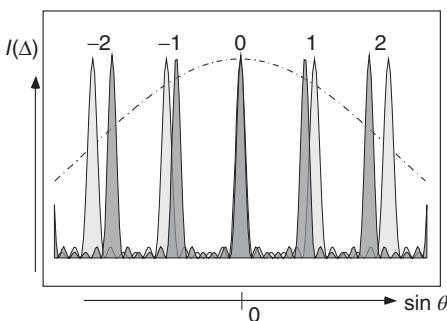
$$\Delta(\theta) = (L/N) \sin \theta \quad (6.7)$$

corresponding to a phase difference

$$\Phi(\theta) = \frac{kL}{N} \sin \theta = \frac{2\pi L}{N\lambda} \sin \theta. \quad (6.8)$$

With homogeneous illumination, the field amplitude is

$$\begin{aligned} E &= E_1 + E_2 + \cdots + E_N \\ &= E_0(1 + e^{-i\Phi} + e^{-2i\Phi} + \cdots + e^{-Ni\Phi}) e^{-i\omega t} \\ &= E_0 \exp \left\{ -i \left[ \omega t + \frac{1}{2}(N-1)\Phi \right] \right\} \frac{\sin(N\Phi/2)}{\sin(\Phi/2)}. \end{aligned}$$



**Figure 6.6** Calculated diffraction pattern of a grating consisting of six single slits at two different wavelengths. The diffractive contribution of the single slit (broken line, width = 0.6 × distance between slits) has been neglected. The light and dark shaded patterns correspond to two different wavelengths.

The spectroscopic resolution is of primary interest. According to the Rayleigh criterion, the main maximum of one wavelength is supposed to fall into the first null of the only just resolvable adjacent wavelength, that is, according to Eq. (6.8):

$$\Phi(\theta + \delta\theta) - \Phi(\theta) \simeq \frac{2\pi}{\lambda} \frac{L}{N} \cos \theta \delta\theta = \frac{2\pi}{N}.$$

The condition for the main maximum varies with the wavelength according to  $m \delta\lambda = (L/N) \cos \theta \delta\theta = \lambda/N$  and so results finally in the resolution:

$$R = \frac{\lambda}{\delta\lambda} = mN.$$

This increases with the number of illuminated slits  $N$  and with the order of interference  $m$ , as can also be easily seen in Figure 6.6 where the diffraction intensity is given for two different wavelengths.

### Example: Resolution of an optical grating

We determine the resolution of a grating with a diameter of 100 mm and number of grooves equal to 800 line mm<sup>-1</sup> at  $\lambda = 600$  nm. We get in first order

$$R = 100 \text{ mm} \times 800 \text{ mm}^{-1} = 0.8 \times 10^5.$$

From that a wavelength can be separated just at a difference of

$$\delta\lambda = \lambda/N \simeq 7 \times 10^{-3} \text{ nm}.$$

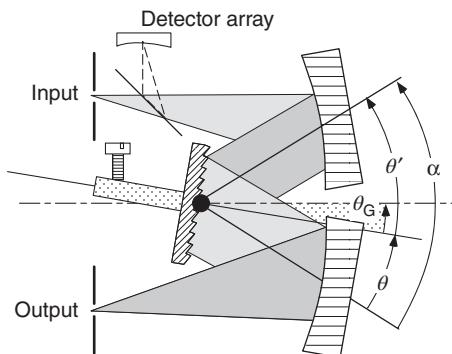
### 6.3.3 Monochromators

Grating monochromators are standard equipment in most optical laboratories, and they play an important role by offering one of the simplest instruments of spectroscopy with high resolution. They all have in common the use of reflective gratings, which are technically superior to transmission gratings. They differ only in those structural details dealing with operation or resolution.

The diffraction pattern (Figure 6.6) of the grating has maxima at  $\Phi = 2m\pi$ , with  $m = 0, \pm 1, \pm 2, \dots$ . There the intensity is calculated from  $I_0 = c\epsilon_0|E_0|^2$  and  $I_{\max} = c\epsilon_0|E(\Phi = 2m\pi)|^2/2 = N^2I_0$ . Diffraction between the intensity maxima is strongly suppressed, and the grating can be used very advantageously as a dispersive element for spectral analysis.

The first minimum appears at  $\Phi = 2\pi/N$ , and the first secondary maximum at  $\Phi = 3\pi/N$ . For large  $N$  the intensity is limited to  $I(\Phi = 3\pi/N) \simeq N^2/(3\pi/2)^2 \approx 0.05I_{\max}$  only. The diffraction grating concentrates the radiation energy into the main maxima.

**Figure 6.7** Principle of the Czerny–Turner monochromator.



As an example we introduce the Czerny–Turner construction (Figure 6.7). Here, the grating has to be completely illuminated to achieve the highest possible resolution, which is why the input light has to be focused on the entrance slit. The grating simultaneously serves as a mirror that is turned with a linear motion drive. One finds according to Eq. (6.8)

$$m\lambda = \frac{L}{N}(\sin \theta - \sin \theta').$$

Because  $\theta = \alpha/2 - \theta_G$  and  $\theta' = \alpha/2 + \theta_G$  (Figure 6.7), one gets

$$\lambda = \frac{2L}{mN} \cos(\alpha/2) \sin \theta_G,$$

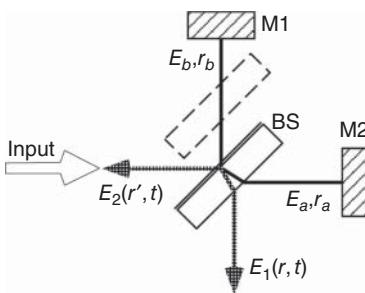
and hence the wavelength at the exit slit only depends on the rotation angle  $\theta_G$ .

Spectral resolution depends on angular resolution in this instrument, and it improves with the distance between the slits and the grating. Thus monochromators are offered with standard lengths of 1/8, 1/4, 1/2 m, and so on, which are a coarse measure of their resolution. Above approximately 1 m, they become large, heavy, and impracticable, so that resolution exceeding  $10^6$  is cumbersome to achieve. Through the advent of laser spectroscopy, which we will discuss in Chapter 12, resolution inconceivable with the conventional methods using grating monochromators has been reached.

## 6.4 Michelson interferometer: longitudinal coherence

The interferometer arrangement given in 1882 for the first time by the American physicist M. Michelson (1852–1931) has become very famous. It was developed to identify experimentally the “ether” postulated in the nineteenth century to be responsible for the spreading of light. If the ether existed, the speed of light should depend on the relative speed of the light source in that medium.

The results by Michelson and Morley were interpreted by assuming that the speed of light was independent of the reference frame – a discovery in full agreement with the descriptions by Poincaré, Lorentz, and finally Einstein who published the theory of special relativity in 1905 [8].



**Figure 6.8** Michelson interferometer. BS = beam splitter; M1, M2 = mirrors.

The heart of a Michelson interferometer (Figure 6.8) is the amplitude beam splitter, mostly consisting of a semitransparent mirror. An incident planar wave  $E = \mathcal{E}_{\text{in}} e^{-i(\omega t - kr)}$  is separated into two partial waves with equal amplitudes  $\mathcal{E}_{a,b} = \mathcal{E}_{\text{in}}/\sqrt{2}$ . Usually the beam splitter consists of a polished glass substrate coated on one side. The reflected and transmitted beams travel along different optical paths. The two paths are different since the substrate acts on one of the two arms only. For compensation, sometimes an additional glass substrate of the same thickness is inserted into the other arm in order to make the interferometer arms as symmetric as possible. Using monochromatic laser light, this does not matter because the difference in the light path lengths can simply be geometrically compensated. If the light is polychromatic, however, dispersion of the glass substrates caused by wavelength-dependent differences in the light paths is also compensated by the additional substrate.

At the end of the two interferometer arms, the two partial waves are reflected and pass through the beam splitter again. The interferometer generates two separate output waves  $E_1$  and  $E_2$ ,

$$\begin{aligned} E_1 &= \frac{1}{\sqrt{2}}(\mathcal{E}_a + \mathcal{E}_b) = \frac{1}{2}\mathcal{E}_{\text{in}} e^{-i(\omega t - kr)}(e^{2ikr_a} + e^{2ikr_b}), \\ E_2 &= \frac{1}{\sqrt{2}}(\mathcal{E}_a - \mathcal{E}_b) = \frac{1}{2}\mathcal{E}_{\text{in}} e^{-i(\omega t - kr')}(e^{2ikr_a} - e^{2ikr_b}), \end{aligned} \quad (6.9)$$

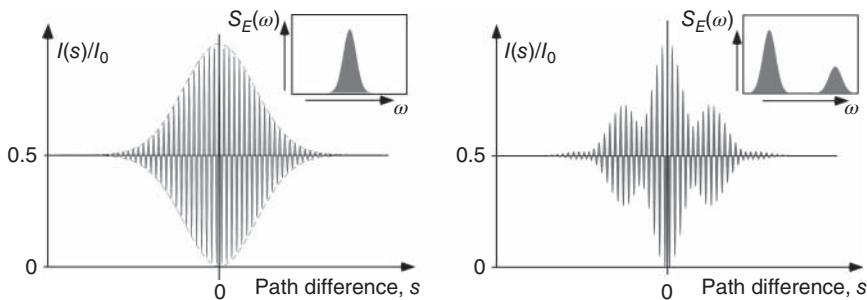
at its exits. We calculate the intensity there and get from  $I = \epsilon_0 c \mathcal{E} \mathcal{E}^*/2$  the results

$$\begin{aligned} I_1 &= \frac{1}{2}I_0\{1 + \cos[2k(r_a - r_b)]\}, \\ I_2 &= \frac{1}{2}I_0\{1 - \cos[2k(r_a - r_b)]\}. \end{aligned} \quad (6.10)$$

According to this, the total intensity is constant but distributed on the two exits  $I_0 = I_1 + I_2$  depending on the difference in light paths  $s = 2(r_a - r_b)$ . Note that, in this arithmetical treatment, the different signs in the sum of the partial beams ( $(E_{1,2} = (\mathcal{E}_a \pm \mathcal{E}_b)/\sqrt{2})$ ) are caused by the reflections off the beam splitter, in one case at the more dense, and in the other case at the less dense, medium. This 90° phase difference is also essential to satisfy energy conservation.

#### 6.4.1 Longitudinal or Temporal Coherence

With the Michelson interferometer, the temporal coherence length  $\ell_{\text{coh}} = c\tau_{\text{coh}}$  is measured by increasing the length of one arm until the interference contrast is decreased to a half. The coherence length is then twice the difference of the two branches,  $\ell_{\text{coh}} = 2|r_1 - r_2|$  in Figure 6.8. Usually, the visibility from Eq. (6.5) is again used as a quantitative measure.



**Figure 6.9** Interferometer signals of a Michelson interferometer for a single and a double spectral line, for example, the yellow D line of the Na vapor lamp. In the upper inset boxes, the associated spectra are shown.

The interference contrast is measured through the field autocorrelation function  $\Gamma_{EE^*}(s/c)$ , according to Eqs. (6.2) and (6.3) with  $\tau = s/c$ . This is linked to the spectral power density

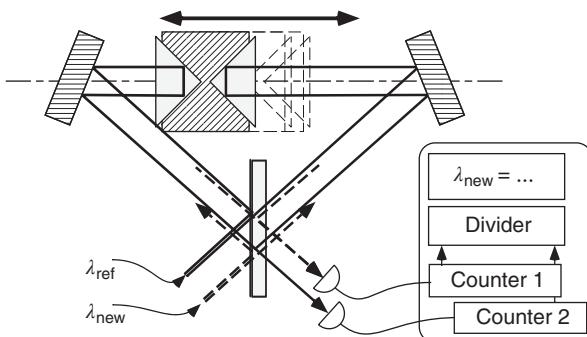
$$S_E(\omega) = \frac{1}{c} \int_0^\infty \Gamma_{EE^*}(s/c) e^{i\omega s/c} ds$$

according to the Wiener–Khinchin theorem (see Eq. (A.9)). So a Fourier transformation of the interferometer's signal as a function of the path difference delivers information about the spectral properties of the light source. Analysis of the light from a sodium vapor lamp with the Michelson interferometer shows this connection very clearly, as we describe qualitatively in Figure 6.9. This relation is also the basis of the Fourier spectrometer, which we mention here for the sake of completeness. Furthermore, the self-heterodyne method from Section 8.1.7.2 can be considered as a variant of the Michelson interferometer. Here the path difference of the arm lengths even has to be so large that no stable interference can be observed in the time average. This method allows determination of the spectral properties of a narrowband laser light source.

### Example: The wavemeter

The *wavemeter*, also known as the lambda meter, is a variant of the Michelson interferometer used in many laser laboratories. Monochromatic laser light sources have a very large coherence length ( $\ell_{coh} > 10\text{m}$ ). During continuous variation of the path length difference of the interferometer arms, they consequently generate a sinusoidal modulation of the interferometer signal with period proportional to the frequency or inverse wavelength of the laser light according to Eq. (6.10). Comparison of the interferometer signal of an unknown wavelength  $\lambda_{\text{new}}$  to a reference laser wavelength  $\lambda_{\text{ref}}$  amounts to the determination of the unknown frequency or wavelength by simple division of the number of fringes counted when the reflector trolley slides along its track.

In the wavemeter arrangement, two retroreflectors are fitted to a mobile carriage (Figure 6.10), so that the incident and reflected beams of the Michelson interferometer are spatially separated. At one exit the reference beam is directed



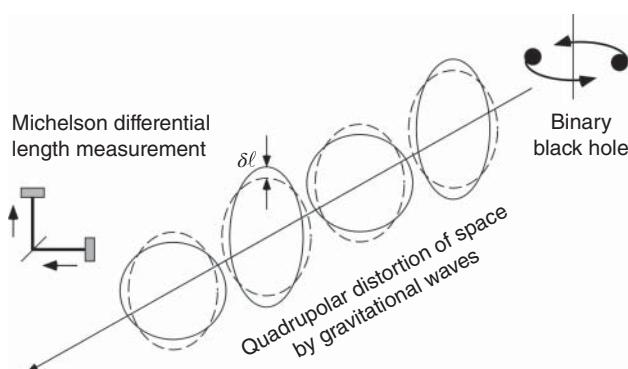
**Figure 6.10** Wavemeter arrangement for the determination of laser wavelengths. For clarity the laser beam to be measured (broken line) is drawn only at the entrance and at the exit.

to a photodiode in order to count the number  $N_{\text{ref}}$  of interferometer fringes for a certain traveling interval. At the other exit it serves as a tracer beam for a differently colored laser beam to be measured. Its interferometer fringes are counted on a second photodiode,  $N_{\text{meas}}$ . Electronic division then yields the unknown wavelength through comparison with the reference laser:  $\lambda_{\text{meas}} = \lambda_{\text{ref}} N_{\text{meas}} / N_{\text{ref}}$ .

#### Excuse: Gravity wave interferometer

A particularly unusual variant of the Michelson interferometer with huge dimensions has been constructed at several places around the world. The gravity wave interferometer at Hannover (Germany) called GEO600 has an arm length of 600 m, while LIGO in the United States and VIRGO in Italy have arm lengths of 4 and 3 km, respectively.

Though gravity waves were predicted in detail by Einstein's theory of general relativity, they were extremely difficult to observe since they only exert an extraordinarily weak force even on big masses. Gravity waves spread like electromagnetic waves; they are transverse but have quadrupolar characteristics (Figure 6.11). A Michelson interferometer is in fact well adapted to detect distortions of space caused by gravity waves by differential length changes in two transverse directions. And for length measurements at long scales and with sensitivity much below the wavelength, interferometry



**Figure 6.11** Gravity waves are causing quadrupolar distortions of space, which induce arm length variations of the Michelson interferometer and hence shift the interferometer fringes.

is the method of choice. Early in 2016 the news broke: The worldwide LIGO–VIRGO collaboration had observed a first signal, which is now interpreted as originating from the merger of two heavy black holes [77]!

For a most sensitive proof of a change of length  $\delta l$  with an interferometer, the instrument itself has to have a length  $\ell$  as large as possible, but

still short compared to the relevant wavelength. According to the theory of general relativity, even for strong astronomical “gravitational wave sources,” for example, supernova explosions, relative sensitivities of  $\delta\ell/\ell \approx 10^{-20}$  are necessary. At a length of 1 km this corresponds to about 100th of a proton radius!

To achieve sufficient signal-to-noise ratio of the interferometer signal, numerous optical technologies are applied: the use of very powerful laser light sources with superb frequency stability is necessary. At the present time neodymium lasers with miser-like stability (Section 8.4.2.2) are preferred for this task. The sensitivity is further increased by folding the light path in each arm; that is, each arm is a Fabry–Pérot resonator (Section 6.5) by itself.

With gravity wave antennas, a new window has been opened for cosmic observations. In the face of these expectations, the plans for the *Laser Interferometer Space Antenna* (LISA) [78] do not seem to be completely eccentric. In this spaceflight project, in about 2034, it is planned to park a Michelson interferometer consisting of three spaceships (two mirrors and a beam splitter with light source) shifted by  $20^\circ$  in Earth’s orbit around the sun. This Michelson interferometer will have an arm length of  $5 \times 10^6$  km!

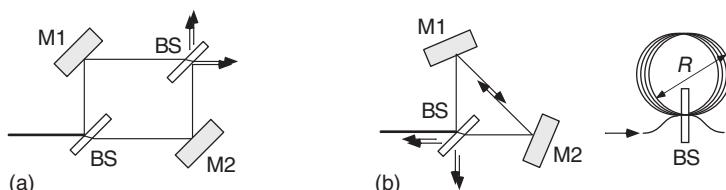
#### 6.4.2 Mach–Zehnder and Sagnac Interferometers

There are numerous variants of the Michelson interferometer that have different methodical advantages and disadvantages. Two important examples are the Mach–Zehnder interferometer (MZI) and the Sagnac interferometer, the latter of which, strictly speaking, forms a class of its own.

##### 6.4.2.1 Mach–Zehnder Interferometer

The MZI is derived from the Michelson interferometer, in which the reflections at the mirrors are no longer carried out at normal incidence and a second beam splitter is used for the recombination of the beams. The MZI is also used for spatially resolved studies of changes in the wave fronts passing objects of interest [14].

The reflection angle at the beam splitters (BS) and mirrors (M1, M2) in Figure 6.12a must not necessarily be set to  $90^\circ$ . Several times the MZI concept has stimulated ideas for interferometric experiments in particle optics, since there mirrors and beam splitters can often be realized only under grazing



**Figure 6.12** Mach–Zehnder (a) and Sagnac (b) interferometers. BS = beam splitter; M = mirror. The Sagnac interferometer can be realized with mirrors or with an optical wave guide. In the latter case multiple coil-like windings (radius  $R$ ) can be realized.

incidence, with small deflection angles. Neutron beams reflected off the crystal lattice of Si single crystals are an interesting example [79].

#### 6.4.2.2 Sagnac Interferometer

The Sagnac interferometer also derives from the Michelson interferometer: here, the light beams are not reflected back to themselves but run back on diametrically opposed paths that are at first identical. But if the interferometer is rotating around an axis perpendicular to its plane, a phase shift between the opposing beams is predicted by the theory of special (and general) relativity. For the sake of simplicity, we consider a circular light path (radius  $R$ ) in a fiber and with one beam splitter. The round-trip time is  $T = L/c = 2\pi nR/c$  with  $n$  being the refractive index of the fiber. We adopt the result from the theory of special relativity that, in a medium moving at velocity  $v$ , the speed of light as measured in the laboratory frame is modified according to [70]

$$c_{\pm} = c \frac{1 \pm nv/c}{n \pm v/c}. \quad (6.11)$$

In the rotating fiber path (angular velocity  $\Omega = v/R$ ), in one direction, the light travels toward the beam splitter, and in the other direction, away. Hence the effective round-trip time is increased or decreased corresponding to the path  $R\Omega T = vT$  traveled by the beam splitter. It is taken into account by the condition  $T_{\pm} = L_{\pm}/c_{\pm} = (L \pm vT_{\pm})/c_{\pm}$ . From this implicit equation, we extract  $T_{\pm} = L/(c_{\pm} \mp v)$ , and with a short calculation using the result Eq. (6.11), we find  $1/(c_{\pm}v) \simeq (n/c)(1 \pm (v/nc))$ . Surprisingly, the time difference  $T_+ - T_-$  no longer depends on  $n$ :

$$T_+ - T_- \simeq 2v/c^2 = 2R\Omega/c^2.$$

For light with frequency  $\omega$ , we now directly obtain the difference of the light paths or phase difference  $\Delta\Phi$  at the beam splitter from this:

$$\Delta\Phi = \omega(T_+ - T_-) \simeq \omega \frac{4\pi R^2 \Omega}{c^2} = \Omega \frac{4F}{\lambda c}.$$

According to this, the interference signal is proportional not only to the angular velocity  $\Omega$  but also to the area  $F = \pi R^2$  of the Sagnac interferometer. The effective area and with it the sensitivity can be increased by the coil-like winding of a glass fiber (Figure 6.12).

#### Example: Phase shift in the Sagnac interferometer

We determine the phase shift generated by Earth's rotation ( $2\pi/24 \text{ h} = 1.8 \times 10^{-6} \text{ s}^{-1}$ ) in a Sagnac interferometer. The fiber has a length of 1 km and is rolled up into an area with a diameter of  $2R = 10 \text{ cm}$ . The interferometer is operated with a diode laser at  $\lambda = 780 \text{ nm}$ . Thus

$$\Delta\Phi = 1.8 \times 10^{-6} \frac{\pi \times 4(0.1/2)^2(10^3/\pi \times 0.1)}{(0.78 \times 10^{-6}) \times (3 \times 10^8)} = 0.77 \times 10^{-5} \text{ rad.}$$

This condition requires a high standard of the art of experimenting but can be realized in the laser gyro.

If a laser amplifier is installed in a Sagnac interferometer, one has realized the “laser gyro.” This is widely used since it allows very sensitive detection of rotary motion and acceleration, but for studies of this, we refer the reader to the specialized literature. It should be emphasized, however, that in the laser gyro, the waves running around to the left and to the right, respectively, have to have different frequencies.

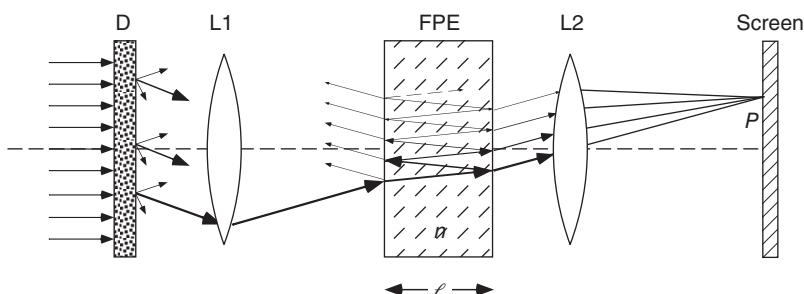
## 6.5 Fabry-Pérot Interferometer

We consider two plane parallel dielectric interfaces illuminated by a light beam at a small angle. Such an optical component can be easily made from a plane parallel glass substrate. In this case it is called a Fabry-Pérot etalon (FPE) (from the French word *étalon*, meaning “calibration spacer” or “gauge”). It is often used for frequency selection in laser resonators or as a simple and very highly resolving diagnostic instrument for laser wavelengths. The light beams are reflected back and forth many times and so exhibit multiple-beam interference in the longitudinal direction analogous to the diffraction grating (Figure 6.13).

The surfaces of an FPE are partly silvered and must be very smooth and planar. Furthermore, their relative tilt, or wedge, has to be very small. For precise measurement, also the distance  $\ell$  of the spacing between the reflecting surfaces must be very well known and controlled. The optical length of the FPE depends on the index of refraction  $n$  of the substrate,

$$\ell_{\text{opt}} = n\ell,$$

which for a material such as glass changes relatively rapidly with temperature ( $dn/dT \simeq 10^{-3} \text{ K}^{-1}$ ). Stable, less sensitive etalons are built with an air gap between glass substrates fixed by spacers with small thermal expansion, for example, quartz rods. If the distance  $\ell$  of the gap can be varied, for example, by a piezo translator, it is called a *Fabry-Pérot interferometer*. This type of instrument was used for the first time by C. Fabry and A. Pérot in 1899.



**Figure 6.13** Multiple-beam interference in the Fabry-Pérot etalon (FPE). A ground glass substrate acts as a diffuser (D) to generate light beams incident onto the etalon in many directions. The second lens L2 induces parallel light rays to interfere on a screen at the focal plane.

The condition for constructive interference can again be determined from the phase difference  $\delta$  between two adjacent beams. One determines the path length  $A-B-C$  in Figure 6.14a and finds, with  $k = 2\pi/\lambda$ ,

$$\delta = k\ell_{\text{opt}} = 2nk\ell \cos \theta = 2\pi N, \quad (6.12)$$

where  $N$  is the *order* of the interference, usually a large number. This result perhaps contradicts our initial expectation, since, because of the geometry, each individual beam in the interferometer travels along an elongated path  $\ell/\cos \theta$  tending to longer wavelengths and smaller frequencies. However, exactly the opposite occurs: tilting of an etalon shifts the interference condition to shorter wavelengths!

Let us now add the individual contributions of each beam, where we now have to account for reflection and transmission. The change of intensity is described by the reflection and transmission coefficients, while the coefficients of the field amplitudes are defined by  $r = \sqrt{R}$  and  $t = \sqrt{T}$ :

$$\begin{array}{ll} r, r' = \text{amplitude reflectivities}, & R, R' = \text{reflection coefficients}, \\ t, t' = \text{amplitude transmissivities}, & T, T' = \text{transmission coefficients}. \end{array}$$

Phase jumps during reflection ( $\pi$  phase shift for reflection off the denser material) are included with the total phase shift accumulated after one round trip and given by  $e^{i\delta}$ . Then the transmitted partial waves contributing to the field amplitude  $E_{\text{tr}}$  at the interference point  $P$  are summed up in a complex geometric series,

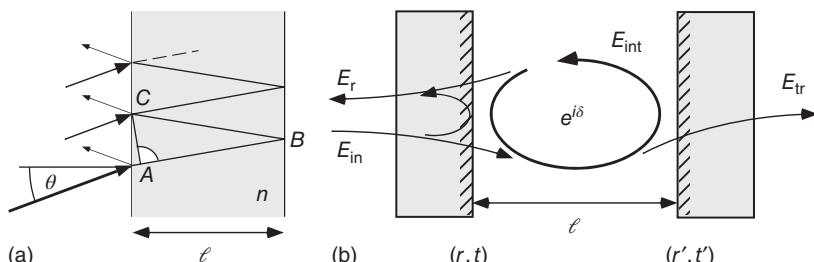
$$E_{\text{tr}} = t'tE_{\text{in}} + rr' e^{i\delta} tt'E_{\text{in}} + (rr')^2 e^{2i\delta} tt'E_{\text{in}} + \dots,$$

yielding the result

$$E_{\text{tr}} = \frac{tt'E_{\text{in}}}{1 - rr' e^{i\delta}}. \quad (6.13)$$

Alternatively, this result can be derived in a clear and efficient manner by considering only the wave circulating within the etalon right after the first mirror (see Figure 6.14), because in equilibrium, it has to be reconstructed by interference of the internal wave after one full round trip and the incident wave:

$$E_{\text{int}} = e^{i\delta} rr'E_{\text{int}} + tE_{\text{in}}.$$



**Figure 6.14** Phase condition for the Fabry-Pérot etalon. (a) Difference in the optical paths of the partial beams. (b) Self-consistent condition for the internal field.

From this, with  $E_{\text{tr}} = t'E_{\text{int}}$ , one again and immediately obtains the first result. Already, to satisfy energy conservation, there has to be a reflected wave. From this consideration, the effect of interference becomes still more transparent:

$$E_{\text{r}} = rE_{\text{in}} - r't'e^{i\delta}E_{\text{int}} = \frac{r - r'e^{i\delta}}{1 - rr'e^{i\delta}} E_{\text{in}}. \quad (6.14)$$

The minus sign occurs here because in this case there was one reflection – and hence one  $\pi$  phase jump – less compared to the circulating wave.

Let us now explore Eq. (6.13) by considering the transmitted intensity. By taking the modulus, we first get

$$I_{\text{tr}} = I_{\text{in}} \frac{TT'}{|1 - \sqrt{RR'}e^{i\delta}|^2}.$$

This can be written more transparently by introducing the *finesse coefficient*,  $F$ ,

$$F = \frac{4\sqrt{RR'}}{(1 - \sqrt{RR'})^2}, \quad (6.15)$$

from which after a short calculation we get the *Airy function*

$$I_{\text{tr}} = I_{\text{in}} \frac{TT'}{(1 - \sqrt{RR'})^2} \frac{1}{1 + F\sin^2(\delta/2)}. \quad (6.16)$$

According to our calculation the transmitted intensity varies over the range

$$\frac{(1-R)(1-R')}{(1 + \sqrt{RR'})^2} \leq \frac{I_{\text{tr}}}{I_{\text{in}}} \leq \frac{(1-R)(1-R')}{(1 - \sqrt{RR'})^2} \quad (6.17)$$

and can even become identical with the incident wave if there are ideal loss-free mirrors with the same reflection coefficients:

$$(R, T) = (R', T') : \quad I_{\text{tr}} = \frac{I_{\text{in}}}{1 + F\sin^2(\delta/2)},$$

$$\delta = N\pi : \quad I_{\text{tr}} = I_{\text{in}}.$$

We will learn more about this case in Section 6.6 when we look at optical resonators.

Now let us determine also the accumulated intensity circulating in the etalon along with the reflected intensity:

$$I_{\text{int}} = \frac{1}{T'} I_{\text{tr}},$$

$$I_{\text{r}} = I_{\text{in}} - I_{\text{tr}}.$$

Real resonators are always affected by losses, which should be as low as possible. If we take the losses per revolution simply into account with a coefficient  $A$ , we get the generalized finesse coefficient

$$F_A = \frac{4\sqrt{RR'(1-A)}}{[1 - \sqrt{RR'(1-A)}]^2}, \quad (6.18)$$

by which we can again calculate the transmitted power according to

$$I_{\text{tr}} = \frac{4TT'(1-A)}{(T+T'+A)^2} \frac{I_{\text{in}}}{1+F_A \sin^2(\delta/2)}.$$

We can find analogous expressions for the reflected and the coupled power.

### Example: Coupling of an optical resonator

Optical resonators, which we shall discuss in more detail in Section 6.6, allow the storage of light energy, albeit only for relatively short times. Thus it is interesting to know the amount of an incident light field that is coupled into the resonator. This can be answered by the recent considerations.

Again, the case of resonance  $\delta = 0$  is particularly important. We find these relations for the reflected and the transmitted fractions of the incident intensity:

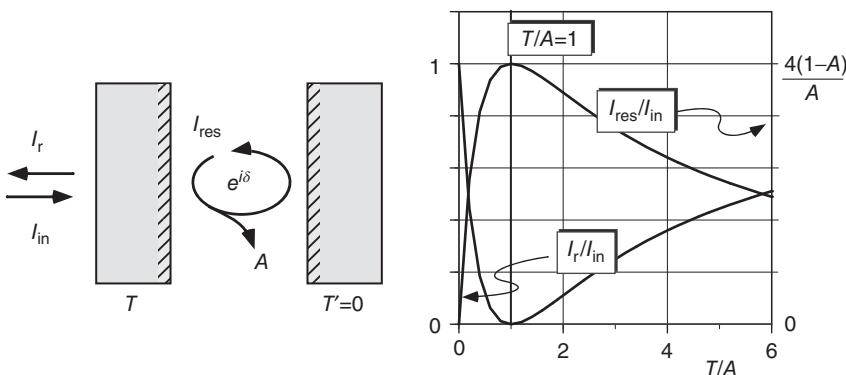
$$\frac{I_r}{I_{\text{in}}} = \left( \frac{T' + A - T}{T' + A + T} \right)^2 \quad \text{and} \quad \frac{I_{\text{tr}}}{I_{\text{in}}} = \frac{4TT'(1-A)}{(T' + A + T)^2}.$$

The power circulating in the resonator can also be easily determined according to  $I_{\text{res}} = I_{\text{tr}}/T$  and is shown in Figure 6.15 as a function of  $T/A$  and for the special but instructive case  $T' = 0$ .

The maximum of the coupled power is reached at  $T/A = 1$ . The power circulating there in the resonator is proportional to  $1/A$  for low  $A$ . In this case the external losses (caused by the coupling mirror) are just equal to the internal losses. This situation is quite well known for resonators: only in the case of perfect “impedance matching” is the full incident power coupled into the resonator; otherwise it is over- or under-coupled.

#### 6.5.1 Free Spectral Range, Finesse, and Resolution

According to Eq. (6.12), the Fabry-Pérot interferometer delivers a periodic series of transmission lines as a function of the frequency  $\omega = ck$  of the incident



**Figure 6.15** Influence of losses on the coupling of a Fabry-Pérot resonator in the case of resonance  $\delta = 0$ . The normalized intensities of the reflected ( $I_r$ ) and stored light field ( $I_{\text{res}}$ ) are displayed for the special case  $T' = 0$ .

light field. The distance of adjacent lines corresponds to successive orders  $N$  and  $N + 1$  and is called the “free spectral range,”  $\Delta_{\text{FSR}}$ :

$$\Delta_{\text{FSR}} = \nu_{N+1} - \nu_N = \frac{c}{2n\ell} = \frac{1}{\tau_{\text{circ}}}.$$
 (6.19)

The free spectral range also just corresponds to the inverse circulation time  $\tau_{\text{circ}}$  of the light in the interferometer. If the gap between the mirrors is empty ( $n = 1$ ), then we simply have  $\Delta_{\text{FSR}} = c/2\ell$ . Typically, Fabry-Pérot interferometers with centimeter distances are used whose free spectral range is calculated according to

$$\Delta_{\text{FSR}} = \frac{15 \text{ GHz}}{\ell/\text{cm}}.$$

They are usually designed for some 100 MHz up to several gigahertz.

The Fabry-Pérot interferometer can only be used for measurements if the periodicity leading to superpositions of different orders is visible. In that case the resolution between two narrowly adjacent frequencies is determined by the width of the transmission maxima. It can approximately be calculated from Eq. (6.16) taking into account that most interferometers have large  $F$  coefficients. Then the sine function can be replaced by the argument

$$I_{\text{tr}} \simeq \frac{I_{\text{in}}}{1 + F(\delta/2)^2}.$$

If two spectral lines are assumed to be separable if their full width at half maximums (FWHMs)  $\delta_{1/2}$  do not overlap, the lowest resolvable frequency difference is determined from  $\delta_{1/2} = 2 \times 2/F^{1/2} = 4/F^{1/2}$ . The required frequency difference  $\Delta_{1/2}$  is calculated from  $2\pi\Delta_{1/2} = c \times \delta k_{1/2}$  and  $\delta_{1/2} = 2n\ell \times \delta k_{1/2} = 2\pi\Delta_{1/2}/\Delta_{\text{FSR}}$ . For a pairs of mirrors with identical reflectivity, one obtains

$$\frac{\Delta_{\text{FSR}}}{\Delta_{1/2}} = \frac{\pi\sqrt{R}}{1-R} =: \mathcal{F}.$$
 (6.20)

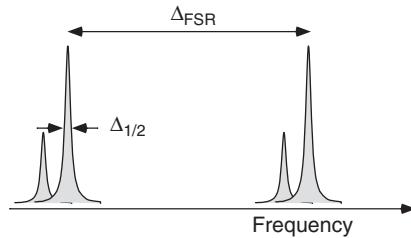
The ratio  $\mathcal{F} = \Delta_{\text{FSR}}/\Delta\nu_{1/2}$  of free spectral range and resolution can be easily read from an oscilloscope screen like that shown in Figure 6.16. This measure is more common than the finesse coefficient  $F$  and  $\mathcal{F}$  is simply called the *finesse*.

$$\mathcal{F} = \frac{\pi}{2}\sqrt{F}.$$
 (6.21)

The interferometric resolution  $\nu/\Delta\nu_{1/2}$  is indeed considerably higher with

$$\mathcal{R} = \nu_N/\Delta_{1/2} = N\Delta_{\text{FSR}}/\Delta_{1/2} = N\mathcal{F}$$

and can easily exceed a value of  $\mathcal{R} > 10^8$ .



**Figure 6.16** Free spectral range (FSR) and full width at half maximum (FWHM) in the Fabry-Pérot resonator.

**Example: Resolution of Fabry–Pérot interferometers**

In Table 6.2 we have compiled some characteristic specifications for typical Fabry–Pérot interferometers, which will play an important role as optical cavities in the next section. In the table it is remarkable that the half-width  $\Delta\nu_{1/2}$  always has a similar order of about 1 MHz. The reason for this is the practical applicability to the continuous laser light sources used in the laboratory, which exhibit typical linewidths of 1 MHz.

**Table 6.2** Characteristics of Fabry–Pérot interferometers.

$\ell$ (mm)	$1 - R = T$	$\Delta_{\text{FSR}}$ (GHz)	$\Delta\nu_{1/2}$ (MHz)	$\mathcal{F}$	Q at 600 nm	$\tau_{\text{res}}$ (ms)
300	1%	0.5	1.7	300	$3 \times 10^8$	0.1
10	0.1%	15	5	3 000	$10^8$	0.03
1	20 ppm	150	1	150 000	$5 \times 10^8$	0.15
100	20 ppm	1.5	0.01	150 000	$5 \times 10^{10}$	15

## 6.6 Optical Cavities

Fabry–Pérot interferometers are very important as optical cavities, which are necessary for the construction of laser resonators or are widely used as optical spectrum analyzers (see details in Section 8.1.7).

### 6.6.1 Damping of Optical Cavities

An electromagnetic resonator stores radiant energy. It is characterized, on the one hand, by the spectrum of its resonant frequencies, also known as modes  $v_{qmn}$ , and, on the other, by their decay or damping times  $\tau_{\text{res}}$ , which are related to the stored energy  $\mathcal{U} \propto E^2$ :

$$\frac{1}{\mathcal{U}} \frac{d\mathcal{U}}{dt} = \frac{2 dE/dt}{E} = -1/\tau_{\text{res}}.$$

We can work out the loss approximately by evenly spreading the mirror reflectivities ( $R = r^2$ ) and other losses over one revolution  $\tau_{\text{circ}} = \Delta_{\text{FSR}}^{-1}$ :

$$\left( \frac{\Delta E}{E} \right)_{\tau_{\text{circ}}} \simeq \frac{1}{2} \ln [(1 - A)RR'] = \ln \sqrt{(1 - A)RR'}.$$

Using the fact that  $R, R', (1 - A)$  have values close to unity and  $\ln (1 + x) \simeq x$ , the relation

$$\tau_{\text{res}} = -\frac{\tau_{\text{circ}}}{\ln \sqrt{(1 - A)RR'}} \simeq \frac{\tau_{\text{circ}}}{1 - \sqrt{(1 - A)RR'}}$$

is obtained. It is again related to the Q value or *quality* factor and the half-width  $\Delta\nu_{1/2}$  by

$$\Delta_{1/2} = \frac{\nu}{Q} = \frac{1}{2\pi\tau_{\text{res}}}.$$

For  $A \rightarrow 0$  and  $R = R'$ , this result reproduces Eq. (6.20). The resonator's damping time  $\tau_{\text{res}}$  rules the transient properties as well as the decay time of optical cavities. In Table 6.2 we have given some  $Q$  values and oscillation damping times  $\tau_{\text{res}}$ . It is assumed that the absorptive losses can be neglected compared to the transmission of the mirrors.

### 6.6.2 Modes and Mode Matching

For stability reasons, resonators no longer use plane mirrors in their construction, but curved ones.<sup>3</sup> With our knowledge of Gaussian beams from Section 2.3, we can understand immediately that an appropriate resonator mode has to be constructed according to the following principle: the surfaces of the mirrors must fit exactly the curvature of the wave fronts (see Figure 6.17).

Whether resonators work stably or unstably can again be investigated by means of the  $ABCD$  law of ray or Gaussian beam optics. A pair of mirrors is completely equivalent to the periodic lens system from Section 1.9.5 if we replace the focal lengths by the radii  $R_1/2$  and  $R_2/2$ . Thus from Eq. (1.28), we obtain the stability diagram (Figure 1.21 on p. 23) for optical cavities according to

$$0 \leq \left(1 - \frac{\ell}{R_1}\right) \left(1 - \frac{\ell}{R_2}\right) \leq 1. \quad (6.22)$$

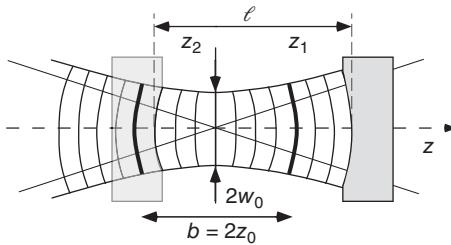
The characteristic parameters of an optical cavity consisting of two mirrors are their radii  $R_1$  and  $R_2$  and their separation  $\ell$ . Between the mirrors, a Gaussian standing wave with confocal parameter  $b = 2z_0$  and beam waist  $w_0$  is excited. The surfaces of the mirrors are at a separation  $\ell$  corresponding exactly to the length of the cavity:

$$\ell = z_1 + z_2.$$

The full solution of the Gaussian modes is described according to Eqs. (2.31) and (2.41):

$$\begin{aligned} E_{mn}(x, y, z) &= \mathcal{E}_0 \frac{w_0}{w(z)} \mathcal{H}_m(\sqrt{2} x/w(z)) \mathcal{H}_n(\sqrt{2} y/w(z)) \\ &\times \exp \{-[(x^2 + y^2)/w(z)]^2\} \exp [ik(x^2 + y^2)/2R(z)] \\ &\times \exp \{-i[kz - (m + n + 1)\eta(z)]\}. \end{aligned} \quad (6.23)$$

In the middle line the geometric form of the Gaussian general solution is given, which is characterized by  $(R(z), w(z))$  and  $(z_0, w_0)$ , respectively. Higher modes cause a transverse modulation  $\mathcal{H}_{m,n}$  of this basic form (upper line). Along the



**Figure 6.17** Gaussian wave and resonator mirrors.

<sup>3</sup> Unstable resonators are also used, for example, for the construction of high-power lasers [6.6]. See also Problem 6.

$z$  axis, the phase is solely determined by the Gouy phase, the last line in Eq. (6.23). That is why we can at first concentrate on the geometric adjustment of the wave fronts, which are described by  $R(z)$  according to Eq. (2.22).

At  $z_{1,2}$  in Figure 6.17, the radii of the wave fronts have to match the radii of curvature of the mirrors exactly,

$$R_{1,2} = \frac{1}{z_{1,2}}(z_{1,2}^2 + z_0^2) = z_{1,2} + \frac{z_0^2}{z_{1,2}}.$$

By means of

$$z_{1,2} = \frac{1}{2}R_{1,2} \pm \sqrt{R_{1,2}^2 - 4z_0^2},$$

we can then express the parameters of the Gaussian wave  $(z_0, w_0)$  by the cavity parameters  $(R_1, R_2, \ell)$ :

$$\begin{aligned} z_0^2 &= \frac{-\ell(R_1 + \ell)(R_2 - R_1 - \ell)}{(R_2 - R_1 - 2\ell)^2}, \\ w_0^2 &= \frac{\lambda z_0}{n\pi}. \end{aligned} \quad (6.24)$$

Exploration of this formula has to take into account that, according to the conventions for  $ABCD$  matrices (p. 18), mirror surfaces with their center to the left and right of the surface, respectively, have different signs.

For the excitation of a cavity mode, the Gaussian beam parameters  $(z_0, w_0)$  have to be precisely tuned to the incident wave. If this, the *mode matching* condition, is not met, only that fraction of the field is coupled, which corresponds to the overlap with the resonator mode.

### 6.6.3 Resonance Frequencies of Optical Cavities

A resonator is characterized by the spectrum of its resonance frequencies. From the Fabry–Pérot resonator, we expect an equidistant pattern of transmission lines at the distance of the free spectral range  $\Delta_{\text{FSR}}$ . For a more exact analysis, we have to take into account the phase factor (the Gouy phase, last line of Eqs. (6.23) and (2.23), respectively). The phase difference must again be an integer multiple of  $\pi$ :

$$\Phi_{mn}(z_1) - \Phi_{mn}(z_2) = q\pi k(z_1 - z_2) - (m + n + 1)[\eta(z_1) - \eta(z_2)]. \quad (6.25)$$

With  $\ell = z_1 - z_2$  and  $\eta(z) = \tan^{-1}(z/z_0)$ , we at first find

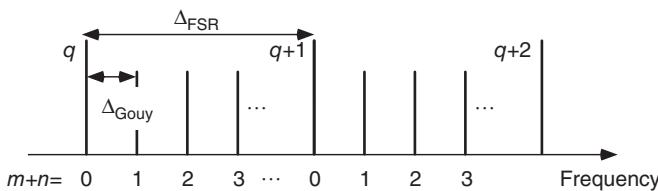
$$k_{qmn}\ell = q\pi + (m + n + 1) \left[ \tan^{-1}\left(\frac{z_1}{z_0}\right) - \tan^{-1}\left(\frac{z_2}{z_0}\right) \right].$$

The resonance frequencies  $v_{qmn}$  are determined from  $k_{qmn}\ell = 2\pi nv_{qmn}\ell/c = \pi v_{qmn}/\Delta_{\text{FSR}}$ . We introduce the resonator Gouy frequency shift

$$\Delta_{\text{Gouy}} = \left[ \tan^{-1}\left(\frac{z_1}{z_0}\right) - \tan^{-1}\left(\frac{z_2}{z_0}\right) \right] \frac{\Delta_{\text{FSR}}}{\pi}, \quad (6.26)$$

which varies between 0 and  $\Delta_{\text{FSR}}$ . We obtain the transparent result

$$v_{qmn} = q\Delta_{\text{FSR}} + (m + n + 1)\Delta_{\text{Gouy}}.$$



**Figure 6.18** Mode or frequency spectrum of a Fabry-Pérot resonator.

It shows a mode spectrum (Figure 6.18) with a rough division into the free spectral range  $\Delta_{FSR}$ . The fine structure is determined by resonance lines at the distance  $\Delta_{Gouy}$ .

#### 6.6.4 Symmetric Optical Cavities

We are now going to investigate the special case of a symmetric optical cavity consisting of two identical mirrors,  $R_2 = R = -R_1$ . In this important special case, the form of Eq. (6.24) is strongly simplified and can be interpreted as

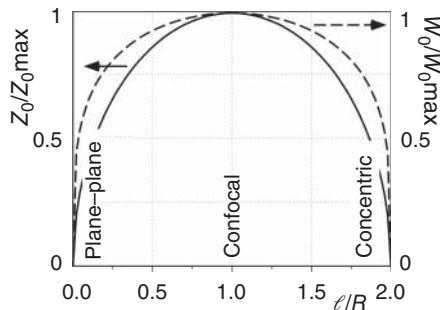
$$z_0^2 = \frac{(2R - \ell)\ell}{4} \quad \text{and} \quad w_0^2 = \frac{\lambda}{2\pi n} \sqrt{(2R - \ell)\ell}. \quad (6.27)$$

The length of the symmetric cavity can be varied from  $\ell = 0$  to  $2R$  before the region of stability is left.

The parameters of the Gaussian wave in a symmetric optical cavity,  $(z_0, w_0)$ , are shown in Figure 6.19, normalized to the maximum values  $z_{0\max} = R/2$  and  $w_{0\max} = (\lambda R/4\pi n)^{1/2}$ . The instability of the plane-plane and the concentric cavity is here also expressed by the sensitive dependence of the mode parameters on the  $\ell/R$  ratio.

In the symmetric cavity the Gouy phase (6.26) depends on the length and the radius of curvature according to

$$\Delta_{Gouy} = \Delta_{FSR} \frac{2}{\pi} \tan^{-1} \left( \frac{\ell}{2R - \ell} \right)^{1/2}. \quad (6.28)$$



**Figure 6.19** Rayleigh length and beam waist for a symmetric optical cavity.

#### 6.6.5 Optical Cavities: Important Special Cases

The three special cases  $\ell/R = 0, 1, 2$  deserve particular attention because they exactly correspond to the plane parallel, confocal, and concentric cavities.

##### 6.6.5.1 Plane Parallel Cavity: $\ell/R = 0$

The Fabry-Pérot interferometer or etalon described in the previous sections exactly corresponds to this extreme case. As we know from Figure 1.21, it is an extreme case in terms of stability. In practical use it is also important

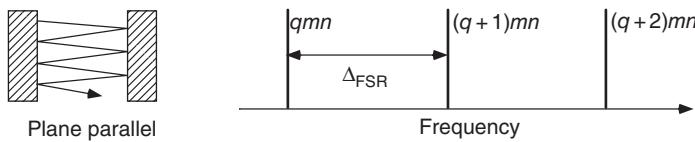


Figure 6.20 Path of rays and resonance frequencies of the plane parallel cavity.

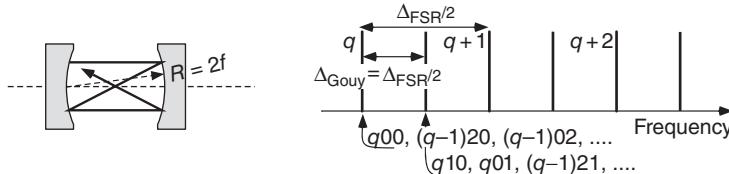


Figure 6.21 Path of rays and resonance frequencies of the confocal cavity.

that polished flat surfaces always have a slight convex curvature for technical reasons, so that an FPE consisting of two plane air-spaced mirrors always tends to instability. The Gouy phase (6.28) of the plane parallel interferometer is  $\Delta_{\text{Gouy}} = 0$ , rendering all transverse modes degenerate as shown in Figure 6.20.

### 6.6.5.2 Confocal Cavity: $\ell/R = 1$

If the focal lengths of the cavity mirrors coincide ( $f_1 + f_2 = R_1/2 + R_2/2 = \ell$ ), the configuration of the confocal cavity is obtained. In the symmetric case we have  $R_1 = R_2 = \ell$ .

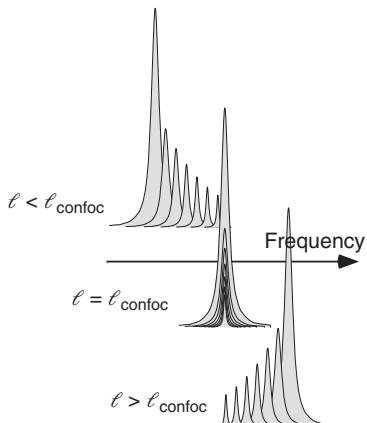


Figure 6.22 When confocal cavities are slightly displaced from their perfect mirror separation at  $\ell/R = \ell_{\text{confoc}}/R = 1$ , the degeneracy of transverse modes is lifted.

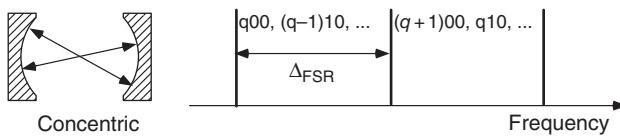
In this case, the Gouy phase is  $\Delta_{\text{Gouy}} = \Delta_{\text{FSR}}/2$ , and the modes are arranged at two highly degenerate frequency positions (Figure 6.21) at a separation of

$$\Delta_{\text{FSR}}^{\text{confoc}} = c/4n\ell. \quad (6.29)$$

The high degeneracy has its ray optical analog in the fact that paraxial trajectories are closed after two revolutions; see Figure 6.21.

If the confocal cavity is irradiated by a laser beam without mode matching, many transverse modes are excited and the frequency separation  $c/4n\ell$  (Eq. (6.29)) can be observed as an effective free spectral range and not as  $c/2nl$ . It is instructive to observe the emergence of transverse modes to the left or right of the fundamental modes if the length of the confocal cavity is slightly displaced from the perfect position  $\ell/R = 1$ , as indicated in Figure 6.22.

The high degeneracy makes the confocal cavity particularly insensitive in terms of handling and convenient for practical spectral analysis (see Section 8.1.7).



**Figure 6.23** Path of ray and resonance frequencies of the concentric cavity showing large degeneracy.

In general, a larger linewidth will be observed than is to be expected according to the simple relation of Eq. (6.20). This broadening is caused by the higher modes, which suffer from stronger damping and show exact degeneracy only within the paraxial approximation.

#### 6.6.5.3 Concentric Cavity: $\ell/R = 2$

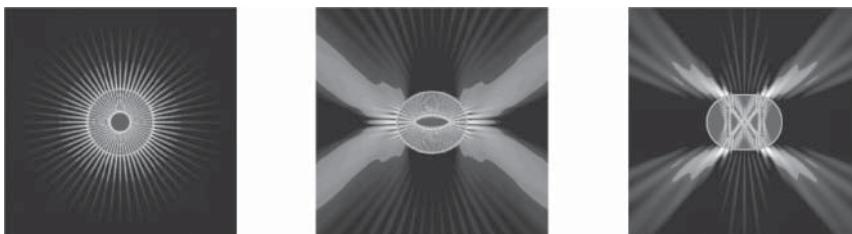
Obviously, this cavity is very sensitively dependent on the exact positions of the mirrors, but it leads to a very sharp focusing, which reaches the diffraction limit. In laser resonators, nearly concentric cavities are frequently used to concentrate the pump laser as well as the laser beam into a small volume where large amplification density is realized (see, e.g., Figure 8.33). In concentric cavities the Gouy phase  $\Delta_{\text{Gouy}} = \Delta_{\text{FSR}}$  causes strong degeneracy of the transverse modes again (Figure 6.23).

#### Excuse: Microcavities

In recent times there has been great interest in miniaturized devices of optical cavities with dimensions of a few micrometers. Since the radiation field is stored in a very small volume, a strong coupling of radiation field and matter can be obtained there.

The external coupling is not simple in such cavities since the direction of the emission is not simply controllable. In this context, there have recently been investigations on oval<sup>4</sup> cavities [80], which help to solve this problem by their shape.

In Figure 6.24 the calculated intensity distributions for a cylindrically symmetric, an elliptical, and an oval cavity are shown. The connection with concepts borrowed from ray optics can be seen in particular for the oval cavity.



**Figure 6.24** Distribution of light in circular, elliptical, and oval microcavities. (By courtesy of [80].)

<sup>4</sup> These are not elliptical cavities, which can be treated analytically and show a discrete spectrum.

## 6.7 Thin Optical Films

Thin optical films play an important role for applications, since dielectric coatings to reduce or enhance optical reflections have found their way into everyday life, for instance, on spectacles. We shall limit ourselves to the interference phenomena associated with thin optical films, and we shall ignore almost completely the important aspects of materials science for their manufacture.

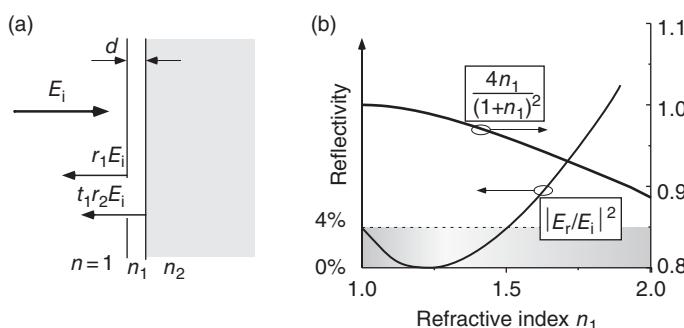
Metallic mirrors cause losses of 2–10% when reflecting visible wavelengths. That is, more than many laser systems can tolerate just to overcome the threshold. With a wealth of transparent materials, dielectric film systems with a structured refractive index can be fabricated, making predictable reflectivities between 0% and 100% possible. For the highest reflectivity, both transmission and absorption must be specified, which in the best case are only a few parts per million!

### 6.7.1 Single-Layer Films

We consider the single film from Figure 6.25 and determine the reflected wave that is the result of the superposition of reflections from the first and second interfaces, that is,  $E_r = r_1 E_i + t_1 r_2 E_i$ . It is straightforward to check using the known formulae for the reflection coefficient from Section 3.1.1 that, for perpendicular incidence, the amplitude of the reflected wave obeys

$$\begin{aligned} E_r &= \left( \frac{1 - n_1}{1 + n_1} + \frac{4n_1}{(1 + n_1)^2} \frac{n_1 - n_2}{n_1 + n_2} e^{i2kn_1 d} e^{i\Phi} \right) E_i \\ &\simeq \left( \frac{1 - n_1}{1 + n_1} + \frac{n_1 - n_2}{n_1 + n_2} e^{i2kn_1 d} e^{i\Phi} \right) E_i. \end{aligned} \quad (6.30)$$

For reflection from the denser medium at the second interface, with  $n_1 < n_2$ , we have  $\Phi = \pi$ , and for  $n_1 > n_2$  (less dense medium),  $\Phi = 0$ . The simplification in the second line of Eq. (6.30) is made possible by the negligible deviation of the transmission factor  $4n_1/(1 + n_1)^2$  from unity in the technically important range between  $n = 1.3$  and  $n = 2$  (see Figure 6.25). Even with a single-layer thin dielectric film, good results in terms of the coating of optical glasses can be obtained. For technically advanced applications, though, systems consisting of many layers are necessary.



**Figure 6.25** Reflection at a single thin film. (a) Film system. (b) Factor  $4n_1/(1 + n_1)^2$  and the effect of a single film on glass with  $n = 1.5$  and optimal film thickness  $d = \lambda/4n_1$ .

### 6.7.1.1 Minimal Reflection: AR Coating, AR Layer, and $\lambda/4$ Film

The thin film is designed as a single-layer  $\lambda/4$  film with  $d = \lambda/4$ . In addition, we choose  $n_1 < n_2$ , so that we have  $\exp(i\Phi = 0) = 1$  because of the reflection at the denser medium and  $\exp(2ikd) = -1$  causing destructive interference of the partial waves. In comparison with the substrate, the film shows low refraction and hence is called an *L-film*. For perfect suppression of optical reflection, the condition

$$\frac{1 - n_1}{1 + n_1} = \frac{n_1 - n_2}{n_1 + n_2}$$

has to be met, which is equivalent to

$$n_1 = \sqrt{n_2}. \quad (6.31)$$

The simple anti-reflection (AR) films used for “coating” of spectacles and windows reduce the reflection of the glass from 4% to typically 0.1–0.5%. A commonly used material is  $MgF_2$ , which quite closely fulfills condition (6.31) when used on glass ( $n = 1.45$ ).

### 6.7.1.2 Reflection: Highly Reflective Films

In this case we first choose a highly refractive film or *H-film* on a substrate with a lower refractive index, that is,  $n_1 > n_2$ . The  $180^\circ$  or  $\pi$  phase jump during the reflection at the less dense medium now causes constructive interference of the two partial waves, and the total reflectivity is enhanced. A single  $TiO_2 \lambda/4$  film on glass, for example (see refractive indices in Table 6.3), increases the reflectivity from 4% to more than 30% (see Figure 6.27).

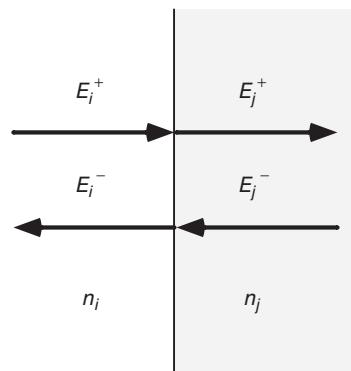
## 6.7.2 Multilayer Films

As a model example of a multilayer film, let us study a periodic film stack consisting of  $N$  identical elements [81]. We have to consider the splitting of the waves at each interface (Figure 6.26):

$$E_j^+ = t_{ij} E_i^+ + r_{ji} E_j^-,$$

$$E_i^- = t_{ji} E_j^- + r_{ij} E_i^+.$$

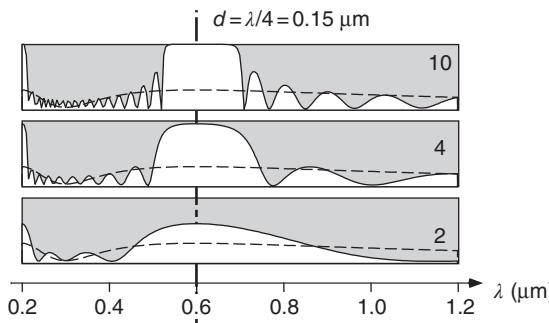
To be specific, we take a wave incident from the left. However, we elaborate the transformation in the opposite direction, since there is no wave traveling to the left behind the last interface in the direction of propagation ( $E_N^- = 0$ ).



**Figure 6.26** Interface within a multilayer optical film.

**Table 6.3** Refractive index of materials for thin dielectric films.

$Na_3AlF_6$	$MgF_2$	$SiO$	$TaO_2$	$TiO_2$
1.33	1.38	1.47	2.05	2.30



**Figure 6.27** Wavelength-dependent reflectivity of multiple films with 2, 4, and 10 layer pairs of films. In this example a stack of  $\text{TiO}_2$  and glass layers each with a thickness of  $0.15 \mu\text{m}$  is assumed. The dashed line marks the reflectivity of a single film. The 10-stack has a reflectivity  $R > 99\%$  between  $0.55$  and  $0.65 \mu\text{m}$ .

The set of equations can be solved and conveniently represented in a matrix if we also use  $r_{ij} = -r_{ji}$  and  $|t_{ij}t_{ji}| + |r_{ij}r_{ji}| = 1$ . Thus

$$\mathbf{E}_i = \mathbf{G}_{ij}\mathbf{E}_j = \frac{1}{t_{ij}} \begin{pmatrix} 1 & r_{ij} \\ r_{ij} & 1 \end{pmatrix} \begin{pmatrix} E_j^+ \\ E_j^- \end{pmatrix}.$$

Before getting to the next interface, the wave undergoes a phase shift  $\varphi = \pm n_i k d$  for the wave running to the right and to the left, respectively. In this case the total transformation from one interface to the other is

$$\mathbf{E}_j = \Phi_{ji}\mathbf{G}_{ji}\mathbf{E}_i = \mathbf{S}_{ji}\mathbf{E}_i \quad \text{with} \quad \Phi_{ji} = \begin{pmatrix} e^{-i\varphi} & 0 \\ 0 & e^{i\varphi} \end{pmatrix},$$

and in particular for  $N$  interfaces

$$\mathbf{E}_1 = \mathbf{S}_{1,2}\mathbf{S}_{2,3} \cdots \mathbf{S}_{N-2,N-1}\mathbf{S}_{N-1,N} \begin{pmatrix} E_N^+ \\ 0 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} E_N^+ \\ 0 \end{pmatrix}. \quad (6.32)$$

Thus the relation between incident, reflected, and transmitted waves is uniquely determined. In particular, the reflectivity can be calculated from  $|R_{21}|^2/|R_{11}|^2$  once  $\mathbf{R} = \mathbf{S}_{1,2}\mathbf{S}_{2,3} \cdots \mathbf{S}_{N-2,N-1}\mathbf{S}_{N-1,N}$  is known. While an analytical solution remains laborious, numerical solution by computer is straightforward. In Figure 6.27 the evolution of reflectivity from a single-layer pair film to a highly reflective multilayer film is shown.

## 6.8 Holography

One of the most remarkable and attractive capabilities of optics is image formation, to which we have already dedicated an entire chapter (Chapter 5). Among the various methods, usually holography (from the Greek word *holo*, meaning “complete” or “intact”) arouses the greatest astonishment. The attraction is mostly caused by the completely three-dimensional (3D) reconstruction of a recorded object! Here, we shall restrict ourselves to the interferometric principles of holography and refer the reader to the specialist literature for more intensive studies [82].

### 6.8.1 Holographic Recording

For a conventional record of a picture, whether by using an old-fashioned film or a modern charge-coupled device (CCD) camera, always the spatial distribution of the light intensity is saved on the film or in digital memory. For a hologram, both the amplitude and the phase of the light field are recorded instead. For this purpose the light field scattered off the object, the *signal wave* with amplitude distribution

$$E_S(x, y) = \frac{1}{2}[\mathcal{E}_S(x, y) e^{-i\omega t} + c.c.],$$

is superposed with a coherent *reference wave*

$$E_R(x, y) = \frac{1}{2}[\mathcal{E}_R(x, y) e^{-i\omega t} + c.c.].$$

The interference pattern is recorded by a film. One thus produces an interferometric record of an object – information about the image is truly contained in the interference pattern! The intensity distribution recording this information results from the superposition of signal and reference waves:

$$2I(x, y)/c^2\epsilon_0 = |E_S + E_R|^2 = |\mathcal{E}_S|^2 + |\mathcal{E}_R|^2 + \mathcal{E}_S \mathcal{E}_R^* + \mathcal{E}_S^* \mathcal{E}_R. \quad (6.33)$$

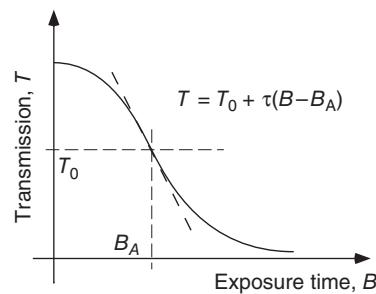
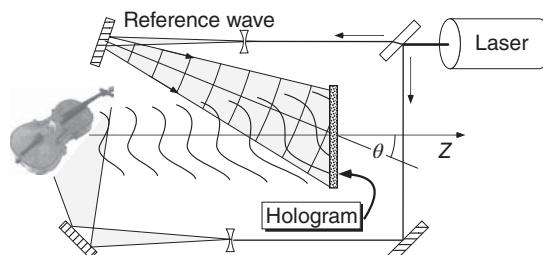
For this we have already assumed that the signal and reference waves have a sufficiently well-defined phase relation, since they originate from the same coherent light source. Otherwise the mixed terms would suffer from prohibitive temporal fluctuations.

The illumination intensity on the film material – which usually has nonlinear properties (see Figure 6.28) – is adjusted such that a linear relation between the transmission and the intensity distribution is obtained, that is,

$$T(x, y) = T_0 + \tau I(x, y). \quad (6.34)$$

Historical experiments in the 1940s by D. Gábor (1900–1979, Nobel Prize winner in 1971) were obtained as *in-line* holograms, since there the requirements for the coherence of the light source are not so stringent. Today, since lasers with a large coherence length are readily available, the holographic record is typically taken according to the *off-axis* method of Leith–Upatnieks shown in Figure 6.29.

**Figure 6.29** Record of a hologram according to the Leith–Upatnieks method.



**Figure 6.28** Holography uses the linear part of the blackening of the film.

The monochromatic signal wave propagates in the  $z$  direction in the arrangement of Figure 6.29, and the transverse phase distribution is caused by the illuminated object:

$$E_S = \mathcal{E}_S e^{-i\omega t} e^{ik_z z} e^{i\phi(x,y)}.$$

The (almost) plane reference wave has identical frequency  $\omega$  and travels at an angle  $\theta$  toward the  $z$  axis. The wave vector  $k$  has components  $k_z = k \cos \theta$  and  $k_y = k \sin \theta$ , and thus

$$E_R = \mathcal{E}_R e^{-i\omega t} e^{ik_z z} e^{ik_y y}.$$

Following Eq. (6.33) at plane  $P$  with  $\phi_0 = k_z z_0$ , we obtain the intensity distribution

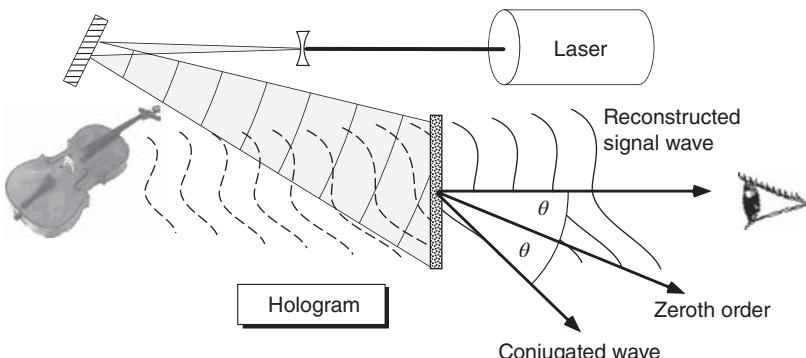
$$I_P(x,y) = I_S + I_R + \mathcal{E}_S \mathcal{E}_R^* e^{i\phi_0} e^{-i[k_y y + \phi(x,y)]} + c.c. \quad (6.35)$$

All contributions cause a blackening of the film material. The reference wave usually corresponds in good approximation to a plane wave and so generates homogeneous blackening. The blackening caused by the signal wave, which for simplification has been assumed to have a constant amplitude, usually generates an inhomogeneous intensity distribution since there are no plane wave fronts emanating from an irregular object. In a different situation, this phenomenon is also known as *laser speckle* and is discussed in more detail in Section 6.9.

### 6.8.2 Holographic Reconstruction

The major fascination of holography is manifest in the actual image reconstruction process, since the holographic film itself – the hologram – does not contain any information for the human eye. For reconstruction, the object is removed, and the hologram is illuminated once again with the reference wave. By diffraction, the secondary waves shown in Figure 6.30 are generated.

Formally we can derive the secondary wave by considering the field distribution immediately after passing through the hologram. We can identify four



**Figure 6.30** Image reconstruction from a hologram with secondary waves (see also Figure 6.29).

different diffracted waves,  $U_0$ ,  $U_0^H$ ,  $U_{+1}$ , and  $U_{-1}$ ,

$$\begin{aligned} E_{\text{recon}} &= T(x, y)E_R \\ &= T_0E_R + \tau E_R I_R + \tau E_R I_S + \tau |E_R|^2 E_S + \tau E_R^2 E_S^* \\ &= U_0(x, y) + U_0^H(x, y) + U_{+1}(x, y) + U_{-1}(x, y), \end{aligned}$$

which we are going to consider in detail. Actually, it is quite complicated to determine the diffraction field of a complex hologram. Fortunately, we can identify every term with a known waveform naturally continued from the local field distribution.

#### 6.8.2.1 Zeroth Order

$$U_0(x, y) = (T_0 + \tau I_R) \mathcal{E}_R e^{-i\omega t} e^{i(k_y y + k_z z)}.$$

This term propagates in zeroth order because its wave vector is identical to that of the incident reference wave, which is continued and plainly attenuated by a constant factor  $(T_0 + \tau I_R) < 1$  due to attenuation.

#### 6.8.2.2 Halo

$$U_0^H(x, y) = \tau I_S \mathcal{E}_R(x, y) e^{-i\omega t} e^{i(k_y y + k_z z)}.$$

As mentioned earlier, the signal wave usually causes inhomogeneous blackening. The secondary wave also propagates in zeroth order, but the diffraction of the speckle pattern (see Section 6.9) leads to broadening compared to the transmitted reference wave and is sometimes called a “halo.”

#### 6.8.2.3 Reconstructed Signal Wave

$$U_{+1}(x, y) = \tau \mathcal{E}_S e^{i\phi(x, y)} \mathcal{E}_R^* \mathcal{E}_R e^{-i\omega t} e^{ikz}.$$

Obviously, with this contribution, the signal wave is exactly reconstructed except for a constant factor! The reconstructed signal wave propagates in the  $z$  direction, which we are going to call the first order in analogy to diffraction by a grating. The virtual image contains all 3D information of the reconstructed object and can therefore be observed – within the light cone – from all sides.

#### 6.8.2.4 Conjugated Wave

$$U_{-1}(x, y) = \tau \mathcal{E}_R^2 \mathcal{E}_S e^{-i\omega t} e^{-i\phi(x, y)} e^{i[2k_y y + (2k_z - k)z]}.$$

In a vector diagram we can determine the propagation direction of the so-called conjugated wave. For small angles  $\theta = k_y/k_z$ , we have  $2k_z - k \simeq k_z$  and  $k_{\text{conj}}^2 = 4k_y^2 + (2k_z - k)^2 \simeq k^2$ . That is why the axis of the conjugated ray runs at angle  $2\theta$  to the  $z$  axis and disappears at  $\theta = \pi/4$  at the latest. Writing it as

$$U_{-1}(x, y) = \tau \mathcal{E}_R^2 (\mathcal{E}_S e^{i\phi(x, y)})^* e^{-i\{\omega t - k[\sin(2\theta)y + \cos(2\theta)z]\}},$$

the “phase-conjugated” form of this ray in comparison with the object wave becomes transparent. From a physical point of view, the curvature of the

wave fronts is inverted, so the wave seems to run backward in time. Again, following the analogy to diffraction by a grating, this wave is also called the  $-1$  order of diffraction.

Compared to an in-line hologram, the three secondary waves of interest can be easily separated geometrically and observed in off-axis holography (Figure 6.30).

### 6.8.3 Properties

Holograms have many fascinating properties, out of which we have selected only a few here.

#### 6.8.3.1 Three-Dimensional Reconstruction

Since the signal wave coming from the object is reconstructed, the virtual image looked at by the observer through the holographic plate appears 3D as well. It is even possible to look behind edges and corners if there exists a line-of-sight connection with the illuminated areas.

#### 6.8.3.2 Partial Reconstruction

The complete object can be reconstructed from each fragment of a hologram. This seems to be inconsistent at first, but becomes clear in direct analogy to diffraction by a grating. There, the diffraction patterns observed from more and more reduced fragments always stay the same as well. However, the width of each diffraction increases, that is, the resolution of the grating is reduced due to the decreasing number of illuminated slits. In a similar way the resolution declines in reconstructing from a holographic fragment. The finer structures of the image disappear, while the gross shape of the signal wave and hence the object is preserved.

#### 6.8.3.3 Magnification

If, in reconstructing an object, light of a different wavelength is used, the scale of the image is correspondingly changed.

## 6.9 Laser Speckle (Laser Granulation)

When a dim wall or a rough object is illuminated with laser light, the observer distinguishes a granular, speckled structure, which does not appear in illumination with a conventional light source and is obviously caused by the coherence properties of the laser light. In fact, coherent phenomena, that is, diffraction and interference, can also be observed using incandescent light sources, but the invention of the laser has really granted us a completely new sensory experience. Newton had already recognized that the “twinkling” of the stars, having been poetically raised by our ancestors, is a coherence phenomenon caused by the inhomogeneities of the atmosphere and thus directly related with speckle patterns.

The granular irregular structure is called *laser granulation* or *speckle pattern*. Reflected off the rough, randomly shaped surface of a large object, a coherent

wave acquires a complex wave front like after passing through a ground glass screen. For simplification, we can imagine that the light beams from a large number of randomly arranged slits or holes interfere with each other. In each plane there is thus a different statistical interference pattern. Indeed, every observer sees a different but spatially stable pattern as well.

Formal treatment of the speckle pattern requires some expense using the mathematical methods of statistics. We briefly discuss this phenomenon at least qualitatively, since it is nearly ubiquitous wherever laser light is used. Although laser granulation at first appears an undesirable consequence of interference, it contains substantial information about the scattering surfaces, and it is even suitable for interferometric application in the measurement techniques for the determination of tiny surface changes [83].

### 6.9.1 Real and Virtual Speckle Patterns

Speckle patterns can be observed, for example, when we expand a laser beam and project it from a diffuse reflector onto a screen. On the wall there is a fixed granular pattern that only changes with a different reflector. This pattern is determined only by the microstructure of the reflector and is called a *real or objective speckle pattern* [22]. It can be recorded by direct exposure of a film.

When it is imaged, however, it is transformed by the imaging process itself. A *subjective or virtual speckle pattern* is generated, with properties determined by the aperture of the imaging optics, for example, the size of the pupil of our eye. This property can be easily understood and observed just with a laser pointer illuminating a white wall. If we form a small hole or some kind of artificial pupil, with our hand, the granulation speckles grow rougher, the smaller the diameter of the hole.

Detailed consideration of the coherent wave field is not usually of interest. We finish this discussion with a short discussion of the intensity distribution in the statistical wave fields, leading to laser granulation and the characteristic dimensions of the speckle grains.

### 6.9.2 Speckle Grain Sizes

The sizes of speckle grains can be estimated by a simple consideration [22]. The lens of an imaging objective is illuminated by a wave field of a granulation pattern with a time-invariable but spatially random phase distribution. The characteristic scale  $d$  of the interference pattern is determined by the resolution of the image and reaches the Rayleigh criterion as in Eq. (2.72). If wave fronts from a large distance are incident on the lens, the beams coming from a certain direction are superimposed in the focal plane at a distance  $f$ . For a circular lens with aperture  $D$ , the diameter of the focal spots cannot become smaller than

$$d = 1.22\lambda f/D.$$

With a decrease in the aperture size, a roughening of the speckle pattern is to be expected. This phenomenon is shown in Figure 6.31, where an effective aperture is formed by focusing of the laser beam onto a ground glass substrate.

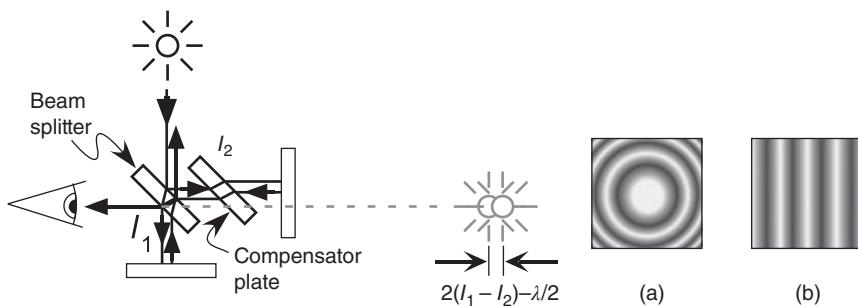


**Figure 6.31** Speckle pattern of a focused helium–neon laser beam after passing through a ground glass substrate, showing the statistical pattern. From left to right, the focus was shifted more and more into the substrate. Stronger focusing leads to coarser interference structures.

A rotating glass disc produces a rotating interference pattern like that shown in Figure 6.31. A light beam cut out by an aperture will show strong intensity fluctuations, resulting in a so-called pseudo-thermal light beam [84]. The quantum properties of such light beams are presented in Section 14.5.4.1.

## Problems

- 6.1 Interference fringes of the Michelson interferometer** The interference pattern at the exit of a Michelson interferometer can be understood with the following construction. An observer looking into the exit sees two virtual images of the light source, which irradiates the input of the interferometer. At the exit the radiation fields of the two virtual light sources interfere with each other. Depending on (mis)alignment, different patterns are observed.
- Verify the path of the rays leading to the two virtual light sources. Explain why the longitudinal difference between the virtual images in Figure 6.32 is  $2(l_1 - l_2 + \lambda/2)$ .
  - Explain the situation that leads to the interference pattern shown in Figure 6.32a.
  - For a slightly misaligned interferometer, one can observe the interference pattern shown in Figure 6.32b. Sketch the path of the rays for this case and give the position of the virtual light sources.
- 6.2 Unitarity of the beam splitter** Consider a beam splitter with reflection and transmission coefficients  $r_{1,2}$  and  $t_{1,2}$  for the two input ports. Show that as a consequence of energy conservation alone, the reflected waves must be  $90^\circ$  out of phase.
- 6.3 Interferences with a ruler** With a simple laser pointer, one can generate beautiful interferences with just a simple student's ruler. Shine the laser pointer at grazing incidence onto that side of the ruler having the tick marks. What interference figures do you get for millimeter and centimeter marks? What is their relation? If instead of centimeter units, inch units



**Figure 6.32** Virtual light sources of the Michelson interferometer.

are used, the rulers frequently have marks at  $1/2, 1/4, 1/8, \dots$ . What is the difference to the standard decimal ruler? Why can we not observe these interferences with a conventional light source?

- 6.4 Fringes of the Fabry-Pérot interferometer** The classic Fabry-Pérot interferometer uses ground glass (Figure 6.13) for illumination. The lens causes ring-shaped interference fringes in the focal plane. Calculate the position of the rings as a function of the focal length  $f$  of the lens.
- 6.5 Quality of the mirrors for a Fabry-Pérot interferometer** Surface roughness of mirrors causes deformations of the phase fronts of the reflected light. These deformations influence the spectral resolution of the Fabry-Pérot interferometer, which is characterized by the finesse (Eq. (6.21)). Consider as a model the influence of a step on the mirror surface of height  $h$ . Show that for  $h < \lambda/2F$  the resolution is not significantly modified.
- 6.6 Unstable confocal resonator** Find the position of the confocal resonator with radii  $\{R_1, R_2\}$  in the stability diagram of Figure 1.21. Show that they are unstable with the exception of the symmetric case  $R_1 = R_2$ . What is the difference between the two branches? How much is the cross section of the beam modified for  $R_1 \neq R_2$  after each round trip? If we interpret the finesse as an effective number of round trips in the resonator, we can define a finesse for the unstable resonator as well. How many round trips are necessary as a function of the ratio of the radii,  $R_1/R_2$ , to widen the cross section to twice its initial value? Unstable resonators play an important role for high-power laser systems. They are discussed in detail in [3].
- 6.7 Cavity ring-down spectroscopy** The storage time for light circulating in an optical resonator is limited by the losses of the resonator. Discuss what contributions leading to losses have to be accounted for in addition to the mirror reflectivity  $R < 1$ . The decay time (“ring-down time”) is further shortened by absorbing samples immersed into the resonator field, for example, atomic or molecular gases. These substances can be

detected by *cavity ring-down spectroscopy*. Show that the storage time can be described by  $\tau = \ell / (c(1 - R) + \alpha s)$  where  $\ell$  is the resonator length,  $R$  the mirror reflectivity,  $\alpha$  the absorption coefficient, and  $s < \ell$  the length of the sample. How large must the mirror reflectivity be chosen in order to obtain reasonable “empty” ring-down times of about  $\tau > 10 \mu\text{s}$ ? What is the sensitivity for detecting atoms with a strong absorption line at resonance? (Hint: Estimate the absorption coefficient  $\alpha$  from Eq. (7.20).)

- 6.8 Computer analysis of multilayer dielectric mirrors** Write a computer program (we suggest using computer algebra programs such as Maple<sup>TM</sup> or Mathematica<sup>TM</sup>) to display Eq. (6.32) graphically as a function of the wavelength or frequency. Study the width of the reflection band centered around  $\lambda_0$  as a function of the number of layers. What is the reflectance at  $2\lambda_0$  and  $3\lambda_0$ ? Extend your program by accounting for material dispersion.

**7**

## Light and Matter

An electromagnetic wave accelerates electrically charged particles in gases, liquids, and solids and, in so doing, generates polarizations and currents. The accelerated charges for their part again generate a radiation field superimposed onto the incident field (Figure 7.1). To understand macroscopic optical properties, it is necessary to describe the polarization properties of matter microscopically, which can only be done by means of quantum theory. Despite that, classical theoretical physics has been able to explain numerous optical phenomena by phenomenological approaches.

The quantum theoretical description of matter has led to the development of “quantum electronics” (see Table 7.1), in which the electromagnetic radiation fields are still taken into account classically, that is, with a well-defined phase and amplitude. This kind of treatment of the radiative interaction is also called “semiclassical.”

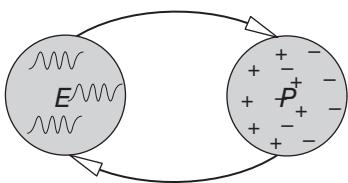
Ultimately also electromagnetic fields have to be dealt with in a quantum theoretical way when phenomena such as the famous “Lamb shift” are to be understood. Today “quantum electrodynamics” (QED) is considered a model case of a modern physical field theory. In “quantum optics” in a narrower sense,<sup>1</sup> in particular the quantum properties of optical radiation fields are dealt with [85, 86], for example, the spectrum of resonance fluorescence or so-called photon correlations. Such topics are treated in Chapter 14.

Since the beginning of the 1980s, it has been possible to influence the motion of atoms by radiation pressure of light, or laser cooling. The kinetic energy in a gas cooled in such a way can be decreased so much that atomic motion can no longer be comprehended like that of classical, or point-like, particles. Instead, their center-of-mass motion has to be dealt with according to quantum theory and can be interpreted in terms of matter waves. In the excursion on p. 187, we have already used this explanation for the diffraction of atomic beams. The hierarchy of theoretical concepts for light–matter interaction is summarized in Table 7.1.

When the effect of a light field on dielectric samples is to be described, generally the electric dipole interaction is sufficient since it is stronger than all other couplings, such as magnetic effects and higher-order multipoles, which

---

<sup>1</sup> The term *quantum optics* is in general not very precisely defined.



**Figure 7.1** An electromagnetic field  $E$  generates a polarization  $P$  in matter consisting of positive and negative charges. The accelerated charges generate a radiation field and so react upon the fields.

**Table 7.1** Treatment of light and matter by theoretical physics.<sup>a)</sup>

	Matter	Light	Atomic motion
Classical optics	C	C	C
Quantum electronics	Q	C	C
Quantum optics	Q	Q	C
Matter waves	Q	Q	Q

a) C = classical physics; Q = quantum theory.

can be neglected. The concepts of optics can also be extended without any problems if such phenomena are to be treated theoretically.

## 7.1 Classical Radiation Interaction

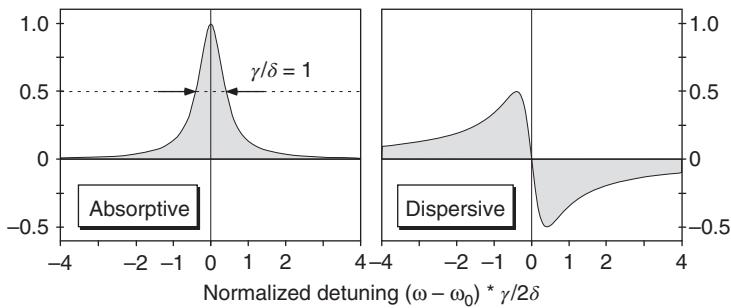
### 7.1.1 Lorentz Oscillators

A simple yet very successful model for the interaction of electromagnetic radiation with polarizable matter goes back to H. Lorentz (1853–1928). In this model, electrons are considered with a spring that are harmonically bound like little planets to an ionic core and oscillate at optical frequencies  $\omega_0$ . The classical dynamics of such a system is well known. The influence of a light field shows up as driving electrical or magnetic forces adding to the binding force  $\mathbf{F}_B = -m\omega_0^2 \mathbf{x}$ .

Additionally, we assume that damping of the oscillator is caused by release of radiation energy. Although this concept cannot be fully explained by classical electrodynamics without some contradictions, in an approximation, it leads to the *Abraham–Lorentz equation*, in which, besides the binding force, a damping force  $\mathbf{F}_R = -m\gamma(d\mathbf{x}/dt)$  occurs, causing weak damping ( $\gamma \ll \omega_0$ ). At this stage the limits of classical electrodynamics become evident [87], because a consistent and correct calculation of  $\gamma$  can only be obtained by means of QED [88]; see Section 14.3.1. For our purposes, however, it is sufficient to consider  $\gamma$  as the phenomenological damping rate.

For simplification we use complex quantities to write the orbit radius,  $\mathbf{x} \rightarrow r = x + iy$ . We consider the equation of motion of the driven oscillator,

$$\ddot{r} + \gamma\dot{r} + \omega_0^2 r = \frac{q}{m} \mathcal{E} e^{-i\omega t}, \quad (7.1)$$



**Figure 7.2** “Quadrature” components of the Lorentz oscillator, which are oscillating in phase ( $x$ , absorptive) and  $90^\circ$  out of phase ( $y$ , dispersive) with the driving field, respectively. The amplitude is normalized to the maximum absorption value in the case of resonance at  $\delta = 0$ .

under the influence of a driving light field  $\mathcal{E} e^{-i\omega t}$ , which is circularly polarized. With the trial function  $r(t) = \rho(t) e^{-i\omega t}$ , the equilibrium solution  $\rho(t) = \rho_0 = \text{const.}$  with

$$\rho_0 = \frac{q\mathcal{E}/m}{(\omega_0^2 - \omega^2) - i\omega\gamma}$$

can be found easily from the secular equation  $\rho(-\omega^2 - i\omega\gamma + \omega_0^2) = q\mathcal{E}/m$ . For the near-resonant approximation,  $\omega \approx \omega_0$ , we can replace  $(\omega_0^2 - \omega^2) \simeq 2\omega_0(\omega_0 - \omega) = -2\omega_0\delta$  with detuning  $\delta$  and introduce the maximum radius  $\rho_{\max} = -q\mathcal{E}/m\omega_0\gamma$  to obtain

$$\rho_0 = \rho_{\max} \frac{\gamma/2}{\delta + i\gamma/2}.$$

For the  $x$  and  $y$  coordinates of the driven oscillator, we have

$$r(t) = x + iy = \rho_{\max} \frac{\gamma}{2} \frac{\delta - i\gamma/2}{\delta^2 + (\gamma/2)^2} e^{-i\omega t}. \quad (7.2)$$

We will see that, in terms of the propagation of light in polarizable matter,  $x$  and  $y$  give exactly the “dispersive” ( $x$ ) and the “absorptive” ( $y$ ) components of the radiation interaction. The shape of the *dispersion curve* and the *Lorentz profile* of absorption are presented in Figure 7.2. Here the term *normal dispersion* refers to the dominant positive slopes of the dispersion curve. This situation is typically found for transparent optical materials, which have electronic resonance frequencies beyond the visible domain in the UV. Negative slopes of dispersion are called “anomalous dispersion.”

It is known that an accelerated charge radiates, and so a charged harmonic oscillator has to lose energy. In classical electrodynamics the damping rate is calculated from the Larmor formula [87] describing the power radiated by a particle with charge  $q$  undergoing an acceleration  $a$ :

$$P_{\text{rad}} = \frac{1}{4\pi\epsilon_0} \frac{2q^2}{3c^3} a^2 \quad (7.3)$$

The damping thus caused is also called “radiation reaction” and has already been accounted for phenomenologically in Eq. (7.1). From a shortened version of the

derivation leading to the well-known Abraham–Lorentz equation, we can infer a damping rate depending on elementary atomic quantities only, which provides further insight into radiative properties. We have to keep in mind, however, that a suitable theory of damping must invoke the full quantum theory of light.

By multiplication with  $\dot{r}$ , we can introduce the Larmor rate of energy transfer from the charged oscillator to the radiation field ( $-P_{\text{rad}}$ ) into Eq. (7.1):

$$\frac{d}{dt} \left( \frac{m\dot{r}^2}{2} + \frac{m\omega_0^2 r^2}{2} \right) + m\gamma\dot{r}^2 = 0 = P_{\text{rad}} + m\gamma\dot{r}^2.$$

If damping is weak ( $\omega_0 \gg \gamma$ ), we can assume that, during one revolution period  $2\pi/\omega_0$ , the amplitude change ( $r$ ) is negligible, and so we can replace  $\ddot{r} = \omega_0^2 r$  and  $\dot{r} = \omega_0 r$ . Then we may identify the radiation power with the power dissipated through friction ( $m\gamma\dot{r}^2$ ), which was introduced phenomenologically before. We obtain

$$\gamma = \frac{q^2 \omega_0^2}{6\pi\epsilon_0 c^3 m} \quad \text{and} \quad \rho_{\text{max}} = \frac{3\epsilon_0 \lambda^3}{4\pi^2 q} \mathcal{E}. \quad (7.4)$$

This result is frequently used to introduce the so-called classical electron radius [87, 89]:

$$r_{\text{el}} = \frac{e^2 / 4\pi\epsilon_0}{2mc^2} = 1.41 \times 10^{-15} \text{ m} \quad \text{with} \quad \gamma = \frac{4}{3} \frac{r_{\text{el}} c}{\lambda^2}.$$

As far as we know from scattering experiments in high-energy physics, the electron is point-like down to  $10^{-18}$  m, and thus this quantity does not have physical significance.

In this way we can obtain the complex dipole moment of a single particle from  $d = q\rho_0$  according to Eqs. (7.2) and (7.4):

$$d(t) = q\rho_0 = -\frac{3\lambda^3}{4\pi^2} \frac{i - 2\delta/\gamma}{1 + (2\delta/\gamma)^2} \epsilon_0 \mathcal{E} e^{-i\omega t}. \quad (7.5)$$

Often the polarizability  $\alpha$  is used as well. It is defined by

$$d(t) = \alpha \mathcal{E} e^{-i\omega t},$$

and the coefficient  $\alpha$  is easily extracted from comparison with Eq. (7.5).

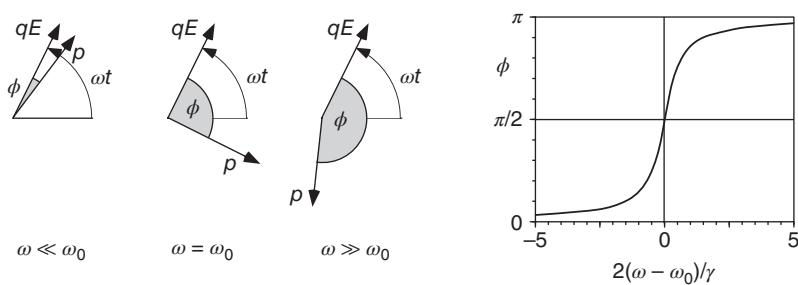
In the  $x$  as well as in the  $y$  component, there is a phase delay  $\phi$  between the electric field and the dipole moment, which is only dependent on the damping rate  $\gamma$  and the detuning  $\delta = \omega - \omega_0$  (Figure 7.3):

$$\phi = \arctan(\gamma/2\delta). \quad (7.6)$$

The so-called phase lag shows the known behavior of a driven harmonic oscillator, that is, in-phase excitation at low (“red”) frequencies, out of phase or  $90^\circ$  following in the case of resonance, and opposite phase at high (“blue”) frequencies.

From Eq. (7.1) we can furthermore infer the time-dependent equation for  $\rho(t)$ . We assume that the oscillation amplitude  $\rho(t)$  changes only slowly in comparison with the oscillation itself, that is,  $\dot{\rho} \ll \omega\rho$ . We then approximately obtain

$$\dot{\rho} + \left( i\delta + \frac{\gamma}{2} \right) \rho = -i \frac{q\mathcal{E}}{2m\omega} \quad (7.7)$$



**Figure 7.3** Phase lag of the Lorentz oscillator in steady state. At low frequencies, the driving field and the dipole oscillate in phase; in the case of resonance, the dipole follows the field out of phase at  $90^\circ$ ; and at high frequencies, it oscillates with opposite phase.

by furthermore applying  $i\omega + \gamma/2 \simeq i\omega$  as well. This complex equation provides an interesting analogy with the result of quantum mechanics discussed in Section 7.2.6. There we will find Bloch vector components exhibiting strong formal similarity with dipole quadrature components  $(u, v)$ , which we introduce here by letting  $\rho = u + iv$ . Decomposition of the complex equation of motion (7.7) into a system of real equations yields

$$\begin{aligned}\dot{u} &= \delta v - \frac{\gamma}{2} u, \\ \dot{v} &= -\delta u - \frac{\gamma}{2} v - \frac{q\mathcal{E}}{2m\omega}.\end{aligned}\tag{7.8}$$

We can furthermore complement this equation by

$$\frac{d}{dt}(u^2 + v^2) = -\gamma(u^2 + v^2) - \frac{q\mathcal{E}}{m\omega} v\tag{7.9}$$

and thus obtain a relation describing the excitation energy of the system. This is analogous to the third optical Bloch equation for the  $w$  component of the difference of the occupation numbers (see Eq. (7.36)).

#### Excuse: Lorentz oscillator in a magnetic field

If a magnetic field influences the motion of a charge, a Lorentz force is added to the equation of motion (7.1), which is  $\mathbf{F}_{\text{Lor}} = q\dot{\mathbf{x}} \times \mathbf{B}$  and results in an extra term  $iq\dot{\rho}\mathbf{B}$  when the replacement  $\mathbf{x} \rightarrow \mathbf{r} = \mathbf{x} + i\mathbf{y} \rightarrow \rho e^{-i\omega t}$  is carried out. If its influence on the dynamics is low,  $|q\mathbf{B}|/m \ll \omega_0$ , then the components of the magnetic field in the  $xy$  plane cause a rotation of the orbital plane, while the  $z$  component modifies the eigenfrequency of the oscillator. The complete equation of motion is now

$$\ddot{\rho} + \gamma\dot{\rho} + \omega_0^2\rho = \frac{q}{m}(\mathcal{E} + i\dot{\rho}\mathbf{B}_z) e^{-i\omega t}.\tag{7.10}$$

We seek solutions using the same procedures as before, and, with the *Larmor frequency*

$$\omega_L = q\mathbf{B}_z/2m,$$

(Continued)

**Excuse: (Continued)**

we obtain the equilibrium solution

$$\rho_0 = \frac{\rho_{\max}\gamma/2}{(\omega_0 - \omega_L - \omega) - i\gamma}. \quad (7.11)$$

In Eq. (7.1), the eigenfrequency  $\omega_0$  has only to be replaced by the modified value  $\omega_0 - \omega_L$ . Otherwise, the results can be taken over. With this theory, H. Lorentz was able to interpret the Zeeman effect, the shift and splitting of atomic resonance lines by external magnetic fields.

We finish by studying the effect of a transverse magnetic field on the motion of an electron. For this purpose we take the vector product of Eq. (7.10) by  $\mathbf{x} \times$  (replacing  $\rho \rightarrow \mathbf{x}$ ) and obtain a new equation for the electronic angular momentum  $\mathbf{L} = m\mathbf{x} \times \dot{\mathbf{x}}$ . Strictly speaking, this should be  $m\mathbf{x} \times (\dot{\mathbf{x}} \times \mathbf{B}) = \mathbf{L} \times \mathbf{B} + m\dot{\mathbf{x}} \times (\mathbf{x} \times \mathbf{B})$ , but in static fields the second term vanishes, and in alternating fields it is equivalent to a relativistic correction of first order,  $(v/c)\mathbf{d} \times \mathbf{E}$ , and can be neglected. So

$$\frac{d}{dt}\mathbf{L} + \gamma\mathbf{L} = \mathbf{d} \times \mathbf{E} + \frac{q}{m}\mathbf{L} \times \mathbf{B}.$$

It can be recognized from this equation that a circularly polarized electric light field as well as a transverse static field ( $\mathbf{B} \perp \mathbf{L}$ ) can cause rotation of the electronic angular momentum. The former case is usually called "optical pumping" in spectroscopy [90], and the latter case occurs in the Hanle effect [91]. See also Problem 7.3.

### 7.1.2 Macroscopic Polarization

The macroscopic polarization  $\mathbf{P}(\mathbf{r}, t)$  has already been introduced in Section 2.1.2 in order to describe the propagation of electromagnetic waves in a dielectric medium. From the microscopic point of view, a sample consists of the microscopic dipole moments of atoms, molecules, or lattice elements. The "near field" of the microscopic particle does not play a role in the propagation of the radiation field, which is always a "far field." If there are  $N_{\text{at}}$  atomic or other microscopic dipoles in a volume  $V$ , the macroscopic polarization is obtained from

$$\mathbf{P} = \frac{N_{\text{at}}}{V}\mathbf{p} = \frac{N}{V}\mathbf{d}(u + iv). \quad (7.12)$$

Here the volume  $V$  is chosen much larger than molecular length scales, for example,  $d_{\text{mol}} < 5 \text{ \AA}$ , and the average volume of a single particle as well. If the microscopic polarization density  $\mathbf{p}(\mathbf{r})$  is known, there is the more exact form:

$$\mathbf{P}(\mathbf{r}, t) = \frac{N_{\text{at}}}{V} \int_V \mathbf{p}(\mathbf{r} - \mathbf{r}', t) d^3 r'.$$

In our classic model the Fourier amplitudes of the polarization  $\mathcal{P} = \mathcal{F}\{\mathbf{P}\}$  and of the driving field  $\mathcal{E}$  are linearly connected,

$$\mathcal{P}(\omega) = \epsilon_0 \chi(\omega) \mathcal{E}(\omega), \quad (7.13)$$

and the susceptibility  $\chi(\omega) = \chi'(\omega) + i\chi''(\omega)$  can be given using the results of Eq. (7.5):

$$\begin{aligned}\chi'(\delta) &= \frac{N_{\text{at}}}{V} \frac{3\lambda^3}{4\pi^2} \frac{2\delta/\gamma}{1 + (2\delta/\gamma)^2}, \\ \chi''(\delta) &= \frac{N_{\text{at}}}{V} \frac{3\lambda^3}{4\pi^2} \frac{1}{1 + (2\delta/\gamma)^2}.\end{aligned}\quad (7.14)$$

Since the temporal behavior of the polarization is also characterized by transient processes, it usually depends on the field intensity also at earlier times. This becomes more apparent in the time-domain expression

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^{\infty} \chi(t - t') \mathbf{E}(\mathbf{r}, t') dt', \quad (7.15)$$

which requires  $\chi(t - t') = 0$  for  $(t - t') < 0$  in order not to violate causality. Here as well the literal meaning of “susceptibility” or “aftereffect” shows up. But for our purposes we assume that we are allowed to neglect relaxation processes occurring in solid materials within picoseconds or less, and therefore we can restrict our treatment to an instantaneous interaction.<sup>2</sup> According to the convolution theorem of Fourier transformation, the relation is, however, much simpler in the frequency domain following Eq. (7.13).

To be more exact, the “dielectric function” (Eq. (2.4))  $\epsilon_0 \kappa(\omega) = \epsilon_0 [1 + \chi(\omega)]$  and the susceptibility are second-rank tensors, for example,  $\chi_{ij} = \partial P_i / \partial E_j$ , and reflect the anisotropy of real materials. The magnetic polarization can mostly be neglected for optical phenomena ( $\mu_r \sim 1$ ), since the magnetic field  $\mathbf{B}$  and the  $\mathbf{H}$  field are identical except for a factor,  $\mathbf{H} = \mathbf{B}/\mu_0$ .

Only in an isotropic ( $\nabla \cdot \mathbf{P} = 0$ ) and, according to Eq. (2.4), linear medium does the wave equation take on a simple form. This is, however, an important and often realized special case where the polarization obviously drives the electric field:

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \mathbf{E} = \frac{1}{\epsilon_0 c^2} \frac{\partial^2}{\partial t^2} \mathbf{P}. \quad (7.16)$$

### 7.1.2.1 Linear Polarization and Macroscopic Refractive Index

If the dielectric polarization depends linearly on the field intensity according to Eq. (7.13), then the modification of the wave velocity within the dielectric,  $c^2 \rightarrow c^2/\kappa(\omega)$ , can be taken into account using the macroscopic refractive index  $n^2(\omega) = \kappa(\omega)$  (see Eq. (3.1)):

$$\nabla^2 \mathbf{E} - \frac{n^2(\omega)}{c^2} \frac{\partial^2}{\partial t^2} \mathbf{E} = 0. \quad (7.17)$$

According to Eq. (7.13) we have  $\mathcal{E} + \mathcal{P}/\epsilon_0 = [1 + \chi(\omega)]\mathcal{E} = n^2(\omega)\mathcal{E}$  with

$$n^2(\omega) = \kappa(\omega) = 1 + \chi(\omega).$$

Here the relation between the complex index of refraction  $n = n' + i n''$  and the susceptibility  $\chi$  becomes simpler in a significant way, if, for example, in optically thin (dilute) matter such as a gas, the polarization is very low,  $|\chi(\omega)| \ll 1$ :

$$n' \simeq 1 + \chi'/2 \quad \text{and} \quad n'' \simeq \chi''/2$$

<sup>2</sup> The methods of femtosecond spectroscopy developed in the 1990s now also allow us to study such fast relaxation phenomena with excellent time resolution.

or

$$n \simeq 1 + \frac{N}{V} \frac{3\lambda^3}{8\pi^2} \frac{i + 2\delta/\gamma}{1 + (2\delta/\gamma)^2}. \quad (7.18)$$

Thus, by measuring the macroscopic refractive index, the microscopic properties of the dielectric requiring theoretical treatment by quantum mechanics can be determined. Using  $(N/V)3\lambda^3/(8\pi^2) \geq 0.1$ , we can also estimate the density of particles where we ultimately leave the limiting case of optically thin media. For optical wavelengths ( $\lambda \simeq 0.5 \mu\text{m}$ ), this transition occurs already at the relatively low density of  $N/V \approx 10^{14} \text{ cm}^{-3}$ , which at room temperature for an ideal gas corresponds to a pressure of only  $10^{-2} \text{ mbar}$ .

The solution for a planar wave according to Eq. (7.17) is then

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{-i(\omega t - n' \mathbf{k} \cdot \mathbf{r})} e^{-n'' \mathbf{k} \cdot \mathbf{r}}.$$

Propagation not only takes place with a modified phase velocity  $v_{\text{ph}} = c/n'$  but also is exponentially damped according to Beer's law in the  $z$  direction with absorption coefficient  $\alpha = 2n''k_z$ :

$$I(z) = I(0) e^{-2n''k_z z} = I(0) e^{-\alpha z}. \quad (7.19)$$

We have chosen  $n'', \chi'' > 0$  for normal dielectrics according to Eq. (7.18); as we will see, in a "laser medium," one can create  $n'', \chi'' < 0$  as well, realizing amplification of an optical wave.

Let us briefly study the question of whether a single microscopic dipole can generate a refractive index, that is, whether it could cause noticeable absorption or dispersion of an optical wave. For this consideration we again rewrite the absorption coefficient as

$$\alpha = 2n''k = \frac{N}{V} \frac{3\lambda^2}{2\pi} \frac{1}{1 + (2\delta/\gamma)^2} = \frac{N}{V} \frac{\sigma_Q}{1 + (2\delta/\gamma)^2}. \quad (7.20)$$

Therefore, the effect of a single atom is determined by a resonant cross section of

$$\sigma_Q = 3\lambda^2/2\pi \quad \text{at} \quad \delta = 0, \quad (7.21)$$

which is much larger than the atom itself. If we succeed in limiting a single atom to a volume with this wavelength as diameter ( $V \simeq \lambda^3$ ), then a laser beam focused on this volume will experience strong absorption. Such an experiment has in fact been carried out with a stored ion [92]. Dispersion is observed for nonzero detuning only, but for small values  $\delta = \pm\gamma/2$  a single atom is predicted to cause a measurable phase shift  $\delta\Phi = \pm 1/(8\pi)$  as well.

### 7.1.2.2 Absorption and Dispersion in Optically Thin Media

Sometimes it is useful to consider directly the effect of polarization on the amplitude of an electromagnetic wave propagating in a dielectric medium. For this, we take the one-dimensional form of the wave equation (7.16),

$$\left( \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathcal{E}(z) e^{-i(\omega t - kz)} = \frac{1}{\epsilon_0 c^2} \frac{\partial^2}{\partial t^2} \mathcal{P}(z) e^{-i(\omega t - kz)},$$

we fix the frequency  $\omega = ck$  for  $\mathbf{E}(z, t) = \mathcal{E}(z) e^{-i(\omega t - kz)}$ , and additionally we assume that the amplitude changes only slowly (on the scale of a wavelength)

during propagation. Thus

$$\left| \frac{\partial^2 \mathcal{E}(z)}{\partial z^2} \right| \ll k \left| \frac{\partial \mathcal{E}(z)}{\partial z} \right|.$$

Then with  $\partial^2/\partial z^2 [\mathcal{E}(z) e^{ikz}] \simeq e^{ikz} [2ik \partial/\partial z - k^2] \mathcal{E}(z)$ , the wave equation is approximated by

$$\left[ 2ik \frac{\partial}{\partial z} - k^2 + \frac{\omega^2}{c^2} \right] \mathcal{E}(z) = -\frac{\omega^2}{\epsilon_0 c^2} \mathcal{P}(z),$$

which with  $k = \omega/c$  further simplifies to

$$\frac{\partial}{\partial z} \mathcal{E}(z) = \frac{ik}{2\epsilon_0} \mathcal{P}(z). \quad (7.22)$$

Now we consider the electromagnetic wave with a real amplitude and phase,  $\mathcal{E}(z) = \mathcal{A}(z) e^{i\Phi(z)}$ , and calculate

$$\begin{aligned} \mathcal{E}(z) \frac{d\mathcal{E}^*(z)}{dz} &= \mathcal{A} \frac{d\mathcal{A}}{dz} + i\mathcal{A}^2 \frac{d\Phi}{dz} = \frac{-ik}{2\epsilon_0} \mathcal{P}^*(z) \mathcal{E}(z), \\ \mathcal{E}^*(z) \frac{d\mathcal{E}(z)}{dz} &= \mathcal{A} \frac{d\mathcal{A}}{dz} - i\mathcal{A}^2 \frac{d\Phi}{dz} = \frac{ik}{2\epsilon_0} \mathcal{P}(z) \mathcal{E}^*(z). \end{aligned}$$

From this we can determine the change of the intensity  $I(z) = \frac{1}{2} c \epsilon_0 \mathcal{A}^2$  of an electromagnetic wave while propagating within a polarized medium according to

$$\frac{d}{dz} I(z) = \frac{\omega}{2} \Im \{ \mathcal{E}(z) \mathcal{P}^*(z) \}$$

and the phase shift according to

$$\frac{d}{dz} \Phi(z) = \frac{\omega}{2I(z)} \Re \{ \mathcal{E}(z) \mathcal{P}^*(z) \}.$$

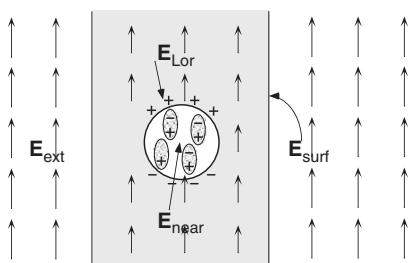
The absorption coefficient  $\alpha$  and the real part of the refractive index  $n'$  can be calculated in an obvious way from

$$\begin{aligned} \alpha &= \frac{1}{I(z)} \frac{dI(z)}{dz} = \frac{\omega}{2I(z)} \Im \{ \mathcal{E}(z) \mathcal{P}^*(z) \}, \\ n' - 1 &= \frac{1}{k} \frac{d\Phi(z)}{dz} = \frac{c}{2I(z)} \Re \{ \mathcal{E}(z) \mathcal{P}^*(z) \}. \end{aligned} \quad (7.23)$$

We naturally reproduce the results from the section on the linear refractive index if we assume the linear relation according to Eq. (7.13). The form developed here also allows us to investigate nonlinear relations and will be useful in the chapter on nonlinear optics (Chapter 15).

### 7.1.2.3 Dense Dielectric Media and Near Fields

Certainly, in a dilute, optically thin medium, we do not make a big mistake by neglecting the field additionally generated in the sample by polarization. But this is no longer the case in the liquid or solid states. In order to determine the “local field” of the sample, we cut out a fictitious sphere with a diameter  $d_{\text{atom}} \ll d_{\text{sph}} \ll \lambda$  with “frozen” polarization from the material (Figure 7.4).



**Figure 7.4** Contributions to the local electric field in an optically dense medium. For a transverse wave the contribution of the surface vanishes in the case of normal incidence.

To determine the microscopic local field  $\mathbf{E}_{\text{loc}}$  at the position of a particle, we decompose it into various contributions,  $\mathbf{E}_{\text{loc}} = \mathbf{E}_{\text{ext}} + \mathbf{E}_{\text{surf}} + \mathbf{E}_{\text{Lor}} + \mathbf{E}_{\text{near}}$ , which depend on the different geometries and structures of the sample and are in total called the “depolarizing field” since they usually weaken the external field  $\mathbf{E}_{\text{ext}}$ :

$\mathbf{E}_{\text{surf}}$ : The field of the surface charges generated by the surface charge density  $\rho_{\text{surf}} = \mathbf{n} \cdot \mathbf{P}(\mathbf{r}_{\text{surf}})$ . It vanishes for a wave at normal incidence.

$\mathbf{E}_{\text{Lor}}$ : The field of the surface of a fictitious hollow sphere cut out from the volume (also known as the *Lorentz field*). For homogeneous polarization, one finds  $\mathbf{E}_{\text{Lor}} = \mathbf{P}/3\epsilon_0$ .

$\mathbf{E}_{\text{near}}$ : The field of the electric charges within the sphere. In the case of isotropic media, this contribution vanishes,  $\mathbf{E}_{\text{near}} = 0$ .

From  $\mathbf{P} = \epsilon_0\chi\mathbf{E}_{\text{loc}} = \epsilon_0\chi(\mathbf{E} + \mathbf{P}/\epsilon_0)$ , we then obtain by insertion of  $\mathbf{E}_{\text{loc}} = \mathbf{E}_{\text{Lor}} = \mathbf{P}/3$  the macroscopic volume susceptibility  $\chi^V$  of an isotropic and linear but dense material:

$$\chi_{ij}^V(\omega) = \frac{1}{\epsilon_0} \frac{P_i}{E_j} = \frac{\chi}{1 - \chi/3}.$$

From this by rearrangement, the *Clausius–Mossotti equation*, can be obtained, which describes the influence of the depolarizing field on the refractive index (density  $\mathcal{N} = N/V$ ):

$$3 \frac{n^2 - 1}{n^2 + 2} = \chi = \frac{\mathcal{N}q^2}{\epsilon_0 m} \frac{1}{(\omega_0^2 - \omega^2) - i\omega\gamma}. \quad (7.24)$$

For small polarizations,  $\chi/3 \ll 1$ , Eq. (7.24) again turns into Eq. (7.18).

Realistic polarizable substances, though, do not have just one degree of freedom like the Lorentz oscillator described here but lots of them. We can extend the Lorentz model for a not too strong field by linearly superimposing many oscillators with different resonance frequencies  $\omega_k$  and damping rates  $\gamma_k$  and weighting them with their relative contribution, their “oscillator strength”  $f_k$ :

$$3 \frac{n^2 - 1}{n^2 + 2} = \frac{\mathcal{N}q^2}{\epsilon_0 m} \sum_k \frac{f_k}{(\omega_k^2 - \omega^2) - i\omega\gamma_k}.$$

Even if the field intensity becomes quite large, we can still use the concepts described here if we introduce a nonlinear susceptibility. This case is dealt with in Chapter 15 on nonlinear optics.

The dimensionless oscillator strength allows a simple transition to the quantum mechanically correct description of the microscopic polarization [93]. For this,

only the matrix element of the dipole transition between the ground state  $|\phi_g\rangle$  and excited states  $|\phi_k\rangle$  of the system,  $q\mathbf{r}_{kg} = q\langle\phi_k|\mathbf{r}|\phi_g\rangle$ , has to be used:

$$f_{kg} = \frac{2m\omega_{kg}|\mathbf{r}_{kg}|^2}{\hbar}.$$

We do not need to require anything specific about the nature of these states. They can be atomic or molecular excitations but also, for example, optical phonons or polaritons within solid states. Strictly speaking, the success of the classical Lorentz model for single atoms is justified by this relation. In atoms the oscillator strengths follow the Thomas–Reiche–Kuhn sum rule  $\sum_k f_{kg} = 1$ ; already for low atomic resonance lines, for example, the well-known doublets of the alkali spectra, we have  $f \sim 1$ ; therefore the other resonance lines have to be significantly weaker.

## 7.2 Two-Level Atoms

### 7.2.1 Are There Any Atoms with Only Two Levels?

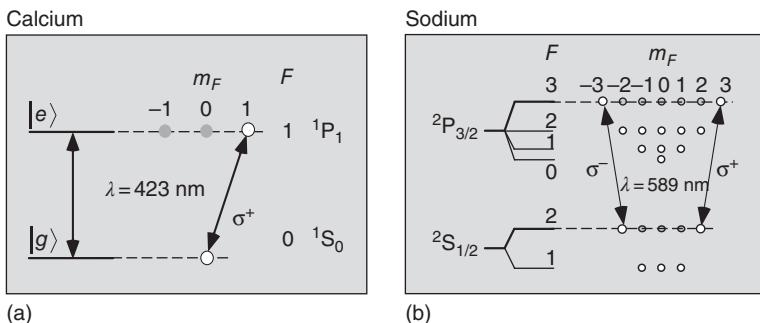
In quantum mechanics we describe atoms by their states. In the simplest case a light field couples a ground state  $|g\rangle$  to an excited state  $|e\rangle$ . This model system can be theoretically dealt with well and is particularly useful for understanding the interaction of light and matter. However, even simple atoms such as the alkali and alkaline earth atoms, which are technically easy to master and widely used for experimental investigations, present a complex structure with a large number of states even in the ground level.<sup>3</sup>

Occasionally, though, it is possible to prepare atoms in such a way that no more than two states are effectively coupled to the light field. The calcium atom, for instance, has a nondegenerate singlet ground state ( ${}^1S_0, \ell = 0, m = 0$ ). By using a light field with a wavelength of 423 nm and proper choice of the polarization ( $\sigma^\pm, \pi$ ), three different two-level systems can be prepared by coupling to the ( ${}^1P_1, \ell = 1, m = 0, \pm 1$ ) states (Figure 7.5).

The famous yellow doublet of the sodium atom ( $\lambda = 589$  nm) is another example that has played a central role in experimental investigations, though it has large total angular momenta  $F = 1, 2$  even in the  ${}^2S_{1/2}$  ground-state doublet due to its nuclear spin of  $I = 3/2$  and presents a wealth of magnetic substructure. By so-called optical pumping [90] with  $\sigma^+$ -polarized light, all the atoms in a gas can be prepared in, for example, the state with quantum numbers  $F = 2, m_F = 2$ . This state is then coupled only to the  $F' = 3, m_{F'} = 3$  substrate of the excited  ${}^2P_{3/2}$  state by the light field.<sup>4</sup>

<sup>3</sup> The wealth of structure is generated by the coupling of the magnetic orbital and spin momenta of electrons and core. For low states the splittings are about 100–1000 MHz. Details can be found in textbooks about quantum mechanics [94] or atomic physics [95].

<sup>4</sup> In reality, the circular polarization is never perfect. Small admixtures of  $\sigma^-$  light to the  $\sigma^+$  light cause, for example, occasional excitations with  $\Delta m_F = -1$  and therefore limit the “quality” of the two-level atom.



**Figure 7.5** Abstract and realistic two-level atoms. (a) Calcium atom. A  $\sigma^+$ -polarized light field couples only states with angular momentum quantum numbers  $|g\rangle = |F, m_F\rangle = |0, 0\rangle$  and  $|e\rangle = |1, 1\rangle$ . (b) Sodium atom. A circularly polarized light field ( $\sigma^\pm$ ) is “pumping” the sodium atoms to the outer  $|F, m_F\rangle = |2, \pm 2\rangle$  states, which with  $\sigma^\pm$  light are coupled only to the  $|3, \pm 3\rangle$  states.

These effective “two-level atoms,” the list of which can easily be extended, play an enormously important role in physical experiments since they provide the simplest models of a polarizable physical system and radiative interaction is reduced to its most fundamental case.

### 7.2.2 Dipole Interaction

The “free” two-level atom with total mass  $M$  is now reduced to a Hamiltonian  $H_{\text{at}}$  having only a ground state  $|g\rangle$  and an excited state  $|e\rangle$ .<sup>5</sup> To complete the picture, we allow for an arbitrary center-of-mass energy  $E_0 = P^2/2M$ . Thus

$$H_{\text{at}} = \frac{P^2}{2M} + \frac{\hbar\omega_0}{2}(|e\rangle\langle e| - |g\rangle\langle g|). \quad (7.25)$$

The energy of the atom is  $E_e = \langle e|H_{\text{at}}|e\rangle = E_0 + \hbar\omega_0/2$  in the excited state and  $E_g = E_0 - \hbar\omega_0/2$  in the ground state. The resonance frequency presents the energy separation of the two states,  $\omega_0 = (E_e - E_g)/\hbar$ .

The dipole operator  $\hat{V}_{\text{dip}}$  is obtained by an analogy with classical electrodynamics, that is, by converting the classical energy of a dipole subject to an electric field into an operator. For the electron position operator  $\hat{\mathbf{r}}$ , we obtain<sup>6</sup>

$$\hat{V}_{\text{dip}} = -q\hat{\mathbf{r}}\mathbf{E}.$$

In a realistic experiment, we always have to take the exact geometric orientation of atom and electric field into account. For the consideration of the two-level

<sup>5</sup> We assume that the reader is familiar with the basic principles of quantum mechanics. Quantum states are given in Dirac notation, where state vectors  $|i\rangle$  are associated with complex wave functions  $\psi_i(\mathbf{r})$ . The expectation values of an operator  $\hat{\mathcal{O}}$  are thus calculated from

$\langle \hat{\mathcal{O}} \rangle = \langle f | \hat{\mathcal{O}} | i \rangle = \int_V dV \psi_f^*(\mathbf{r}) \hat{\mathcal{O}} \psi_i(\mathbf{r})$ .

<sup>6</sup> Rigorous analysis according to quantum mechanics results in the product of electron momentum and electromagnetic vector potential  $\hat{\mathbf{p}}\mathbf{A}$ , but it can be shown that in the vicinity of resonance frequencies,  $\hat{\mathbf{r}}\mathbf{E}$  leads to the same result [96].

atom, however, we neglect this geometric influence and restrict the problem to one dipole coordinate  $\hat{d} = q\hat{r}$  only:

$$\hat{V}_{\text{dip}} = -\hat{d}\mathcal{E}_0 \cos \omega t.$$

Using the completeness theorem of quantum mechanics, we can project the position operator onto the states involved ( $\langle i|\hat{d}|i\rangle = 0$ ):

$$\hat{d} = |e\rangle\langle e|\hat{d}|g\rangle\langle g| + |g\rangle\langle g|\hat{d}|e\rangle\langle e|.$$

We use the matrix element  $d_{eg} = \langle e|\hat{d}|g\rangle$  of the dipole operator. Using the definition of atomic raising and lowering operators,  $\sigma^\dagger = |e\rangle\langle g|$  and  $\sigma = |g\rangle\langle e|$ , we write

$$\hat{d} = d_{eg}\sigma^\dagger + d_{eg}^*\sigma. \quad (7.26)$$

With those operators we can already express the atomic Hamiltonian and the dipole operator very compactly:

$$\begin{aligned} H_{\text{at}} &= \frac{P^2}{2M} + \hbar\omega_0 \left( \sigma^\dagger\sigma - \frac{1}{2} \right), \\ \hat{V}_{\text{dip}} &= -(d_{eg}\sigma^\dagger + d_{eg}^*\sigma)\mathcal{E}_0 \cos \omega t. \end{aligned} \quad (7.27)$$

From linear combinations of the atomic field operators, Pauli operators can be generated, which are known to describe a spin-1/2 system with only two states:

$$\begin{aligned} \sigma_x &= \sigma^\dagger + \sigma, \\ \sigma_y &= -i(\sigma^\dagger - \sigma), \\ \sigma_z &= \sigma^\dagger\sigma - \sigma\sigma^\dagger = [\sigma^\dagger, \sigma]. \end{aligned}$$

We will see that we can interpret the expectation values of  $\sigma_x$  and  $\sigma_y$  as components of the atomic polarization and  $\sigma_z$  as the difference of occupation numbers or "inversion." With  $\sigma^\dagger\sigma - 1/2 = (\sigma_z + 1)/2$ , we find that the Hamiltonian operator from Eq. (7.27) is equivalent to a spin-1/2 system subject to a homogeneous magnetic field:

$$H_{\text{at}} = \frac{P^2}{2M} \frac{\hbar\omega_0}{2} (\sigma_z + 1). \quad (7.28)$$

In general any two-level atom or any two-level quantum system can be described in terms of a *pseudo-spin system*, exhibiting fully analogous dynamics.

The Pauli operators have the form

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

in the matrix representation and follow the generally useful relation  $[\sigma_i, \sigma_j] = 2i\sigma_k$  on cyclic permutation of the coordinates  $x, y, z$ . In addition we have

$$\begin{aligned} \sigma^\dagger &= \frac{1}{2}(\sigma_x + i\sigma_y), \\ \sigma &= \frac{1}{2}(\sigma_x - i\sigma_y). \end{aligned}$$

The operators' equation of motion is obtained from the Heisenberg equation

$$\dot{\sigma}_i = \frac{\partial}{\partial t} \sigma_i + \frac{i}{\hbar} [H, \sigma_i].$$

For this, the Hamiltonian is usefully written in the form

$$H = \frac{1}{2}\hbar\omega_0\sigma_z - \frac{1}{2}(d_{eg} + d_{eg}^*)\mathcal{E}_0 \cos(\omega t)\sigma_x - \frac{1}{2}i(d_{eg} - d_{eg}^*)\mathcal{E}_0 \cos(\omega t)\sigma_y.$$

Often real values can be chosen for  $d_{eg}$ . Then the third ( $\sigma_y$ ) term is omitted, and it can simply be written as

$$H = \frac{1}{2}\hbar\omega_0\sigma_z - d_{eg}\mathcal{E}_0 \cos(\omega t)\sigma_x.$$

If the operators are not explicitly time dependent, the result is an equation system known as *Mathieu's differential equations*:

$$\begin{aligned}\dot{\sigma}_x &= -\omega_0\sigma_y, \\ \dot{\sigma}_y &= \omega_0\sigma_x - \frac{2d_{eg}\mathcal{E}_0}{\hbar} \cos(\omega t)\sigma_z, \\ \dot{\sigma}_z &= \frac{2d_{eg}\mathcal{E}_0}{\hbar} \cos(\omega t)\sigma_y.\end{aligned}\tag{7.29}$$

It can easily be shown that only the orientation, but not the magnitude, of the angular momentum is changed under the effect of the light field; we have  $\sigma_x^2 + \sigma_y^2 + \sigma_z^2 = 1$  as for the Pauli matrices.

### 7.2.3 Optical Bloch Equations

Until now we have considered the development of atomic operators under the influence of a light field. For the semiclassical consideration, we can replace them by expectation values<sup>7</sup>  $S_i = \langle \sigma_i \rangle$  and again obtain the equation system (7.29), only now for classical variables [97]. To produce transparent solutions, it is advantageous to consider the evolution of variables in a new coordinate system rotating with the light frequency  $\omega$  around the  $z$  axis, that is, with the polarization,

$$\begin{aligned}S_x &= u \cos \omega t - v \sin \omega t, \\ S_y &= u \sin \omega t + v \cos \omega t, \\ S_z &= w.\end{aligned}$$

This often used approximation is called the “rotating wave approximation” (RWA). The variables  $(u, v)$  describe the sine and cosine components of the induced electric dipole moment, and  $w$  is the difference in occupation numbers. The close relation of these variables with the classical Lorentz model in Section 7.1.1 and their physical interpretation will be explained in more detail in Section 7.2.7.

With detuning  $\delta = \omega - \omega_0$ , we obtain after some algebra

$$\begin{aligned}\dot{u} &= \delta v - \frac{d_{eg}\mathcal{E}_0}{\hbar} \sin(2\omega t)w, \\ \dot{v} &= -\delta u - \frac{d_{eg}\mathcal{E}_0}{\hbar} [1 + \cos(2\omega t)]w, \\ \dot{w} &= \frac{d_{eg}\mathcal{E}_0}{\hbar} \sin(2\omega t)u + \frac{d_{eg}\mathcal{E}_0}{\hbar} [1 + \cos(2\omega t)]v.\end{aligned}$$

---

<sup>7</sup> There are no operator products that could cause typically quantum mechanical signatures due to noncommutativity.

For typical optical processes the contributions oscillating very rapidly with  $2\omega t$  play only a small role (they cause the so-called Bloch–Siegert shift) and are therefore neglected. We introduce the *Rabi frequency*

$$\Omega_R = |d_{eg} \mathcal{E}_0 / \hbar| \quad (7.30)$$

and get the undamped optical Bloch equations

$$\begin{aligned} \dot{u} &= \delta v, \\ \dot{v} &= -\delta u + \Omega_R w, \\ \dot{w} &= -\Omega_R v, \end{aligned} \quad (7.31)$$

originally found for magnetic resonance by F. Bloch (1905–1983, Nobel Prize 1952) in order to describe there the interaction of a magnetic moment with spin-1/2 in a strong homogeneous magnetic field exposed to a high-frequency field.

The system of equations (7.31) can also be written in a shorter way by introducing the Bloch vector  $\mathbf{u} = (u, v, w)$  and  $\boldsymbol{\Omega} = (\Omega_R, 0, -\delta)$ :

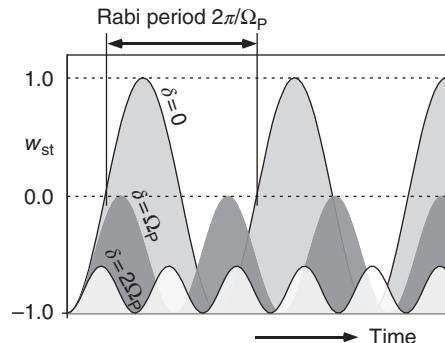
$$\dot{\mathbf{u}} = \boldsymbol{\Omega} \times \mathbf{u}. \quad (7.32)$$

It describes the behavior of a magnetic dipole transition, for example, between the hyperfine states of an atom, in an excellent approximation. In Section 7.2.4 we explore in more detail the situation of perfect resonance, that is, at zero detuning  $\delta = 0$ . There, the occupation number difference, or  $w$  component of the Bloch vector, oscillates sinusoidally with Rabi frequency  $\Omega_R$  and maximum amplitude. If the detuning  $\delta$  does not vanish, there is still a sinusoidal oscillation with generalized Rabi frequency

$$\Omega = \sqrt{\delta^2 + \Omega_R^2}. \quad (7.33)$$

In Figure 7.6 occupation number oscillations are shown for initial conditions  $(u, v, w)(t = 0) = (0, 0, -1)$  and  $\delta = 0, \Omega_R, 2\Omega_R$ . With increasing detuning the generalized Rabi frequency increases, while the amplitude of the occupation number oscillations decreases. Full modulation is obtained only at perfect resonance.

Since optical two-level systems obey an identical set of equations, almost all the concepts of coherent optics are borrowed from electron and nuclear spin resonance. In contrast to magnetic systems, spontaneous emission (see Sections 7.3.2 and 14.3.1) plays a dominant role for optical processes, however. Hence the optical Bloch equations (7.31) need to be extended by introducing damping rates (Section 7.2.6).



**Figure 7.6** Solutions for Eq. (7.31) with initial condition  $(u, v, w)(t = 0) = (0, 0, -1)$ : occupation number oscillation for different detunings  $\delta$ . See also Figure 7.8.

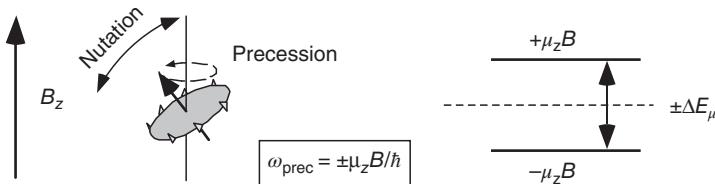
### 7.2.4 Pseudo-spin, Precession, and Rabi Nutation

Every quantum mechanical two-level system can be understood in analogy with the magnetic spin-1/2 system where we only have the “up” and “down” state; it is thus a pseudo-spin-1/2 system. In a classical vector model for the spin and its magnetic moment, a gyromagnet, the frequency corresponding to the energy splitting caused by the magnetic dipole energy in an external magnetic field is also associated with the precession frequency of the dipoles around the external field axis (Figure 7.7). Quantum transitions between the two states are equivalent to additional forces (a transverse radiofrequency  $B$  field in magnetic resonance experiments), resulting in nutation of the spin vector in the classical analogue (Table 7.2).

Let us now consider special solutions of the optical Bloch equations (7.31) to analyze the dynamics of the two-level system. The resonance case is particularly easy to determine: here the detuning vanishes,  $\delta = 0$ . The occupation number  $w$  and the  $v$  component of the polarization perform an oscillation with the *Rabi frequency* according to Eq. (7.31). At thermal ambient energies, an atom with an optical excitation frequency usually resides in its ground state, and therefore we normally have  $w(t=0) = -1$  and

$$\begin{aligned} v(t) &= -\sin(\Omega_R t), \\ w(t) &= -\cos(\Omega_R t). \end{aligned}$$

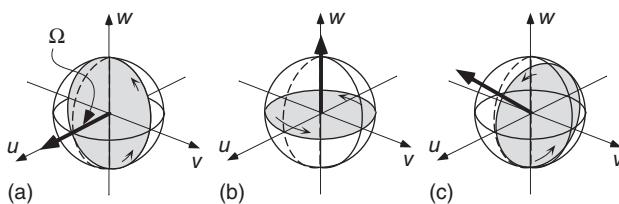
The dynamical Bloch vector evolution for resonant excitation of a system with initial value  $w(t=0) = -1$  is shown in Figure 7.8a. Rotations around the  $w$  axis should be called *precession*, around any other axis (Rabi) *nutation*, in analogy with classical gyroscopes. For  $\Omega_R t = \pi$ , the atomic system rotates from  $w = -1$  into the completely “inverted” state with  $w = 1$ ! An electric field pulse driving a transition with  $\Omega_R t = \pi$  is therefore called a  $\pi$ -pulse;  $\pi/2$  pulses rotate the vector to the



**Figure 7.7** Precession and nutation of a magnetic dipole moment associated with a gyro. The pseudo-spin of the two-level atom performs an analogous motion in which the longitudinal  $z$  direction is associated with the difference in occupation numbers  $w$ , while the transverse directions are identified with the polarization components  $u$  and  $v$ .

**Table 7.2** Comparison of the spin-1/2 system and Bloch vector components.

	Spin-1/2	Bloch vector
Transverse components	$x, y$	$u, v$
Longitudinal components	$z$	$w$
		Polarization
		Occupation number difference



**Figure 7.8** Dynamics of Bloch vectors: (a)  $\Omega = (-\Omega_R, 0, 0)$ , (b)  $\Omega = (0, 0, -\delta)$ , and (c)  $\Omega = (-\Omega_R, 0, -\delta)$ . The torque vector  $\Omega = (\Omega_R, 0, -\delta)$  defines the normal vector to the plane in which the Bloch vector rotates.

equatorial plane. They play an important role for the generation of quantum-state superpositions, for example, in Ramsey spectroscopy (see Problem 12.2).

### 7.2.5 Microscopic Dipoles and Ensembles

Extending the classical Lorentz model of Section 7.1.1, we have to find again the transition from the microscopic model to macroscopically measurable physical quantities. From quantum mechanics it is known that the calculation of expectation values predicts the distribution of measured values for an ensemble. Thus the  $w$  component of the Bloch vector is measured from the number of atoms in the excited state  $N_e$  and in the ground state  $N_g$ :

$$w = \frac{N_e - N_g}{N_e + N_g} = \frac{\Delta N}{N}. \quad (7.34)$$

The  $w$  component equals the normalized occupation number difference, and more strictly speaking inversion occurs for  $w > 1$  only, when more particles are in the excited than in the ground state. More loosely  $w$  and  $\Delta N$  are called inversion irrespective of their values.

In a physical system at thermal equilibrium, the occupation number  $n_{\text{th}}$  decreases with increasing energy of the quantum state,  $n_{\text{th}} e^{-E/kT}$ . In this case the occupation number difference  $N_e - N_g$  and hence  $w$  are always negative. Optical transitions are furthermore very energetic in comparison with thermal energies (2 eV vs 1/40 eV); thus the thermal equilibrium value of an unperturbed sample is typically  $w = -1$ .

For the determination of the macroscopic polarization density of a sample, we use the particle density  $N/V$ . For a system of identical particles, we obtain, in close analogy to the classical case,

$$P = \frac{N}{V} d_{eg}(u + iv). \quad (7.35)$$

In contrast to Eq. (7.12)  $d_{eg}$  is now the quantum mechanical transition dipole moment, and  $(u, v)$  depend in a nonlinear way on the intensity of the driving field.

### 7.2.6 Optical Bloch Equations with Damping

We have learned that the motion is undamped according to (7.31), whereas an optical atomic excitation is damped by multiple processes. Among them there is

not only the radiative decay but also collisions and other phenomena. For now we introduce phenomenological relaxation rates, which are justified more deeply in Section 14.3.1.

The *longitudinal relaxation rate*  $\gamma = 1/T_1$  describes the energy loss of the two-level system characterized by the difference in occupation numbers and the  $w$  coordinate of the Bloch vector, respectively. In equilibrium without a driving light field, the stationary thermal value of the inversion  $w_0 = -1$  must be reproduced.

The *transverse relaxation rate*  $\gamma' = 1/T_2$  describes the damping of polarization, that is, of the  $u$  and  $v$  components of the Bloch vector. In an ensemble the macroscopic polarization can also get lost because each particle precesses with a different speed, and so the particles lose their original phase relation (in the precession angle). For pure radiation damping, we have  $T_2 = 2T_1$ . The polarization vanishes as well, when the light field is switched off.

The complete set of *optical Bloch equations* including damping is

$$\begin{aligned}\dot{u} &= \delta v - \gamma' u, \\ \dot{v} &= -\delta u - \gamma' v + \Omega_R w, \\ \dot{w} &= -\Omega_R v - \gamma(w - w_0).\end{aligned}\tag{7.36}$$

Its similarity to the classical equations (7.8) and (7.9), a result of the Lorentz model, cannot be overlooked anymore. Apparently the ratio of the Rabi frequency  $\Omega_R$  to the damping rates  $\gamma$  and  $\gamma'$  determines the dynamics of the system. We expect oscillatory properties as in the undamped system only when

$$\Omega_R \gg \gamma, \gamma'.\tag{7.37}$$

This limiting case where the driving forces are much stronger than the damping forces is called *strong coupling* and is of great importance in the interaction of matter with strong laser fields (see Section 13.1).

Often the optical Bloch equations are written in the more compact complex notation using the language of the density matrix theory from quantum mechanics (see Section B.2). With  $\rho_{eg} = u + iv$  and  $w = \rho_{ee} - \rho_{gg}$ , we find

$$\begin{aligned}\dot{\rho}_{eg} &= -(\gamma' + i\delta)\rho_{eg} + i\Omega_R w, \\ \dot{w} &= -\Im\{\rho_{eg}\}\Omega_R - \gamma(w - w_0).\end{aligned}\tag{7.38}$$

### 7.2.7 Steady-State Inversion and Polarization

#### 7.2.7.1 Steady-State Inversion and Saturation Intensity

We consider the situation when transients have settled, that is, a time  $t \gg T_1, T_2$  has passed since switching on the light field or  $\dot{u} = 0$  in (7.36). The stationary solution  $w_{st}$  is related to the inversion equation (7.34),  $w_0 = \Delta N_0/N$ , without a driving light field:

$$w_{st} = \frac{w_0}{1 + \frac{\Omega_R^2}{\gamma\gamma'} \frac{1}{1 + (\delta/\gamma')^2}} = \frac{w_0}{1 + s} = \frac{\Delta N_0}{N} \frac{1}{1 + s}.\tag{7.39}$$

In a light field with intensity  $I$ , the “saturation parameter”

$$s = \frac{s_0}{1 + (\delta/\gamma')^2} \quad \text{with} \quad s_0 = \frac{I}{I_0} = \frac{\Omega_R^2}{\gamma\gamma'} \quad (7.40)$$

determines the significance of coherent processes with dynamics determined by the Rabi frequency and in comparison with incoherent damping processes determined by the relaxation rates  $\gamma$  and  $\gamma'$ . Owing to  $\Omega_R^2 = |-d_{eg}E_0/\hbar|^2 = |-d_{eg}/\hbar|^2(2I/c\epsilon_0)$ , the saturation intensity  $I_0$  can be calculated as

$$I_0 = \frac{c\epsilon_0}{2} \frac{\hbar^2\gamma\gamma'}{d_{eg}^2}. \quad (7.41)$$

From Figure 7.9 it is clear that the saturation intensity sets the typical scale for the onset of coherent processes.

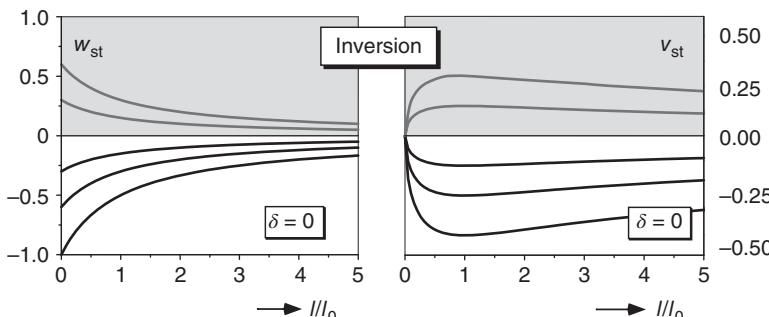
Using the known result for spontaneous emission  $\gamma = d_{eg}^2\omega^3/3\pi\hbar\epsilon_0c^3$  (see Eq. (7.49)), the saturation intensity can be determined just by knowing the resonance wavelength  $\lambda$  and the transverse relaxation rate  $\gamma'$ . We get a useful correlation with the resonant cross section of the absorption  $\sigma_Q$  from Eq. (7.21),

$$I_0 = \frac{2\pi\hbar c\gamma'}{3\lambda^3} = \frac{\hbar\omega\gamma'}{\sigma_Q}, \quad (7.42)$$

which can be interpreted in the following way. Apparently, at the saturation intensity, the energy of just one photon flows through the resonant absorption cross section  $\sigma_Q$  during the transverse coherence time  $T' = 1/\gamma'$ .

If only radiative decay is possible, as, for example, in dilute gases or atomic beams, then the saturation intensity with  $\gamma' = \gamma/2$  depends only on the properties of the free atom and is given by

$$I_0 = \frac{\pi\hbar c\gamma}{3\lambda^3}.$$



**Figure 7.9** Effect of a light field on the equilibrium values of the difference in occupation numbers (“inversion”)  $w_{st}$  and the polarization components  $u_{st}$  and  $v_{st}$  as functions of the saturation parameter  $s(\delta = 0) = I/I_0$  according to Eqs. (7.39) and (7.44). With the light field vanishing, the values  $w_0 = -1, -0.6, -0.3, +0.3$ , and  $+0.6$  have been used.

**Table 7.3** Saturation intensity of some important atomic resonance lines.

Atom	H	Na	Rb	Cs	Ag	Ca	Yb
Transition	1S→2P	3S→3P	5S→5P	6S→6P	5S→5P	4S→4P	6S→6P
$\gamma/2\pi$ (10 <sup>6</sup> s <sup>-1</sup> )	99.5	9.9	5.9	5.0	20.7	35.7	0.18
$\lambda$ (nm)	121.6	589.0	780.2	852.3	328.0	422.6	555.8
$I_0$ (mW cm <sup>-2</sup> )	7242	6.34	1.63	1.06	76.8	61.9	0.14

As an example, we present the saturation intensity for several important atoms. They can be realized technically using continuous-wave laser light sources without special effort except for the case of the hydrogen atom. The “strength” of the transition is characterized by the decay rate  $\gamma/2\pi = \Delta_{1/2}$ , given in Table 7.3 in units of the natural linewidth. It is perhaps strange at first glance that the saturation intensity becomes smaller with decreasing linewidth and therefore weaker lines, but it has to be taken into account that coherent coupling needs more and more time to reach excitation.

According to Eq. (7.39) increasing driving field intensities reduce the occupation number difference  $\Delta N$  and inversion  $w_{st}$ , which can be expressed conveniently by its dependence on the saturation intensity:

$$w_{st} = w_0 \frac{\delta^2 + \gamma'^2}{\delta^2 + \gamma'^2(1 + I/I_0)}.$$

The evolution of  $w_{st}(I/I_0)$  is shown in Figure 7.9 for  $\delta = 0$  and different unperturbed or *unsaturated* values of the inversion. Note that the sign of  $w_{st}$  depends on the sign of the unsaturated inversion only!

### 7.2.7.2 Steady-State Polarization

The simplest experimental quantities of light–matter interaction are the index of refraction  $n$  and absorption coefficient  $\alpha$ , which by means of Eq. (7.20) and using Eqs. (7.23) and (7.35) are related to the polarization or  $v$  component of the Bloch vector:

$$\alpha = 2n''k = -\frac{2\pi}{\lambda} \frac{N}{V} \frac{v}{\epsilon_0 \mathcal{E}} d_{eg}. \quad (7.43)$$

For a more detailed exploration of the properties of the equilibrium absorption coefficient in Sections 7.4.3 and 12.2.1, we must therefore evaluate the stationary ( $u, v$ ) values and find

$$v_{st} = \frac{w_0}{1+s} \frac{\gamma' \Omega_R}{\delta^2 + \gamma'^2}, \quad (7.44)$$

$$u_{st} = \frac{\delta}{\gamma'} v_{st}.$$

Sometimes it is technically more convenient to express the Rabi frequency according to (7.40) again by the saturation intensity and  $w$  by the occupation

numbers (7.34). For  $u_{\text{st}}$  and  $v_{\text{st}}$  we then obtain the more transparent form

$$\begin{aligned} v_{\text{st}} &= \frac{\Delta N_0 / N}{1 + I/I_0 + (\delta/\gamma')^2} \sqrt{\frac{I}{I_0}} \frac{(\gamma/\gamma')^{1/2}}{1 + (\delta/\gamma')^2} \\ &\rightarrow \frac{\Delta N_0 / N}{1 + I/I_0} \sqrt{\frac{I}{I_0}} \left( \frac{\gamma}{\gamma'} \right)^{1/2} \quad \text{for } \delta \rightarrow 0. \end{aligned} \quad (7.45)$$

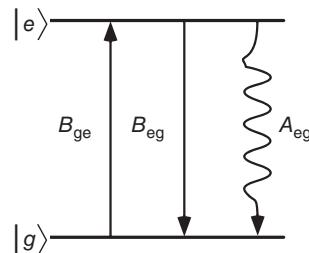
The intensity dependence of the polarization is presented in Figure 7.9 as a function of the normalized intensity  $I/I_0$  in the special case of perfect resonance at  $\delta = 0$ . It increases very rapidly and decreases at high intensities as  $1/\sqrt{I/I_0}$ , that is, with the amplitude of the driving field. For low intensities we find again the limiting classical case. Then  $(u_{\text{st}}, v_{\text{st}})$  correspond with the  $(u, v)$  coordinates of the Lorentz oscillator from Eq. (7.8) and of course exhibit as well the frequency characteristic from Figure 7.2.

The treatment of an ensemble of two-level atoms according to quantum mechanics thus predicts that as in the classical case a light field will always be absorbed for arbitrary intensities because according to (7.45)  $v_{\text{st}} \propto \Delta N_0 < 0$ : the polarization then lags the field and always causes absorption. If, however, an initial inversion  $w > 0$  can be realized, we expect a change of sign for  $v_{\text{st}}$  (Eq. (7.44)) and as a consequence also for  $\alpha$  (Eq. (7.43)). A light field traveling in such a medium will no longer be attenuated but amplified – the condition for the laser.

### 7.3 Stimulated and Spontaneous Radiation Processes

In the previous section we investigated the coupling of an atom to a monochromatic light wave. Three different radiation processes were identified (Figure 7.10):

- 1) By coupling to the driving field, an atom can be promoted from the ground state to the excited state. This process is called “stimulated absorption” and can only take place if there is an applied external field.
- 2) An analogous process takes place as well from the excited to the ground state and is called “stimulated emission.” The stimulated processes describe the coherent evolution of the atom–field system, that is, phase relations play an important role.
- 3) If an atom is in the excited state, it can decay to the ground state by “spontaneous emission.” This process is incoherent, always takes place (apart from the exceptions in the so-called cavity QED; Section 14.3.3), and has been taken into account phenomenologically in Eq. (7.36) by introducing the damping constants.



**Figure 7.10** Two-level quantum system with Einstein coefficients, stimulated and spontaneous radiation processes.

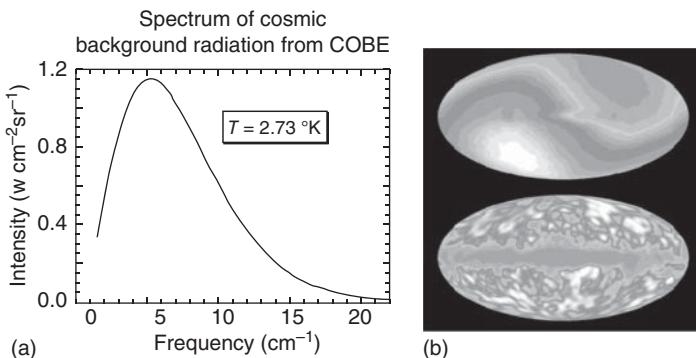
### Excuse: The spectrum of blackbodies

Just before the end of the nineteenth century, the spectrum of blackbodies was very carefully studied at the Physikalisch-Technische Reichsanstalt, Berlin (the historic German National Laboratory for Standards and Technology). At that time, light bulbs for public lighting had only very recently been introduced, and the intention was to control their output and increase their efficiency. This blackbody spectrum has since played an outstanding role in modern physics in general and our understanding of light sources in particular. During these investigations, it turned out that the formula given by Wien for low frequencies,  $S_E(\omega) \propto \omega^3 \exp(-\hbar\omega/kT)$ , no longer matched the experimental results. At the same time, in England, Lord Rayleigh gave a different, more appropriate radiation formula for low frequencies,  $S'_E(\omega) \propto \omega^2 T$ .

Max Planck arrived at his famous radiation formula by a clever interpolation, here in the modern notation of the spectral density  $S_E(\omega)$  of the blackbody radiation field,

$$S_E(\omega) = \frac{8\pi}{c^3} \frac{\hbar\omega^3}{\exp(\hbar\omega/kT) - 1}. \quad (7.46)$$

Today we know that this formula is derived from the product of the density of states of the radiation field at frequency  $\omega$  and the occupation probability according to Bose-Einstein statistics. This formula, published for the first time by Planck in Berlin on December 14, 1900, was the beginning of a sequence of ideas leading to modern physics. Thermal light sources, the concepts of optics, and a problem of truly applied research – the efficiency of light bulbs – all played an important role in the birth of quantum physics!



**Figure 7.11** (a) Spectrum of the 2.7 K cosmic background radiation. (b) Celestial maps of the intensity fluctuations. Top: Dipole asymmetry. Bottom: Residual fluctuations with maximum  $\Delta T/T \simeq 10^{-5}$ . (After [98].)

The unbroken fascination of radiation physics has recently also been confirmed by radio-astronomical measurements. It is remarkable that the most exact measurement of the universal blackbody spectrum (7.46) is now obtained from the spectrum of the cosmic background radiation. The difference between the measured values and the

theoretical curve in Figure 7.11 is in fact not visible! The average temperature of this radiation, often interpreted as the “afterglow” of the now very much cooled-down Big Bang, can be determined as  $T = 2.726 \pm 0.005$  K.

The measurements by the *Cosmic Background Explorer* (COBE) satellite [98, 99] are so exact that the temperature fluctuations of radiation related to the average detected from a certain direction can be mapped on a celestial map. The spectacular result shows a dipole-like asymmetry of the order  $\Delta T/T \simeq 10^{-3}$ , which can be explained by the proper motion of our galaxy relative to a homogeneous radiation background. Beyond that, the microwave radiation is isotropic except for smaller spatial fluctuations of about  $\Delta T/T \simeq 10^{-5}$ . It is assumed [99] that those small fluctuations reflect the density fluctuations of the early universe and have acted as seeds for the observable matter, which is not homogeneously distributed across the universe.

The terms *stimulated emission* and *spontaneous emission* were developed by Einstein in relation to thermal broadband light sources, since both types were necessary for thermodynamic reasons. Coherent coupling of light fields and atoms was neither conceptually nor experimentally conceivable at that time.

### 7.3.1 Stimulated Emission and Absorption

Let us now investigate how we can obtain the limiting case of a broadband incoherent light field from the Bloch equations. For this purpose we use the complex form in Eq. (7.38) and assume  $|\rho_{eg}| \ll w \simeq -1$ . With the equilibrium value  $\rho_{eg}$ , we obtain without difficulty

$$\dot{w} = -\frac{\gamma' \Omega_R^2}{\gamma'^2 + \delta^2} w - \gamma(w - w_0).$$

We are interested in the first term, containing the stimulated processes (emission and absorption) because of  $\Omega_R^2 = d_{eg}^2 \mathcal{E}_0^2 / \hbar^2 \propto I$ , and we take the broadband spectrum into account by integrating over all detunings  $\delta$  and defining  $\mathcal{E}_0$  to be the mean quadratic field amplitude:

$$\dot{w} = \pi(d_{eg}^2/3)\mathcal{E}_0^2/\hbar^2 - \gamma(w + 1).$$

The coupling of unpolarized field and atomic dipole generates a factor 1/3 by averaging over the space directions, and with  $\rho_{ee} = (w + 1)/2$  we find the form

$$\dot{\rho}_{ee} = \frac{\pi d_{eg}^2}{3\epsilon_0 \hbar^2} u(\nu_0)w - \gamma(w + 1)/2 = B_{eg}u(\nu_0)(\rho_{gg} - \rho_{ee}) - \gamma\rho_{ee}. \quad (7.47)$$

Here we call  $u(\nu_0) = \epsilon_0 \mathcal{E}_0^2/2$  the energy density at the resonance frequency  $\nu_0$ . The coefficient

$$B_{eg} = \frac{\pi d_{eg}^2}{3\epsilon_0 \hbar^2} \quad (7.48)$$

is called the *Einstein B coefficient* and determines the rate of stimulated emission and absorption, respectively. With these coefficients we can also give the rate

equations for an ensemble of absorbers subject to this radiation field:

$$\begin{aligned}\dot{\rho}_{ee} &= B_{eg} u(v_0)(\rho_{gg} - \rho_{ee}) - \gamma \rho_{ee}, \\ \dot{\rho}_{gg} &= B_{eg} u(v_0)(\rho_{ee} - \rho_{gg}) + \gamma \rho_{ee}.\end{aligned}$$

### 7.3.2 Spontaneous Emission

Rigorous calculation of the spontaneous emission rate requires a treatment according to the rules of QED, that is, with the help of a quantized electromagnetic field. In fact the calculation of the spontaneous emission rate by Weisskopf and Wigner [88] in 1930 was the first major success of this, then very new, theory. It will be made explicit in Section 14.3.1.

We here choose a much shorter way by using the result of the Larmor formula (7.3) known from classical electrodynamics. It says that the radiation power of an accelerated electric charge is proportional to its squared acceleration:

$$P = \gamma h v_0 = \frac{2}{3c^3} \frac{e^2}{4\pi\epsilon_0} \ddot{x}^2.$$

We assume that, during the characteristic decay time  $\gamma^{-1}$ , just the excitation energy  $h\nu_0$  is emitted. From quantum mechanics we adopt the result  $\ddot{x} = x\omega_0^2$ , and by *ad hoc* multiplication with the factor 2 delivered only by QED, we obtain the result

$$\gamma = A_{eg} = 2 \times \frac{P}{h\nu_0} = \frac{d_{eg}^2 \omega_0^3}{3\pi\hbar\epsilon_0 c^3} \quad (7.49)$$

for the *Einstein A coefficient*.

By comparison with Eq. (7.47), we confirm the result that Einstein obtained from purely thermodynamic reasoning

$$\frac{A}{B} = \frac{\hbar\omega_0^3}{\pi^2 c^3} = \hbar\omega \frac{\omega^2}{\pi^2 c^3}. \quad (7.50)$$

With  $\rho(\omega) = \omega^2/\pi^2 c^3$ , the latter form contains just the state density of the radiation field for the frequency  $\omega$  (see Section B.3). If the driving field contains photons in a certain mode  $\bar{n}_{ph}$ , then the ratio of spontaneous and stimulated emission rate in this mode has to be

$$A : B = 1 : \bar{n}_{ph}. \quad (7.51)$$

The occurrence as well as the rate of spontaneous emission can be derived from fundamental thermodynamical arguments. Nevertheless it is accessible to experimental modifications, as will be described in Section 14.3.3.

## 7.4 Inversion and Amplification

The treatment of light–matter interaction in Section 7.2.7 predicts the preparation of an *amplifying* medium only if we succeed in populating an excited state more intensely than a lower-lying ground state, that is,  $N_e > N_g$ . Formally speaking this corresponds to a *negative temperature*, since the Boltzmann factor

$N_e/N_g = \exp{-(E_e - E_g)/kT}$  can only become larger than unity if  $T < 0$  – an indication that light amplifying media are not in thermal equilibrium.

#### 7.4.1 Four-, Three-, and Two-Level Laser Systems

It is impossible to obtain inversion by optically exciting a two-level system, as outlined in Section 7.2.7. In a system with four states, however, we can build a dynamic equilibrium that generates a stationary inversion between two of the four levels by supplying energy and so fulfills the requirement for running a laser. An inversion (and thus the requirement for laser operation) can be obtained with three levels as well. However, the four-level system causes a strict separation of the states directly contributing to the pumping process and the laser process. That is why it is preferred for a transparent treatment.

The idealized system is presented in Figure 7.12. The pumping process promoting (*pumping*) particles from the ground state  $|0\rangle$  to the pump level  $|p\rangle$  at a total rate  $R = V\mathcal{R}$  can be driven by electron impact in a discharge, by absorption from the light of an incandescent lamp or a laser, or by other mechanisms. We will get to know some of them in the chapter about lasers.

Our focus is on the two levels  $|e\rangle$  and  $|g\rangle$ , which from now on are to be referred to as *laser levels*. Inversion (and hence the condition for laser operation) can also be obtained with three levels if, for example, the pump level  $|p\rangle$  and the upper laser level  $|e\rangle$  are identical.

The four-level system provides strict separation of all levels by their role in the laser process. A simplified laser model (see Section 9.1) is restricted to the laser levels  $|e\rangle$  and  $|g\rangle$  (two-level laser model). In this model, the contribution of the other auxiliary levels ( $|p\rangle$  and  $|0\rangle$  in Figure 7.12) is implemented by taking into account the pump rate  $R$  and the depopulation rate  $\gamma_{\text{dep}}$ .

#### 7.4.2 Generation of Inversion

We consider the rate equations for the occupation numbers  $n_0$ ,  $n_p$ ,  $n_e$ , and  $n_g$ . We focus on weak pumping processes, where most of the atoms remain in the ground state, and we may keep  $n_0 \approx 1$  to a good approximation. By a short consideration or calculation, it can be found that for these conditions the rate equation system can effectively be limited to the laser states  $|e\rangle$  and  $|g\rangle$ . The dynamics is determined by the population rate  $R$  of the upper state, by its decay rate  $\gamma$ , by the partial transition rate  $\gamma_{eg} \leq \gamma$  falling to the lower laser level, and finally by the depopulation rate of the lower laser level  $\gamma_{\text{dep}}$ :

$$\begin{aligned}\dot{n}_e &= R - \gamma n_e, \\ \dot{n}_g &= \gamma_{eg} n_e - \gamma_{\text{dep}} n_g.\end{aligned}\tag{7.52}$$

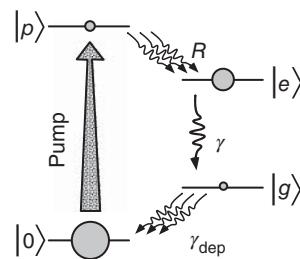


Figure 7.12 Four-level system with inversion between the upper ( $|e\rangle$ ) and lower ( $|g\rangle$ ) levels. The circles indicate the population of the levels in dynamic equilibrium.

The stationary solutions  $n_e^{\text{st}} = R/\gamma$  and  $n_g^{\text{st}} = \gamma_{\text{eg}}R/\gamma\gamma_{\text{dep}}$  are found, and the difference in the occupation numbers can be calculated in equilibrium, but in the absence of any light field, that could cause stimulated emission:

$$n_0 = n_e^{\text{st}} - n_g^{\text{st}} = \frac{R}{\gamma} \left( 1 - \frac{\gamma_{\text{eg}}}{\gamma_{\text{dep}}} \right). \quad (7.53)$$

If the depopulation rate  $\gamma_{\text{dep}}$  of the lower state is larger than the decay rate of the upper state, then apparently an *inversion*,  $n_0 > 0$ , is maintained in this system because  $\gamma_{\text{eg}}/\gamma_{\text{dep}} < 1$ . The inversion is a nonequilibrium situation from the thermodynamic point of view and requires an energy flow through the system.

Since the imaginary part of the polarization is now also *positive* (Eq. (7.44)), we expect the polarization not to be absorbed by the field causing it but in contrast to be intensified! A field growing stronger, though, reduces this inversion according to Eq. (7.45) but maintains the amplifying character (Figure 7.9). With this system, the requirements for an optical amplifier are met. It is known that an amplifier excites itself by feedback and works as an oscillator. We call these devices “lasers.”

### 7.4.3 Optical Gain

If inversion occurs ( $w_0 > 0$ ), then, because of the positive  $v_{\text{st}} > 0$ , a negative absorption coefficient is caused (Eq. (7.19)), also known as the *optical gain coefficient*. Its unit is  $\text{cm}^{-1}$  as well. By evaluation of Eq. (7.43) with the stationary values for  $v_{\text{st}}$ , we obtain

$$\alpha = \frac{2\pi N}{\lambda V} \left[ \frac{w_0}{1+s} \frac{-d_{\text{eg}}^2 \mathcal{E}/\hbar\gamma'}{\epsilon_0 \mathcal{E}(1+(\delta/\gamma')^2)} \right].$$

The result is more transparently expressed with the help of Eqs. (7.21) and (7.34):

$$\alpha = \frac{N}{V \sigma_Q} \frac{-\Delta N_0/N}{(1+s)(1+(\delta/\gamma')^2)} .$$

Especially for  $\delta = 0$  we find the straightforward relation

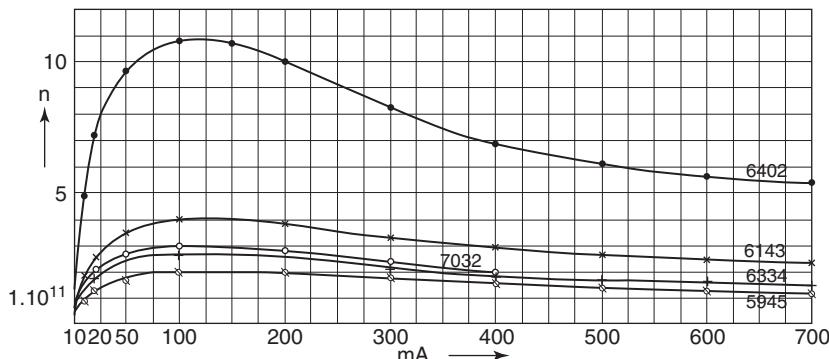
$$\alpha = \frac{N}{V \sigma_Q} \frac{-\Delta N_0/N}{1+I/I_0}. \quad (7.54)$$

According to Figure 7.9, it is clear that the inversion – and so the gain – is reduced under the influence of a light field. For laser operation, we will call this *saturated gain*. In the case of very low intensities  $I/I_0 \ll 1$ , the gain is constant, and it is called *small signal gain*. This value is usually given for a laser material.

We can express Eq. (7.54) also with  $\Delta N_0 = N_e - N_g$ :

$$\alpha = \frac{N_g}{V \sigma_Q} \frac{1 - N_e/N_g}{1+I/I_0}. \quad (7.55)$$

In this version the contribution of *negative dispersion* introduced in 1925 by Kramers and Heisenberg [102] is recognized. This negative dispersion was already observed in 1928 by Ladenburg and Kopfermann [101] (Figure 7.13), more than 25 years before the first description of the optical maser or laser by Schawlow and Townes [103].



**Figure 7.13** Negative dispersion was observed already in 1930 by Ladenburg [100, 101]. The diagram shows the “N”-number, which is evaluated from a measurement of the index of refraction of the strongest transitions (wavelengths in Ångstrom) of the neon 1s–2p system (compare Figure 8.5). It is a measure of the number  $N_g(1 - N_e/N_g)$  in Eq. (7.55). Initially the number of neon atoms in the lower 1s levels  $N_g$  rises rapidly. With increasing current, 2p levels become excited more efficiently, causing a reduction of absorption and dispersion.

#### 7.4.4 The Historical Path to the Laser

With Table 7.4 we highlight important milestones leading to the invention and the application of the laser.

**Table 7.4** Selected milestones on the route to the laser [104, 105].

Year	Event	References
1917	A. Einstein publishes the <i>Quantentheorie der Strahlung</i> ( <i>Quantum Theory of Radiation</i> ) and introduces the $A$ and $B$ coefficients	[106]
1925	H. A. Kramers and W. Heisenberg publish a theoretical manuscript on <i>negative dispersion</i> in atomic gases	[102]
1928	R. Ladenburg and H. Kopfermann demonstrate negative dispersion of a neon discharge at the Kaiser-Wilhelm-Institut for physical chemistry in Berlin	[101]
1951	W. Paul visits C. Townes and colleagues at Columbia University in New York. He reports his work on magnetic hexapole lenses, which can be used to focus atomic and molecular beams	[5, 105]
1954	J. Gordon, H. Zeiger, and C. Townes use Paul's focusing method to generate inversion in an ammonia beam and realize the first maser	[107]
1954	N. Basov and A. Prokhorov publish a theoretical manuscript on <i>molecular amplifiers</i> at the Lebedev Institute in Moscow	[108]
1957	Two patents are filed: for a maser at optical frequencies (C. Townes and A. Schawlow) and for an optically pumped laser amplifier (G. Gould)	
1958	A. Schawlow and C. Townes describe the laser or <i>optical maser</i> , respectively, in an extensive publication	[103]
1960	T. Maiman operates the ruby laser as the first pulsed and visible laser at Hughes Research Laboratories	[109]

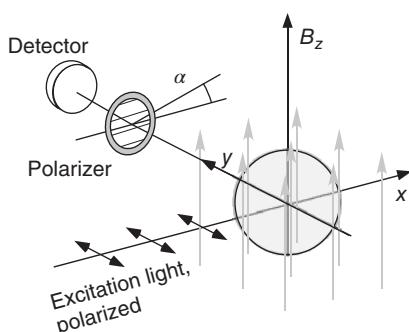
(Continued)

**Table 7.4** (Continued)

Year	Event	References
1960	A. Javan, W. R. Bennett, and D. R. Herriott realize the first continuous-wave (cw) He-Ne gas laser at Bell Laboratories	[110]
1962	A. White and J. Rigden operate the best-known visible laser line, the He-Ne laser at 633 nm wavelength	[111]
1962	Homojunction semiconductor lasers are realized simultaneously by the teams of R. N. Hall of GEC Schenectady, M. I. Nathan of IBM, N. Holonyak of GEC Syracuse, and T. M. Quist of Lincoln Labs at MIT	[112]

## Problems

- 7.1 Doppler-broadened cross section** For gas lasers a realistic absorption or gain coefficient, respectively, is determined by the Doppler-broadened cross section. Use the velocity distribution from Section 12.3.2 to estimate an effective cross section for an atomic resonance line, that is, a transition with moment  $d_{eg} = ea_0$ .
- 7.2 Classical Bloch equations** Compare the optical Bloch equations (Eq. (7.36)) and the equations for the classical dipole oscillator (Eq. (7.8)). Contrast the significance of the parameters. If the dynamics of the (undamped) Bloch vector can be described as motion on the Bloch sphere, which geometric form is appropriate for the classical dipole? Why is it an approximation for the Bloch sphere?
- 7.3 Hanle effect** An oscillating linear electric dipole reradiates its excitation energy according to the radiation pattern from Figure 2.6. For instance, in free space, no radiation field is expected to travel in the direction of the detector, if as in the drawing the linearly polarized  $E$  field is directed toward the detector (Figure 7.14). A dipole exhibiting a magnetic moment  $\mu$ , too, precesses around the axis of an external magnetic field  $B_z$  with Larmor frequency  $\omega_L = \mu_\perp B_z / \hbar$  where  $\mu_\perp$  is the component of the dipole orthogonal to the magnetic field. This

**Figure 7.14** Schematic for observing the Hanle effect?

situation occurs frequently for atoms, and it modifies the radiation distribution because the orientation of the electric dipole can rotate. Consider a simplified classical model that was proposed as long ago as 1924 by Hanle [113]. At  $t = 0$  electric dipoles are excited, which radiate at  $B_z = 0$  the amplitude  $E(t) = E_0 \exp(-i\omega_0 t) \exp(-\gamma t/2)$  into the direction of the detector. For  $B_z \neq 0$  assume that the dipoles precess with frequency  $\omega_L$  around the axis of the  $B_z$  field. The average intensity at the detector is calculated by integration over all times,  $I_D = \int_0^\infty I(t) dt$ .

- Linear polarization  $E \parallel B_z$ ,  $\pi$ -polarization. Show that the magnetic field does not influence the fluorescence intensity registered at the detector,  $I_D(t) = \text{const.}$
- Linear polarization  $E \perp B_z$ ,  $\sigma_+$  and  $\sigma_-$  polarization. Show that for the polarization analyzer position  $\alpha$ , the intensity at the detector is  $I_D(t) = I_0 \exp(-\gamma t) \cos^2[\omega t - \alpha]$ . Calculate and sketch the average intensity at the detector as a function of the external magnetic field for analyzer positions  $\alpha = 0, \pi/4, \pi/2, 3\pi/4$ .

**7.4 Can a sodium atom be saturated with sunlight?** The sodium  $D$  line at  $\lambda = 589$  nm has a natural linewidth of 10 MHz corresponding to a saturation intensity of  $I_0 = 63.4 \text{ W m}^{-2}$ . (a) On Earth a sodium atom is separated by  $D_{S-E} = 1.5 \times 10^8$  km from the sun. Is it possible to saturate the  $D$  line with sunlight? (b) Is the sodium  $D$  line saturated if the atom is placed directly on the surface of the sun (radius  $r_S = 7 \times 10^5$  km)? (c) Determine the temperature at which on the surface of the sun the atom is just saturated. (Hint: Consider the sun to be an ideal blackbody and use Planck's formula for the spectral energy density (Eq. (7.46).) The temperature of the sun is 5 700 K.)

**7.5 Bloch equations: magnetic resonance and optical transitions** The optical Bloch equations used to describe light-matter interaction at optical frequencies are identical to the Bloch equations applied to the methods of magnetic resonance, of which nuclear magnetic resonance (NMR) is the most important. Discuss the differences of the optical and magnetic systems in the solutions by studying the relevant time constants. Estimate the spontaneous decay rate of a magnetic transition by extending Eq. (7.49) to a typical magnetic moment.

**7.6 Kramers–Kronig relations** An important result of theoretical electrodynamics are the *Kramers–Kronig relations*, according to which the real and imaginary parts of the susceptibility are connected with each other by

$$\chi''(\omega) = -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi'(\omega')}{\omega' - \omega} d\omega' \quad \text{and} \quad \chi'(\omega) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi''(\omega')}{\omega' - \omega} d\omega'.$$

Here  $\mathcal{P}$  denotes the principal value of the integral. Show that the susceptibility according to Eq. (7.14) fulfills the Kramers–Kronig relation.

**8**

## The Laser

The laser concentrates light energy in space, in time, and in the spectral domain. These properties have rendered the laser one of the most important instruments not only for physical research but also for almost all fields of everyday life. The laser must be counted among the most significant inventions of the twentieth century and has not yet stopped to open more fields of application.

In this chapter we introduce the basic principles of lasers by presenting systems of central importance and with some technical details. An introduction to the theoretical description of the most important dynamic physical properties of lasers is given in Section 9.1. A special chapter (Chapter 10) is devoted to semiconductor lasers, which because of their compact design and inexpensive production play a central role in the still growing scientific as well as economic relevance of lasers and laser-based optical instruments.

**Laser = Light Amplification by Stimulated Emission of Radiation**

*Laser* has become a well-recognized word in everyday language. It first appeared in a 1957 patent filed by J. Gordon for an “optically pumped laser amplifier” and is derived from its predecessor, the *maser* (the acronym “maser” stands “microwave amplification by stimulated emission of radiation”).

*Natural maser and laser sources.* A laser is fundamentally a source of intense coherent light fields. Laser light appears absolutely artificial to us, and our ancestors certainly never happened to experience the effect of a coherent light beam.<sup>1</sup> In the cosmos, though, there exist several natural sources of coherent radiation. Their wavelengths are usually too long to be considered a laser and are thus identified as masers [114]. They occur in the vicinity of hot stars, where inversion can be sustained in, for instance, molecular gases. An example with a relatively short wavelength is the hydrogen gas surrounding a star named MWC349 in the Cygnus constellation, which is excited to luminescence by the ultraviolet (UV) radiation of this hot star. The hydrogen gas arranged in a

---

<sup>1</sup> Though interference and coherence phenomena can be observed even in our everyday environment – for example, take a piece of thin, fine fabric and watch some distant, preferably colored, lights through it, for example, the rear lights of a car. You will observe a diffraction pattern. The twinkling stars have fascinated humankind all along.

disc amplifies the far-infrared radiation of the star at the wavelength of  $169\text{ }\mu\text{m}$  several million times, such that it can be detected on Earth.

Of course very different conditions regarding density and temperature prevail in space in comparison with environments on Earth. Of the 130 cosmic maser lines derived from 10 different molecules known today, only 2 are observed in the laboratory. Apart from vibrational transitions of the HCN molecule [115], line of the ammonia molecule that was used by Townes and coworkers [107] to operate the first maser in 1954 is interesting enough. Cosmic lasers provide astronomers with interesting data on the dynamics of large interstellar molecular clouds.

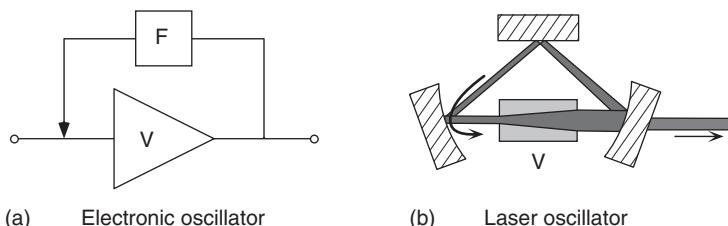
*Laser amplifiers and oscillators.* The laser has historical roots in high-frequency and gas discharge physics. It was known from the maser that it was possible to construct an amplifier and an oscillator for electromagnetic radiation with an inverted molecular or atomic system. In a famous publication [103] A. Schawlow (1921–1999, Nobel Prize 1981) and C. Townes (1915–2015, Nobel Prize 1964) had theoretically predicted the properties of an “optical maser.”

The optical properties of atomic gases had already been studied in discharges for a long time. The question was raised whether an inversion and thus amplification of light could be achieved by a suitable arrangement. So it becomes understandable that the first continuous-wave (cw) laser realized by the American physicist Ali Javan (born 1928) in 1960 [110] with an infrared wavelength of  $1.152\text{ }\mu\text{m}$  was a surprisingly complex system consisting of a gaseous mixture of helium and neon atoms.

The laser bears a close analogy to an electronic amplifier that is excited to oscillations by positive feedback. Its oscillation frequency is determined by the frequency characteristic of gain and feedback (Figure 8.1). It is known that an amplifier oscillates with positive feedback if the gain becomes greater than the losses:

Oscillation condition:  $\text{gain} \geq \text{losses}$ .

Once oscillation has started, the amplitude grows more and more until all the pump current supplying the gain is used up to just compensate the losses and outcoupling of the laser or oscillator circuit, respectively. The effective gain then equals the losses. This is called a *saturated gain* (see Section 9.1.3.1).



**Figure 8.1** (a) Analogy between a laser and an electronic amplifier ( $V$ ), which becomes an oscillator by feedback coupling. The oscillator frequency can be selected, for example, by a filter ( $F$ ) in the feedback path. For the laser, the feedback is achieved using resonator mirrors. (b) For clarification, a ring resonator with three mirrors has been chosen. The spectral properties of the amplifying medium as well as the wavelength-dependent reflectivity of the resonator mirrors determine the frequency of the laser.

**Table 8.1** Laser types.

	Gaseous	Liquid	Solid state
Fixed frequency	Neutral atoms	Rare-earth ions 3d ions	
	Ions		
Multiple frequency	Molecules		
Tunable		Dyes 3d Ions Semiconductor	

As we already know from the chapter about light and matter, an inversion of the laser medium is necessary to achieve an intensification of a light wave. If the loss coefficient is  $\alpha_v$ , then the gain condition according to Eq. (7.54) reads

$$\frac{N}{V} \sigma_Q \frac{-\Delta N_0/N}{1 + I/I_0} > \alpha_v.$$

Using the simplest picture, there always have to be many more atoms in the upper excited state than in the lower one. If this condition is not fulfilled, the laser oscillations die or do not even start. An ideal laser is supposed to deliver a gain as large as possible and independent of the frequency. Since no such a system has yet been found, a multitude of laser systems is used. The most important variants roughly divided into classes (Table 8.1) will now be introduced with their technical concepts, strengths, and weaknesses.

## 8.1 The Classic System: The He–Ne Laser

The helium–neon laser (He–Ne laser) has played an unsurpassed role in scientific research on the physical properties of laser light sources, for example, experimental investigations of coherence properties. Just for this alone it is the “classic” of all laser systems. We shall introduce several important laser features using this system as an example.

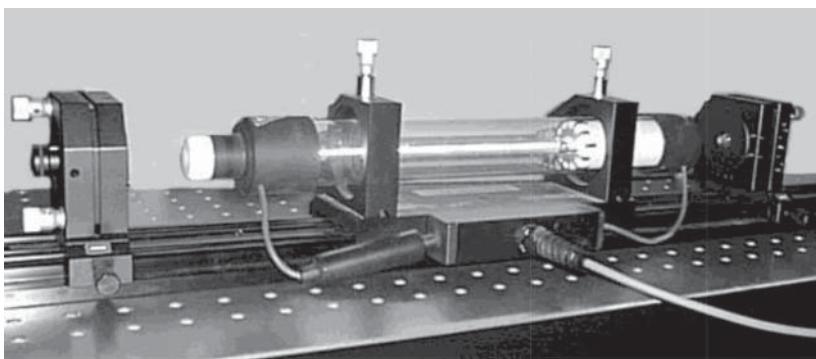
### 8.1.1 Construction

The helium–neon laser obtains its gain from an inversion in the metastable atomic excitations of the Ne atom (the luminescence of Ne atoms is also known due to the proverbial neon tubes).

#### 8.1.1.1 Amplifier

In Figure 8.3 the relevant atomic levels with some important features and some selected laser wavelengths (“lines”) are presented. Since the gas mixture is quite dilute, we can easily understand the He–Ne laser using the picture of independent atoms.

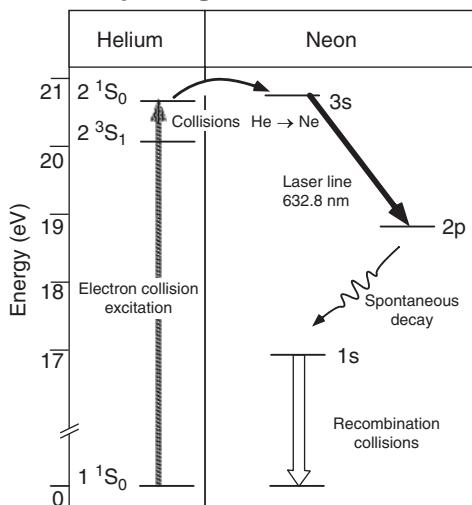
The Ne atoms are excited not directly by the discharge but by energy transfer from He atoms, which are excited to the metastable  $^1S_0$  and  $^3S_1$  levels by electron



**Figure 8.2** Helium–neon laser in an open experimental setup. The current is supplied to the discharge tube by the two cables. The resonator mirrors and the laser tube are mounted on finely tunable bearings. Compare Figure 8.4.

impact. The Ne atom has nearly resonant energy levels so that an efficient energy transfer is enabled by resonant impacts. In the He–Ne laser, the excitation and the laser transition are split up into two different atomic systems, which are helpful for the realization of the desirable four-level system, though there is a problem at the lower laser level of the Ne atoms (Figure 8.3), which is metastable as well and cannot be emptied by radiative decay. In a narrow discharge tube, collisions with the wall lead to efficient depopulation of the lower laser level.

#### 8.1.1.2 Operating Conditions

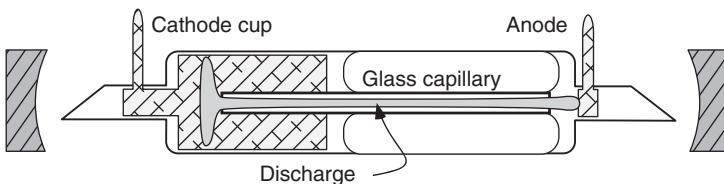


**Figure 8.3** Energy levels of He and Ne atoms with the most prominent optical transition at 632.8 nm. For nomenclature, the spectroscopic terms are used.

the gas filling pressure being slightly different depending on the application. Optimum conditions for the product of pressure  $p$  and diameter  $d$  are empirically

The inversion can only be maintained in a rather dilute gas mixture compared to the atmospheric environment. The He pressure  $p$  is some 10 mbar, and the He:Ne mixing ratio is about 10 : 1. The He discharge is operated at a current of several millamps and a voltage of 1–2 kV, and it burns in a capillary tube with a diameter of  $d \leq 1$  mm. At its walls the metastable Ne atoms (Figures 8.2 and 8.4) fall back to the ground state again due to collisional relaxation and are available for another excitation cycle. The discharge is ignited by a voltage pulse of 7–8 kV (Figure 8.4).

All He–Ne lasers have this construction principle in common, only the construction length and



**Figure 8.4** Schematic representation of a helium–neon laser. The big cathode cup prevents fast erosion caused by the discharge. The Brewster windows at the ends of the laser tube reduce reflection losses at the windows and uniquely determine the laser polarization.

found at

$$p \cdot d \simeq 5 \text{ mbar mm.}$$

The output power of commercial He–Ne lasers varies between 0.5 mW, which will just not damage the eyes, and 50 mW. The power depends on the discharge current and the length of the tube. Both can only be increased to a certain amount. The gain is proportional to the density of inverted Ne atoms, but this already reaches a maximum at a few tens of milliamps since increasing electron collisions de-excite the atoms. The length of the tube cannot be significantly expanded over  $\ell = 1 \text{ m}$  for the following reasons. On the one hand, the diameter of the Gaussian modes grows with increasing mirror distance and does not fit into the capillary tube anymore. Furthermore, with a larger construction length, the  $3.34 \mu\text{m}$  line starts oscillating as a superradiator even without mirrors and thus withdraws energy from laser lines competing for the same reservoir of excited atoms.

#### 8.1.1.3 The Laser Resonator

The resonator mirrors can be integrated into the discharge tube and may be once and forever adjusted during manufacturing. Especially for experimental purposes, an external resonator with manually adjustable mirrors is used. The tube has windows at its ends. In the simplest case the resonator only consists of two (dielectric) mirrors and the discharge tube. To avoid losses, the windows are either antireflection-coated or inserted at the Brewster angle.

#### Example: Radiation field in the He–Ne laser resonator

The laser mirrors determine the geometry of the laser radiation field according to the rules of Gaussian optics (see Section 2.3). They have to be chosen such that the inverted Ne gas in the capillary tube is used as optimally as possible. For a symmetric laser resonator with mirror radii  $R = 100 \text{ cm}$  (reflectivity 95% and 100%), and separated by  $\ell = 30 \text{ cm}$ , one obtains for the red 633 nm line a  $\text{TEM}_{00}$  mode with the following parameters:

Confocal parameter	$b = 2z_0 = 71 \text{ cm}$ ,
Beam waist	$2w_0 = 0.55 \text{ mm}$ ,
Divergence	$\Theta_{\text{div}} = 0.8 \text{ mrad}$ ,
Power inside/outside	$P_i/P_o = 20/1 \text{ mW}$ .

There are no problems of fitting the laser beam over the complete length to the typical cross section of the plasma tube of about 1 mm. Even at a distance of 10 m, it has just a cross section of about 4 mm.

### 8.1.2 Mode Selection in the He–Ne Laser

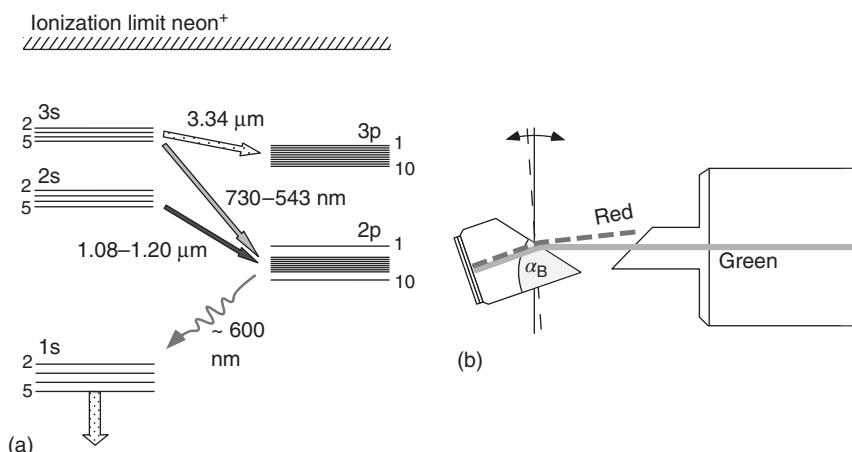
We devote the next two sections to the physical properties of the He–Ne laser (mode selection and spectral properties) because it is a good model to present the most important general laser properties and since for the He–Ne laser the physical properties have been investigated particularly thoroughly.

The aim of mode selection in every cw laser is the preparation of a light field oscillating both in a single spatial or transverse mode and in a single longitudinal mode, that is, at just a single optical frequency  $\omega$ . The methods used with the He–Ne laser can be applied to all other laser types with slight modifications. The desirable transverse mode is mostly a  $\text{TEM}_{00}$  or closely related mode. It has fewer losses due to its comparatively small cross section and thus is often intrinsically preferred anyway. In case of doubt the relevant spatial mode can be selected by using a suitable adjustment of the resonator or insertion of an additional aperture.

#### 8.1.2.1 Laser Line Selection

If several laser lines of the neon atom have a common upper laser level (e.g., 2s; Figure 8.5), only that one with the highest gain can be observed. If the laser line couples completely different levels (e.g., 2s–2p and 3s–3p), then the lines can be activated simultaneously.

Owing to the helium discharge in the neon gas, the 2s state and the 3s state are populated, with the occupation of the uppermost  $3s_2$  substate dominating in the



**Figure 8.5** Wavelength selection in the He–Ne laser. (a) Part of the energy level scheme of the neon atom showing the important laser transitions. The common notation does not follow the customary singlet/triplet convention according to the  $LS$  coupling scheme. The notation used here goes back to Paschen, who simply numbered the levels consecutively. An s level splits into 4, and a p level into 10 angular momentum states. (b) Littrow prism as a dispersive end mirror for wavelength selection.

3s group. The largest gain factors are obtained at the wavelengths of 0.633, 1.152, and 3.392 μm. Transitions with a low gain can be excited if the feedback coupling by the resonator selectively favors or suppresses certain frequencies by means of some suitable optical components. In general, all dispersive optical components – such as optical gratings, prisms, and Fabry-Pérot etalons – are appropriate. One of the simplest methods is the installation of a Littrow prism as shown in Figure 8.5. The Littrow prism is a Brewster prism divided in half so that the losses for p-polarized light beams are minimized. The backside of the Littrow prism is coated with dielectric layers to make a highly reflective mirror. Since the refraction angle depends on the wavelength, the laser line can be selected by tilting the Littrow prism.

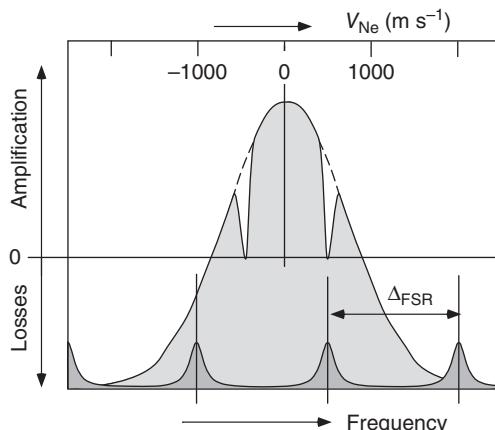
Another specialty is the extremely high gain coefficient of the infrared 3.34 μm transition (typically  $10^3 \text{ cm}^{-1}$ ), causing the line to start oscillating almost every time. It can be suppressed by using infrared-absorbing glass and limiting the length of the plasma tube. The latter fact is unfortunate since it imposes a technical limit to the output power, which otherwise increases with length.

### 8.1.3 Gain Profile, Laser Frequency, and Spectral Holes

For the example of the He-Ne laser, we are now going to ask how the oscillation frequency of a laser line depends on the combined properties of the gas and the resonator. The spectral width of the fluorescence spectrum (the width of the optical resonance line) is determined by the Doppler effect caused by the neon atoms moving at thermal speed of several  $100 \text{ m s}^{-1}$  in a gas at room temperature. This broadening is called *inhomogeneous* (see Section 12.3.2), since atoms with different velocities have different spectra.

With regard to the laser process, this especially means that the coupling of the neon atoms to the laser light field depends very strongly on their velocity. For the red laser line at 633 nm, the Doppler linewidth at room temperature is about  $\Delta\nu_{\text{Dopp}} = 1.5 \text{ GHz}$  according to Eq. (12.9) and can just be resolved with a high-resolution spectrometer (e.g., Fabry-Pérot).

In Figure 8.6 the gain profile and its significance for the laser frequency are presented. It makes the laser begin oscillating if the gain is higher than the losses. Within the gain profile the laser frequency is determined by the resonance



**Figure 8.6** Gain profile of a He-Ne laser in operation. Without the laser field, the small signal gain corresponds to the neon Doppler profile (dashed line); the laser operation modifies the gain profile by the so-called spectral or Bennett holes (see text). Two holes symmetric to  $v = 0 \text{ m s}^{-1}$  occur because in a standing wave the Doppler shift acts in two directions.

frequencies of the laser resonator (here indicated by the transmission curve showing maxima at frequencies separated by the free spectral range  $\Delta_{\text{FSR}}$ ). At these “eigenfrequencies” the laser may start oscillating, as we shall investigate more deeply in Section 9.1. True lasers are slightly shifted off the resonances of the empty resonator, an effect that is called *mode pulling*.

At the eigenfrequencies of the He–Ne laser, one can observe spectral holes (the so-called Bennett holes). The atoms make up the difference between their rest-frame frequency  $v_0$  and the laser frequency  $v_L$  by their velocity  $v_z$  in the direction of the resonator axis, and the atoms of a gas laser contribute to the gain only within their homogeneous, that is, their natural, linewidth. This situation is called *saturated gain*. In equilibrium the effective gain is reduced at these frequencies to that value just corresponding to the losses (including the decoupling at the resonator mirrors).

The small signal gain from Figure 8.6 can be measured by sending a very weak tunable probe beam through the He–Ne laser and measuring the gain directly. Since in a resonator with standing waves atoms can couple to the light field in both directions, two spectral holes can be observed at

$$v_L = v_0 \pm kv_z.$$

This observation also indicates that two different velocity groups of atoms contribute to the gain of the backward- and forward-running intra-cavity wave, respectively. Thus a very interesting case occurs when both holes are made to coincide by, for example, changing the resonator frequency by length variations with the help of a piezo mirror. At  $v_z = 0$  a lower gain than outside the overlap region of the holes is available, and the output power of the laser decreases. This collapse is called *Lamb dip* after Willis E. Lamb (born 1913),<sup>2</sup> and it initiated the development of Doppler-free saturation spectroscopy.

#### 8.1.4 The Single-Frequency Laser

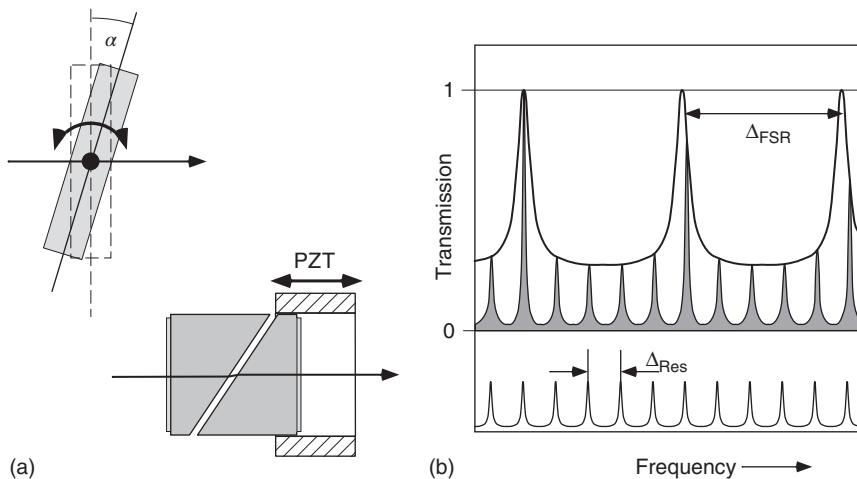
In Figure 8.6 only one resonator frequency lies within the gain profile such that the laser threshold is exceeded. Since the free spectral range  $\Delta_{\text{FSR}} = c/2\ell$  of the He–Ne laser exceeds the width of the Doppler profile at  $\lambda = 633 \text{ nm}$  below 10 cm, for typical, that is, larger, construction lengths, generally 2–4 frequencies start to oscillate because in the inhomogeneous gain profile, there is no competition between the modes about the available inversion. But we can still insert additional (and low-loss) dispersive elements into the resonator, which modulate the spectral properties of the gain profile in a suitable way to filter the desired laser frequency from the available ones. To discriminate between adjacent resonator modes, highly dispersive elements such as Fabry–Pérot etalons are required.

#### Example: Gain modulation with intra-cavity etalons

Etalons cause a modulation of the effective laser gain that is periodic with  $\Delta_{\text{FSR}}$ , the free spectral range (Eq. (6.19)) separating the adjacent transmission maxima.

---

<sup>2</sup> W. E. Lamb has become immortal in physics mainly by the discovery of the *Lamb shift*, also named after him, for which he received the Nobel Prize in 1955. He also contributed significantly to the pioneering days of laser spectroscopy.



**Figure 8.7** Frequency selection by intra-cavity etalons. (a) “Thin” etalon and “thick” etalon. For rough tuning the thin etalon is tilted where the walk-off caused by the tilt is tolerable. The thick etalon is constructed with an air gap at the Brewster angle, which can be varied in length by a piezo translator. (b) Combined effect of an etalon ( $\Delta_{\text{FSR}}$ ) and the laser resonator ( $\Delta_{\text{Res}}$ ) on the transmission (and thus on the gain). For this example the length of the etalon is about one-fifth of that of the resonator.

The periodicity can be chosen by the etalon thickness or length  $\ell$  and refractive index  $n$ . Tilting the etalon causes the transmission maxima to shift, and from geometrical considerations following Section 6.5, one obtains

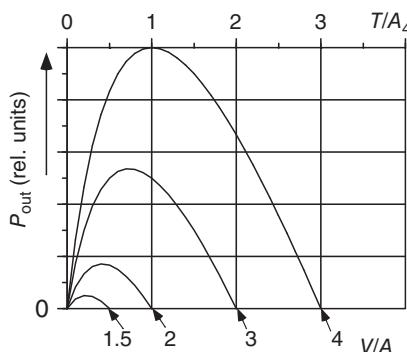
$$\nu_{\max} = N\Delta_{\text{FSR}}(\alpha) = N \frac{c}{2\ell\sqrt{n^2 - \sin^2\alpha}} \simeq N \frac{c}{2\ell n} \left[ 1 - \frac{1}{2} \left( \frac{\alpha}{n} \right)^2 \right].$$

Small tilts by  $\alpha$  change the free spectral range only slightly but are enough due to the high-order  $N$  to tune the center frequency efficiently. The monolithic thin etalon presented in Figure 8.7 is simple and intrinsically stable, but when tilted, it leads to a *walk-off* and thus increased losses. The air-spaced etalon (the variable air gap is cut at the Brewster angle to keep the resonator losses as small as possible) has a more complex mechanical construction but avoids the walk-off losses of the tilted device.

The intra-cavity etalons often do not need any additional coating. Even using only the glass–air reflectivity of 4%, a modulation depth of the total gain of about 15% is obtained according to Eq. (6.17). With regard to the small gain, in many laser types, this is completely sufficient.

### 8.1.5 Laser Power

We shall now investigate how to optimize the laser output power, that is, for practical reasons, generally to maximize it. The properties of the amplifier medium are physically determined and thus can be influenced only by a suitable choice, length, density, and so on. The losses can be kept as small as possible by design and choice of the components of the resonator. Finally there is only the choice



**Figure 8.8** Output power of a laser as a function of the transmission  $T$  of the outcoupling mirror and the gain  $V$ , normalized to the resonator losses  $A$ .

of the mirror reflectivity as the remaining free parameter, which also acts as a loss channel.

We consider a Fabry-Pérot interferometer with gain as a model. For this we make use of the considerations about dissipative resonators from Section 6.5. The laser is always operated in the resonance case. We here anticipate the relation between gain  $V$ , losses  $A$ , and transmission  $T$  of the only outcoupling mirror more accurately dealt with in Section 9.1 (Eq. (9.18)),

$$I_{\text{out}} = I_0 \frac{T(V - A - T)}{A + T},$$

and investigate it graphically as a function of the transmission  $T$  in Figure 8.8.

### 8.1.6 Spectral Properties of the He-Ne Laser

#### 8.1.6.1 Laser Linewidth

Until now we have taken monochromatic optical light fields for granted, that is, we have assumed that the optical wave can be described by a single exactly defined frequency  $\omega$ . In the chapter about laser theory, we will see that laser light comes closer to this deeply classical idea of a perfect harmonic oscillation than almost any other physical phenomenon. The physical limit for the spectral width of a laser line measured according to the so-called Schawlow-Townes limit (Section 9.4.4) amounts to several hertz or even less! This physical limit is imposed due to the quantum nature of the light field. It has been mentioned already by the authors in that paper proposing the laser in 1958. According to this the linewidth of the laser is (see Eq. (9.33))

$$\Delta\nu_L = \frac{N_2}{N_2 - N_1} \frac{\pi h v_L \Delta\nu_c^2}{P_L},$$

where  $v_L$  is the laser frequency,  $\Delta\nu_c = \gamma_c$  the damping rate or linewidth of the laser resonator,  $P_L$  the laser power, and  $N_{1,2}$  the occupation numbers of the upper and lower laser level, respectively.

#### Example: Schawlow-Townes linewidth of the He-Ne laser

We consider the He-Ne laser from the previous example. The laser frequency is  $v_L = 477 \text{ THz}$ , the linewidth of the resonator is  $\Delta\nu_c = 8 \text{ MHz}$  (according to the data from the example on p. 253), while all internal resonator losses are neglected according to Eq. (6.20). The He-Ne laser is a four-level laser so that we have  $N_1 \approx 0$ . For an output power of 1 mW, we calculate a laser linewidth of just

$$\Delta\nu_L \approx \frac{\pi h \times 477 \text{ THz} (8 \text{ MHz})^2}{1 \text{ mW}} = 0.07 \text{ Hz}.$$

The extremely small Schawlow–Townes linewidth of the red He–Ne line corresponds to a  $Q$  value  $v/\Delta v \simeq 10^{15}!$  Even today laser physicists think of this limit as a thrilling challenge because it promises to make the laser the ultimate precision instrument wherever a physical quantity can be measured by means of optical spectroscopy.

From the beginning the He–Ne laser has played an extraordinary role for precision experiments, and it is indeed a challenge even just to measure this linewidth! It is thus useful to illustrate the methods used to measure the linewidth of a laser.

### 8.1.7 Optical Spectral Analysis

The spectrum of a laser oscillator, like that of any other oscillator, can be investigated using several methods. A Fabry–Pérot interferometer is used as a narrowband filter. Its mid-frequency is tuned over the area of interest. A photodiode measures the total power transmitted within the filter passband.

Alternatively the laser beam can be superimposed with a second coherent light field (*local oscillator*) onto a photodiode. The photodiode generates the difference frequency or “heterodyne” beat signal. The superposition signal in turn can be analyzed by radio-frequency methods or Fourier analysis.<sup>3</sup>

#### 8.1.7.1 The Fabry–Pérot Spectrum Analyzer

In the simplest, and therefore very often used, method, a tunable optical filter is used, also called in brief a “scanning Fabry–Pérot,” usually a confocal optical resonator. One of its mirrors can be displaced (“scanned”) by several  $\lambda/4$  corresponding to several free spectral ranges with the help of a piezo translator. The resolution of the optical filter usually reaches some megahertz and therefore can be used only for rough analysis or as a laser with a large linewidth (e.g., diode lasers; see Chapter 10).

If the linewidth of the laser is smaller than the width of the transmission curve of the Fabry–Pérot interferometer, some information about the frequency fluctuations can still be obtained by setting its frequency to the wing of the filter curve and using this as a frequency discriminator (Figure 8.9). So any frequency variations are converted to amplitude fluctuations, which in turn can be analyzed by means of radio-frequency techniques or by Fourier transformations.

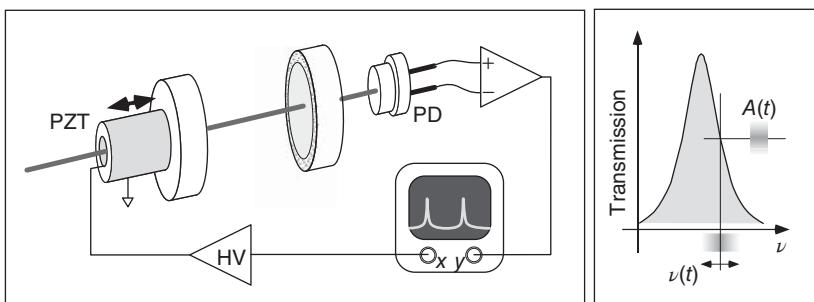
#### 8.1.7.2 The Heterodyne Method

When applying the heterodyne method, it is important to closely match the wave fronts of both light fields and enter the photodiode with excellent flatness so that the detector is exposed to the same phase everywhere. Otherwise the beating signal is strongly reduced.

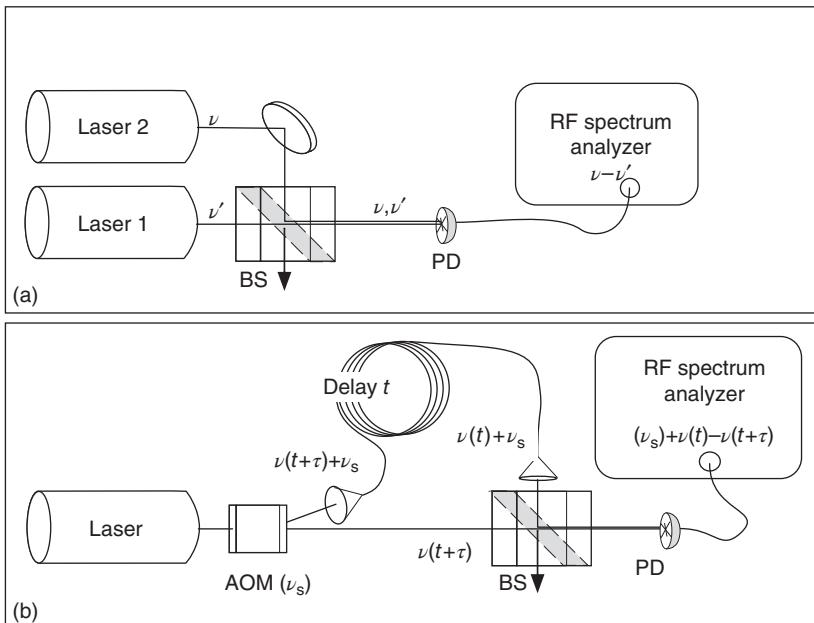
In Figure 8.10 the scheme is presented according to which a heterodyne signal can be achieved at radio frequencies. A second laser can be used as a local oscillator. Its frequency should be much more “stable” than the laser to be tested.

---

<sup>3</sup> Superposition of two electromagnetic waves with different frequencies is usually called “heterodyning,” while superposition of identical frequencies is called “homodyning.” Heterodyning is generally preferred since the noise properties of both detectors and receivers are favorable at higher frequencies.



**Figure 8.9** Scanning Fabry-Pérot interferometer for the spectral analysis of laser radiation.



**Figure 8.10** Heterodyne method for the determination of the laser linewidth.

(a) Superposition with a reference laser used as a local oscillator. (b) Superposition according to the autocorrelation method. PD = photodiode; BS = beam splitter; AOM = acousto-optical modulator, driver frequency  $\nu_s$ .

Furthermore, it must not deviate too far from the test frequency, since above 1–2 GHz high-speed photodiodes become more and more unwieldy (the active area shrinks more and more to avoid parasitic capacitances and to gain bandwidth) and expensive.

The autocorrelation method is an alternative in which the laser in a way “pulls itself out of the mud.” One part of the laser light is split off by an acousto-optical modulator (AOM; see Section 3.5.4) and at the same time is shifted by its driver frequency  $\nu_s$ , which is typically some 10 MHz. One of the two light beams is now delayed over a long optical fiber so that there is no longer any phase correlation (“coherence”) between the light waves. Both light waves are

superposed onto a photodiode as before, and the mixed signal is investigated using a radio-frequency spectrum analyzer.

The method is completely analogous to a Michelson interferometer. It is operated at a path difference larger than the coherence length. There is no visibility (i.e., the average of the interference signal vanishes) but a fluctuating instantaneous beat signal that delivers a good measure of the spectral properties of the laser.

### 8.1.8 Applications of the He–Ne Laser

For the production of He–Ne laser tubes, the manufacturing technology of radio tubes was readily available and very suitable. Radio tubes were being replaced by transistors in the 1960s, so a high production capacity was available when the He–Ne laser was developed. Historically, its rapid distribution was significantly supported by this fact.

The best-known wavelength of the He–Ne laser is the red laser line at 632 nm, which is used in countless alignment, interferometric, and reading devices. The application of the red He–Ne laser declined rapidly since mass-produced and hence cheap red diode lasers (Chapter 10) have become available, which can be operated with normal batteries, are very compact, and yet offer very acceptable  $\text{TEM}_{00}$  beam profiles as well. The He–Ne laser still plays an important role in metrology (the science of precision measurements). The red line is used, for example, to realize length standards. The infrared line at 3.34  $\mu\text{m}$  constitutes a secondary frequency standard if it is stabilized on a certain vibrational resonance transition of the methane molecule [116].

## 8.2 Other Gas Lasers

Stimulated by the success of the helium–neon laser, many other gas systems have been investigated for their suitability as a laser medium. Gas lasers have a small gain bandwidth and are *fixed-frequency lasers* when their small tunability within the Doppler bandwidth is neglected. Like the He–Ne laser, they play a role as instrumentation lasers, provided they have reasonable physical and technical properties, for example, good beam quality, high frequency stability, and a low energy consumption. Some gas lasers are in demand because they deliver large output power, not in pulsed but in cw operation. Listed in Table 8.2 are those gas lasers that nowadays have practical significance. It is technically desirable to have a substance already gaseous at room temperature. That is why the rare gases are particularly attractive. Among them argon- and, even more so, krypton-ion lasers have achieved technical significance.

### 8.2.1 The Argon Laser

The argon-ion laser plays an important role since it is among the most powerful sources of laser radiation and is commercially available with output powers of several tens of watts. However, the technical conversion efficiency, that is, the

**Table 8.2** Overview: gas lasers.

Laser	Short form	cw/P <sup>a)</sup>	Laser lines	Power
<i>Neutral-atom gas lasers</i>				
Helium–neon	He–Ne	cw	633 nm 1.152 nm 3.391 nm	50 mW 50 mW 50 mW
Helium–cadmium	He–Cd	cw	442 nm 325 nm	200 mW 50 mW
Copper vapor	Cu	P	511 nm 578 nm	60 W 60 W
Gold vapor	Au	P	628 nm	9 W
<i>Noble-gas ion lasers</i>				
Argon-ion laser	Ar <sup>+</sup>	cw	514 nm 488 nm 334–364 nm	10 W 5 W 7 W
Krypton-ion laser	Kr <sup>+</sup>	cw	647 nm 407 nm	5 W 2 W
<i>Molecular gas lasers</i>				
Nitrogen	N <sub>2</sub>	P	337 nm	100 W
Carbon monoxide	CO	cw	4–6 μm	100 W
Carbon dioxide	CO <sub>2</sub>	cw	9.2–10.9 μm	10 kW
Excimers	F <sub>2</sub> , ArF, KrF, XeCl, XeF	P	0.16–0.35 μm	250 W

a) cw = continuous wave; P = pulsed.

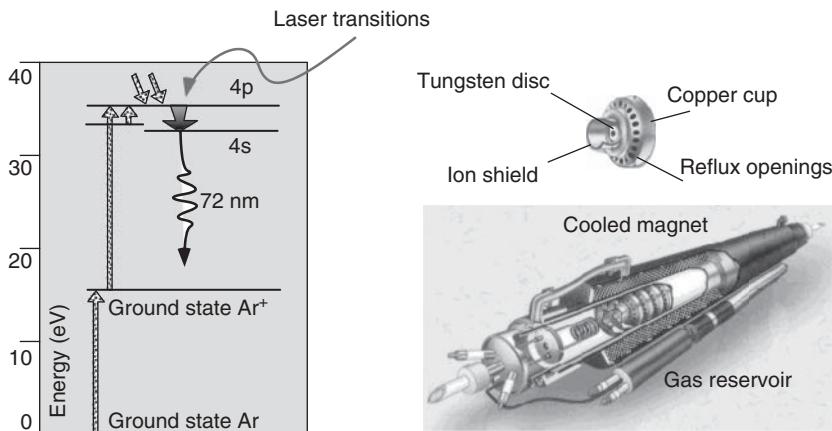
ratio of electric power consumption and effective optical output power, is typically 10 kW : 10 W. For many applications this is absolutely unacceptable. In addition, it is also burdened with the necessity to annihilate most of the energy spent by a costly water cooling system. Therefore, the frequency-doubled solid-state lasers (e.g., Nd:YAG; see Section 8.4.2) have widely replaced the argon laser. In the UV range, however, there is no competitor to the Ar-ion laser in sight yet.

### 8.2.1.1 The Amplifier

Excitation of the high-lying Ar<sup>+</sup> states is obtained by stepwise electron impact. That is why there is a very much higher current density necessary than in a He–Ne laser. The upper laser level can be populated from the Ar<sup>+</sup> ground state as well as from other levels above or below it. The krypton laser follows a quite similar concept, but it has achieved less technical importance.

### 8.2.1.2 Operating Conditions

In the 0.5–1.5 m long tubes, a discharge is operated maintaining an argon plasma. The cross section through the plasma tube shown in Figure 8.11 indicates the



**Figure 8.11** Lasing process of the argon-ion laser and a cross section of the plasma tube. A magnetic field supports concentration of the plasma current near the axis. The copper discs have additional bores to allow reflux of the argon ions.

elaborate technology that is necessary due to the high plasma temperatures. The inner bores of the plasma tube are protected by robust tungsten discs, which are inserted into copper discs in order to rapidly remove the heat. A magnetic field additionally focuses the plasma current onto the axis to protect the walls against erosion. Since, as a result of diffusion, the argon ions move to the cathode, the copper discs have holes for the compensating current. An argon laser consumes gas since the ions are implanted into the walls. Therefore, commercial ion lasers are equipped with an automatic reservoir. The gas pressure is 0.01–0.1 mbar.

### 8.2.1.3 Features and Applications

Most laser laboratories cannot be imagined without high-power visible laser sources because they are used to “pump” other tunable lasers, for example, dye lasers and Ti-sapphire lasers. Ar ions have several optical transitions up to the UV spectral range, which can be operated with a high output power. For this reason they used to dominate the market for fixed-frequency lasers with high pump power until about the year 2000, but frequency-doubled solid-state lasers (see Section 8.4.2.1) have proven to be much more economic.

## 8.2.2 Metal-Vapor Lasers

The copper- and gold-vapor lasers were commercially successful because they offer attractive specifications for high-power (quasi-)cw applications: they are pulsed lasers but with very high repetition rates of about 10 kHz – to our eyes, they are perfectly continuous. The pulse length is some 10 ns, and the average output power can be 100 W. The most important wavelengths are the yellow 578 nm line and the green 510 nm line ( $^2P_{1/2,3/2} \rightarrow ^2D_{3/2,5/2}$ ) of the copper atom.

The physical reasons for this success, which cannot necessarily be expected given the high operating temperature of the metal vapor of about 1500 °C, are the large excitation probability by electron impact (the discharge is supported,

e.g., by a neon buffer gas) and the high coupling strength of the dipole-allowed transitions.

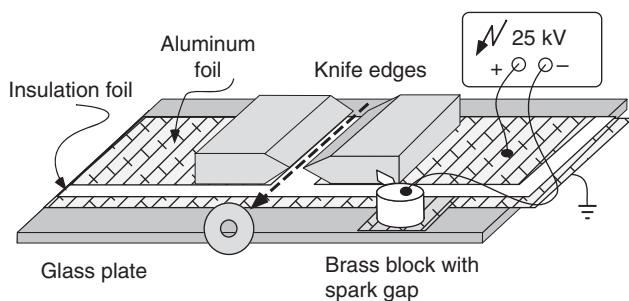
### 8.2.3 Molecular Gas Lasers

In contrast to atoms, molecules have vibrational and rotational degrees of freedom and thus a much richer spectrum of transition frequencies, which in principle also results in a complex spectrum of laser lines. The electronic excitations of many gaseous molecules are at very short wavelength, where the technology is involved, and in the interesting visible spectral range, not very many systems have been realized. Exceptions include the sodium-dimer laser ( $\text{Na}_2$ ), which however has not achieved any practical significance since sodium vapor contains a reasonable density of dimers only at very high temperatures, and the nitrogen laser, which is today used only for demonstration purposes. However, it is quite simple to construct!

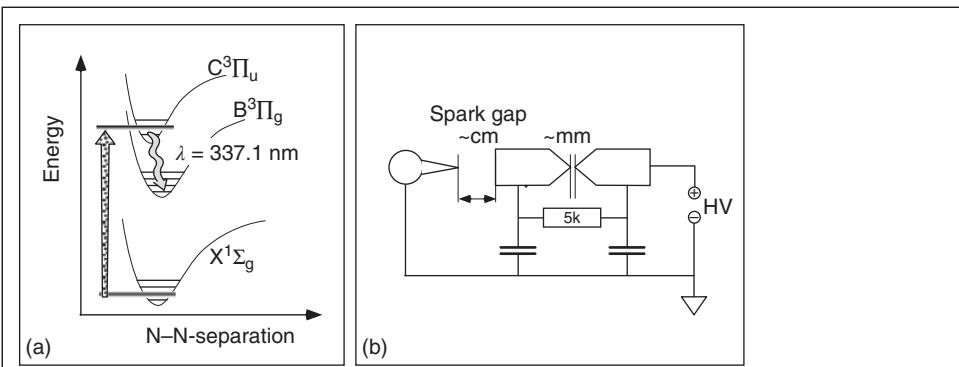
#### Excuse: Can a laser be operated just with air?

The short answer is Yes! The 78% nitrogen fraction of the air can be used as a laser amplifier. And even better, a primitive "air laser" is so simple to construct that with some skill (and caution because of the high voltage!), it can be copied in school or in scientific practical training. The original idea of a simple nitrogen laser was already presented with instructions in *Scientific American* in 1974 [117]. It is still costly insofar as a vacuum apparatus is necessary for the control of the nitrogen flow. In a practical project carried out by high school students [118], the laser – with slightly reduced output power – was operated directly with the nitrogen from the air.

In the simplest version a spark discharge along the knife edges shown in Figure 8.12 is employed. The spark is generated according to the circuit diagram in Figure 8.13. First, the knife edges are charged to the same high voltage potential. The breakdown of the air then takes place at the sharp tip of the spark gap so that between the two knife edges, the full voltage is abruptly switched on. The discharge runs along the knife edges and also turns off rapidly again due to the high-voltage sources used with large source impedance. Suitable high-voltage sources are available in many institutions, but a small voltage multiplier can also be built by oneself without any great expense. The central experimental challenge is a reproducible and stable discharge as experience shows.



**Figure 8.12** Simple homemade "air laser." Good parallel alignment of the knife edges is essential for successful operation.



**Figure 8.13** (a) Molecular potentials in the nitrogen molecule (schematic). The discharge generates inversion in the  $C^3\Pi_u/B^3\Pi_g$  system. (b) Circuit diagram of the air laser.

Between the knife edges a line-shaped occupation inversion of nitrogen molecules is generated, which leads to laser emission on the 337.1 nm UV line. In Figure 8.13 relevant molecular levels are shown along with their designations. The lower laser level is emptied only very slowly since the two involved states belong to the triplet system of the molecule (parallel electronic spins of paired electrons), which has no dipole transitions to the singlet ground state. Therefore, the inversion of energy levels cannot be maintained by a continuous discharge either, and laser operation breaks off after few nanoseconds.

Strictly speaking, the “mirror-less air laser” is not a laser but a so-called superradiator. Here, superradiance occurs as a result of spontaneous emission, which is amplified along the pencil-like inversion zone (*amplified spontaneous emission (ASE)*). It is emitted as a coherent and well-directed flash of light.

### 8.2.3.1 The $\text{CO}_2$ Laser

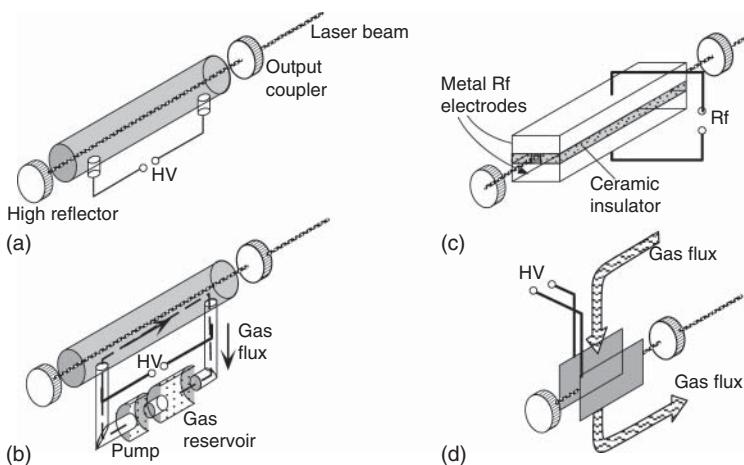
The most important examples of the molecular gas laser are the carbon monoxide and carbon dioxide ( $\text{CO}$  and  $\text{CO}_2$ ) lasers, which are based on infrared transitions between vibration–rotation energy levels. The  $\text{CO}_2$  laser is one of the most powerful lasers in general and thus plays an important role for materials processing with lasers [119].

### 8.2.3.2 Gain

The molecular states involved in the laser process of the  $\text{CO}_2$  laser can be found in Figure 8.15: a symmetric ( $v_1$ ) and an antisymmetric ( $v_3$ ) stretching vibration as well as a bending vibration ( $v_2$ ). Vibrational quantum states of the  $\text{CO}_2$  molecule are identified by quantum numbers ( $v_1, v_2, v_3$ ).

The (001) level decays by dipole transitions, which are very slow due to the  $\omega^3$  factor of the Einstein  $A$  coefficient, however, allowing convenient buildup of inversion in this level. The most important laser transition takes place between the (001) and the (100) level (Figure 8.14).

The  $\text{CO}_2$  lasers are excited by a discharge. The occupation of the upper laser level is possible directly but is much more favorable with the addition of nitrogen.

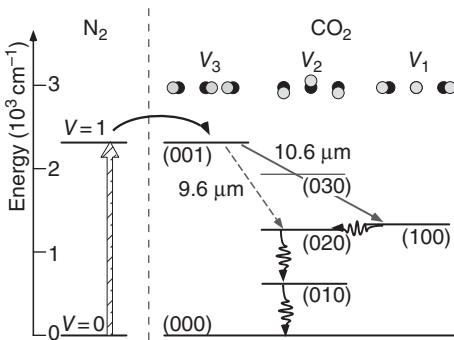


**Figure 8.14** Important configurations of CO<sub>2</sub> lasers. The conventional laser (a) is operated with a sealed tube and longitudinal discharge. To increase the output power, a longitudinal gas flow (b) or a radio-frequency wave guide laser (c) can be used. The highest power can be achieved if the gas flow and the discharge are operated transversely to the laser beam (transversely excited (TE) laser) (d).

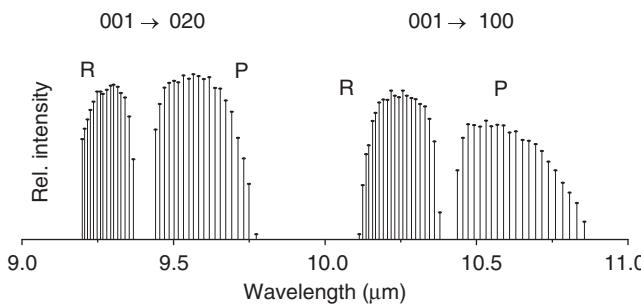
Metastable N<sub>2</sub> levels not only can be excited very efficiently in a discharge but also can transfer the energy to the CO<sub>2</sub> molecules very profitably as well.

The (100) level is emptied very rapidly by collisional processes. In addition, it is energetically adjacent to the (020) level, which itself ensures quick relaxation to thermal equilibrium also with the (000) and (010) levels. For this, the so-called vv relaxation plays an important role, which is based on processes of the (020) + (000) → (010) + (010) type. The heating of the CO<sub>2</sub> gas associated with these processes is not desirable because it increases the occupation of the lower laser level. It can be significantly reduced by adding He as a medium for thermal conductivity.

In Figure 8.15 we have completely neglected the rotational levels of the molecule, though they cause a fine structure of the vibrational transitions, leading to many closely adjacent laser wavelengths (Figure 8.16). A typical CO<sub>2</sub> laser makes available about 40 transitions from the P and S branches of the rotation-vibration spectrum. The gain bandwidth of each line (50–100 MHz) is very small since the Doppler effect no longer plays a significant role at low infrared frequencies. The laser lines of a CO<sub>2</sub> laser can be selected by a grating used as one of the laser mirrors.



**Figure 8.15** Transitions relevant for the CO<sub>2</sub> laser. Numbers in parentheses give quantum numbers ( $v_1, v_2, v_3$ ).



**Figure 8.16** Emission lines of the  $\text{CO}_2$  laser on the  $9.6$  and  $10.6\ \mu\text{m}$  lines. If one resonator mirror is replaced by a grating (with mirrorlike reflection in –first order), tuning can be quite easily achieved by rotating the grating. The terms “R” and “P” branch are taken from molecular spectroscopy. At the R branches of the spectrum, the rotational quantum number  $J$  of the molecule is decreased by 1, and it is increased by 1 at the P branches,  $J \rightarrow J \pm 1$ .

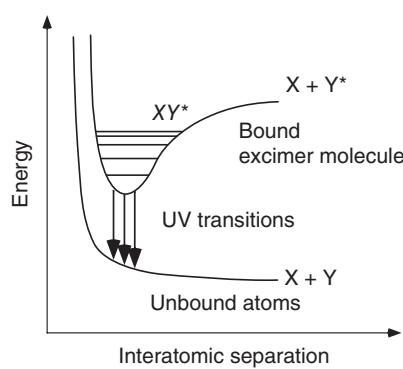
### 8.2.3.3 Operating Conditions

The  $\text{CO}_2$  laser is among the most powerful and robust of all laser types. It makes available a high and focusable energy density that is highly favorable for contactless materials processing and laser machining. Owing to the strong application potential, multiple technically different  $\text{CO}_2$  laser types have been developed, a selection of which is displayed in Figure 8.14. The operation of the  $\text{CO}_2$  laser is disturbed by induced chemical reactions. Thus the laser gas needs to be regenerated, either by maintaining a continuous flow through the laser tube or by adding some suitable catalysts to the gas, for example, a small amount of water, which oxidizes the undesired  $\text{CO}$  molecules back to  $\text{CO}_2$ . Output powers of some  $10\ \text{kW}$  are routinely achieved in larger laser systems.

### 8.2.3.4 The Excimer Laser

Excimer lasers play an important role for applications since they offer very high energy and furthermore the shortest UV laser wavelengths, although in the pulsed mode only. The term *excimer* is a short form of “excited dimer,” which means unusual diatomic molecules (dimers) that exist in an excited state only. Today the term has been transferred to all molecules that exist only excitedly, for example,  $\text{ArF}$  or  $\text{XeCl}$ , to mention just two examples important for laser physics.

The level scheme and the principle of the excimer laser are presented in Figure 8.17. Since the lower state is intrinsically unstable, the inversion condition is always fulfilled once the excimer molecules have come into existence. In order to generate them, the gas is pre-ionized with UV light to increase the conductivity and thus to increase the



**Figure 8.17** Laser process within the excimer laser.

efficiency of excitation in the following discharge. The lifetime of the excimer molecules is typically about 10 ns, which also determines the pulse period of this laser type.

Generation and handling of a gas of excimer molecules are quite costly. The gas is corrosive, and the laser medium ages after some thousands or millions of pulses (with typical repetition rates of 10–1000 pulses per second). That is why selected materials and sophisticated gas exchange systems are used. The strong demand for excimer lasers for medical applications and their increasing use as a light source for optical lithography in the semiconductor industry (see “Excursion” in Section 5.4.1.2) have already made the KrF laser at 248 nm a mature product. In the near future, the ArF (193 nm) laser as well as even the laser with the shortest commercially available wavelength, the  $F_2$  laser, will probably follow.

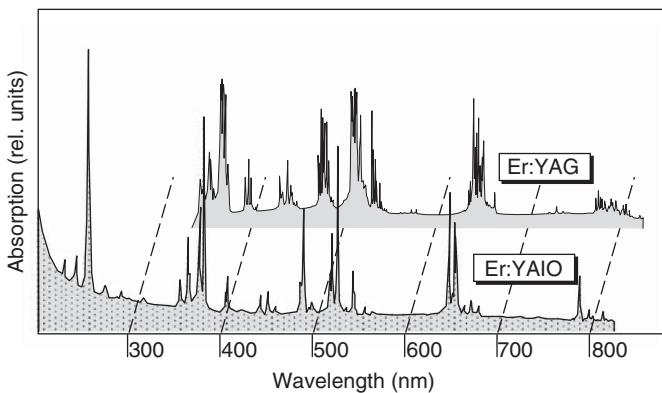
## 8.3 The Workhorses: Solid-State Lasers

The world’s first laser, constructed by Maiman (born 1927) [109], was a pulsed ruby laser. Its red light ( $\lambda = 694.3$  nm) is emitted by the chromium dopant ions of the Cr:Al<sub>2</sub>O<sub>3</sub> crystal, and therefore it was a solid-state laser. Though today the ruby laser plays a role for historical reasons only, solid-state lasers have received increasing attention since many types can be excited efficiently by diode lasers. Furthermore, electrical power inserted into the system can be converted into light power with an efficiency up to 20%. Solid-state lasers are thus among the preferred laser light sources because of their robust construction and economical operation.

### 8.3.1 Optical Properties of Laser Crystals

Optically active ions can be dissolved in numerous host lattices, and such systems can be considered as a frozen gas if the concentration of the former is not higher than a few percent at most. Nevertheless the density of these impurity ions within the crystal is much higher than the particle density within a gas laser and thus allows a higher gain density if there are suitable optical transitions. Of course, the host lattices have to have a high optical quality since losses by absorption and scattering impair laser oscillation. Impurity ions can be inserted into a host crystal particularly easily if they can replace a chemically similar element. Therefore many host materials contain yttrium, which can be replaced very easily by rare-earth metals.

Another important property of the host crystals is their thermal conductivity, since a large amount of the excitation energy is always converted into heat within the crystal. Inhomogeneous temperature distributions within the laser crystal cause, for example, lensing effects due to the temperature sensitivity of the index of refraction, which may alter the properties of the Gaussian resonator mode significantly. Since few laser media fulfill all requirements simultaneously, the growth of new and improved laser crystals is still an important field of research in laser physics.



**Figure 8.18** Absorption spectrum of the erbium ion  $\text{Er}^{3+}$  in the host materials YAG (top) and  $\text{YAlO}_3$  (bottom) according to [121].

In the simplest case the properties of the free ions are modified only slightly when the dopant ions are dissolved in the host material. The energy levels of the erbium ion in different systems may be taken as an example (Figure 8.18). This laser can be described very well using the concept of a “frozen” glass laser.

A most important group of dopant elements are the rare-earth ions. Their unusual electron configuration makes them very suitable for laser operation. Another group is formed by the ions of simple transition metals, which allow one to build laser systems tunable over large wavelength ranges. These are the so-called vibronic lasers including color center and Ti-sapphire lasers.

### 8.3.2 Rare-Earth Ions

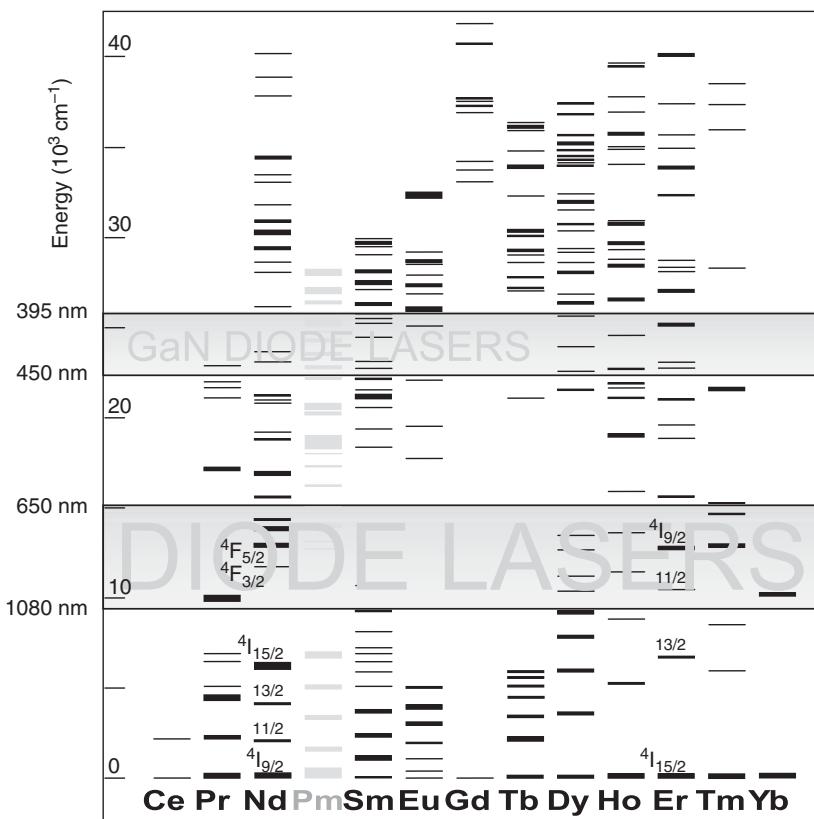
The 13 elements following lanthanum (La, atomic number 57) with  $N = 58$  (cerium, Ce) to 70 (ytterbium, Yb) are called the *lanthanides* or rare-earth metals.<sup>4</sup> Being impurity ions, the lanthanides are usually triply ionized, with electron configuration  $[\text{Xe}]4\text{f}^n$  with  $1 \leq n \leq 14$  for the  $n$ th element after lanthanum. The optical properties of an initially transparent host crystal are determined by the 4f electrons, which are localized within the core of these ions and thus couple only relatively weakly to the lattice of the host crystal.

To a good approximation the electronic states are described by  $LS$  coupling and Hund’s rules [95]. Because of the large number of electrons that each contribute orbital angular momentum  $\ell = 3$ , there are in general a multitude of fine-structure states, which lead to the wealth of levels shown in Figure 8.19.

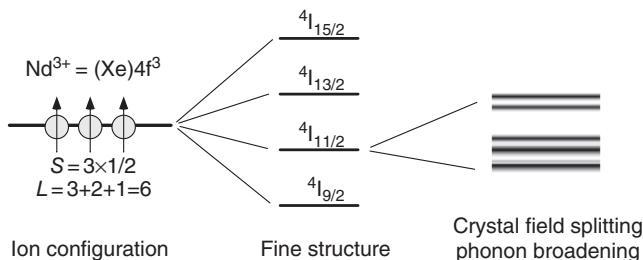
#### Example: Energy levels of the neodymium $\text{Nd}^{3+}$ ion

The  $\text{Nd}^{3+}$  ion has three electrons in the 4f shell. According to Hund’s rules, they couple in the ground state to the maximum total spin  $S = 3/2$  and total orbital angular momentum  $L = 3 + 2 + 1 = 6$ . From the  $^4\text{I}$  multiplet the ground state

<sup>4</sup> The rare-earth metals are not rare at all within the Earth’s crust. Since their chemical properties are very similar, it was difficult for quite a long time to produce them with high purity. The element promethium (Pm) cannot be used because of its strong radioactivity.



**Figure 8.19** Energy levels of the rare-earth metals with selected designations. The range currently (2006) accessible by diode laser radiation for optical excitation is marked. (Energy levels from [122].)



**Figure 8.20** Energy levels of neodymium  $\text{Nd}^{3+}$  ions within the solid state. The details of the splitting depend on the host lattice.

is expected to be at  $J = 9/2$  due to the less than half-filled shell. Unlike the free atom or ion, the magnetic degeneracy in the  $m$  quantum number is lifted by anisotropic crystal fields in the local vicinity. The coupling to the lattice oscillations (“phonons”) leads eventually to the homogeneously broadened multiplets in Figure 8.20.

The rigorous dipole selection rules of the free atom ( $\Delta\ell = \pm 1$ ) are lifted by the (weak) coupling of the electronic states to the vicinity of the electrical crystal field that causes a mixture of  $4f^n$  and  $4f^{n-1}5d$  states. The energy shift by this interaction is quite small, but for the radiative decay, the dipole coupling is now predominant and reduces the lifetimes of the states dramatically to the range of some microseconds (see Problem 8.5). Therefore, intensive absorption and fluorescence of the rare-earth ions can be observed on transitions between the fine-structure levels.

On the other hand, fluorescence cannot be observed from every level since there are competing relaxation processes caused by the coupling of the ionic states to the lattice oscillations, or phonons, of the host lattice, which can lead to fully radiation-free transitions. The more probable those processes are, the more the fine-structure levels are lying next to each other. In Figure 8.18 the fluorescence lines are quite narrow, showing that the atomic character of the ions is largely maintained.

## 8.4 Selected Solid-State Lasers

From Figure 8.19 it can easily be imagined that there are countless laser media containing rare-earth ions [123]. We have selected special solid-state lasers that play an important role as efficient, powerful, or low-noise fixed-frequency lasers. Those lasers are used, for example, as pump lasers for the excitation of tunable laser systems or for materials processing that demands intensive laser radiation with good spatial coherence properties. Tunable lasers, which increasingly employ solid-state systems, will also be discussed in a subsequent section about vibronic lasers (Section 8.5).

### 8.4.1 The Neodymium Laser

The neodymium laser is among the systems already developed in the very early days of the laser. It was originally excited with high-pressure noble-gas lamps. Only a small amount of their light energy was absorbed, while the larger part was removed as heat and dissipated. Although the idea to excite neodymium lasers by laser diodes arose quite early, it could not be realized until the advent of reliable high-power laser diodes by the end of the 1980s. Nowadays, neodymium lasers are primarily excited by efficient, and indispensable, laser diodes, and no end to this successful development is in sight even after more than 10 years. In Table 8.3 we have already presented hosts that have great significance for practical applications.

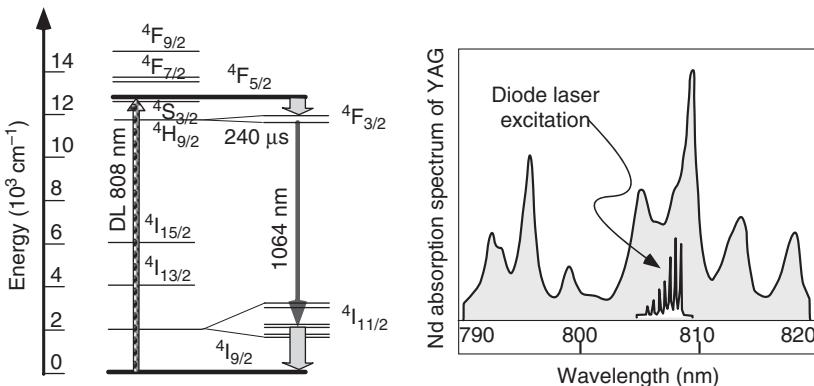
#### 8.4.1.1 The Neodymium Amplifier

The energy level structure of neodymium ions has already been presented (Section 8.3.2). We have already mentioned as well that with ions within the solid state, a much higher density of excited atoms can be achieved than within the gas laser. In most of the host crystals, this is valid for concentrations up to a few percent. Above this level the ions interact with each other, causing detrimental non-radiative relaxations. But there are also special materials, for example,

**Table 8.3** Selected host materials [120].

Host	Formula	Thermal conductivity (W cm <sup>-1</sup> K <sup>-1</sup> )	$\partial n / \partial T^a)$ (10 <sup>-6</sup> K <sup>-1</sup> )	Ions
Garnet	YAG	Y <sub>3</sub> Al <sub>5</sub> O <sub>12</sub>	0.13	Nd, Er, Cr, Yb
Vanadate	YVO	YVO <sub>4</sub>		Nd, Er, Cr
Fluoride	YLF	LiYF <sub>4</sub>	0.06	-0.67(o) -2.30(e) Nd
Sapphire	Sa	Al <sub>2</sub> O <sub>3</sub>	0.42	13.6(o) 14.7(e) Ti, Cr
Glass		SiO <sub>2</sub>	Typ. 0.01	3–6 Nd, Yb

a) (o), (e): Ordinary (extraordinary) index of refraction in birefringent materials.



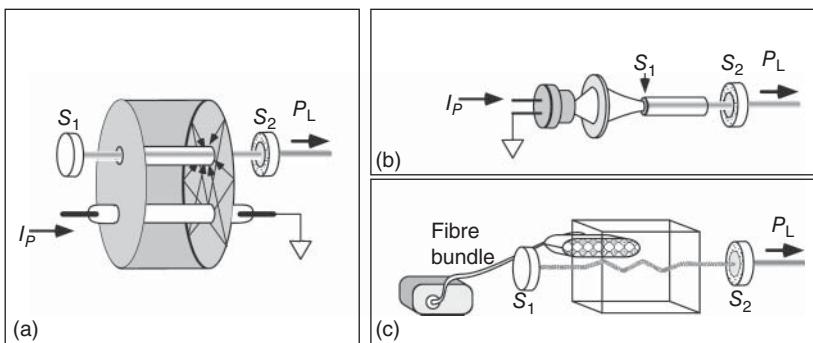
**Figure 8.21** Laser transitions of neodymium lasers and absorption spectrum.

Nd:LSB (Nd:LaScB), which stoichiometrically contain 25% neodymium. Owing to the extremely high gain density of these materials, remarkably compact, intense laser light sources can be built.

The  $^4\text{I}_{9/2} \rightarrow ^4\text{F}_{5/2}$  transition of the Nd<sup>3+</sup> ion can be excited very advantageously by diode lasers at the wavelength of 808 nm where the upper  $^4\text{F}_{3/2}$  laser level is populated very rapidly by phonon relaxation. Since the lower  $^4\text{I}_{11/2}$  level is emptied just as quickly, the neodymium laser makes an excellent four-level laser system (Figure 8.21).

#### 8.4.1.2 Configuration and Operation

Because of its widespread application potential, there are numerous technical variants of the neodymium laser. Before diode laser pumps were available with sufficient quality, the crystal within a cw laser was generally excited by a high-pressure Xe lamp placed at the second focus of an elliptical cavity in order to achieve a high coupling efficiency (Figure 8.22a).



**Figure 8.22** Configurations of neodymium lasers.  $P_L$ : laser output;  $S_{1,2}$ : laser end mirrors;  $I_p$ : drive currents. (a) Pump lamp and laser bar located at the two foci of an elliptical resonator. (b) Neodymium laser longitudinally pumped with a diode laser. (c) In the slab laser, the pump energy is supplied transversely. The laser beam is guided by total internal reflection within the crystal.

Using diode lasers, life has become much more easy in this respect. In Figure 8.22b such a linear laser rod pumped from the end (“end-pumped laser”) is presented. One of its end mirrors is integrated into the laser rod. With this arrangement the pump power is inhomogeneously absorbed so that the gain also varies strongly along the laser beam. That is why a Z-shaped resonator (see Figure 8.23) is often used to allow symmetrical pumping from both ends.

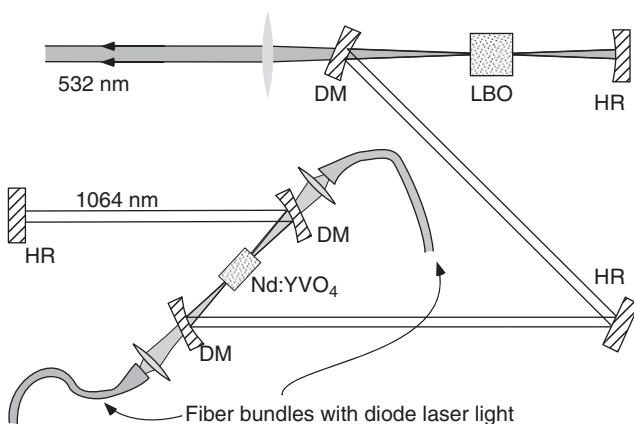
Further enhanced output power can be achieved by using the so-called slab geometries in which the pump energy is supplied transversely. With this layout the light of several pumping diode lasers can be used at the same time. It is technically advantageous to operate the heat-producing laser diodes spatially separated from the laser head. The pump light is then transported to the laser amplifier by fiber bundles in a literally flexible way, and optimal geometric pump arrangements can be used. Even with considerable output power of several watts, the laser head itself measures no more than  $50 \times 15 \times 15 \text{ cm}^3$ . The end of technological developments in this field is still not in sight.

#### 8.4.2 Applications of Neodymium Lasers

Neodymium lasers have been used in countless applications for a long time, and the more recent advent of efficient diode laser pumps at the wavelength of 808 nm has lent additional stimulus. Here we present two recent examples that symbolize the large range of possible applications: one is the powerful frequency-doubled neodymium laser, which has replaced the expensive argon-ion laser technology more and more, and the other is the extremely frequency-stable monolithic *miser*.

##### 8.4.2.1 Frequency-Doubled Neodymium Lasers

In Figure 8.23 we have presented a neodymium laser concept that allows the generation of very intense visible laser radiation at  $1064/2 = 532 \text{ nm}$ . The pump energy is applied to the Nd:YVO<sub>4</sub> material through fiber bundles, and the Z-shaped resonator offers a convenient geometry to combine the power of



**Figure 8.23** Powerful frequency-doubled neodymium laser. The light of the laser diodes is supplied through fiber bundles over dichroic mirrors (DM). Within the resonator, 1064 nm light circulates at high intensity. The nonlinear crystal (Lithium Triborate  $\text{LiB}_3\text{O}_5$ , LBO) is used for the frequency doubling. HR = high reflector [124].

several diode lasers and generate very high power at the fundamental wavelength at 1064 nm. In one arm of the laser, the light is focused into a nonlinear crystal (here LBO; see Section 15.4), which causes efficient frequency doubling.

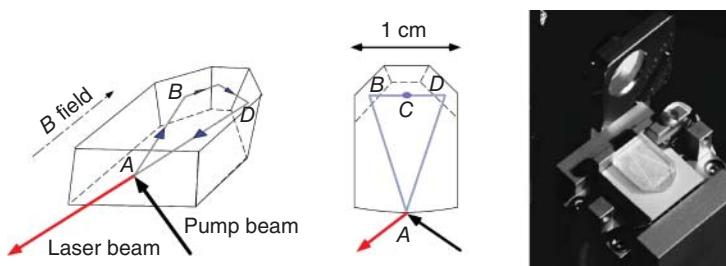
In principle, it has been clear for a long time that intense visible laser radiation can be generated with the concept presented here. Before it could be used for producing commercial devices, however, not only technological problems caused by the large power circulating in the resonator had to be solved but also physical issues such as the so-called greening problem. This is caused by mode competition [125], leading to very strong intensity fluctuations. It can be solved by operating the laser either at a single frequency or at a large number of simultaneously oscillating frequencies.

#### 8.4.2.2 The Monolithically Integrated Laser (Miser)

The passive frequency stability of any common laser (i.e., in the absence of active control elements) is predominantly determined by the mechanical stability of the resonator, which undergoes length variations due to acoustic disturbances such as environmental vibrations, sound, and so on.

It is therefore advantageous to build laser resonators as compact and also as light as possible, since devices with a small mass have higher mechanical resonance frequencies, which can be excited less easily by environmental acoustic noise. In the extreme case, the components of a ring laser (see Section 8.6) – laser medium, mirrors, and optical diode – can even be integrated into one single crystal. Kane and Byer [126, 127] realized this concept in 1985 and called it *miser*, a short form of the term *monolithically integrated laser*.

The miser is pumped by diode laser light. The ring resonator is closed by using total internal reflection at suitably ground and polished crystal planes. An interesting and intrinsic optical diode is also integrated into the device. The so-called out-of-plane configuration of the resonator mode (in Figure 8.24 the trajectory



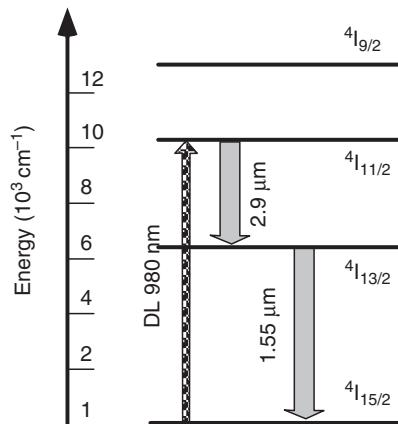
**Figure 8.24** Monolithic neodymium ring laser. The beam is led out of the plane between *B* and *D*. In connection with a magnetic field (see text), this “out-of-plane” configuration realizes an optical diode and makes unidirectional operation possible. (With permission by Innolight GmbH.)

*BCD*) causes a rotation of the polarization of the laser field due to the “slant” reflection angles in analogy to a  $\lambda/2$  device. In addition, a magnetic field in the direction of the long axis of the miser causes nonreciprocal Faraday rotation. In one of the directions, the rotations compensate each other, whereas in the other direction, they add. Since the reflectivity of the exit facet depends on the polarization, one of the two directions is strongly favored in laser operation.

#### 8.4.3 Erbium Lasers, Erbium-Doped Fiber Amplifiers (EDFAs)

Erbium (Er) ions can be dissolved in the same host crystals as neodymium ions and are especially interesting for applications at infrared wavelengths near  $1.55\text{ }\mu\text{m}$ . They can be excited by strained quantum well (SQW) laser diodes (see Section 10.3.3.2) at  $980\text{ nm}$  very efficiently. Another long-wavelength laser transition at  $2.9\text{ }\mu\text{m}$  is used mostly for medical applications (Figure 8.25). A favorable feature is the eye-safe operation at these long wavelengths. A significant technological breakthrough was achieved by Payne and Desurvire [128] in 1989 when they were able to demonstrate amplification at the wavelength of  $1550\text{ nm}$  by using Er-doped optical fibers.

**Figure 8.25** Part of the energy level scheme of the erbium laser showing the two important laser transitions. The exact transition wavelength depends on the host crystal.



*Erbium-doped fiber amplifiers* (EDFAs) have very soon become an important amplifier device (gain typically 30–40 dB) for the long-distance transmission of data. It is due to them that today the residual losses of optical fibers in this third telecommunications window (see Figure 4.7 on p. 130) do not impose any constraints on the achievement of even the highest possible transmission rates over large distances. This breakthrough again was not conceivable without the availability of inexpensive and robust diode lasers for excitation, as we will discuss in more detail in the following section about fiber lasers.

#### 8.4.4 Fiber Lasers

The total gain of an optical wave in a laser medium is determined by the inversion density (which determines the gain coefficient) and the length of the amplifying medium. Snitzer already mentioned in 1961 [129] that optical wave guides or fibers with suitable doping of the core should offer the best qualifications to achieve high total gains.

Although the attractive concept of fiber lasers had already been recognized quite early, the advent of robust and convenient diode lasers was instrumental in making fiber lasers attractive devices. Even the mediocre transverse coherence properties of an array of laser diodes (see Section 10.6) are far superior to conventional lamps as used in the conventional neodymium laser configuration in Figure 8.22 with regard to focusability and can be used for efficient excitation of the small active fiber volumes.

Fiber lasers are a field of active technological development that is ongoing, and an excellent account of the state of art is given in [130]. We will limit ourselves to the presentation of a few specific concepts since the layout of a fiber laser does not differ basically from other laser types – one might say that it just has a very long and thin amplifying medium [131].

##### 8.4.4.1 Cladding Pumping

An interesting trick to make the application of the pump energy more efficient has been developed with the so-called cladding pumping (Figure 8.26). It is quite obvious that monomode fibers with a narrow core should be used as the active medium for the sake of obtaining good-quality transverse laser modes. But then efficient coupling of the pump laser radiation from high-power laser diodes becomes difficult, since direct concentration or focusing of their power into the small active fiber core volume is difficult.

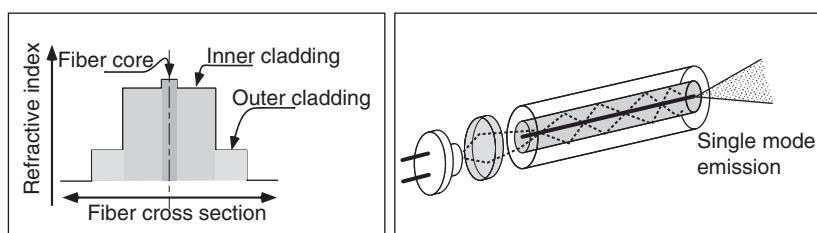


Figure 8.26 Cladding pumping in the fiber laser.

**Table 8.4** Elements and wavelengths of selected fiber lasers.

Wavelength ( $\mu\text{m}$ )	Element	Wavelength ( $\mu\text{m}$ )	Element
3.40	Er	0.85	Er
2.30	Tm	0.72	Pr
1.55	Er	0.65	Sm
1.38	Ho	0.55	Ho
1.06	Nd	0.48	Tm
1.03–1.12	Yb	0.38	Nd
0.98	Er		

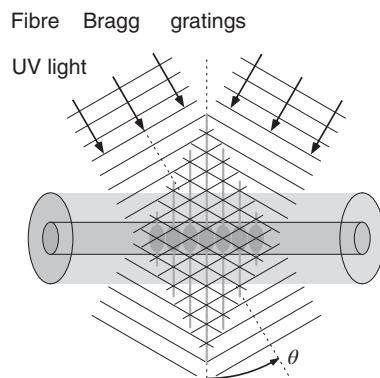
This problem can be overcome by using a double fiber cladding that generates a multimode wave guide around the active fiber core. The pump power is coupled into this multimode fiber, and again and again, it scatters into the core and is absorbed there. To optimize the scattering, the core is given, for instance, a slightly star-shaped structure instead of a purely cylindrical one.

Table 8.4 contains a number of available wavelengths for fiber lasers widely spread across the infrared and visible spectra. With fiber laser media not only very low threshold values for lasing are achieved but also a remarkable output power of several tens of watts. Fiber lasers have not reached the end of their development by any means. Continued interest is also being shown in the development of light sources for blue wavelengths. There are several concepts, for example, the so-called “up-conversion” lasers, which can emit blue or even shorter-wave radiation from higher energy levels excited by stepwise absorption of several pump photons.

#### 8.4.4.2 Fiber Bragg Gratings

To make fiber lasers practicable, many relevant components for the control of a light beam, for example, mirrors, output couplers, and modulators, have been directly integrated into the fiber. For their detailed discussion, we refer the reader to the specialist literature [132] and restrict ourselves to the example of fiber Bragg gratings (FBGs) [133] used as efficient mirrors and spectral filters.

The Bragg grating is realized by a periodic modulation of the refraction coefficient along the direction of propagation. For this, the Ge-doped fiber core is exposed to two intense UV beams crossing each other at an adjustable angle  $\theta$  (Figure 8.27). The UV light induces changes that can be of



**Figure 8.27** Production of a Bragg reflector (qualitatively).

chemical or photorefractive nature,<sup>5</sup> and are proportional to the local intensity of the standing-wave field. The period  $\Lambda$  of the Bragg mirror can be determined by the choice of the crossing angle and the UV wavelength  $\lambda$ ,  $\Lambda = \lambda/(2 \sin \theta)$ .

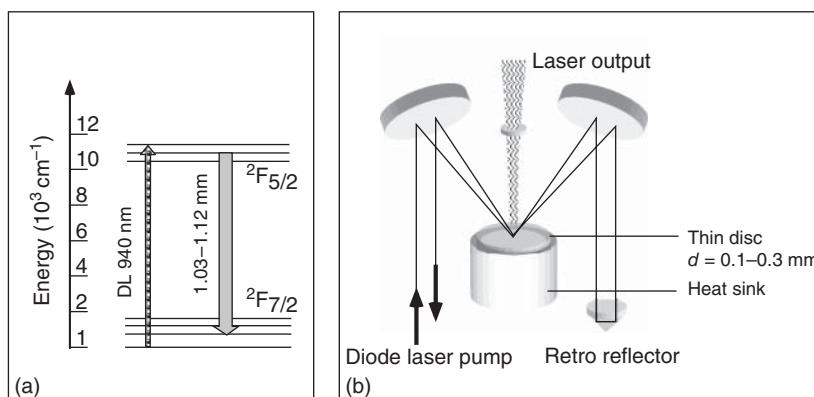
#### 8.4.5 Ytterbium Lasers: Higher Power with Thin-Disc and Fiber Lasers

In a more recent development, the dominance of the Nd lasers has been challenged by Yb lasers, which can be dissolved within the host materials listed in Table 8.3 as well as rare-earth ions. Many physical and technical details about Yb can be found in [130].

This relatively new laser material offers putative advantages in the quest for ever more output power: Yb-ion doping can reach 25% and thus strongly exceeds the 1–2% limit for Nd ions, offering higher gain density. Also, the Yb ions are excited at 940 nm (Nd: 808 nm), while lasing takes place typically between 1030 and 1120 nm. Therefore less heat is generated, which generally impairs the laser process. Finally, Yb suffers less from excitation into non-lasing states and reabsorption of fluorescent light than Nd.

Technological breakthroughs have furthermore supported the advancement of Yb lasers. The thin-disc technology and improved performance of fiber lasers reduce the problems of heat dissipation associated with laser rods in high-power applications. Commercially available output power exceeds 4 kW, while excellent coherence properties are preserved.

The most important advantage of the thin-disc concept (Figure 8.28) compared to conventional laser rods is the much improved removal of heat generated in the excitation process. The thin disc is mounted on a heat sink and has a favorable surface-to-volume ratio. Heat gradients occur in the longitudinal rather than in the transverse direction, and hence the Gaussian resonator is much less disturbed



**Figure 8.28** (a) Part of the energy level scheme of the ytterbium laser showing the pump and the laser transition. The exact transition wavelength depends on the host crystal. (b) Cavity and pump geometry for a thin-disc laser. The pump is sent through the crystal in multi-pass configuration for efficient absorption.

5 For the photorefractive effect, charges are released by illumination and transported within the crystal lattice, which causes a spatial modification of the refractive index up  $10^{-3}$ .

by heat gradients, resulting in laser beams with excellent mode purity. As a consequence of the thin-disc arrangement, single-pass absorption of the diode laser pump light is relatively small. However, the multi-pass configuration (several tens of times in applications) allows the pump energy to be delivered with up to 90% efficiency to the small laser volume.

## 8.5 Tunable Lasers with Vibronic States

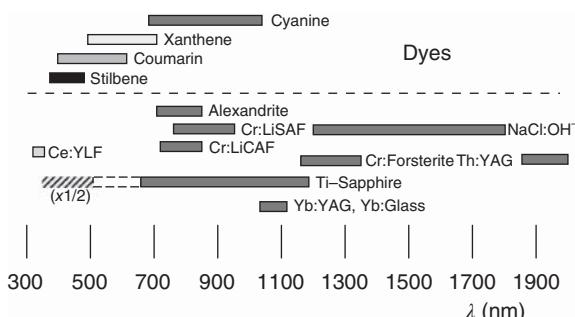
Until around 1990, the market for tunable laser light sources was dominated by dye lasers due to their convenience and – using multiple chemicals – tunability across the visible spectrum. Since then technical development favored solid-state systems, which are particularly interesting if they can be excited by diode lasers.

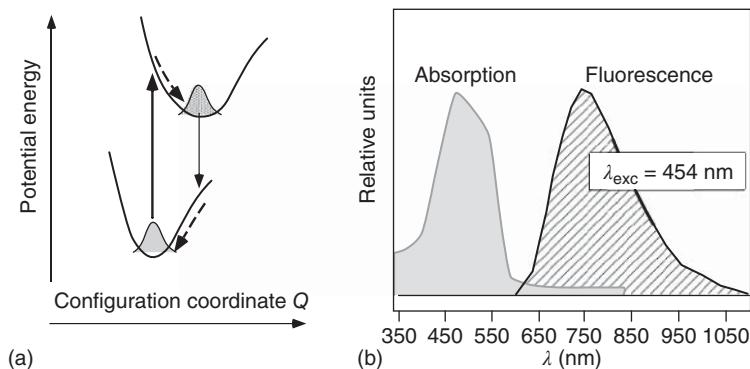
The tunability of so-called vibronic laser materials comes from the strong coupling of electronic excited states of certain ions (especially 3d elements) to the lattice oscillations. In principle, also semiconductor or diode lasers are among them, and we shall devote an extra chapter to them due to their special significance. Even the dye laser can be conceptually assigned to this class since their band-like energy scheme is generated by the oscillations of large molecules. Figure 8.29 offers an overview about important tunable laser materials. Here we present some important systems with their physical properties, and we explain the technical concepts of widely tunable ring lasers in which these laser materials are normally used.

### 8.5.1 Transition-Metal Ions

The 3d transition metals lose their outer 4s electron in ionic solid states and additionally some 3d electrons, their configuration being  $[Ar]3d^n$ . Often the third as well as the fourth ionization state of these ions can be found. Crystal fields have much more effect on the 3d electrons than on the 4f electrons of the rare-earth metals, since those form the outermost shell of electrons. The coupling to the lattice oscillations (which are described by a configuration coordinate  $Q$ ) leads to a band-like distribution of states (Figure 8.30). The transitions are called “vibronic.” On the one hand, these transitions have a large bandwidth, which accounts for broadband absorption as well as for fluorescence. No less important is the very short relaxation time, which leads to the thermal

**Figure 8.29** Tuning ranges of selected laser systems (pulsed and continuous wave). The frequency-doubled range of the Ti-sapphire laser is hatched.

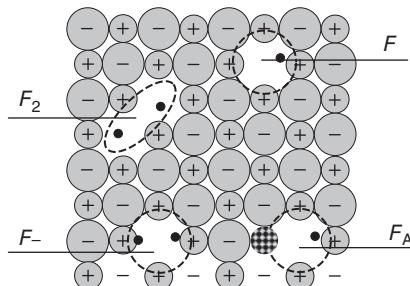




**Figure 8.30** (a) Vibronic states of solid-state ions. The shaded curves indicate the (quasi-)thermal distributions in the configuration coordinate  $Q$ . If the equilibrium positions at the ground and the excited state do not coincide, absorption and emission wavelengths are well separated and offer optimum conditions for a four-level laser system. Relaxation to a thermal or quasi-thermal distribution takes a few picoseconds only. (b) Absorption and fluorescence spectra of a Ti-sapphire crystal. The fluorescence spectrum was excited at a pump wavelength of 454 nm.

equilibrium position of the vibronic states within picoseconds. The chromium ions and especially the titanium ions have belonged to this important class of laser ions since the first demonstration in the 1980s [134]. The extraordinary position of the Ti–sapphire laser can clearly be recognized in Figure 8.29, too.

### 8.5.2 Color Centers



**Figure 8.31** Models of some color centers.

In contrast to the optical impurities of rare-earth and transition metals, color centers are generated not by impurity atoms but by vacancy lattice sites. They have been investigated for a long time. In an ionic crystal such vacancies have an effective charge relative to the crystal to which electrons or holes can be bound. Different types are collected in Figure 8.31. Like the transition-metal ions, the electronic excitations have a broadband vibronic structure and are well suitable for the generation of laser radiation.

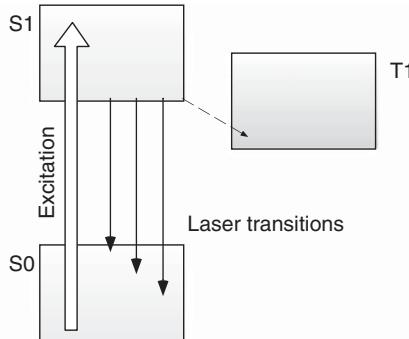
Color center lasers have to be held at the temperature of liquid nitrogen (77 K), and some of them even require an auxiliary light source. By this means color centers are brought back from parasitic states in which they can drop through spontaneous transitions and do not take part in the laser cycle. Since the operation is thus technologically quite costly, color center lasers have almost vanished. They have been replaced by alternatives including quantum cascade lasers (Section 10.3.4) and efficient optical parametric oscillators (see Chapter 15).

### 8.5.3 Dyes

Especially for wavelengths of 550–630 nm, the dye laser is still a tunable light source with little competition. In this range of the visible spectrum, our color sensual perception changes quickly from green to yellow to red.

For this reason the light of dye lasers is superior to all solid-state lasers so far developed with regard to aesthetic and emotional quality. Dyes are organic molecules with a carbon–carbon double bond, that is, with a pair of electrons. In Figure 8.32 the typical energy level scheme of a dye is presented. The paired ground state ( $S_0$ ) consists of a  $^1S_0$  state, that is, orbital angular momentum and total spin vanish. The dye molecules are dissolved (in alcohol or, if they are ejected from a nozzle into free space, in liquids with a higher viscosity such as glycol). The electronic states have a vibration–rotation fine structure that is broadened to continuous bands because of the interaction with the solvent similar to the vibronic ions. After absorption, the molecules relax rapidly to the upper band edge where the laser emission takes place. Some classes of dye molecules are also shown in Figure 8.29.

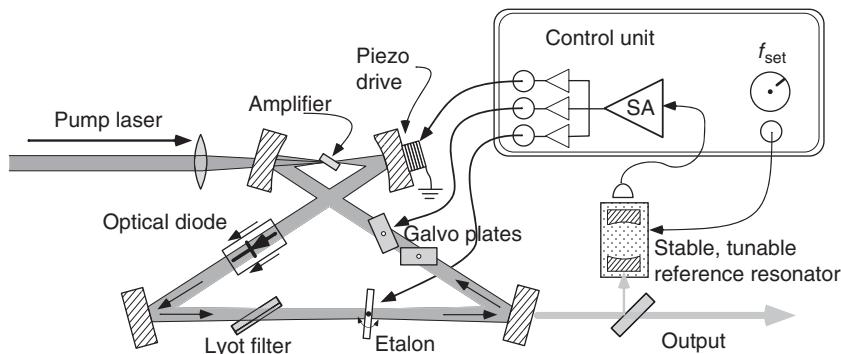
In complete analogy to two-electron atoms like helium, there are a singlet and a triplet system in dye molecules [95], only the transitions between them (intercombination lines) are not as strongly suppressed. The lifetime of the triplet states is very long, however, so that the molecules accumulate there after several absorption–emission cycles and no longer take part in the laser process. Pulsed dye lasers can be operated in an optical cell, which is stirred, but with cw laser operation, glass cells rapidly alter. Instead, creation of a jet stream expanding freely into air has been successful. The liquid is ejected into a jet stream from a flat “nozzle” into the focus of a pump laser and laser resonator. The surface of the jet stream has optical quality. One of the most robust dye molecules is rhodamine 6G, which delivers an output power of up to several watts. Furthermore, it can be used for a long time, in contrast to many other dyes, which age rapidly.



**Figure 8.32** Laser process in the dye laser (schematically).  $S_0$ ,  $S_1$ ,  $T_1$ : molecular electronic singlet and triplet states, respectively.

## 8.6 Tunable Ring Lasers

The success of vibronic laser materials is closely related to the success of the ring laser, which allows user-friendly setting of a wavelength or frequency. It is quite remarkable that with this device, the fluorescence spectrum of these materials, which has a spectral width of some 10–100 nm or 100 THz, can be narrowed to some megahertz, that is, up to eight orders just by a few optical components!



**Figure 8.33** Ring laser system with optical components for frequency control. SA = servo amplifier.

In contrast to the linear standing-wave laser, a traveling wave propagates in the ring laser. In the linear laser the so-called hole burning occurs since the amplification has no effect in the nodes of the standing-wave field (see Problem 8.1). For this reason the gain profile is periodically modulated and makes the oscillation of another spectrally adjacent mode possible, which fits to the periodic gain pattern and the resonator. In the ring laser the entire gain volume contributes to a single laser line. Therefore it is the preferred device for spectroscopic applications with high spectral resolution.

In Figure 8.33 we show one of numerous proven concepts of a ring laser. This layout is usually called a *bowtie* resonator. The foci of pump laser beam and laser mode are tightly overlapped between two spherical mirrors, which in the rest of the laser generate a Gaussian mode propagating with low divergence, which is also coupled out by one of the partially transparent resonator mirrors. To avoid losses the amplifier (Ti-sapphire rod, color center crystal, or dye laser jet) and other optical components are inserted into the resonator at the Brewster angle. An optical diode (see Section 3.5.6) allows unidirectional operation.

For wavelength control in general, several optical components with hierarchical spectral resolution (free spectral range) are used, which we list here: a Lyot or birefringent filter (p. 108) ensures rough spectral narrowing, and one or two etalons (thin and thick etalon; see Section 6.5) with different free spectral ranges select a single resonator mode. For tuning from the megahertz to gigahertz scale, the resonator length can be varied with different elements: a pair of so-called galvo plates varies the light path by a small synchronous rotation of the glass plates mounted at the Brewster angle, albeit at relatively low scanning speed. More rapid tuning is required for frequency stabilization. For this purpose the resonator length is adjusted by means of a light mirror mounted on a piezo translator, which allows implementation of 100 kHz bandwidth servo loops; even higher actuation speed may be achieved by phase modulators (EOMs; see Section 3.5.1) inserted into the resonator.

In experiments a voltage-controlled variation of the laser wavelength is desirable. For this purpose the so-called *feed-forward* values are applied to the optical components of the ring laser. Simultaneously, the laser frequency is also

compared with the voltage-controlled reference frequency of a passive optical resonator (e.g., according to p. 260) and is servoed to this value by suitable electronic feedback circuitry. With this method, typical continuous tuning ranges of 30 GHz or  $1\text{ cm}^{-1}$  are achieved, which offer excellent conditions for experiments in high-resolution spectroscopy.

## Problems

- 8.1 Spatial hole burning** Consider a laser crystal with homogeneously broadened gain profile. The crystal is 1 cm long and mounted at the center of a 15 cm long, linear resonator. The center wavelength of emission is  $1\text{ }\mu\text{m}$ .  
 The standing wave of the operating laser causes spatially inhomogeneous reduction of amplification so that further modes can start oscillation. What gain bandwidth is necessary to sustain oscillation of more than one mode? Sketch the intensity distribution and the inversion density for the laser crystal. What is the difference if the crystal is mounted directly in front of one of the mirrors?
- 8.2 Single-mode operation** Even for a gain medium with very large band width  $\Delta_G$ , single-mode operation of the laser can in principle be obtained by construction of a correspondingly short resonator. However, this is technically not very practicable.  
 Consider the example of a Ti-sapphire laser with  $\Delta_G = 47\text{ THz}$ . Instead of using a short resonator, an etalon is introduced into the linear resonator of length  $L$  and with a free spectral range that satisfies  $\Delta_{\text{FSR}} \geq \Delta_G/2$ . Show that in this case the condition  $L \leq cF/\Delta_G$  suffices to obtain single-mode operation.
- 8.3 Gain of the laser** The losses of a resonator built from two mirrors with reflectivities  $R_1 = 100\%$  and  $R_2 = 99\%$  are dominated by mirror transmission. The laser line transition of the gaseous medium has a cross section of  $\sigma = 10^{-12}\text{ cm}^2$ , the gas pressure is 1 mbar, and the length of the gas tube is 10 cm. (a) Calculate the inversion density necessary for laser operation above threshold. (b) How large is the density of particles in the upper laser level if the lower one is emptied instantaneously? (c) How large is the amplification at threshold?
- 8.4 Properties of the Nd:YAG laser** The total lifetime of the upper  ${}^4\text{F}_{3/2}$  laser level (Figure 8.21) is dominated by radiative decay (98% probability) and is  $240\text{ }\mu\text{s}$ . The branching ratio for the most important line at  $1.06\text{ }\mu\text{m}$  is about 14%. At 300 K it is homogeneously distributed across the bandwidth of 200 GHz. (a) Calculate the fluorescence rate of the  $1.06\text{ }\mu\text{m}$  line. (b) Calculate the absorption cross section at  $1.06\text{ }\mu\text{m}$ . (The index of refraction for YAG is  $n = 1.82$ .) (c) Suppose that in a laser crystal 1% of all  $\text{Y}^{3+}$  ions have been replaced with the  $\text{Nd}^{3+}$  laser ion. What is the single-pass amplification if

the inversion is 1%? (The density of YAG is  $4.56 \text{ g cm}^{-3}$ , and the length of the crystal 1 cm.) Neglect occupation of the lower laser level.

- 8.5 Transition matrix element of rare-earth ions** The laser transitions of the rare-earth ions occur between fine-structure energy levels of the electronic 4f state with  $\Delta\ell = 0$ , that is, they are not allowed according to the dipole selection rules of quantum mechanics for radiative transitions. Owing to the nonspherical symmetry, however, a bit of 5d character is mixed into the 4f wave function. Estimate the mixing coefficient by comparing a strong dipole transition (dipole moment = charge  $e$   $\times$  Bohr radius  $a_0$ ) at the Nd:YAG 1.06  $\mu\text{m}$  line with the observed radiative lifetime of 250  $\mu\text{m}$ .
- 8.6 Which laser for which problem?** You are planning an experiment in atomic physics or quantum optics and want to buy a laser to efficiently excite the atoms. Which type of laser do you get in order to work with the atoms from Table 7.3. Justify your selection when several alternatives exist.
- 8.7 Compensation of astigmatism** The ring laser construction of Figure 8.33 causes astigmatism. Thus, for the field components parallel and orthogonal to the plane of the resonator, there exist different conditions for stability (see Section 1.9.5). Identify the origin of the astigmatism and make suggestions for optical elements compensating the astigmatism. (Hint: See Figure 1.15.)
- 8.8 Servo speed of piezo elements in stabilized lasers** Consider the piezo mirror of the tunable ring resonator of Figure 8.33. The piezo tube with expansion coefficient  $0.8 \mu\text{m}/100 \text{ V}$  has capacity  $15 \text{ nF}$ . The weights of the tube and the attached mirror are 2 and 4 g, respectively. Estimate how rapidly the length of the resonator can change if the current of the piezo driver is limited to 10 mA.

**9**

## Laser Dynamics

In this chapter we shall take a closer look at the dynamical properties of laser light sources, for example, the response of a laser system to changes in the operating parameters or to fluctuations of amplitude and phase of its electromagnetic field. For this we first have to investigate theoretically the correlation between the microscopic properties of the laser system and macroscopically measurable quantities such as intensity and phase.

### 9.1 Basic Laser Theory

In Section 7.2 we studied the response of a simplified polarizable system with only two states to an external driving field. There we found that this polarization can amplify a light field and thus become a source of electromagnetic fields.

We know the relation between polarization and electric field already from the wave equation:

$$\left( \nabla^2 - \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{E} = -\frac{1}{\epsilon_0 c^2} \frac{\partial^2}{\partial t^2} \mathbf{P}. \quad (9.1)$$

Here we have already taken into account that laser radiation is often generated by particles diluted in a host material with refraction coefficient  $n$ . The electric field  $\mathbf{E}$  contains the dynamics of the laser field, while the polarization  $\mathbf{P}$  contains the dynamics of the atoms or other excited particles, which is determined in the simplest approximation according to the Bloch equations (7.36).

#### 9.1.1 The Resonator Field

In general, multiple eigenfrequencies can be excited in a laser resonator so that we expect a complicated time evolution of field and polarization. This situation though is mostly undesirable for applications. Therefore we concentrate on the special case of only one single mode of the resonator being excited. In many cases this situation is in fact routinely achieved for practical laser operation. Formally speaking, we decompose the field into its eigenmodes labeled by index  $k$ , factorizing thereby spatial and temporal dependence:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2} \sum_k [E_k(t) e^{-i\Omega_k t} + \text{c.c.}] \mathbf{u}_k(\mathbf{r}). \quad (9.2)$$

The amplitudes  $E_k(t)$  correspond to an average of the amplitude in the resonator volume  $V$ . The spatial distributions  $\mathbf{u}_k$  obey an orthogonality relation,

$$\frac{1}{V} \int_V \mathbf{u}_k \mathbf{u}_l \, dV = \delta_{kl}, \quad (9.3)$$

and  $\Omega_k$  is the passive eigenfrequency of the resonator (without a polarizable medium), so that the Helmholtz equation (Eq. (2.14)) is valid:

$$\nabla^2 \mathbf{u}_k(\mathbf{r}) = -\frac{n^2 \Omega_k^2}{c^2} \mathbf{u}_k(\mathbf{r}).$$

The polarization can be expanded within the same set of functions  $\mathbf{u}_k(\mathbf{r})$ :

$$\mathbf{P}(\mathbf{r}, t) = \frac{1}{2} \sum_k [P_k(t) e^{-i\Omega_k t} + c.c.] \mathbf{u}_k(\mathbf{r}).$$

Owing to (9.3) Eq. (9.1) decomposes into a set of separate equations, of which we use only one but for the very important special case of the *single-mode* or *single-frequency laser*:

$$\left( \Omega^2 + \frac{d^2}{dt^2} \right) E(t) e^{-i\omega t} = -\frac{1}{n^2 \epsilon_0} \frac{d^2}{dt^2} P(t) e^{-i\omega t}.$$

From this equation, among other things, we have to determine the “true” oscillation frequency of the light field.

### 9.1.2 Damping of the Resonator Field

A rigorous theory of the damping of the resonator field cannot be presented here but can be found in, for example, [135, 136]. As for the Bloch equations, we limit ourselves to a phenomenological approach and assume that the energy of the stored field relaxes with rate  $\gamma_c$ . The field amplitude then has to decay with  $\gamma_c/2$ :

$$\frac{d}{dt} E_n(t) = -\frac{\gamma_c}{2} E_n(t).$$

Note that the stored energy (Eq. (2.15)) thus decays with rate  $\gamma_c$ . Another frequently used measure of resonator damping is the *Q factor* (derived from *quality*). For a resonator with eigenfrequency  $\Omega$ , it is given by

$$Q = \Omega/\gamma_c.$$

Damping of the field is caused not only by the outcoupling of a usable light field  $E_{out}$ ,

$$E_{out}(t) = \frac{1}{2} \gamma_{out} E_n(t),$$

but also by scattering or absorption losses within the resonator,

$$\gamma_c = \gamma_{out} + \gamma_{loss}. \quad (9.4)$$

We now insert the damping term into the wave equation (9.1) as well. The spatial dependence is eliminated, and we find

$$\left( \Omega^2 + \gamma_c \frac{d}{dt} + \frac{d^2}{dt^2} \right) E e^{-i\omega t} = -\frac{1}{\epsilon_0 c^2} \frac{d^2}{dt^2} P e^{-i\omega t}.$$

We are interested first of all in the change of the amplitude, which is slow compared to the oscillation with the light frequencies  $\omega$  or  $\Omega$ . In the *slowly varying*

*envelope approximation* (SVEA), which has already been used several times, we neglect contributions scaling like

$$\left[ \frac{d}{dt} E(t), \gamma_c E(t) \right] \ll \omega E(t)$$

and obtain

$$(-\Omega^2 + \omega^2)E(t) + 2i\omega \frac{d}{dt}E(t) + i\gamma_c \omega E = -\frac{\omega^2}{n^2 \epsilon_0} P(t).$$

In the customary approximation  $(-\Omega^2 + \omega^2) \simeq 2\omega(\omega - \Omega)$  for  $\omega \simeq \Omega$ , we get the *simplified amplitude Maxwell equations*

$$\frac{d}{dt}E(t) = i \left( \omega - \Omega + i \frac{\gamma_c}{2} \right) E(t) + \frac{i\omega}{2n^2 \epsilon_0} P(t). \quad (9.5)$$

In the absence of polarizable matter, that is,  $P(t) = 0$ , we recover a field oscillating with the frequency  $\omega = \Omega$ , which is damped with rate  $\gamma_c/2$  exactly as we expected. The macroscopic polarization is already known from (7.12), and its dynamics is described through the optical Bloch equations (7.36). There the occupation number difference  $w(t)$  occurs, which we replace by the inversion density  $\mathcal{N}$  and the total inversion  $n$ , respectively, with the definition

$$\mathcal{N}(t) = \frac{n(t)}{V} = \frac{N_{\text{at}}}{V} w(t).$$

The entire system of atoms and light field is then described by the *Maxwell–Bloch equations*:

$$\begin{aligned} \frac{d}{dt}E(t) &= i \left( \omega - \Omega + i \frac{\gamma_c}{2} \right) E(t) + \frac{i\omega}{2n^2 \epsilon_0} P(t), \\ \frac{d}{dt}P(t) &= (-i\delta - \gamma')P(t) - i \frac{d_{\text{eg}}^2}{\hbar} E(t) \mathcal{N}(t), \\ \frac{d}{dt}\mathcal{N}(t) &= -\frac{1}{\hbar} \Im \{ P(t)^* E(t) \} - \gamma[\mathcal{N}(t) - \mathcal{N}_0]. \end{aligned} \quad (9.6)$$

Lasing can only start if the inversion is maintained by an appropriate pumping process generating the unsaturated inversion density  $\mathcal{N}_0 = n_0/V$  (see Eq. (7.53)). All in all there are *five* equations since field strength  $E$  and polarization  $P$  are complex quantities.

Using this system of equations, several important properties of laser dynamics can be understood. Let us introduce another transparent form of the equations that can be obtained when we normalize the intensive quantities field amplitude  $E(t)$ , polarization density  $P(t)$ , and inversion density  $\mathcal{N}(t)$  to the extensive quantities field strength per photon  $a(t)$ , number of dipoles  $\pi(t)$ , and total inversion  $n(t)$ . For this we use the average “field strength of a photon”  $\langle \mathcal{E}_{\text{ph}} \rangle = \sqrt{\hbar\omega/2\kappa\epsilon_0 V_{\text{mod}}}$  (already introduced in Section 2.1.8). Here  $\kappa = n^2$  ( $n$ : refractive index) accounts for the dielectric properties of the laser material:

$$\begin{aligned} a(t) &:= E(t) \sqrt{\frac{2\kappa\epsilon_0 V_{\text{mod}}}{\hbar\omega}}, \\ \pi(t) &:= N(u + iv) = VP(t)/d_{\text{eg}}. \end{aligned} \quad (9.7)$$

Furthermore, for the Rabi frequency  $\Omega_R$  and the detuning  $\delta$  (between electric field frequency and the eigenfrequency of the polarized medium), it is advantageous to use normalized quantities:

$$g := -\frac{d_{eg}}{\hbar} \sqrt{\frac{\hbar\omega}{2\kappa\epsilon_0 V}} \quad \text{and} \quad \alpha := \frac{(\omega - \omega_0)}{\gamma'} = \frac{\delta}{\gamma'}. \quad (9.8)$$

The coupling factor  $g$  describes the rate (or Rabi frequency) with which the internal excitation state of the polarizable medium is changed at a field strength corresponding to just one photon. The alpha parameter  $\alpha$  is the detuning normalized to the transverse relaxation rate  $\gamma'$ . It will again be of interest in the section about semiconductor lasers, where it has considerable influence on the linewidth.

With the normalized quantities, Eqs. (9.6) have a new transparent structure:

$$\begin{aligned} \text{(i)} \quad \dot{a}(t) &= i \left( \Omega - \omega + \frac{1}{2}i\gamma_c \right) a(t) + \frac{1}{2}ig\pi(t), \\ \text{(ii)} \quad \dot{\pi}(t) &= -\gamma'(1 + i\alpha)\pi(t) - ig(a(t)n(t)), \\ \text{(iii)} \quad \dot{n}(t) &= -g \operatorname{Im}\{\pi(t)^*a(t)\} - \gamma[n(t) - n_0]. \end{aligned} \quad (9.9)$$

The field amplitude  $a(t)$ , the polarization  $\pi(t)$ , and the inversion  $n(t)$  are coupled by the single-photon Rabi frequency  $g$ . At the same time, there is damping with the relaxation time constants  $\gamma_c$ ,  $\gamma'$ , and  $\gamma$ , respectively. The dynamical properties of the laser system are determined by the ratio of these four parameters, which we have compiled for important laser types in Table 9.1.

Equations (9.9) already have great similarity to the quantum theory of the laser field, which we introduce in Chapter 14. By analogy, for instance, normalized amplitudes may be simply promoted to field operators,  $a(t) \rightarrow v(t)$ , to obtain the correct quantum equations.

### 9.1.3 Steady-State Laser Operation

We are now interested in stationary values  $a^{st}$ ,  $\pi^{st}$ , and  $n^{st}$  and begin first by using Eq. (9.9 (ii)):

$$\pi^{st} = -i \frac{ga^{st}n^{st}}{\gamma'(1 + i\alpha)} = -i \frac{\kappa n^{st}}{g} (1 - i\alpha)a^{st}. \quad (9.10)$$

Here we have already introduced the quantity

$$\kappa := \frac{g^2}{\gamma'(1 + \alpha^2)}, \quad (9.11)$$

**Table 9.1** Typical time constants of important laser types.

Laser	Wavelength $\lambda$ ( $\mu\text{m}$ )	Rates			
		$\gamma_c$ ( $\text{s}^{-1}$ )	$\gamma$ ( $\text{s}^{-1}$ )	$\gamma'$ ( $\text{s}^{-1}$ )	
Helium–neon	0.63	$10^7$	$5 \times 10^7$	$10^9$	$10^4$ – $10^6$
Neodymium	1.06	$10^8$	$10^3$ – $10^4$	$10^{11}$	$10^8$ – $10^{10}$
Diode	0.85	$10^{10}$ – $10^{11}$	$3$ – $4 \times 10^8$	$10^{12}$	$10^8$ – $10^9$

which plays the role of the Einstein  $B$  coefficient, as we will see more clearly in the relation with Eq. (9.20). Also,  $\kappa n^{\text{st}}$  may be interpreted as the rate of stimulated emission.

### 9.1.3.1 Saturated Gain

We insert the result into (9.9 (i)), sort into real and imaginary parts, and find a very transparent equation with

$$\dot{a}(t) = \left\{ i \left[ \omega - \Omega - \frac{1}{2} \kappa n(t) \alpha \right] - \frac{1}{2} [\gamma_c - \kappa n(t)] \right\} a(t). \quad (9.12)$$

This equation describes dynamical properties of the amplitude of the resonator field in a good approximation if the damping rate of the polarization  $\gamma'$  dominating in Eq. (9.9) is much larger than all the other time constants. In that case  $\pi(t)$  can always be replaced by its quasi-stationary value.

For the moment we are only interested in the stationary values for the inversion  $n^{\text{st}}$  and the amplitude  $a^{\text{st}}$ :

$$0 = \left[ i \left( \Omega - \omega - \frac{1}{2} \kappa n^{\text{st}} \alpha \right) - \frac{1}{2} (\gamma_c - \kappa n^{\text{st}}) \right] a^{\text{st}}. \quad (9.13)$$

If a laser field already exists ( $a^{\text{st}} \neq 0$ ), the real and imaginary parts of Eq. (9.13) have to be satisfied separately. Especially the real part clearly illustrates that the rate of stimulated emission  $\kappa n^{\text{st}}$  corresponds exactly to the gain rate  $G_s$ , for it has to compensate exactly the loss rate  $\gamma_c$ :

$$n^{\text{st}} = \gamma_c / \kappa \quad \text{or} \quad G = \gamma_c = \kappa n^{\text{st}}. \quad (9.14)$$

Once the laser oscillation has started, the gain no longer depends on the pumping rate but only on the loss properties of the system. This case is called “saturated gain”  $G = G_s = \gamma_c$ . When the laser has not started yet, the (small signal) gain increases linearly with the inversion according to Eq. (7.53),  $G = \kappa n_0$ . This relation is presented in Figure 9.1.

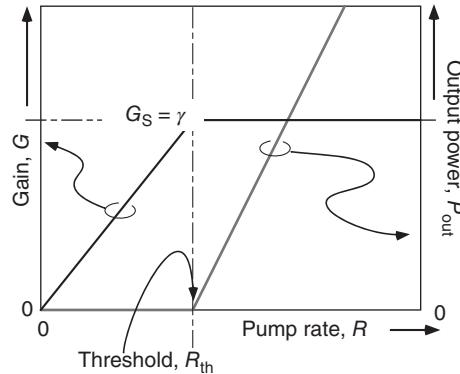


Figure 9.1 Saturated gain and laser power.

### 9.1.3.2 Mode Pulling

The imaginary part of Eq. (9.12) delivers the “true” laser frequency  $\omega$  with which the combined system of resonator and polarized medium oscillates. We replace  $\kappa n^{\text{st}} = \gamma_c$ , use  $\alpha = (\omega - \omega_0)/\gamma'$  according to Eq. (9.8), and get the result

$$\omega = \frac{\gamma' \Omega + \gamma_c \omega_0 / 2}{\gamma' + \gamma_c / 2}.$$

According to this, the eigenfrequencies of either component are weighted with the damping rates for the polarization of the respective other part. The true oscillation frequency always lies between the frequencies of the amplifying medium ( $\omega_0$ ) and the resonator ( $\Omega$ ).

### 9.1.3.3 Field Strength and Number of Photons in the Resonator

According to Eq. (9.7) the photon number and the normalized field strength are connected by  $n_{\text{ph}}(t) = |\alpha(t)|^2$ . Therefore from the third equation of (9.9), it can be derived that

$$\bar{n}_{\text{ph}} = |\alpha^{\text{st}}|^2 = \frac{\gamma}{\gamma_c} (n_0 - n^{\text{st}}). \quad (9.15)$$

Only when the unsaturated inversion  $n_0$  meets the saturated value  $n^{\text{st}}$  does laser oscillation start because the photon number must be positive. Below that threshold we here get the result  $\bar{n}_{\text{ph}} = 0$ . In a refined theoretical description, we will see in Section 9.3 that even below the threshold, stimulated emission leads to an increased photon number in the resonator.

### 9.1.3.4 Laser Threshold

Above the threshold, the unsaturated inversion  $n_0$  must be larger than the inversion at steady-state operation,  $n^{\text{st}}$ , and thus delivers a value for the pumping power or rate  $R_{\text{th}}$  at the laser threshold according to  $n_0 \geq n^{\text{st}}$  and Eq. (7.53). A transparent form is obtained by using the coupling parameter  $g$  according to Eqs. (9.8) and (9.14):

$$R_{\text{th}} = \frac{\gamma_c \gamma}{\kappa} \frac{1}{1 - \gamma/\gamma_{\text{dep}}} = \gamma n_e^{\text{st}}. \quad (9.16)$$

In this model at threshold, apparently the entire pumping energy is still lost due to spontaneous processes since the laser field has not started lasing yet. Above the threshold we can now describe the number of photons within the laser resonator (9.15) by means of the pumping rate,

$$\bar{n}_{\text{ph}} = \frac{1 - \gamma/\gamma_{\text{dep}}}{\gamma_c} (R - R_{\text{th}}) \xrightarrow{\gamma/\gamma_{\text{dep}} \rightarrow 0} \frac{1}{\gamma_c} (R - R_{\text{th}}), \quad (9.17)$$

which has a simple form especially for the “good” four-level laser ( $\gamma/\gamma_{\text{dep}} \rightarrow 0$ ).

For the interpretation of Eq. (9.16), it can also be taken into account that most lasers are operated in an open geometry. Then the coupling constant according to (9.8) is connected to the natural decay constant according to (7.49) by  $g^2 = \gamma(3\pi c^3/\omega^2 e V_{\text{mod}})$  with  $V_{\text{mod}}$  for the mode volume of the resonator field. Using (9.11) one calculates

$$R_{\text{th}} = \gamma_c \gamma' \frac{1 + \alpha^2}{1 - \gamma/\gamma_{\text{dep}}} \frac{\omega^2 e V_{\text{mod}}}{3\pi c^3}.$$

It is intuitively clear that a smaller outcoupling (low  $\gamma_c$ ; see Eq. (9.4)) reduces the laser threshold. According to this relation there are also advantages for small transition strengths (low  $\gamma, \gamma'$ ), fast depopulation rates for the lower laser level (large  $\gamma_{\text{dep}}$ ), and good correspondence of laser frequency and resonance frequency of

the amplifying medium ( $\alpha = 0$ ). The attractive construction of UV lasers suffers, among other things, from the influence of the transition frequency  $\omega$  visible here. On the other hand, the concentration of the resonator field onto a small volume  $V_{\text{mod}}$  is favorable. We are going to follow this path further in Section 9.3 under the heading “Threshold-Less Lasers” and “Micro-lasers.”

#### 9.1.3.5 Laser Power and Outcoupling

The outcoupled laser power is directly connected to the number of photons in the resonator according to

$$P_{\text{out}} = h\nu\gamma_{\text{out}}\bar{n}_{\text{ph}} = h\nu\gamma_{\text{out}}\frac{\gamma}{\gamma_c}(n_0^{\text{st}} - n^{\text{st}}). \quad (9.18)$$

It is furthermore worthwhile to consider the influence of outcoupling on resonator damping:

$$P_{\text{out}} = h\nu\gamma_{\text{out}} \left( \frac{R}{\gamma_{\text{out}} + \gamma_{\text{loss}}} - \frac{\gamma}{\kappa} \right). \quad (9.19)$$

For very small outcoupling ( $\gamma_{\text{out}} \ll \gamma_{\text{loss}}$ ), the output power increases with  $\gamma_{\text{out}}$  and passes through a maximum, and at  $R/(\gamma_{\text{out}} + \gamma_{\text{loss}}) = \gamma/\kappa$  laser oscillation dies out. In order to achieve an output power as high as possible,  $\gamma_{\text{out}}$  has to be controlled by the reflectivity of the resonator mirrors. With the example of the helium–neon laser in Figure 8.8 on p. 258, we have already investigated this question in slightly different terms.

## 9.2 Laser Rate Equations

The Maxwell–Bloch equations (9.6) and (9.9) describe the dynamical behavior of each of the two components of the electric field  $E(t)$  and  $a(t)$ , respectively, and the polarization density  $P(t)$  and the dipole number  $\pi(t)$ , respectively. Furthermore, the inversion has to be taken into account through its density  $\mathcal{N}(t)$  or the total inversion  $n(t)$ . The equations raise the expectation of, in principle, a complicated dynamical behavior that finds its special expression in the isomorphy of the laser equations with the Lorentz equations of nonlinear dynamics that literally lead to “chaos.”

However, most conventional lasers behave dynamically in a very well-natured way or in good approximation according to the stationary description that we just have dealt with intensively. They owe their stability to a fact that also simplifies the mathematical treatment of the Maxwell–Bloch equations enormously. The relaxation rate of the macroscopic phase between laser field and polarization,  $\gamma'$ , is typically very much larger than the relaxation rates of inversion ( $\gamma$ ) and resonator field ( $\gamma_c$ ). Under these circumstances the polarization density follows the amplitude of the electric field nearly instantaneously and therefore according to Eq. (9.10) can always be replaced by its instantaneous ratio to field strength  $a(t)$  and inversion density  $n(t)$ :

$$\pi(t) \simeq -\frac{-iga(t)n(t)}{\gamma'(1+i\alpha)}.$$

Once the polarization density has been “eliminated adiabatically,” it is not worth further investigating the phase dependence of the electric field because it is only interesting in relation to the polarization. Instead of this we investigate the time-varying dynamics of the photon number according to

$$\frac{d}{dt}|\alpha(t)|^2 = \alpha(t) \frac{d}{dt}\alpha^*(t) + \alpha^*(t) \frac{d}{dt}\alpha(t).$$

We obtain the simplified *laser rate equations* where we use the pumping rate  $R \simeq n_0/\gamma$  instead of the unsaturated inversion  $n_0$ :

$$\begin{aligned} \text{(i)} \quad \frac{d}{dt}n_{\text{ph}}(t) &= -\gamma_c n_{\text{ph}}(t) + \kappa n_{\text{ph}}(t)n(t), \\ \text{(ii)} \quad \frac{d}{dt}n(t) &= -\kappa n_{\text{ph}}(t)n(t) - \gamma n(t) + R. \end{aligned} \quad (9.20)$$

Unlike common linear differential equations, these equations are connected nonlinearly by the coupling term  $\kappa n_{\text{ph}}(t)n(t)$ , which is the rate of stimulated emission:

$$R_{\text{stim}} = \kappa n_{\text{ph}}(t)n(t). \quad (9.21)$$

Because of stimulated emission, the rate of change of the photon number depends on the number of photons already present.

At first we again study the equilibrium values  $\bar{n}_{\text{ph}}$  and  $n^{\text{st}}$ . Equation (9.20) (i)) yields two solutions, the first of which,  $\bar{n}_{\text{ph}} = 0$ , describes the situation below the laser threshold. There the inversion grows linearly with the pumping rate according to (9.20) (ii)) and (7.53) (we again assume the case of a “good” four-level laser with  $\gamma/\gamma_{\text{dep}} \ll 1$ ):

$$\bar{n}_{\text{ph}} = 0 \quad \text{and} \quad n^{\text{st}} = n_0 \simeq R/\gamma.$$

When laser oscillation has started ( $\bar{n}_{\text{ph}} > 0$ ), then according to (9.20) (i)), the inversion in equilibrium must be clamped at the saturation value  $n^{\text{st}}$ , and Eq. (9.14) is again found.

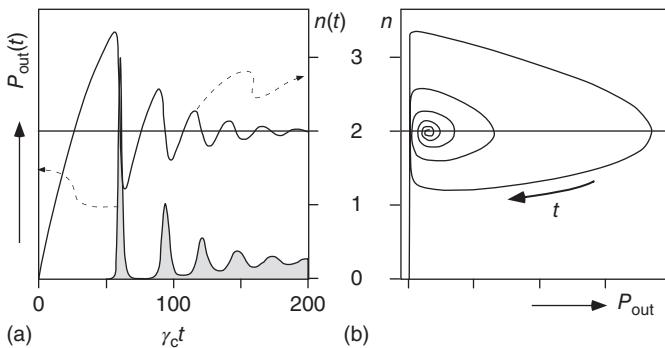
As expected we recover the behavior of Figure 9.1. The gain only grows until laser threshold is reached and then becomes saturated, that is, a constant value due to *gain clamping*. From (9.20) (ii)) we find that at the same time the steady-state number of photons increases with pump rate  $R$  according to

$$\bar{n}_{\text{ph}} = \frac{1}{\gamma_c} \left( R - \frac{\gamma\gamma_c}{\kappa} \right) = \frac{1}{\gamma_c}(R - R_{\text{th}}). \quad (9.22)$$

The value  $R_{\text{th}} = \gamma\gamma_c/\kappa$  just corresponds to the pumping power at the threshold. A linear dependence of laser power ( $\propto n_{\text{ph}}$ ) on the pumping rate  $R$  is predicted, which is indeed a frequent property of common laser types, as shown for diode lasers, for example, in Figure 10.12.

### 9.2.1 Laser Spiking and Relaxation Oscillations

The laser rate equations (9.20) are nonlinear and can in general only be investigated by numerical analysis. In Figure 9.2 we present two examples where the laser is switched on suddenly. For  $t < 0$  we have  $R = 0$ , and the switching is instantaneous, at least compared to one of the two relaxation rates  $\gamma$  (inversion density) or  $\gamma_c$  (resonator field). The numerical simulation can easily be carried out with



**Figure 9.2** Numerical evaluation of relaxation oscillations. (a) Inversion  $n(t)$  (proportional to the gain) and output power  $P_{\text{out}}(t)$ . (b) Phase-space representation. The system parameters in Eq. (9.20) are chosen to be  $\kappa = 1$ ,  $\gamma_c = 2$ ,  $\gamma = 0.02$ ,  $R = 0.1$ .

many programs of computer algebra and shows very well the phenomenon of “laser spiking,” being observed, for example, at fast turn-on (nanoseconds or faster) of neodymium or diode lasers (see Figure 9.4).

Relaxation oscillations in a narrower sense occur when the gain (or the loss rate) changes suddenly. Fluctuations of gain are induced by the variations of the pumping processes, for example, by switching on or off an optical pump laser. For many purposes, for example, for the stability analysis of frequency and amplitude of a laser oscillator, it is sufficient to consider small deviations of the photon numbers and the inversion from their equilibrium values:

$$n_{\text{ph}}(t) = \bar{n}_{\text{ph}} + \delta n_{\text{ph}}(t) \quad \text{and} \quad n(t) = n^{\text{st}} + \delta n(t).$$

We insert into Eq. (9.20), neglect products of the type  $\delta n \delta n_{\text{ph}}$ , and obtain the linearized equations:

$$\begin{aligned} \text{(i)} \quad & \frac{d}{dt} \delta n_{\text{ph}} = \kappa \bar{n}_{\text{ph}} \delta n = \left( \frac{\kappa R}{\gamma_c} - \gamma \right) \delta n, \\ \text{(ii)} \quad & \frac{d}{dt} \delta n = -(\gamma + \kappa \bar{n}_{\text{ph}}) \delta n - \gamma_c \delta n_{\text{ph}}. \end{aligned} \quad (9.23)$$

For simplicity we introduce the normalized pumping rate  $\rho = R/R_{\text{th}} = \kappa R/\gamma \gamma_c$ , which has the value 1 at threshold, and for both  $x = \{\delta n_{\text{ph}}, \delta n\}$ , we obtain the standard equation of the damped harmonic oscillator:

$$\ddot{x} + \gamma \rho \dot{x} + \gamma \gamma_c (\rho - 1)x = 0. \quad (9.24)$$

From this we infer without further difficulties that the system can oscillate for

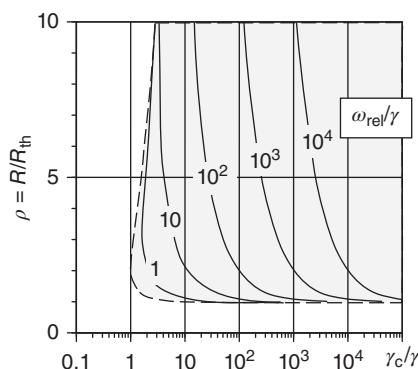
$$(\gamma_c/\gamma) \left[ 1 - \sqrt{1 - (\gamma/\gamma_c)} \right] < \rho/2 < (\gamma_c/\gamma) \left[ 1 + \sqrt{1 - (\gamma/\gamma_c)} \right]$$

with normalized frequency

$$\omega_{\text{rel}}/\gamma = \sqrt{(\gamma_c/\gamma)(\rho - 1) - (\rho/2)^2}, \quad (9.25)$$

and is damped with the rate

$$\gamma_{\text{rel}} = \gamma \rho/2 = \gamma R/2R_{\text{th}}. \quad (9.26)$$



**Figure 9.3** Relaxation oscillations as a function of  $\gamma_c/\gamma$  and  $\rho$ .

Especially solid-state lasers typically have long lifetimes in the excited laser level and thus large  $\gamma_c/\gamma$  ratios, for example,  $10^3\text{--}10^4$  for semiconductor lasers and  $10^4\text{--}10^5$  for Nd lasers. In Figure 9.3 it can be seen that in this case relaxation oscillations are triggered immediately above the laser threshold at  $\rho = 1$ . They can also be driven by external forces, for example, by modulating the pumping rate appropriately, and they play an important role for the amplitude and frequency stability of laser sources (see Section 9.4), since they are induced by noise sources of all kinds.

### Example: Relaxation oscillations in the Nd:YAG laser

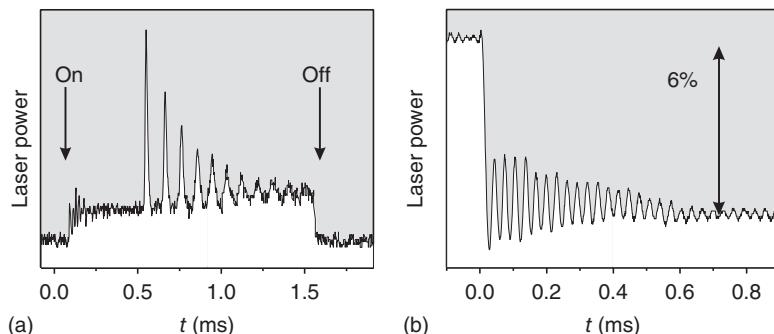
We consider the 1064 nm line of the Nd:YAG laser with the following characteristic quantities:

$$\begin{array}{lll} \text{Natural lifetime} & \tau = 240 \mu\text{s}, & \gamma = 4.2 \times 10^3 \text{ s}^{-1}, \\ \text{Resonator storage time} & \tau_c = 20 \text{ ns}, & \gamma_c = 5.0 \times 10^7 \text{ s}^{-1}, \\ \text{Normalized pump rate} & R/R_{\text{th}} = 1.0\text{--}1.5. & \end{array}$$

The properties of the relaxation oscillations observed in experiment (Figure 9.4) correspond to the theoretical estimates. For the Nd:YAG parameters the second term in Eq. (9.25) can be neglected in calculating the oscillation frequency due to  $\gamma \ll \gamma_c$ ,

$$\omega_{\text{rel}} \approx \sqrt{\gamma\gamma_c} \sqrt{\rho - 1} \approx 72 \text{ kHz} \sqrt{\rho - 1},$$

and according to (9.26) the damping rate is  $\gamma_{\text{rel}} \approx 2 \times 10^3 \text{ s}^{-1} R/R_{\text{th}}$ .



**Figure 9.4** Spiking and relaxation oscillations in the Nd:YAG laser. The power of the pumping laser diode is modulated by a square signal. Complete modulation (a) causes spiking; partial modulation (b, 6%) causes relaxation oscillations. Compare this with Figure 9.2.

### 9.3 Threshold-Less Lasers and Micro-lasers

We have already seen in the section about spontaneous emission that a reflecting environment changes the rate of spontaneous emission. In principle, this effect occurs in every laser resonator, though it is mostly so small that it can be neglected without any problems. The influence is so small because in open resonator geometry (Figure 9.5), the more or less isotropic spontaneous radiation of an excited medium, for example, of an atomic gas, is emitted only with a small fraction into that solid angle, which is occupied by the electric field modes of the laser resonator.

These changes though can no longer be neglected if the resonator becomes very small or if, as a result of large steps of the refraction coefficient of the laser medium, the emitted power is more and more confined to the resonator. For this case, the modified effect of the spontaneous emission is often taken into account by the so-called spontaneous emission coefficient  $\beta$ . The  $\beta$  factor indicates which geometrical part of the radiation field couples to the laser mode (rate  $\beta\gamma$ ) and which part is emitted into the remaining volume (rate  $(1 - \beta)\gamma$ ).

Spontaneous emission can be considered as stimulated emission by a single photon, and therefore we set the coupling coefficient  $\beta\gamma = \kappa(n_{ph=1})$ . With this trick we can account for spontaneous emission in the laser rate equation (9.20 (i)), and by replacing  $n_{ph} \rightarrow n_{ph} + 1$ , we obtain

$$(i) \quad \frac{d}{dt}n_{ph}(t) = -\gamma_c n_{ph}(t) + \beta\gamma n_{ph}(t)n(t) + \beta\gamma n(t),$$

$$(ii) \quad \frac{d}{dt}n(t) = -\beta\gamma n_{ph}(t)n(t) - \beta\gamma n(t) - (1 - \beta)\gamma n(t) + R.$$

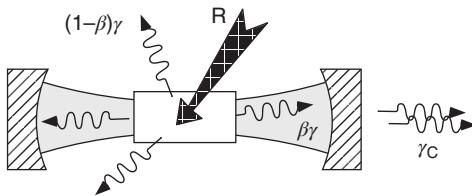
For the steady-state situation, the equations can immediately be simplified to

$$(i) \quad 0 = -\gamma_c \bar{n}_{ph} + \beta\gamma n^{st}(\bar{n}_{ph} + 1),$$

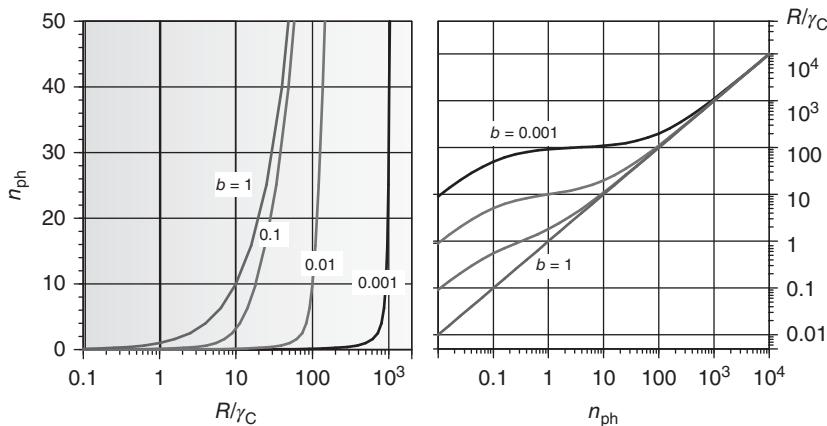
$$(ii) \quad 0 = R - \beta\gamma \bar{n}_{ph} n^{st} - \gamma n^{st},$$

where spontaneous emission is especially prominent through the factor  $\bar{n}_{ph} + 1$  in (i). In order to solve this system of equations, it is convenient to express the pumping rate as a function of the photon number in the resonator. We substitute  $n^{st}$  in (ii) by means of (i) and obtain (Figure 9.6)

$$\frac{R}{\gamma_c} = \left( \frac{1}{\beta} + \bar{n}_{ph} \right) \frac{\bar{n}_{ph}}{\bar{n}_{ph} + 1}. \quad (9.27)$$



**Figure 9.5** Relaxation and pumping rates of the laser. The  $\beta$  coefficient of spontaneous emission is a coarse measure for the fraction of spontaneous emission coupled to the laser mode ( $\beta$ , shaded area) and to the other solid angle ( $1 - \beta$ ).



**Figure 9.6** Threshold behavior of laser oscillators. Photon number in the resonator as a function of the pumping rate.

Far above the laser threshold, that is, for  $\bar{n}_{\text{ph}} \gg 1/\beta \geq 1$ , the relation between pumping rate and photon number obviously turns again into the result (9.17), as expected. According to the condition (9.27), the laser threshold is reached when the photon number in the resonator meets or exceeds the value  $1/\beta$ . So in a common laser ( $\beta \ll 1$ ) at threshold, there are already so many photons present in the laser mode. To be more exact, there are so many that the rate of stimulated emission into the laser mode precisely equals the total spontaneous decay rate. Above this threshold, additional pumping power is used predominantly to increase the photon number and thus to build up the coherent radiation field.

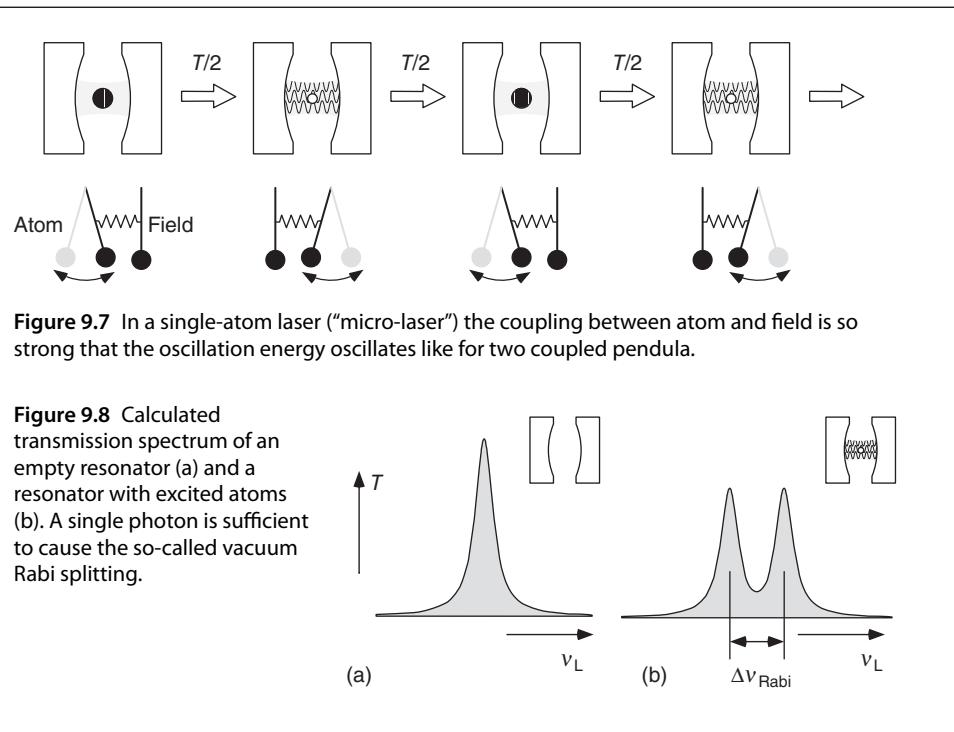
#### Excursion: Micro-maser, micro-laser, and single-atom laser

Experiments first with the so-called micro-maser, and later on with the micro-laser, have had a strong stimulating effect on the concept of the “threshold-less” laser. The term *micro-* does not refer so much to the miniaturized layout but rather to the microscopic character of interaction. The coupling between the field of a micro-laser and micro-maser is so strong that an excited atom does not forego its energy once and forever to the electromagnetic field like in a common laser. For this so-called strong coupling regime, which is described in more detail in Section 13.1, the rate  $g$  of Eq. (9.8) has to be larger than every other time constant (see Table 9.1):

$$g \gg \gamma, \gamma_c, \gamma'.$$

Then the resonator field stores the emitted energy, and the atom (the polarized matter) can reabsorb the radiation energy. So the energy oscillates between atom and resonator (Figure 9.7). This situation can be realized already – or even particularly well – with a single atom, thus the term *single-atom maser*, which was very often used at first.

In order to realize the situation of a micro-maser experimentally, resonators with extremely long radiation storage times have to be used. Since superconducting resonators for microwaves have been available for a longer time, the micro-maser was



**Figure 9.7** In a single-atom laser (“micro-laser”) the coupling between atom and field is so strong that the oscillation energy oscillates like for two coupled pendula.

**Figure 9.8** Calculated transmission spectrum of an empty resonator (a) and a resonator with excited atoms (b). A single photon is sufficient to cause the so-called vacuum Rabi splitting.

realized before the micro-laser. The description of the micro-maser requires a joint treatment of atom and field according to quantum theory within the framework of the so-called Jaynes–Cummings model, which goes far beyond the scope of this excursion. However, it is intuitively clear that the transmission of the combined system of resonator and atom exhibits a different spectral behavior from the empty resonator following the ordinary Lorentz curve (Figure 9.8).

Threshold-less lasers are extraordinarily interesting for applications in integrated optics. For example, semiconductor components may be designed where single electron–hole pairs are directly converted into single photons. Current research follows different routes to construct radiation fields confined to a small mode volume, with long storage times, and intense coupling to the excited medium. At optical wavelengths a small mode volume also means using miniaturized resonators. For the traditional layouts following the linear resonator, the integration of highly reflective mirrors to achieve large storage densities though is difficult. A solution is offered by the appropriate use of total reflection. Tiny electrical resonators from a monolithic substrate with very high quality have already been realized. At the rim of mushroom- or mesa-shaped semiconductor lasers and dielectric spheres made from silica, circulating field modes, the so-called whispering-gallery modes, have been prepared and shown to be long-lived. Recently also micro-resonators with an oval geometry have been discussed for micro-laser applications because they allow a particularly strong coupling of laser medium and radiation field (see Section 6.6.4) [80].

## 9.4 Laser Noise

All physical quantities are subject to fluctuations, and the laser light field is no exception: the perfect harmonic wave with fixed amplitude and phase remains a fiction! But the laser light field approaches this ideal of a harmonic oscillator more closely than any other physical phenomenon. According to an old estimate by Schawlow and Townes, the coherent laser light field shows extremely small fluctuations of amplitude and phase. Not the least for this reason, it has continued to inspire wide areas of experimental physics to this day. The “narrow linewidth” (sub-hertz) has already been introduced in Section 8.1.6.1 with the example of the He–Ne laser. It promises extremely long coherence times ( $> 1$  s) or enormous lengths ( $> 10^8$  m), which can be used for high-precision measurements for a wide variety of phenomena.

Usually the so-called Schawlow–Townes limit of the linewidth is hidden by technical and generally much bigger fluctuations. If this fundamental limit is realized, however, it offers information about the physical properties of the laser system. In this section we investigate what physical processes impair ideal oscillator performance.

### 9.4.1 Amplitude and Phase Noise

The stationary values of the laser light field (Section 9.1.3) have been determined through the photon number  $\bar{n}_{\text{ph}}$  and the true laser frequency  $\omega$  on p. 289. There we assumed that the phase evolution of the field behaves like a perfect oscillator according to classical electrodynamics:

$$E(t) = \Re\{E_0 \exp[-i(\omega t + \phi_0)]\}.$$

The coupled system of polarized laser medium and resonator field though is also coupled to its environment, for example, by the spontaneous emission, causing stochastic fluctuations of the field strength and the other system quantities.<sup>1</sup> More realistically we thus introduce noise terms:

$$E(t) \rightarrow \Re\{[E_0 + e_N(t)] \exp[-i(\omega t + \phi_0 + \delta\phi(t))]\}.$$

Here we assume that we can distinguish contributions to amplitude noise ( $e_N(t)$ ) and to phase noise ( $\delta\phi(t)$ ), which are slow in comparison with the oscillation frequency, that is,  $(de_N/dt)/e_N, d\delta\phi/dt \ll \omega$ . This separation is not unambiguous, but the two types yield rather different noise spectra.

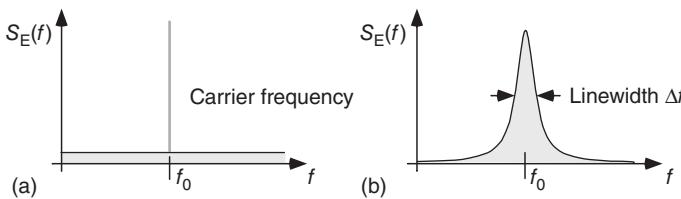
In Figure 9.9 the effect of white, that is, frequency-independent, noise of the amplitude and phase, respectively, on the power spectrum of the electromagnetic field (for the definition see Section A.1) is presented. The exact calculation requires information about the spectral properties of the noise quantities.

#### 9.4.1.1 Amplitude Fluctuations

We begin with the amplitude fluctuations and first assume perfect phase evolution ( $\delta\phi(t) = 0$ ). If the fluctuations of the noise amplitude are entirely random, that is, very “fast” even during the integration time  $T$  of the analyzer, they are only

---

<sup>1</sup> In the micro-maser (see p. 296) though the aim is to eliminate exactly this coupling to the environment.



**Figure 9.9** Field spectrum for (a) white amplitude noise and (b) white phase noise, respectively. The spectral width of the carrier frequency in (a) is limited only by the resolution of the spectrum analyzer.

correlated at all at delay time \$\tau = 0\$ (“delta correlated”), and we can describe the correlation function of the noise amplitude using the mean square value \$e\_{\text{rms}}^2 = \langle |e\_N(t)|^2 \rangle\$:

$$\langle e_N(t) \rangle = 0 \quad \text{and} \quad \langle e_N(t)e_N^*(t + \tau) \rangle = e_{\text{rms}}^2 T \delta(\tau).$$

With this information we can calculate the correlation function of an electromagnetic field with amplitude fluctuations where we take advantage of the Poynting theorem (see Section A.2):

$$\begin{aligned} C_E(\tau) &= \langle \Re{E(t)} \Re{E(t + \tau)} \rangle = \frac{1}{2} \langle \Re{E(t)E^*(t + \tau)} \rangle \\ &= \frac{1}{2T} \int_{-T/2}^{T/2} \Re{[E_0 + e_N(t)][E_0^* + e_N^*(t + \tau)]} dt \\ &= \frac{1}{2} |E_0|^2 + e_{\text{rms}}^2 T \delta(\tau). \end{aligned}$$

The finite integration interval causes errors of magnitude \$\mathcal{O}(1/\omega T)\$, which can be neglected since at optical frequencies \$\omega T\$ is always very large. Using the Wiener–Khinchin theorem (Eq. (A.9)), the spectrum

$$S_E(f) = \frac{1}{2} E_0^2 \delta(f) + e_{\text{rms}}^2 / \Delta f$$

can be obtained. The “Fourier frequencies” \$f\$ give the distance to the much larger optical carrier frequency \$\omega = 2\pi\nu\$. The second contribution causes a “white noise floor,” and we have already replaced \$T = 1/\Delta f\$ to indicate that in an experiment the filter bandwidth always has to be inserted here. The first contribution represents the carrier frequency like for a perfect harmonic oscillation. The delta function indicates that the entire power in this component can always be found in one channel of the spectrum analyzer, and so its width is always limited by the filter bandwidth.

#### 9.4.1.2 Phase Fluctuations

In order to study the influence of a fluctuating phase, we follow the presentations of Yariv [137] and Loudon [85] and calculate the correlation function of an electromagnetic field \$E(t) = \Re{E\_0 e^{-i[\omega t + \theta(t)]}}\$ with a slowly fluctuating phase \$\theta(t)\$ again using the Poynting theorem (see Section A.2):

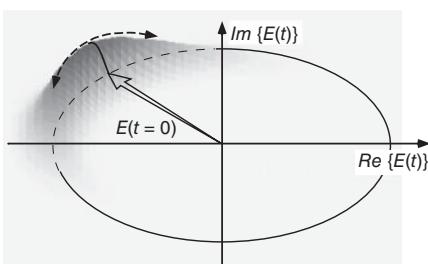
$$\begin{aligned} C_E(\tau) &= \frac{1}{2} |E_0|^2 \langle \Re{e^{i[\omega\tau + \Delta\theta(t, \tau)]}} \rangle \\ &= \frac{1}{2} |E_0|^2 \Re{e^{i\omega\tau} \langle e^{i\Delta\theta(t, \tau)} \rangle}. \end{aligned} \tag{9.28}$$

The average extends only over the fluctuating part with  $\Delta\theta(t, \tau) = \theta(t + \tau) - \theta(t)$ . Though we do not know the exact variation in time (i.e., just the nature of noise), we assume it to exhibit stationary behavior so that properties such as the frequency spectrum do not depend on time itself. If the statistical distribution of the average phase deviations  $\Delta\theta(\tau)$  is known, we can use the ensemble average over the probability distribution  $p(\Delta\theta(\tau))$  instead of the time average to calculate the average in (9.28). For symmetric distributions we have to take only the real part into account:

$$\langle e^{i\Delta\theta(\tau)} \rangle = \langle \cos \Delta\theta(\tau) \rangle = \int_{-\infty}^{\infty} \cos \Delta\theta p(\Delta\theta(\tau)) d\Delta\theta. \quad (9.29)$$

The sought-after probability distribution is completely characterized when  $p(\Delta\theta(\tau))$  is explicitly given or, for a known type such as the normal distribution, one of its so-called statistical moments, for example, the mean square deviation  $\Delta\theta_{\text{rms}}^2$ , is given.

How do we obtain the required statistical information? From the point of view of experimental physics, one would simply measure the phase fluctuations, for example, by heterodyning the laser field under investigation with a stable reference wave to determine the macroscopic phase. Theoretical models are, however, necessary to establish a connection with the microscopic physical properties of the laser system.

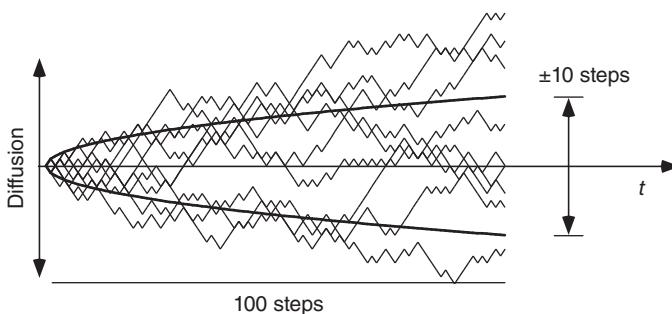


**Figure 9.10** Phasor model of the laser field with phase diffusion.

Let us first concentrate on the widely known phase diffusion model. For this we study the phasor model of the amplitude of the laser field in Figure 9.10. By the Maxwell–Bloch equations, only the amplitude of the laser field is fixed, but not the phase, since there is no restoring force binding the phase to a certain value. Thus the phase diffuses unobstructedly away from its initial value. We will see that for this – if technical disturbances can be excluded – especially spontaneous emission processes are responsible.

The phase can change in one dimension only, and therefore our model is one-dimensional as well. We assume that the phase is subject to small leaps occurring at a rate  $R$  still to be determined. The leaps are completely independent of each other, that is, in every single case the direction of the next step is entirely random. This results in stochastic motion known from the Brownian motion of molecules. Therefore it is also called a *random walk* and is often compared with the walk of a drunk who is not aware of his next step.

Using the random-number generator of the computer, such motions can easily be simulated. In Figure 9.11 several trajectories are presented, and additionally the time-dependent expectation value of the square fluctuation is shown. The root mean square (r.m.s.) deviation of the Gaussian normal distribution after  $N$  steps is  $(\Delta N_{\text{rms}}^2)^{1/2} = (\langle N^2 \rangle - \langle N \rangle^2)^{1/2} = \sqrt{N}$ , as is well known. Since the number



**Figure 9.11** One-dimensional random walk. Every single step is randomly set in either the  $\pm$  direction. The thick line marks the mean square deviation.

of steps increases in proportion with time  $\tau$ , the r.m.s. deviation has to be proportional to  $\sqrt{\tau}$ . Therefore we can construct the normal distribution

$$p(\Delta\theta(\tau)) = \frac{\exp(-\Delta\theta^2/2\Delta\theta_{\text{rms}}^2)}{\sqrt{2\pi}\Delta\theta_{\text{rms}}} \quad \text{with} \quad \int_{-\infty}^{\infty} p(\Delta\theta) d\Delta\theta = 1,$$

with mean square value  $(\Delta\theta_{\text{rms}})^2 = \theta_0^2 R \tau$  and  $\theta_0$  for the length of a single step. Now we can evaluate the integral of Eq. (9.29), obtaining the simple result

$$\langle \cos \Delta\theta(\tau) \rangle = \exp(-\Delta\theta_{\text{rms}}^2/2) = \exp(-\theta_0^2 R \tau / 2).$$

The complete correlation function ( $\omega = 2\pi\nu$ ) reads

$$C_E(\tau) = \frac{1}{2} \langle E(0)E^*(\tau) \rangle = \frac{1}{2} |E_0|^2 e^{i2\pi\nu\tau - \theta_0^2 R \tau / 2}.$$

The correlation function can also be interpreted as the average projection of the field vector onto its initial value at the time  $\tau = 0$ . Its form is identical with the time dependence of a damped harmonic oscillator. We calculate the spectrum again according to the Wiener–Khinchin theorem (Eq. (A.9)) and find that white phase noise (Figure 9.9) leads to a Lorentz-shaped line with width  $\Delta\omega = 2\pi\Delta\nu_{1/2} = \theta_0^2 R$  centered at the carrier frequency  $\nu = \omega/2\pi$ :

$$S_E(f) = \frac{|E_0|^2}{T} \frac{\theta_0^2 R / 2}{(2\pi f)^2 + (\theta_0^2 R / 2)^2}. \quad (9.30)$$

#### 9.4.2 The Microscopic Origin of Laser Noise

The considerations of the previous section are generally valid for oscillators of every kind. We now have to correlate the macroscopically observed properties to the specific microscopic properties of the laser. A rigorous theory (i.e., a consequent theoretical calculation of correlation functions as in Eq. (9.28)) requires a treatment according to quantum electrodynamics, for which we have to refer to the relevant literature. The theories by Haken [138] and by Lax and Louisell [139, 140] are among the important successes in the quantum theory of “open systems” and were presented shortly after the invention of the laser. We have to limit ourselves here to simplified models, but we can put forward some reflections about the nature of the noise forces.

The fluctuations of laser light field reflect several noise sources. The best-known process is caused by the spontaneous emission out of the amplifying medium into the environment. These radiation processes do not contribute to the laser field but cause stochastic fluctuations of the inversion and the (dielectric) polarization. Since the amplitudes of resonator field and polarization relax back to their steady state, amplitude and phase fluctuations result.<sup>2</sup> Other noise processes are caused because the resonator field also suffers from random losses or because the pumping process transfers its noise properties to the stimulated emission. It is normally “incoherent,” that is, the excitation states are produced with a certain rate but with a random, typically Poisson, distribution. In a semiconductor laser, electron–hole pairs are injected into the amplification zone. For large current density the charge carriers repel each other, and successive arrival times are more evenly spaced out. It has been shown that this “regularization” of the pumping process also gives rise to a decrease of the intensity fluctuations [141]!

Many processes can be heuristically interpreted through the “grainy” structure of the quantized light field. Let us therefore study the changes of amplitude and phase of the laser field when “photons” are added to or taken away from it.

### 9.4.3 Laser Intensity Noise

The time evolution of the laser amplitude was investigated in Section 9.2.1 for the system reacting to sudden changes of the pumping rate, for example, through deterministic switching events. In our simple model such changes are now caused by small random changes of the photon number  $n_{\text{ph}}(t) = \bar{n}_{\text{ph}} + \delta n_{\text{ph}}(t)$  fluctuating around the mean value  $\bar{n}_{\text{ph}}$ .

#### 9.4.3.1 Quantum Limit of the Laser Amplitude

Let us estimate the mean square deviation  $\langle \delta n_{\text{ph}}^2 \rangle$  of the photon number, and hence the field amplitude, without ascertaining the distribution more exactly. For this we rewrite the linearized equation (9.24) for the photon number by inserting the stationary photon number  $\bar{n}_{\text{ph}}$  from Eq. (9.22):

$$\frac{d^2}{dt^2} \delta n_{\text{ph}} + (\kappa \bar{n}_{\text{ph}} + \gamma) \frac{d}{dt} \delta n_{\text{ph}} + \gamma_c \kappa \bar{n}_{\text{ph}} \delta n_{\text{ph}} = 0.$$

We multiply this equation by  $\delta n_{\text{ph}}$  and arrive at

$$\frac{1}{2} \frac{d^2}{dt^2} \delta n_{\text{ph}}^2 - \frac{1}{2} (\kappa \bar{n}_{\text{ph}} + \gamma) \frac{d}{dt} \delta n_{\text{ph}}^2 - \frac{1}{2} \left( \frac{d}{dt} \delta n_{\text{ph}} \right)^2 + \gamma_c \kappa \bar{n}_{\text{ph}} \delta n_{\text{ph}}^2 = 0. \quad (9.31)$$

When we search for the steady-state solution of the mean value  $\langle \delta n_{\text{ph}}^2 \rangle$ , we can eliminate the average of the derivatives  $\langle (d/dt) \delta n_{\text{ph}} \rangle$  but not that of the square of

---

<sup>2</sup> In another formulation it is often said that spontaneous emission radiates “into the laser mode.” In this interpretation, polarization and laser field both separately have to relax back again to their equilibrium relation. Since in the theoretical description used here the coupling of resonator field and polarization is already completely included, the interpretation chosen here appears to be physically more conclusive.

the fluctuation rate  $[(d/dt)\delta n_{\text{ph}}]^2$ :

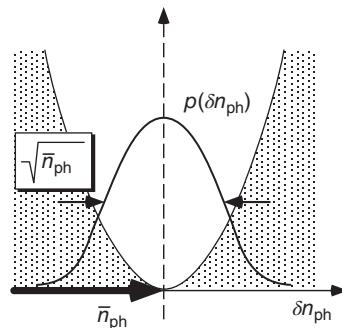
$$-\frac{1}{2}\langle[(d/dt)\delta n_{\text{ph}}]^2\rangle + \gamma_c \kappa \bar{n}_{\text{ph}} \langle\delta n_{\text{ph}}^2\rangle = 0. \quad (9.32)$$

We cannot give a rigorous theoretical description here, but for an intuitive treatment, we can use Eq. (9.23 (i)),  $[(d/dt)\delta n_{\text{ph}}]^2 = (\kappa \bar{n}_{\text{ph}} \delta n)^2$ . It is reasonable to assume an inversion undergoing random fluctuations and hence obeying Poisson statistics induced by both spontaneous emission and the stochastic pumping process, yielding mean square value  $\delta n^2 = n^{\text{st}}$ . We can now evaluate Eq. (9.31) and find with  $\gamma_c = \kappa n^{\text{st}}$  from (9.14)

$$\langle\delta n_{\text{ph}}^2\rangle = \langle(d/dt)\delta n_{\text{ph}}^2\rangle / \gamma_c \kappa \bar{n}_{\text{ph}} = \bar{n}_{\text{ph}}.$$

Most importantly, we find that the number of photons in the resonator fluctuates by an amount proportional to  $\sqrt{\bar{n}_{\text{ph}}}$ . A more exact analysis shows that the distribution indeed again has the shape of a Poisson distribution (which for large numbers is essentially a Gaussian distribution) (Figure 9.12).

The investigation of the photon number distribution offers an intuitive picture, which we study in a bit more detail in Figure 9.13. The total number of photons is proportional to the field energy ( $E^2 \propto h\nu\bar{n}_{\text{ph}}$ ). Removal or addition of one “photon” changes the field energy by the amount  $h\nu$ .



**Figure 9.12** Distribution of the photon number of the laser field. The fluctuations of the photon number are stabilized by an effective potential (see Eq. (9.32)).

#### 9.4.3.2 Relative Intensity Noise (RIN)

The fluctuations of the external laser power  $P(t) = P_0 + \delta P(t)$  are measured in an experiment. Using Eq. (9.22), the fluctuations of the photon number can be converted into the r.m.s. deviation of the laser power  $\delta P_{\text{rms}} = \langle\delta P^2\rangle^{1/2}$ . Thus,  $\delta P_{\text{rms}} = \sqrt{h\nu\gamma_{\text{out}}} \sqrt{P}$ .

Intensity fluctuations of the idealized laser are caused by quantum fluctuations only, the fundamental physical limit. According to the results of the previous section, their relative significance decreases with increasing laser power because  $\sqrt{\delta n_{\text{ph}}^2}/\bar{n}_{\text{ph}} = 1/\sqrt{\bar{n}_{\text{ph}}}$ . Moreover, many laser types show noise contributions that are not always exactly identified but increasing proportionally to the output power.

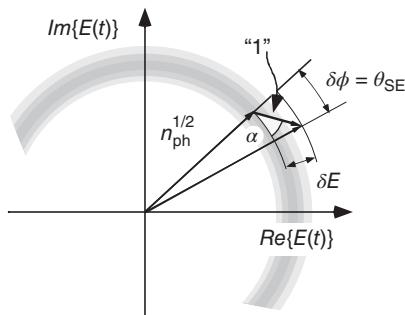
For the quantitative characterization of the amplitude noise, the *relative intensity noise* (RIN) has been introduced:

$$\text{RIN:} = \frac{\delta P_{\text{rms}}^2}{P^2},$$

which is a phenomenological quantity and can be straightforwardly measured. For a more exact analysis of the intensity noise, again its spectral distribution has to be determined. In the simplest case of completely random fluctuations,

it shows a flat spectrum, that is, white noise of Figure 9.9.<sup>3</sup> In the section about semiconductor lasers, we will find that in the intensity spectrum, for example, the relaxation oscillations play a role as well.

#### 9.4.4 Schawlow–Townes Linewidth



**Figure 9.13** Effect of a “photon” on the time evolution of the laser field. See also Figure 9.10.

When all technical disturbances are eliminated – mechanical, temperature, ambient air pressure fluctuations, and so on – the laser linewidth is determined only through spontaneous emission processes. It is then called *Schawlow–Townes linewidth*  $\Delta_{\text{ST}}$ . Here we use the phase diffusion model following Eq. (9.30) to calculate its value. For this derivation we have to evaluate the variance of the phase fluctuations caused by individual spontaneous emission events occurring at the rate  $R_{\text{spont}}$ :

$$\Delta v_{\text{ST}} = \langle \theta_{\text{SE}}^2 \rangle R_{\text{spont}}.$$

In Figure 9.13 a simple phasor model is shown, which illustrates the influence of a randomly emitted photon onto the field amplitude of the laser field.

The length of the electric field vector of the laser field is proportional to  $\sqrt{\bar{n}_{\text{ph}}} \gg 1$ . In these units, spontaneous emission causes a field contribution of length unity, which has a random phase with respect to the laser field. Thus in Figure 9.13, it has a random direction. The resulting field, the sum of the original laser field and the field of the spontaneously emitted photon, is slightly modified in both amplitude and phase. The small phase shift of an individual event is  $\delta\phi = \theta_{\text{ST}} \simeq \cos \alpha / \sqrt{\bar{n}_{\text{ph}}}$ . The variance, the r.m.s. value, then has the value  $\langle \cos^2 \alpha / \bar{n}_{\text{ph}} \rangle = 1/2\bar{n}_{\text{ph}}$ .

Spontaneous processes contribute at a ratio  $1 : \bar{n}_{\text{ph}}$  to the stimulated processes, the ratio of the Einstein  $A$  and  $B$  coefficients (Eq. (7.51)), with regard to the evolution of the resonator field. The rate is proportional to the number of excited particles  $n_e^{\text{st}}$ , so that drawing on Eqs. (9.14) and (9.21) we can write

$$R_{\text{spont}} = R_{\text{stim}} / \bar{n}_{\text{ph}} = \kappa n_e^{\text{st}} = \gamma_c n_e^{\text{st}} / n^{\text{st}}.$$

On the other hand, according to Eq. (9.18), we can connect the photon number with the output power,  $\bar{n}_{\text{ph}} = P/h\nu\gamma_{\text{out}} \simeq P/h\nu\gamma_c$ . So finally we arrive at the Schawlow–Townes linewidth (empty cavity linewidth  $\Delta v_{\text{cav}=\gamma_c}$ )

$$\Delta v_{\text{ST}} = \frac{n_e^{\text{st}}}{n^{\text{st}}} \frac{\pi h\nu}{P} \Delta v_{\text{cav}}^2. \quad (9.33)$$

<sup>3</sup> We should be aware of the fact that even “white noise” has an upper limit frequency – otherwise the r.m.s. value of the fluctuation would be unbounded according to Eq. (A.6)!

In a “good” four-level laser, the first factor is  $n_e^{\text{st}}/n^{\text{st}} \simeq 1$ . This surprising formula was presented by Schawlow and Townes as long ago as 1958 [103] and is called the *Schawlow–Townes linewidth*. As we already calculated in the section about He–Ne lasers, an extremely small linewidth of a few hertz or less is expected even for conventional lasers. Larger linewidths are only observed for small resonators with low mirror reflectance, for example, in semiconductor lasers. They are also subject to an additional broadening mechanism caused by amplitude–phase coupling (see Section 10.4.2).

## 9.5 Pulsed Lasers

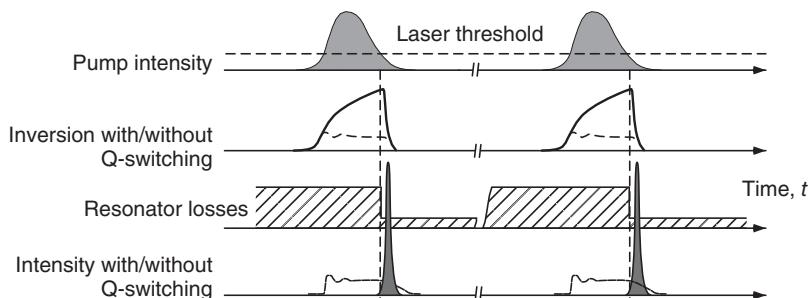
In Section 9.2.1 on relaxation oscillations, we found (see Figure 9.2) that switching processes can induce short laser pulses with intensities much higher than average. With pulsed lasers, a large amount of radiation energy, in common systems up to several joules, can be delivered within a short period of time. Its peak power depends on the pulse length.

One important method for generating short and very intense laser pulses is realized by the so-called “Q-switch” concept. Another method creates a coherent superposition of very many partial waves (“mode locking”), resulting in a periodic sequence of extremely short laser pulses.

### 9.5.1 “Q-Switch”

Pulsed neodymium lasers are among the most common systems offering very high peak powers. In such pulsed lasers, the pump energy is supplied through an excitation pulse, for example, from a flash lamp. The pump pulse (Figure 9.14) builds up the inversion until the laser threshold is passed. Then stimulated emission starts and the system relaxes to the equilibrium value. In the neodymium laser the amplitude damping occurs so fast that the output power follows the excitation pulse with small relaxation oscillations.

Alternatively, lasing can initially be suppressed by increasing the resonator losses with a Q-switch. If the accumulation time is short compared to the decay period of the upper laser level (for the neodymium laser, e.g., 0.4 ms), the laser medium acts as an energy storage device, and the inversion continues to



**Figure 9.14** Time chart of pulsed laser oscillation with and without (dashed lines) Q-switch.

increase. If the Q-switch triggered by an external impulse is again set to high Q factor or low-loss mode, stimulated emission begins, and now, by fast exhaustion of the accumulated energy, a laser pulse is generated that is short compared to the non-switched operation with much higher peak power. The repetition rate of such a laser system usually lies between 10 Hz and 1 kHz.

### 9.5.1.1 Technical Q-Switches

Q-Switches have to fulfill two conditions: in the open state the resonator Q factor has to be reduced efficiently, whereas in the closed state its insertion loss has to be small compared with other losses. Typical systems for a Q-switch are presented in Figure 9.15 and described as follows:

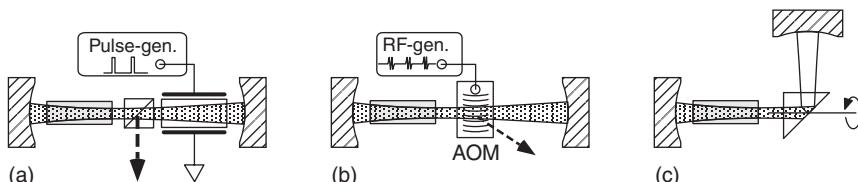
- 1) **Pockels cell.** The Pockels effect already described in Section 3.5.1 makes a voltage-driven retarder plate available. In combination with a polarizer (Figure 9.15), the resonator transmission can be modulated very efficiently. The switching time of a Pockels cell is in the nanosecond domain. It is primarily limited by the capacitance of the crystal electrodes and the resistance of the electrical leads.
- 2) **Acousto-optical modulator (AOM).** In the AOM (see also Section 3.5.4), a radio-frequency (r.f.) generator induces an acoustic wave, causing a periodic variation of the refraction coefficient. Laser radiation is deflected by diffraction off this grating out of the resonator and frequency-shifted at the same time. The r.f. power can be switched by suitable semiconductor components with nanosecond rise times.
- 3) **Rotating prism.** The Q-switch can also be realized by a mechanical rotating prism, which allows the laser to start only in a narrow acceptance angle range.

### 9.5.1.2 Cavity Dumping

The Pockels cell and the AOM of Figure 9.15 provide a second output port. This may be used for the so-called *cavity dumping* method. For this, in the closed laser oscillator, a strong oscillation builds up within the resonator first. Through an external pulse triggering the AOM or Pockels cell, this energy is then dumped out of the resonator. The method can also be combined with the mode locking concept of the following section in order to achieve particularly high peak powers.

### 9.5.2 Mode Locking

Even the simplest superposition of two laser beams with different frequencies  $\omega$  and  $\omega + \Omega$  causes periodic swelling up and down, as is well known from



**Figure 9.15** Q-switch and cavity dumping: (a) electro-optical (Pockels cell), (b) acousto-optical (Bragg cell), and (c) mechanical (rotating prism). See text.

amplitude modulation. For equal partial amplitudes with  $I_0 = c\epsilon_0|E_0|^2$ , we have

$$\begin{aligned} E(t) &= E_0 e^{-i\omega t} + E_0 e^{-i\omega t} e^{-i\Omega t} e^{-i\phi}, \\ I(t) &= \frac{1}{2}c\epsilon_0|E(t)|^2 = I_0[1 + \cos(\Omega t + \phi)]. \end{aligned}$$

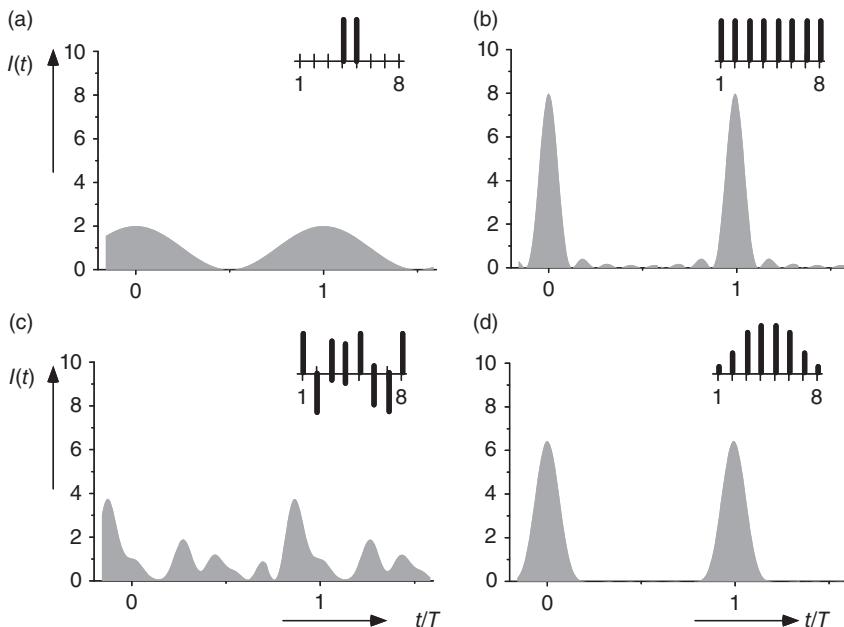
When we neglect the dispersive influence of the optical elements, laser resonators of length  $n\ell$  ( $n$  is refraction coefficient) provide an equidistant frequency spectrum with  $\Omega = 2\pi c/2n\ell$  (Eq. (6.19)) that virtually offers itself for synthesis of time-periodic intensity patterns. *Mode locking* establishes a technical procedure to physically realize Fourier time series consisting of many optical waves.

While for two waves the phase  $\phi$  causes only an overall phase shift of the sinusoidal modulation pattern, the pattern originating from superposition of multiple waves also depends on individual phase positions, as we show in Figure 9.16 with the example of eight superimposed waves. We can calculate the field amplitude in general according to

$$E_N(t) = \frac{E_0}{\sqrt{N}} e^{-i\omega t} e^{iN\Omega t/2} \sum_{n=1}^N \alpha_n e^{-in\Omega t} e^{-i\phi_n},$$

where important characteristic quantities include

$$\begin{array}{ll} \text{Pulse sequence frequency} & f = \Omega/2\pi \\ \text{Pulse period} & T = 2\pi/\Omega. \end{array}$$

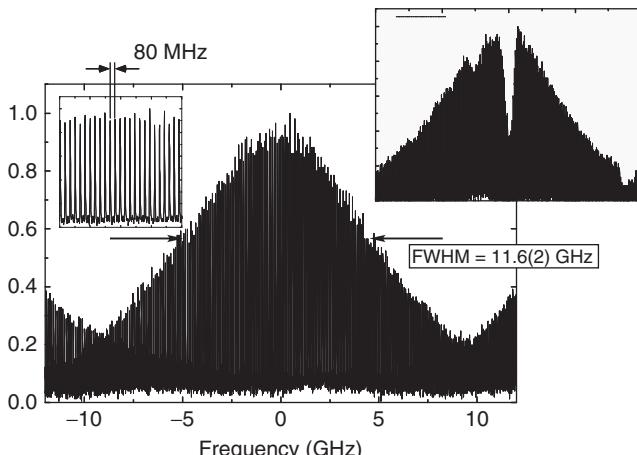


**Figure 9.16** Intensity variation in time for the superposition of up to eight harmonic waves. The vertical bars indicate the relative strength and phase position of the partial waves. (a) 2 partial waves with equal amplitudes cause sinusoidal amplitude modulation; (b) 8 partial waves with equal amplitudes and initial phases identical; (c) with equal amplitudes and random initial phases; (d) with gaussian amplitude distribution and initial phases identical.

The mid-frequency is called the *carrier frequency*  $\omega_0 = \omega - N\Omega/2$ , and different waves with frequency differences  $\Delta f = nf = n\Omega/2\pi$  contribute to the total wave with phases  $\phi_n$ . The partial amplitudes have been chosen in such a way that the intensity  $I_0 = (c\epsilon_0/2)E_0^2 \sum_n \alpha_n^2$  is distributed among partial amplitudes with  $\alpha_n E_0$  and  $\sum_n \alpha_n^2 = 1$ . Thus the intensity distributions in Figure 9.16 are comparable to each having the same mean power. In Figure 9.16 three characteristic situations are presented:

- 1) In the upper right part, all partial waves have identical amplitudes  $\alpha_n = \sqrt{1/n}$  and are in phase with  $\phi_n = 0$  for all  $n$ . For this situation, very sharp periodic maxima with a small peak width  $\Delta t \approx 2\pi/(N\Omega) = T/N$  occur. The secondary maxima are characteristic for an amplitude distribution with a sharp boundary.
- 2) In the lower right part, the partial waves are in phase as well, though the amplitudes have been chosen  $\alpha_n \propto \exp\{-[(2n - N - 1)/2]^2/2\}$ , that is, following a Gaussian distribution, which is symmetrical to the carrier frequency  $\omega_0$ . By this distribution, the side lobes that occur between the maxima in the previous example are suppressed very efficiently, and the laser power is concentrated to the maxima. The achievable peak power though is slightly lower. This situation resembles closely the conditions of a real laser resonator. In Figure 9.17 a frequency spectrum of a periodic train of 27 ps Ti-sapphire laser pulses measured in a Fabry-Pérot resonator is shown.
- 3) In Figure 9.16c, for comparison, the situation for random phases  $\phi_n$  of the partial waves is presented, which makes a noisy but periodic pattern.

Let us now study the relation between pulse length and bandwidth (see Section 3.3) and therefore consider a periodic series of Gaussian-shaped pulses with  $E(t) = \sum_n E_0 \exp\{-[(t - nT)/\Delta t]^2/2\}e^{-i\omega t}$ . According to the theory of Fourier



**Figure 9.17** Frequency spectrum of the 27 ps pulses of a mode-locked Ti-sapphire laser recorded with a Fabry-Pérot resonator with 7.5 mm distance between the mirrors or  $\Delta_{\text{FSR}} = 20$  GHz. The smaller pictures show an enlarged detail and the absorption of a cesium vapor cell in the ray trajectory [142], respectively.

series, we can obtain the  $n$ th Fourier amplitude for  $n\Omega$  from

$$\mathcal{E}_n = E_0 \int_{-\tau/2}^{\tau/2} e^{-(t/\Delta t)^2/2} e^{-in\Omega t} dt \approx E_0 e^{-(n\Omega\Delta t)^2/2}.$$

For this, to a good approximation, for very sharp pulses, only that single pulse centered at  $t = 0$  is taken into account, and the integration limits are extended to  $\pm\tau/2 \rightarrow \pm\infty$ . We define a bandwidth by  $2\pi f_B = \Omega_B = 2N\Omega$  with  $2N$  being the effective number of participating laser modes. The contribution of the modes to the total power drops to  $1/e$  of the central mode at  $n = N$ . The bandwidth  $f_B$  and pulse length  $2\Delta t$  (measured at relative amplitude value  $1/e$ ) are connected to each other and to the peak width  $\Delta t_{\text{FWHM}}$ :

$$\Delta t = \frac{1}{N\Omega} = \frac{\Delta t_{\text{FWHM}}}{\sqrt{8 \ln 2}} = \frac{\Delta t_{\text{FWHM}}}{2.35}.$$

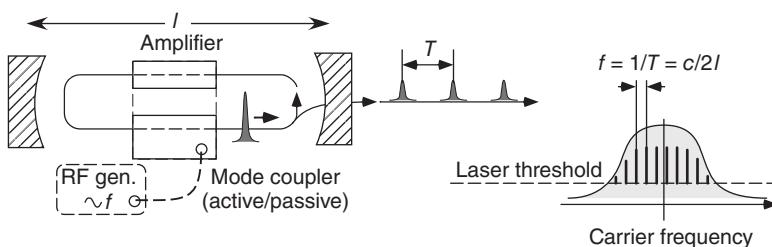
### 9.5.3 Methods of Mode Locking

In order to achieve pulses as short as possible, it is first important to use a laser amplifier with a very large bandwidth. For a sufficiently long lifetime of the upper laser level, the excitation can conveniently be generated by a continuous-wave laser. The stored energy is withdrawn from the laser medium by pulses separated typically by 12.5 ns, a time that is short compared with, for example, the lifetime of 4  $\mu$ s of the upper Ti-sapphire laser level. For other systems such as the dye laser, also the so-called “synchronous pumping” excitation scheme is used. In that case and owing to the short lifetime of the upper laser level, the pumping laser delivers a periodic and exactly synchronized sequence of short pulses. As a certain special case, which we skip here, we just mention the diode laser. By suitable modulation of the injection current (Section 10.4.1), it directly delivers very short pulses down to 10 ps. It has been intensively studied because of its significance for optical communication.

Table 9.2 contains important examples of lasers used for the generation of extremely short pulses and, for comparison, the limited potential of the helium–neon classic. The typical repetition rate of mode-locked lasers is 80 MHz and 12.5 ns pulse distance, which is determined by the characteristic construction lengths  $\ell$  setting the repetition rate at  $T = 2n\ell/c$ .

**Table 9.2** Mode locking and bandwidth.

Laser	Wavelength $\lambda$ (nm)	Bandwidth $f_B$ (THz)	Pulse duration $2\Delta t$	Pulse length $\ell_p = 2c\Delta t$
Helium–neon	633	0.001	150 ps	–
Nd:YLF	1047	0.4	2 ps	0.6 mm
Nd:glass	1054	8	60 fs	18 $\mu$ m
GaAs diode	850	2	20 ps	6 mm
Ti–sapphire	900	100	6–8 fs	2 $\mu$ m
NaCl–OH <sup>–</sup>	1600	90	4 fs	1.5 $\mu$ m



**Figure 9.18** Laser with mode locking. In the resonator, a spatially well-localized light pulse is circulating. Mode locking is achieved actively, for example, by modulation of the cavity  $Q$  factor, or passively by saturable absorbers or Kerr lens mode locking. On the lower right a spectrum of the contributing modes is shown.

Mode locking within the laser resonator is achieved by modulation of the resonator losses synchronized to the pulse circulation. In Figure 9.18 the mode coupler is set to transmission only when the pulse passes and to opaque otherwise. This modulation can be controlled either actively by the Q-switch components of Figure 9.15 or by passive nonlinear elements. Among them is the so-called saturable absorber, which is mainly used for dye lasers. A saturable absorber (optical saturation of an electric dipole transition is treated in Section 12.2.1) has an absorption coefficient that dies away at intensities above the so-called saturation intensity  $I_{\text{sat}}$ :

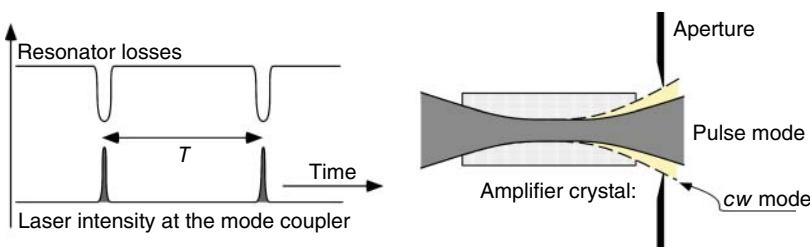
$$\alpha(I(t)) = \frac{\alpha_0}{1 + I(t)/I_{\text{sat}}}.$$

By means of an intense laser pulse circulating in the resonator (Figure 9.18), the absorber is easily saturated, and hence resonator losses are rapidly reduced during pulse passage. This passive modulation leads to self-locking of the laser modes. A variant of the passive mode locking not studied intensively anymore (colliding pulse mode locking (CPM) laser) uses two pulses circulating in the resonator that hit each other exactly in the saturable absorber.

The most successful method in technical applications at this time is the so-called Kerr lens mode locking (KLM), which causes a time-dependent variation of the resonator geometry due to the intensity dependence of the refraction coefficient

$$n = n_0 + n_2 I(t).$$

KLM is an example of the application of self-focusing and will be discussed in more detail in the section on nonlinear optics (Section 16.2.1). The dispersive nonlinearity reacts extremely fast, essentially instantaneously, to variations of the intensity and therefore is advantageous for very short pulses. At the center of a Gaussian-shaped beam profile (for positive  $n_2$ ), the refractive index is increased more strongly than in the wings and hence causes self-focusing, which changes the beam geometry as presented in Figure 9.19. Since the resonator losses depend on the beam geometry (the alignment of the resonator!), this phenomenon has the same effect as a saturable absorber and can be used for mode locking.



**Figure 9.19** Time dependence of resonator losses and the influence of a Kerr lens on the beam geometry. The mode locking effect can be supported by the use of an additional aperture.

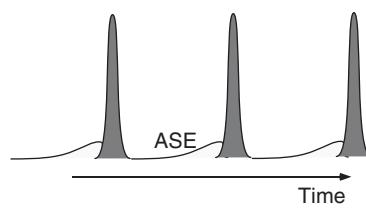
KLM was discovered by Sibbett *et al.* in the Ti–sapphire laser in 1991 [143]. It has led to revolutionary simplifications for the generation of ultrafast pulses due to its particularly simple application, since the nonlinear passive mode locker, the Kerr lens, is intrinsic to the Ti–sapphire amplifier crystal.

The KLM method alone though is not sufficient to generate pulses shorter than about 1 ps. In Section 3.3.1 we investigated the influence of dispersion and group velocity dispersion (GVD) on the shape of propagating light pulses, which naturally play an important role when the shortest light pulses are to be generated in a laser resonator containing several dispersive elements. The GVD can be compensated through the arrangement of prisms of Figure 16.6 on p. 489. The prism combination is traversed twice per round trip in the resonator. In a ring resonator two pairs of prisms have to be supplied to recombine the beams again. Another technique for dispersion control is offered by dielectric mirrors with specially designed coatings (*chirped mirrors*). Very compact femtosecond oscillators can be built with them.

Here we have considered the mode-locked lasers only in their simplest situation, that is, for steady-state conditions. The operation of mode-locked lasers though raises many interesting questions about laser dynamics, for which we refer the reader to the specialized literature. Such questions include the starting behavior. How does the passively locked laser get to this state at all? From a naive point of view, we can make, for example, intensity fluctuations responsible for this, which may always be triggered by slight mechanical vibrations.

Another phenomenon is the amplified spontaneous emission (ASE), which sometimes causes annoying side effects in experiments. It occurs because, during the pumping phase between the pulses, the amplifier already emits radiation energy, which is intensified in the direction of the desired laser beams due to the geometry (Figure 9.20).

The ASE can be suppressed by, for example, saturable absorbers that transmit light only above a certain threshold intensity or separate it from the laser beam by external spatial filtering since the ASE in principle has much larger divergence.



**Figure 9.20** Amplified spontaneous emission, preceding each laser pulse.

### 9.5.4 Measurement of Short Pulses

The measurement of the temporal properties (especially pulse duration) of short pulses is limited to about 100 ps by common photodiodes and oscilloscopes due to their limited bandwidth (several gigahertz). On the electronic side, the so-called streak camera can be used, a channel plate generating an electron beam that is deflected rapidly similar to an oscilloscope. It leaves a trace on the camera and so converts the time dependence into a local variation. With recent models a time resolution down to 100 fs can be achieved.

A purely optical standard method is offered by the autocorrelator, for example, realized by the schematic setup of Figure 9.21: a pulsed laser beam is split into two partial beams and superimposed in a nonlinear crystal (NLC) in such a way that a frequency-doubled signal (details about frequency doubling will be presented in Section 15.4) occurs. A signal is only registered on the photodiode if the partial pulses are superimposed correctly. The voltage signal as a function of the displacement  $\Delta x = c\Delta t$  of one arm relative to the other one,

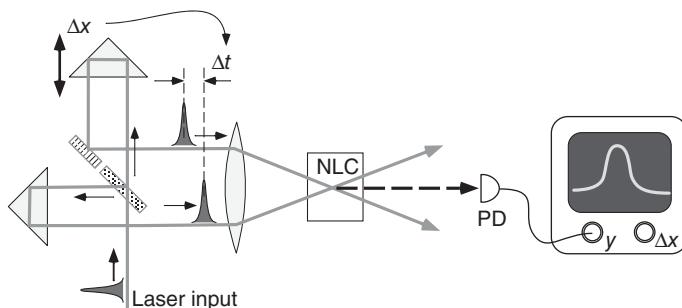
$$I_{PD}(\Delta t) \propto E(t)E(t + \Delta t),$$

also has pulse shape, but is the result of a convolution of the pulse with itself (therefore autocorrelation), from which the pulse shape has to be deduced by some suitable transformations or models.

### 9.5.5 Tera- and Petawatt Lasers

The new potential for generating extremely short laser pulses has also opened a window to the generation of extremely intense laser “flashes,” at least for a very short period in time. The field intensities are so large that matter is transferred to completely new states, which at best can be expected in special stars.

Even with a “common” femtosecond oscillator (Ti-sapphire laser, 850 nm,  $f = 80$  MHz,  $\langle P \rangle = 1$  W, medium power), using appropriate components for the compensation of GVD [144], pulses can be generated with a duration of only  $2\Delta t = 10$  fs. Even though such pulses only contain small amounts of energy



**Figure 9.21** Autocorrelator for the measurement of the time dependence of very short laser pulses. In the direction of the photodiode (PD), a signal only occurs when the laser pulses are correctly superimposed in both time and space in the nonlinear crystal (NLC).

$E_{\text{pulse}}$ , they already make available considerable peak power  $P_{\text{max}}$  and peak field intensities  $E_{\text{max}}$ :

$$\begin{aligned} E_{\text{pulse}} &= 1 \text{ W}/80 \text{ MHz} = 12.5 \text{ nJ}, \\ P_{\text{max}} &\approx E_{\text{pulse}}/(2\Delta t) \simeq 1 \text{ MW}, \\ E_{\text{max}} &\approx 2P_{\text{max}}/(\pi w_0^2 c \epsilon_0) = 7 \times 10^7 \text{ V cm}^{-1}. \end{aligned}$$

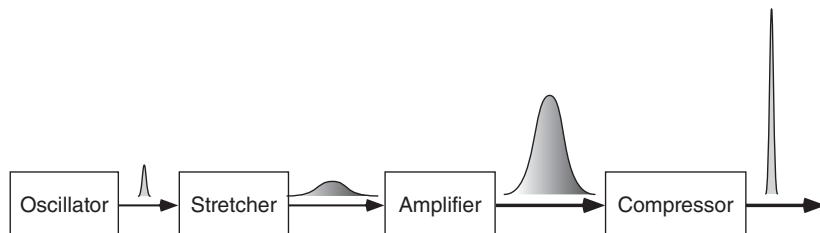
For the calculation of the field intensity, we have assumed the laser power to be concentrated onto a focal spot with a diameter of  $10 \mu\text{m}$ . Besides, there, even a 1 mW He–Ne laser reaches field intensities of about  $1 \text{ kV cm}^{-1}$ ! According to this an increase of the pulse energy to 1 J, which can be achieved today using tabletop equipment, promises a power of about 100 TW, and even the petawatt range is in sight. For this, field intensities of up to  $10^{12} \text{ V cm}^{-1}$  are achieved, about 1000 times the “atomic field intensity”  $E_{\text{at}} = e/4\pi\epsilon_0 a_0^2 = 10^9 \text{ V cm}^{-1}$  experienced by an electron in the lowest hydrogen orbit!

However, the generation and use of such intense laser pulses are hindered by this highly interesting strong interaction with matter. In common materials (initiated by multiphoton ionization), dielectric optical breakdown occurs and destroys the amplifier. An elegant solution for this situation is offered by the method of chirped pulse amplification (CPA) (see Figure 9.22), for which the short pulse is first stretched (in space and time) to decrease the peak power. The stretched pulse is amplified, and the stretching is reversed immediately before the application to recover the original pulse shape.

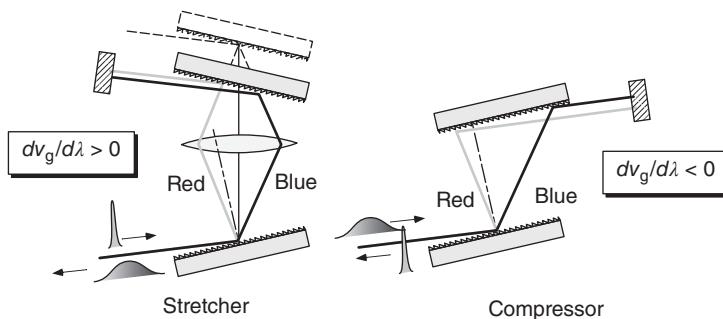
Optical gratings have proven to be very suitable components for achieving stretching as well as compression [145]. The concept of a grating stretcher and compressor is presented in Figure 9.23. The grating deflects red and blue parts of an incident pulse in different directions. In the stretcher two gratings are combined with 1 : 1 imaging properties. In a completely symmetric layout (dashed upper grating on the left in Figure 9.23), the upper grating would not change the shape of the impulses at all, only at the drawn position.

### 9.5.6 Coherent White Light

It seems contradictory at first to speak of white light in this context since the bias from classical optics suggests an absolutely incoherent light source with this term. White light covering the full spectral range of visible colors can, however,



**Figure 9.22** Chirped pulse amplification. By stretching, the peak power is decreased far enough that amplification without damage becomes possible.



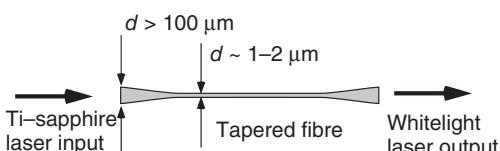
**Figure 9.23** Grating stretcher and compressor for femtosecond pulses.

be generated from ultrashort pulse lasers. White laser light is dispersed by a grating like common white light and exhibits the full range of colors. In contrast to conventional light sources, however, this light field shows well-modulated interferences when superposed with itself, as shown in Figure 9.24. It is thus coherent and truly laser light!



**Figure 9.24** Interference pattern observed when a white light laser is superposed with itself. The light beam was dispersed by a prism to demonstrate simultaneous interference of all contributing colors. (With permission from Harald Telle and Jörn Stenger.)

version processes. Optical fibers offer good transverse confinement and long interaction lengths. Conventional optical fibers (see Section 4.1), however, are operated in the *weak guiding limit* where small steps in the index of refraction of order 1% provide guiding but also cause the optical wave to be spread out over still a relatively large cross section.



**Figure 9.25** White light generation with a tapered optical fiber.

tapered section. Due to the large index of refraction step glass fiber versus air, the light field is now propagating in the strong guiding limit and confined to a cross section of about 1  $\mu\text{m}$ .

In contrast to continuous-wave lasers, ultrashort, intense laser pulses have a wide spectrum to begin with (see Section 9.5.3), although even Ti-sapphire lasers are limited to about 10% of the visible spectrum by their gain spectrum.

The generation of coherent white light – covering the entire visible spectrum or more – thus needs efficient nonlinear conversion processes.

One example of a scheme for white light generation uses a *tapered fiber* [146] as shown in Figure 9.25. The fiber is drawn out to extremely narrow cross sections (down to  $\simeq 500$  nm) from a conventional fiber (initial diameter 125  $\mu\text{m}$ ), providing efficient coupling into the

Several processes of the nature described in Chapters 15 and 16 on nonlinear optics such as self-phase modulation are responsible for the spectral broadening occurring during propagation of the short light pulse through the tapered section. Strong guiding is also provided by *photonic fibers* (see Section 4.2.6) [147, 148] where the optical wave is confined to the narrow core with diameter 1–2  $\mu\text{m}$ , much stronger than in any weakly guiding fiber, by large index of refraction steps resembling to the glass–air interface.

In photonic crystal fibers (PCFs), nonlinear conversion is not only supported by the large intensities obtained in the narrow core but can furthermore be controlled by the unusual dispersion properties of the fibers [149]. An *octave spanning* broadening (i.e., a spectrum spanning from a lower frequency  $f$  to at least its twofold value at  $2f$ ) has been realized and was instrumental to the successful operation of frequency combs.

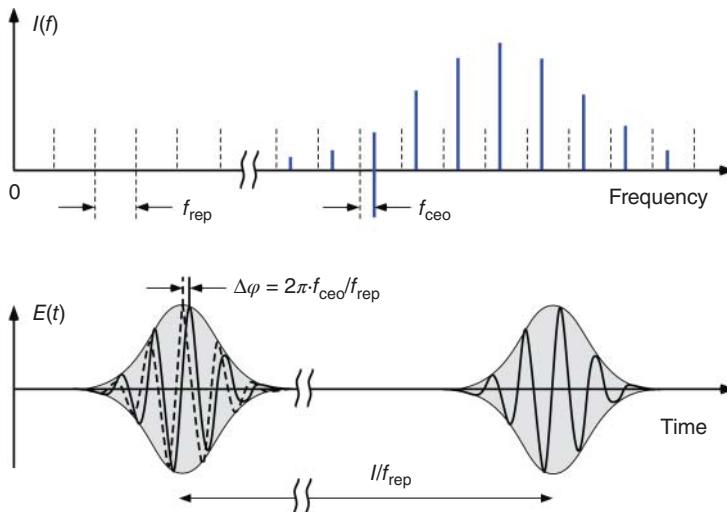
### 9.5.7 Frequency Combs

A frequency comb provides an array (a “comb”) of discrete frequencies where every individual line is an integer multiple of the repetition rate  $f_{\text{rep}}$  added to  $f_{\text{ceo}}$ , the so-called carrier envelope offset frequency,

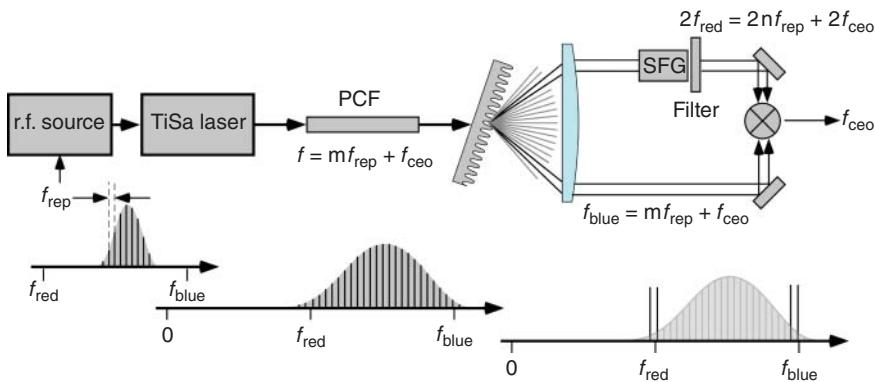
$$f_n = n \cdot f_{\text{rep}} + f_{\text{ceo}} . \quad (9.34)$$

Typical frequencies for  $f_{\text{rep}}$  and  $f_{\text{ceo}}$  are in the r.f. regime  $10^8$ – $10^{10}$  Hz. Thus the frequency comb links r.f. oscillators with a laser optical oscillator around  $10^{14}$ – $10^{15}$  Hz in a single step:  $n$  is typically a large number of order  $10^6$ .

In time domain, the complementary picture is composed of a series of short pulses with period  $1/f_{\text{rep}}$  as discussed already in Section 9.5.2 and Figure 9.16. A nonvanishing  $f_{\text{ceo}}$  accounts for the difference of phase and group velocity, the carrier phase “slips” with respect to the pulse envelope as illustrated in Figure 9.26.



**Figure 9.26** Spectral and time domain properties of a frequency comb.



**Figure 9.27** Self-referencing frequency comb.

While the large bandwidth of pulsed lasers seems to generally suggest short temporal coherence, it was in fact realized early on in 1978 [150] that the discrete mode structure of mode-locked lasers on the contrary offers a very high degree of temporal coherence for every individual spectral line, which could rival the most frequency-stable continuous-wave lasers. The full potential, however, was only realized around 2000 when – contrary to expectations – it was found that the spectrum of a mode-locked laser could be broadened to a much wider spectrum with the help of special PCFs without corrupting coherence properties! [149].

In order to fully exploit the full precision potential of the frequency comb, both  $f_{\text{rep}}$  and  $f_{\text{ceo}}$  must be known and controlled. While  $f_{\text{rep}}$  can be straightforwardly taken from an atomic clock-controlled r.f. oscillator, the offset frequency can be measured with the scheme shown in Figure 9.27 for a so-called self-referencing comb where the spectrum covers low frequencies  $f_{\text{red}}$  on the red wing of the comb spectrum as well as their second harmonic on the blue wing,  $f_{\text{blue}} \approx 2 \cdot f_{\text{red}}$ . Since  $f_{\text{ceo}}$  is related to the dispersion properties of the laser (phase vs group velocity), controlling this frequency (even to zero) can be achieved by actuators modifying the dispersion, for instance, tilting the end mirror of a KLM laser (Figure 16.6) or varying the pump power.

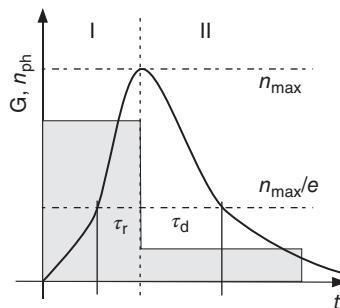
The frequency comb is a unique instrument for precision measurements: it offers to count the cycles of any optical frequency simply by comparing it to a nearby comb line. The Nobel Prize (shared with Roy Glauber) was awarded in 2005 “for their contributions to the development of laser-based precision spectroscopy, including the optical frequency comb technique” to J.L. Hall (born 1934) and T.W. Hänsch (born 1941). Revolutionary applications include the emergence of optical (rather than microwave) atomic clocks, efficient trace gas spectroscopy, calibration of astronomical spectrographs, controlled attosecond pulse generation, and more [151].

## Problems

- 9.1 Rate equations and Q-switch** Write a computer program in order to numerically study the rate equations (9.20). How do you set initial conditions? Study the influence of pump and loss rates. Make a model for a Q-switch.

- 9.2 Schawlow–Townes limit** (a) Give a qualitative explanation why the Schawlow–Townes laser linewidth is inversely proportional to the laser intensity,  $\Delta\nu_{\text{ST}} \propto 1/P$ . What is the origin of the dependence on the resonator damping rate  $\gamma_c^2$ ? (b) Compare the Schawlow–Townes linewidth of a He–Ne laser at 633 nm and a GaAs semiconductor laser at 850 nm for similar output powers of  $P = 1 \text{ mW}$ . Determine for this purpose the damping rates of resonators with  $\ell = 20 \text{ cm}$ ,  $n_{\text{ref}} = 1$ , and  $R_1 = 100\%$ ,  $R_2 = 99\%$  for the He–Ne laser, and  $\ell = 300 \mu\text{m}$ ,  $n_{\text{ref}} = 3.5$  for the GaAs laser. The mirrors of the semiconductor laser are formed by the cleaved facets of the laser crystal. (c) Assume that the He–Ne laser is used for precision frequency measurements. If the precision is limited only by the Schawlow–Townes limit, what are the maximum tolerable length fluctuations? What variation of the index of refraction inside the resonator is allowed?
- 9.3 Q-switching, pulse length, and peak power** Use the laser rate equations to make a simple model for a Q-switched laser pulse. Spontaneous emission is completely neglected for these intense pulses. We divide the pulse into two phases (Figure 9.28): the increasing part I with rise time  $\tau_r$  and the decreasing part II with decay time  $\tau_d$ .
- For an estimate of the rise time, we neglect the variation of the inversion. How does the photon number  $n_{\text{ph}}$  grow? Give the rise time as a function of  $n/n_{\text{st}}$ , where  $n$  describes the non-saturated and  $n_{\text{st}}$  the saturated inversion.
  - Once the gain is exhausted, the photon number starts to decay in the second phase. Which time constant is relevant, and how large is the entire pulse length? Nd:YAG and Nd:YLF laser ions have the same absorption and emission cross section, respectively, but the lifetime in the YAG crystal is only half as long as in the YLF crystal. Which one is more advantageous for large peak powers?
  - Estimate the photon number by neglecting pump rate and inversion decay during the duration of the pulse. Assume an initial inversion  $n(t_0) > n_{\text{st}}$  and stop the growth of the photon number  $n_{\text{ph}}$  when  $n(t) = n_{\text{st}}$ .

**Figure 9.28** Description of the laser pulse and the time evolution of amplification.



# 10

## Semiconductor Lasers

Immediately after the demonstration of the ruby laser (1960) and the helium–neon laser (1962), the lasing of diodes, or “semiconductor lasers,” was also predicted, and a little bit later was realized experimentally.<sup>1</sup> However, it took more than 20 years for those components to become commercially successful products, since numerous technological problems had to be overcome. The first laser diodes, for example, could operate only at cryogenic temperatures, while applications in general require operating temperatures close to room temperature. Moreover, GaAs was the first relevant material for the manufacture of laser diodes, and not silicon, which, then as now, otherwise dominates semiconductor technologies.

Today, laser diodes belong to the most important “optoelectronic” devices because they allow the direct transformation of electrical current into (coherent!) light. Therefore there are countless physical, technical, and economic reasons to dedicate a chapter of its own to these components and related laser devices.

### 10.1 Semiconductors

For a detailed description of the physical properties of semiconducting materials, we refer the reader to the known literature [38]. Here we summarize the properties of importance for the interaction with optical radiation.

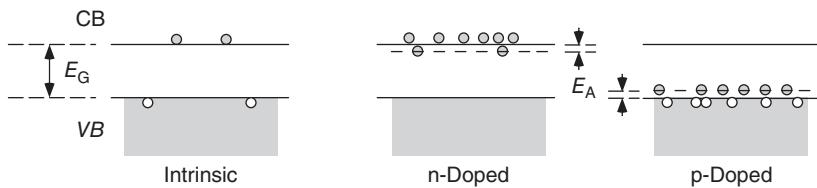
#### 10.1.1 Electrons and Holes

In Figure 10.1 the valence and conduction bands of a semiconducting material are presented. Free electrons with negative charge carry the current in the conduction band, whereas holes with positive charge<sup>2</sup> do so in the valence band.

---

<sup>1</sup> John von Neumann (1903–1957) carried out the first documented theoretical consideration of a semiconductor laser in 1953. This unpublished manuscript was reproduced in [152].

<sup>2</sup> It should not be forgotten that the term *hole* is only an – albeit very successful – abbreviation for a basically very complex physical many-particle system. Most of the physical properties (conductivity, Hall effect, etc.) of the electrons of the valence band can be very well described as if there were free particles with a positive charge and a well-defined effective mass.



**Figure 10.1** The band model for semiconductors. Electrons and holes can move freely and independently from each other. CB: conduction band; VB: valence band;  $E_g$ : bandgap energy;  $E_A$ : excitation energy of the dopant impurities.

The distribution of the electrons into the existing states is described by the Fermi function  $f(E)$ ,

$$f_{\text{el}}(E, \varepsilon_F) = [1 + e^{(E - \varepsilon_F)/kT}]^{-1}, \quad (10.1)$$

which is determined by the Fermi energy for electrons of the conduction band  $\varepsilon_F = \varepsilon_{\text{CB}}$  and temperature  $T$ . Especially at  $T = 0$  all energy states below the Fermi energy are completely filled and above it completely empty. The distribution of holes – missing electrons – is described in analogy by

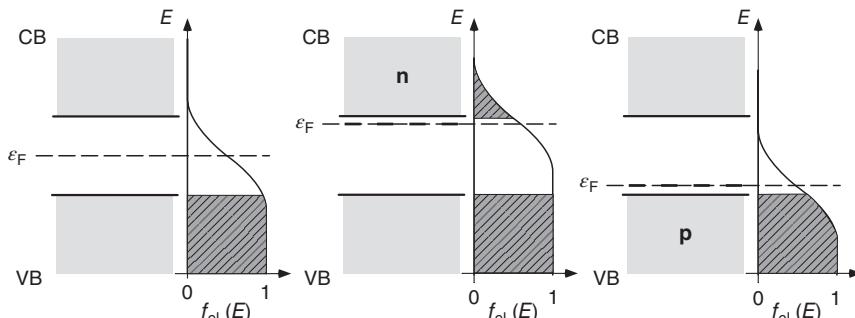
$$f_h = 1 - f_{\text{el}} = [1 + e^{(\varepsilon_F - E)/kT}]^{-1}. \quad (10.2)$$

Without external bias voltage, the equilibrium occupation numbers of electrons and holes are characterized by a common Fermi energy  $\varepsilon_F = \varepsilon_{\text{CB}} = \varepsilon_{\text{VB}}$ . In forward-biased operation at a pn junction a nonequilibrium, current-carrying situation relevant for laser operation arises with different Fermi energies for electrons and holes,  $\varepsilon_{\text{CB}} \neq \varepsilon_{\text{VB}}$ .

Some important situations of the Fermi distribution in intrinsic and doped (see Section 10.1.2) semiconductors are presented in Figure 10.2. At  $T = 0$  the Fermi energy gives exactly the energy up to which the energy levels are occupied.

### 10.1.2 Doped Semiconductors

An *intrinsic* semiconductor consists of a pure crystal, for example, the technologically most important material Si from main group IV in the periodic table or the III–V compound GaAs. In such a material the Fermi energy is found close to the middle of the bandgap. The occupation probability of the states can then



**Figure 10.2** Fermi distribution in intrinsic, n- and p-doped semiconductors with dopant energy levels, respectively.  $\varepsilon_F$ : Fermi energy.

approximately be described according to Boltzmann's formula:

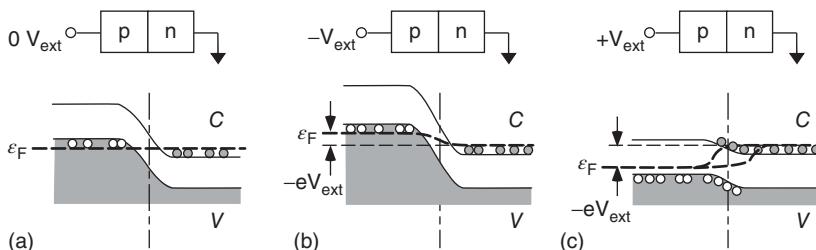
$$f_{\text{el}}(T) \simeq e^{-E_g/kT}.$$

The bandgap energy  $E_g$  depends on the material and is of the order of a few electronvolts; therefore at room temperature ( $kT \approx 1/40$  eV), there are only very few electrons in the conduction band. The revolutionary significance of semiconductors arises in principle from the possibility to increase the conductivity dramatically via doping (e.g., in Si, with impurity ions from main groups III or V) and even via different concentrations for holes and electrons (Figure 10.1). The deficit or excess of electrons of the impurity atoms generates energy states near the band edges, which are easy to excite at thermal energies. Electron charge carriers are generated in this way in an *n-doped* semiconductor, and holes in a *p-doped* system. The Fermi energy lies in this case near the acceptor (p doping) or the donor (n doping) level (Figure 10.2). Already at room temperature such a doped semiconductor exhibits a large conductivity caused by electrons in n-type and holes in p-type material.

### 10.1.3 pn Junctions

If electrons and holes collide with each other, they can “recombine,” emitting dipole radiation at optical or near-optical frequencies. Such processes are facilitated by having an interface between p- and n-doped semiconducting materials (a pn junction), which is the heart of every semiconductor diode. The properties of a pn junction that depend crucially on the sign of the bias voltage are summarized in Figure 10.3.

- *Zero-bias equilibrium (a)*: At the interface, electrons diffuse into the p-doped area and holes into the n-doped one, where they can recombine. At the barrier, a layer called *depletion zone* is emptied of charge carriers, and an electrical field is generated, counteracting any further diffusion.
- *Reverse bias (b)*: The depletion zone is enlarged.
- *Forward bias (c)*: A current flows through the junction, and electrons and holes flood the barrier layer and cause recombination radiation. Within the



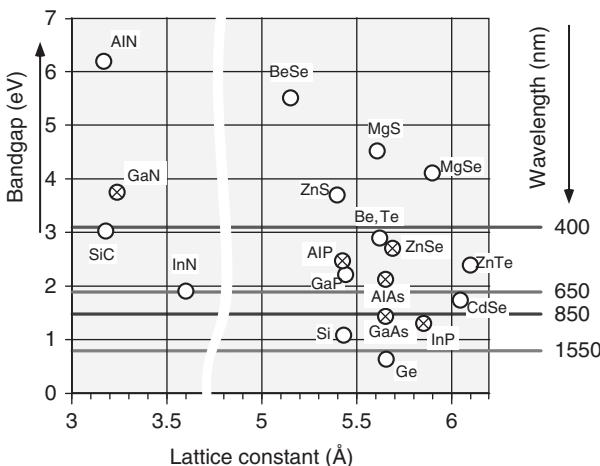
**Figure 10.3** A pn junction with free charge carriers. The dashed line indicates the Fermi energy  $\epsilon_F$ . (a) Zero-bias equilibrium. At the interface, electrons diffuse into the p-doped area and holes into the n-doped one, where they can recombine. At the barrier, a layer is depleted of charge carriers and an electrical field is generated, counteracting any further diffusion. (b) At reverse bias, the depletion zone is enlarged. (c) At forward bias, a current flows through the junction, electrons and holes flood the barrier layer and cause recombination radiation. Within the conduction and valence bands, there is thermal equilibrium characterized by two different Fermi energies for electrons and holes.

conduction and valence bands, there is thermal equilibrium characterized by two different Fermi energies for electrons and holes.

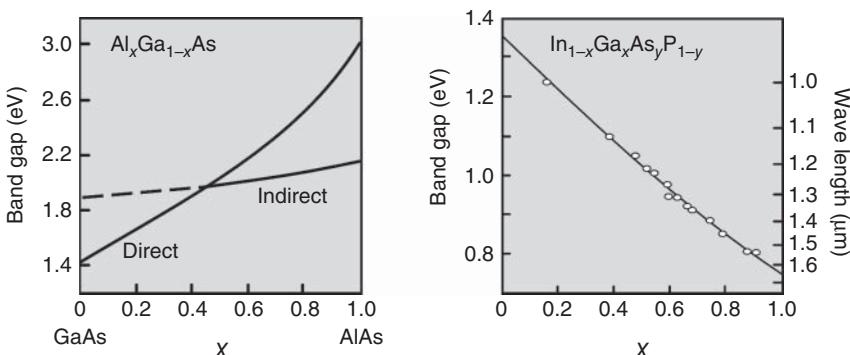
## 10.2 Optical Properties of Semiconductors

### 10.2.1 Semiconductors for Optoelectronics

From the optoelectronic point of view, the energy gap at the band edge is the most important physical quantity, since it determines the wavelength of the recombination radiation. It is presented in Figure 10.4 for some important optoelectronic semiconductors as a function of the lattice constants, which have technological meaning for the formation of compound crystals. A particular gift of nature for this is the extremely small difference of the lattice constants of GaAs and AlAs. Because of the excellent lattice match, the bandgap can be controlled over a wide range by the mixing ratio  $x$  in  $(\text{Al}_x\text{Ga}_{1-x})\text{As}$  compound crystals (Figure 10.5).



**Figure 10.4** Bandgap energy of some important semiconducting materials. Materials for which lasing has been realized already are marked with a cross. At the right-hand side, some technically relevant laser wavelengths are given.



**Figure 10.5** Bandgap energy in AlGaAs and InGaAsP as a function of the mixing ratio.

Other compound crystals have been in use as well for quite a long time. Especially, the wavelength of  $1.55\text{ }\mu\text{m}$  that is most important for optical telecommunications can be obtained from a quaternary InGaAsP crystal. Silicon, the economically most significant semiconducting material, does not play any role, since it does not have a direct bandgap but only an indirect one (see Section 10.2.4).

### Excuse: Blue luminescent gallium nitride, a scientific fairytale

The development of laser diodes experienced rapid progress in the 1980s and 1990s, but 1996 will go down in history as a very special year. In that year, Shuji Nakamura, with the Japanese company Nichia Chemical Industries Ltd, was able to present the world's first blue laser diodes to an astonished audience. He had made the devices based on GaN, which had been considered completely unsuitable for optoelectronics! This research was supported neither commercially nor academically, and success would not have been possible without the confidence of his boss, Nobuo Ogawa. With no experience at all on this topic in his company and not very much in touch with semiconductor lasers, since 1989 he had allowed the then 36-year-old engineer to pursue a research program that went against all established opinions about the potential of gallium nitride [153].

In fact, there had been commercial interest in blue luminescence long before the interest in blue laser radiation, since only with blue light sources was there the hope to produce fully colored screens based on semiconductors. Worldwide large sums had been invested in research on ZnSe, which was supposed to have the biggest chance of success. In textbooks, it could be read that GaN was unsuitable in spite of its well-known and attractive physical properties ((Al,Ga,In)N has a direct bandgap of  $1.95\text{--}6.2\text{ eV}$ ), since it could not be p-doped. This assertion though could not be maintained anymore after 1988, when Akasaki *et al.* were successful with the preparation of such crystals, though at first with a costly electron-beam technique. S. Nakamura succeeded crucially in the thermal treatment of GaN samples by replacing the  $\text{NH}_3$  atmosphere by  $\text{N}_2$ . He found that the ammonia atmosphere dissociated and the released hydrogen atoms passivated the acceptors in GaN.

Twenty years after this discovery, blue LEDs and laser diodes are widely available and instrumental in changing lighting appliances from cars to everybody's home. It comes as no surprise that this story from the scientific book of fairytales earned I. Akasaki (born 1929), H. Amano (born 1960), and S. Nakamura (born 1954) the Nobel Prize in 2014 "for the invention of efficient blue light-emitting diodes which has enabled bright and energy-saving white light sources."

### 10.2.2 Absorption and Emission of Light

In a semiconductor, electrons are excited from the valence band to the conduction band on absorption of light with a wavelength

$$\lambda < (E_g/hc)^{-1},$$

so that electron–hole pairs are generated. Under certain conditions, for example, at very low temperatures, absorption of light can be observed already below the band edge. During this process no free mobile charge carriers are generated, rather pairs bound in "excitonic" states with a total energy slightly below the edge

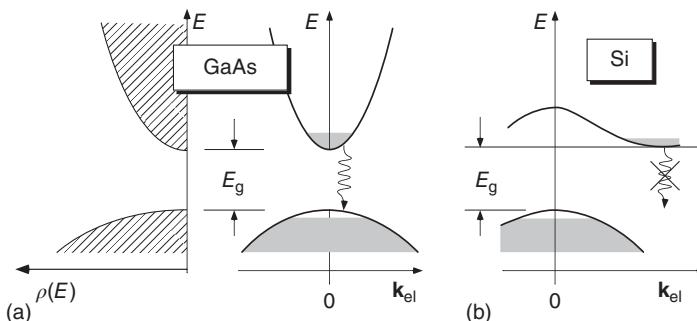
of the conduction band. Excitons, which resemble atoms made from pairs of electrons and holes, will however not play any role in our considerations.

If free electrons and holes are available, they can recombine under emission of light that again has a wavelength corresponding roughly to the band edge due to energy conservation. The “recombination radiation” though has furthermore to fulfill momentum conservation<sup>3</sup> for the electron–hole pair ( $\hbar\mathbf{k}_{el}$ ,  $\hbar\mathbf{k}_h$ ) as well as for the emitted photon ( $\hbar\mathbf{k}_{ph}$ ):

$$\begin{array}{ll} \text{Energy:} & E_{el}(\mathbf{k}_{el}) = E_h(\mathbf{k}_h) + \hbar\omega, \\ \text{Momentum:} & \hbar\mathbf{k}_{el} = \hbar\mathbf{k}_h + \hbar\mathbf{k}_{ph}. \end{array} \quad (10.3)$$

The  $\mathbf{k}$  vectors of the charge carriers are of magnitude  $\pi/a_0$  with  $a_0$ , indicating the lattice constant and therefore very much larger than  $2\pi/\lambda$ . That is why optical transitions only take place if the lowest-lying electronic states in the  $E$ – $\mathbf{k}$  diagram (the “dispersion relation”) are directly above the highest-lying hole states.

In Figure 10.6 the situation for two particularly important semiconductors is schematically presented. In the so-called “direct” semiconductor GaAs, at  $\mathbf{k} = 0$ , a conduction band edge with “light” electrons meets a valence band edge with “heavy” holes (the effective mass of the charge carriers is inversely proportional to the curvature of the bands); there direct optical transitions are possible. Silicon, on the other hand, is an indirect semiconductor. The band edge of the electrons occurs at large  $k_{el}$  values, and that of the holes at  $k = 0$ ; thus silicon cannot radiate! There are however weaker and more complex processes, for example, with the participation of a phonon that supplements a large  $\mathbf{k}$  contribution and thus ensures momentum conservation in Eq. (10.3) at negligible energy expense.



**Figure 10.6** (a) Electronic density of states and simplified dispersion relation for direct semiconductors (GaAs). The different curvatures of the bands are the origin of the different effective masses (see Eq. (10.4)). In equilibrium there are charge carriers only at the band edges (shadowed areas indicate filled states). Optical transitions start and end with hardly any change of the  $\mathbf{k}$  vector since the momentum of the photons is not visible on this scale. They can only take place if electrons of the conduction band meet an unoccupied state, a hole in the valence band. (b) For indirect semiconductors (Si), direct optical transitions are suppressed.

<sup>3</sup> In a crystal it is more exact to speak about *quasi-momentum conservation*.

The recombination radiation is caused by an optical dipole transition with a spontaneous lifetime  $\tau_{\text{rec}}$  of typically

$$\text{Recombination time } \tau_{\text{rec}} \simeq 4 \times 10^{-9} \text{ s.}$$

The recombination rate is also called the “interband” decay rate and is very slow compared to the collision time  $T'$  of the charge carriers with defects and phonons within the conduction and valence bands. This “intraband” scattering takes place on the picosecond time scale

$$\text{Relaxation time } T' \simeq 10^{-12} \text{ s,}$$

and ensures that, owing to relaxation within each of the bands, there is an equilibrium state determined by the crystal temperature.

### 10.2.3 Inversion in the Laser Diode

In a semiconductor, coherent light is generated by stimulated recombination radiation. In the beginning the pn junctions had to be very deeply cooled down to the temperature of liquid helium in order to suppress loss processes competing with luminescence and to generate an adequate inversion density for lasing. The development of the heterostructure laser, which we are going to discuss a little later, has overcome this problem and contributed decisively to the still growing success of semiconductor lasers.

The amplification is determined, among other things, by the number of charge carriers that can emit recombination radiation at a certain energy difference. For this, their density of states (DOS) has to be calculated from the  $(E, \mathbf{k})$  dispersion relations

$$E_{\text{el}} = E_{\text{CB}} + \frac{\hbar^2 \mathbf{k}^2}{2m_{\text{el}}^*} \quad \text{and} \quad E_{\text{h}} = - \left( E_{\text{VB}} + \frac{\hbar^2 \mathbf{k}^2}{2m_{\text{h}}^*} \right) \quad (10.4)$$

With  $E_{\text{CB},\text{VB}}$  we denote the edges of the conduction and valence bands. In the vicinity of the band edges, the dispersion relation is quadratic like for free particles, and the curvature is proportional to the inverse effective mass  $m^*$  (Figure 10.6), which, for example, for GaAs, yields electrons with  $m_{\text{el}}^* = 0.067 m_{\text{el}}$  and heavy holes with  $m_{\text{h}}^* = 0.55 m_{\text{el}}$ . In the three-dimensional (3D) volume, we have  $k_x^2 + k_y^2 + k_z^2 = k^2$ , and using  $\rho_{\text{el,h}}(k) dk = k^2 dk / 2\pi^2$  (see Section B.3), the DOS for electrons and holes are separately calculated according to

$$\rho_{\text{el,h}}(E) dE = \frac{1}{2\pi^2} \left( \frac{2m_{\text{el,h}}^*}{\hbar^2} \right)^{3/2} (E - E_{\text{CB,VB}})^{1/2} dE,$$

with  $E$  for electrons and holes counted from each band edge  $E_{\text{CB,VB}}$ . With this relation we can also determine the density of charge carriers for electrons and holes. We introduce the two quantities  $\alpha_{\text{el}} = (E_{\text{CB}} - E_{\text{CB}})/kT$  and  $\alpha_{\text{h}} = (E_{\text{VB}} - E_{\text{VB}})/kT$  and replace the integration variable by  $x = (E - E_{\text{CB}})/kT$  and  $x = (E_{\text{VB}} - E)/kT$ , respectively, giving

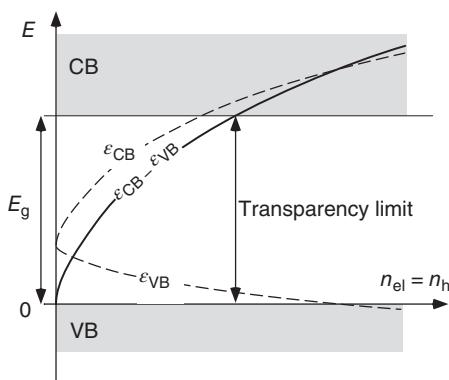
$$\begin{aligned} n_{\text{el,h}} &= \int_{E_{\text{CB,VB}}}^{\infty} \rho_{\text{el,h}} f_{\text{el,h}}(E, \epsilon_{\text{CB,VB}}) dE \\ &= \frac{1}{2\pi^2} \left( \frac{2m_{\text{el,h}}^* kT}{\hbar^2} \right) \int_0^{\infty} \frac{\exp(-\alpha_{\text{el,h}}) \sqrt{x} dx}{\exp(x) + \exp(-\alpha_{\text{el,h}})}. \end{aligned}$$

Estimates can be obtained easily by inserting the characteristic effective masses for GaAs. We obtain after a short calculation for  $T = 300\text{ K}$

$$\left\{ \begin{array}{l} n_{\text{el}} \\ n_{\text{h}} \end{array} \right\} = \left\{ \begin{array}{l} 4.7 \times 10^{17} \text{ cm}^{-3} \\ 1.1 \times 10^{19} \text{ cm}^{-3} \end{array} \right\} e^{-\alpha_{\text{el,h}}} \int_0^\infty \frac{\sqrt{x} dx}{e^x + e^{-\alpha_{\text{el,h}}}}. \quad (10.5)$$

For every charge carrier density, analysis of the implicit equation (10.5) determines a Fermi energy for both conduction and valence bands, ( $\{n_{\text{el}}, n_{\text{h}}\} \leftrightarrow \{E_{\text{CB}}, E_{\text{VB}}\}$ ), where we typically assume that electrons and holes have the same concentration. In a laser diode this density of charge carriers is maintained by the injection current (see p. 329).

### Example: Charge carrier densities in GaAs



**Figure 10.7** Density of charge carriers and Fermi energies.

A case of special interest occurs at  $\alpha_{\text{el,h}} = 0$  since there the Fermi energy just reaches the edges of the valence and conduction bands. This case can even be resolved analytically:

$$n_{\text{el}} = 4.7 \times 10^{17} \text{ cm}^{-3} \int_0^\infty \frac{\sqrt{x} dx}{e^x + 1} = 3.2 \times 10^{17} \text{ cm}^{-3}.$$

In general, Eq. (10.5) has to be evaluated by numerical methods. The result of such an evaluation is presented in Figure 10.7.

Owing to the smaller effective masses, the electron concentration

makes the Fermi energy  $\epsilon_{\text{CB}}$  increase faster than the hole concentration  $\epsilon_{\text{VB}}$ , and it reaches the band edge first. Through strong p doping, though, the Fermi energy in the currentless state (i.e., free of charge carriers) is shifted closer to the valence band, so that  $\epsilon_{\text{VB}}$  gets to the valence band edge at a lower density of charge carriers.

As we will see later (Eq. (10.9)), it is already sufficient for inversion if the difference of the Fermi energies  $\epsilon_{\text{CB}} - \epsilon_{\text{VB}}$  is larger than the bandgap energy  $E_g$ . There the so-called transparency limit is reached since the radiation field is no longer absorbed but amplified.

In terms of lasing we are more interested in which states can contribute to a transition with energy  $E = \hbar\omega > E_g = E_{\text{CB}}(\mathbf{k}) - E_{\text{VB}}(\mathbf{k})$  or where we can expect inversion. We obtain the rate of stimulated emission from the Einstein  $B$  coefficient. A selected  $\mathbf{k}$  vector contributes with a rate

$$R_{\text{CV}}^{\mathbf{k}} = B_{\text{CV}} U(\omega(\mathbf{k})) \{ f_{\text{el}}^{\text{CB}}(E_{\text{el}}(\mathbf{k})) (1 - f_{\text{el}}^{\text{VB}}(E_{\text{h}}(\mathbf{k}))) \}$$

to the total rate of stimulated emission at frequency  $\omega = (E_{\text{el}} - E_{\text{h}})/\hbar$ . In this term the occupation probabilities at the energy difference of the direct transition  $\hbar\omega$

in the valence band ( $f_{\text{el}}^{\text{CB}}(E_{\text{el}}(\mathbf{k}))$ ) and conduction band ( $1 - f_{\text{el}}^{\text{VB}}(E_{\text{h}}(\mathbf{k}))$ ) are taken into account. The energy density of the radiation field is given by  $U(\omega(\mathbf{k}))$ . The rate for absorption is derived accordingly:

$$R_{\text{VC}}^k = B_{\text{VC}} U(\omega(\mathbf{k})) \left\{ f_{\text{el}}^{\text{VB}}(E_{\text{h}}(\mathbf{k})) (1 - f_{\text{el}}^{\text{CB}}(E_{\text{el}}(\mathbf{k}))) \right\}.$$

The total rate of transitions at frequency  $\omega$  has to be determined from the sum  $R_{\text{CV}}(\omega) = \sum_{\mathbf{k}} R_{\text{CV}}^k \delta(\omega - (E_{\text{el}} - E_{\text{h}})/\hbar) \rho(\mathbf{k}) d^3 k$ . From the joint dispersion relations for electrons and holes,

$$E = E_{\text{el}} - E_{\text{h}} = E_g + \frac{\hbar^2 \mathbf{k}^2}{2m_{\text{el}}^*} + \frac{\hbar^2 \mathbf{k}^2}{2m_{\text{h}}^*},$$

we find using the methods from Section B.3 the so-called reduced DOS with reduced mass  $\mu^{-1} = m_{\text{el}}^{*-1} + m_{\text{h}}^{*-1}$  and  $\rho(\omega) = \hbar \rho(E)$ :

$$\rho_{\text{red}}(\omega) = \frac{1}{2\pi^2} \left( \frac{2\mu}{\hbar} \right)^{3/2} (\omega - E_g/\hbar)^{1/2}. \quad (10.6)$$

Then the difference of emission and absorption rates can be calculated with  $B_{\text{CV}} = B_{\text{VC}}$  from

$$\begin{aligned} R_{\text{CV}} - R_{\text{VC}} &= B_{\text{CV}} U(\omega) [f_{\text{el}}^{\text{CB}}(1 - f_{\text{el}}^{\text{VB}}) - f_{\text{el}}^{\text{VB}}(1 - f_{\text{el}}^{\text{CB}})] \rho_{\text{red}} \\ &= B_{\text{CV}} U(\omega) [f^{\text{CB}} - f^{\text{VB}}] \rho_{\text{red}}. \end{aligned} \quad (10.7)$$

The role of inversion, which in conventional lasers is given by the occupation number difference of the excited state ( $N_e$ ) and lower state ( $N_g$ ) of the laser transition, is now taken over by the product

$$(N_e - N_g) \rightarrow (f^{\text{CB}} - f^{\text{VB}}) \rho_{\text{red}} [(E_{\text{CB}} - E_{\text{VB}})/\hbar],$$

with the first factor controlled by the injection current. The existence of inversion clearly depends on the sign of the factor  $(f^{\text{CB}} - f^{\text{VB}})$ , or

$$f_{\text{el}}^{\text{CB}} > f_{\text{el}}^{\text{VB}} \quad \text{and} \quad \frac{1}{1 + e^{E_{\text{el}} - \epsilon_{\text{CB}}}} > \frac{1}{1 + e^{E_{\text{h}} - \epsilon_{\text{VB}}}}.$$

A small transformation uncovers

$$\epsilon_{\text{CB}} - \epsilon_{\text{VB}} < E_g,$$

and we find that frequencies with gain must be smaller than the corresponding difference of the Fermi levels,  $\epsilon_{\text{CB}}$  and  $\epsilon_{\text{VB}}$ . On the other hand, only energy differences above the bandgap energy can be amplified; hence the inversion condition for semiconductor lasers can be written as

$$\epsilon_{\text{CB}} - \epsilon_{\text{VB}} > E_g.$$

For GaAs this situation is typically reached with charge carrier densities  $n_{\text{el}} = n_{\text{h}} = 10^{18} \text{ cm}^{-3}$ , as detailed in the example on p. 326.

#### 10.2.4 Small Signal Gain

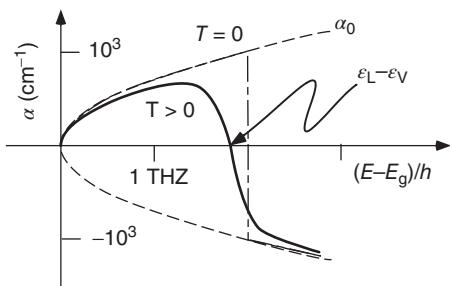
Consider a pulse of light propagating in the  $z$  direction with group velocity  $v_g$  and spectral intensity  $I(\omega) = v_g U(\omega)$ . The change of the intensity by absorption and emission is described according to Eq. (10.7). After a short travel length

$\Delta z = v_g \Delta t$ , we can thus write  $\Delta I = (R_{CV} - R_{VC})\hbar\omega\Delta z$ . Then the absorption and emission coefficients are determined according to Eq. (7.23):

$$\alpha(\omega) = \frac{\Delta I}{I\Delta z} = \frac{(-R_{VC} + R_{CV})\hbar\omega}{v_g U(\omega)}.$$

We use the identity  $B_{CV} = A_{CV}/[\hbar\omega(\omega^2/\pi^2c^3)] = A_{CV}/[\hbar\omega\rho_{ph}(\omega)]$  according to Eq. (7.50) to relate the Einstein coefficient to the microscopic properties of the semiconductor. Then we can write with  $\tau = A_{CV}^{-1}$

$$\alpha(\omega) = \frac{1}{v_g \tau} \frac{\rho_{red}(\omega)}{\rho_{ph}(\omega)} (f^{CB} - f^{VB}) = \alpha_0 (f^{CB} - f^{VB}), \quad (10.8)$$



**Figure 10.8** Absorption and (small signal) gain at a pn junction for a given density of charge carriers at  $T = 0$  K and at elevated temperature.

where we have introduced the maximum absorption coefficient  $\alpha_0(\omega)$ , which is proportional to the reduced DOS.

For an estimate we use specifications characteristic for GaAs lasers: wavelength  $\lambda_L = 850$  nm, reduced effective mass  $\mu = 0.06m_{el}$ , recombination time  $\tau_{rec} = 4 \times 10^{-9}$  s, and group velocity  $v_g \simeq c/3.5$ . For typical separations of the laser frequency from the band edge of  $1$  THz =  $10^{12}$  Hz, corresponding to 2 nm in terms of wavelengths, one can calculate

$$\alpha_0 = 6.8 \times 10^3 \text{ cm}^{-1} \sqrt{(v_L - E_g/h)/\text{terahertz}}$$

The very large gain factors  $\alpha_0$  are somewhat reduced in a room-temperature laser by the Fermi factor from Eq. (10.8).

Like in the gas laser, amplification in the laser diode is achieved when stimulated emission overcomes the losses caused by outcoupling, scattering, and absorption. In Figure 10.8 we have calculated the gain and loss profile for an example. At  $T = 0$  the Fermi distributions are step-like, and therefore the value of the absorption coefficient is exactly at  $\alpha_0(\omega)$ . Moreover, it becomes immediately clear that an inversion of the charge carriers can occur only if there are *different* Fermi energies in the conduction and valence bands:

$$\epsilon_{CB} - \epsilon_{VB} > h\nu > E_g. \quad (10.9)$$

The charge carrier distribution corresponds to a dynamic equilibrium that can only be sustained for forward-biased operation of the diode. The more exact calculation of the semiconductor gain is an elaborate matter since it depends on the details of the technical layout, which is much more complex as we are going to see later on.

### Example: Threshold current of the semiconductor laser

The threshold current density required can easily be determined when the critical density of charge carriers  $n_{el} \geq 10^{18} \text{ cm}^{-3}$  is known. The density of charge carriers

is transported by the injection current to the pn junction and recombines there spontaneously with rate  $\tau_{\text{rec}}^{-1} = 2.5 \times 10^8 \text{ s}^{-1}$ :

$$\frac{dn_{\text{el}}}{dt} = -\frac{n_{\text{el}}}{\tau_{\text{rec}}} + \frac{j}{ed}.$$

Without difficulties we derive the stationary current density for a width of the space charge zone  $d = 1 \mu\text{m}$  of the pn junction:

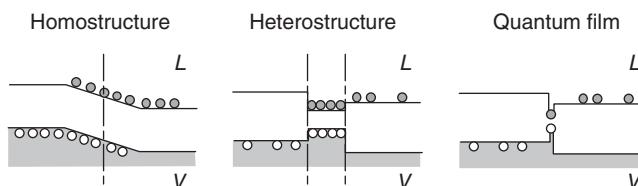
$$j = \frac{n_{\text{el}}ed}{\tau_{\text{rec}}} \geq 4 \text{ kA cm}^{-2}.$$

For an active zone with a typical area of  $0.3 \times 0.001 \text{ mm}^2$ , this current density already corresponds to 12 mA, which has to be concentrated exactly onto this small volume. It is obvious that it is worthwhile to technically reduce the natural width of the diffusion zone of the charge carriers in order to lower the threshold current density. This concept is precisely pursued by heterostructure and quantum film lasers.

### 10.2.5 Homo- and Heterostructures

Although the basic concept for the operation of a semiconductor laser originates from the early days of the laser, it was initially mandatory to cool the pn junction to cryogenic temperatures to obtain lasing at all. The light mobile electrons have a large diffusion length ( $\geq 0.5 \mu\text{m}$ ), so that large threshold currents were required and at room temperature the gain could not overcome the losses caused especially by non-radiative recombination and reabsorption. In the 1970s, however, this problem was solved by the concept of “heterostructures,” and ever since laser diodes have continued their triumphant route as sources for coherent light. In a heterostructure two different materials (e.g., with different composition and different bandgaps; compare Figure 10.4) are adjacent to each other. The interface creates potential steps that inhibit the diffusion of charge carriers across the barrier. For laser materials the bandgap is chosen in such a way that electrons and holes are confined between two layers with a larger bandgap in a zone with a smaller bandgap (“double heterostructure” (DH)). Otherwise the light generated at the center would be absorbed again in the outer areas of the amplification zone.

This advantage of heterostructures compared with simple homostructures is schematically presented in Figure 10.9. The strongly simplified potential scheme indicates that the motion of the charge carriers is now limited to a narrow layer ( $\approx 0.1 \mu\text{m}$ ) in order to realize an accordingly high gain density by their strong



**Figure 10.9** Band structure for electrons and holes: homostructure, heterostructure, and quantum films. The quantum film limit is usually entered at thickness below 200 Å.

confinement. Furthermore, when the refractive index in this area is higher than in the adjacent layers, a favorable wave guide effect is obtained, which in this case is called “index guiding.” Also, the spatial variation of the charge carriers causes changes of the refractive index and wave guiding again, which is called “gain guiding” in this case. With further miniaturization of the active layer, we get to the realm of quantum film systems, which are not just simply smaller but also show qualitatively novel properties (see Section 10.3.3).

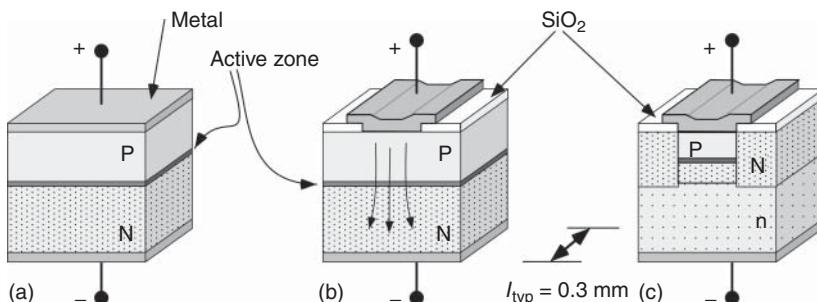
## 10.3 The Heterostructure Laser

The most important material for the manufacture of optoelectronic semiconductors until now has been GaAs. As a direct semiconductor, not only does it offer the necessary microscopic properties, but also, by variation of the  $\text{Ga}_x\text{Al}_{1-x}\text{As}$  compound crystal composition, it offers widespread technical potential to adjust the bandgap and the refractive index to the requirements for applications. The characteristic wavelength at 850 nm has technological significance as well, because it lies in one of three spectral windows (850, 1310, 1550 nm) suitable for the construction of optical networks. Today, the concepts of the AlGaAs laser have been transferred to other systems as well, for example, InAlP.

### 10.3.1 Construction and Operation

#### 10.3.1.1 Laser Crystal

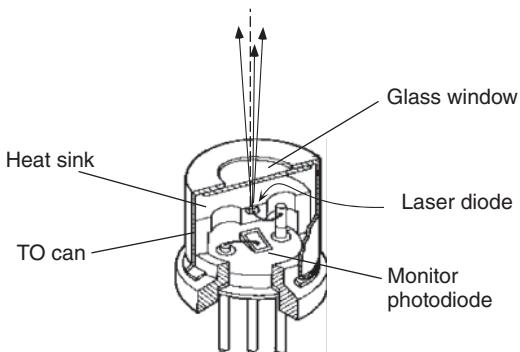
Laser crystals are produced through epitaxial growth.<sup>4</sup> The composition of these layers can be controlled along the growth direction by regulating the precursor flux. The vertical DH is controlled by such growth. The lateral structuring on the micrometer scale is engineered through methods known from microelectronics, for example, optical lithography processes (Figure 10.10).



**Figure 10.10** Layer systems for laser diodes. (a) Plain homostructure. (b) The current flux is narrowed by insulating oxide layers and causes a concentration of the inversion density. The inhomogeneous amplification, or charge carrier density, furthermore generates a wave guide that leads the light field along the gain zone (“gain guiding”). (c) Double heterostructures generate a precisely controlled amplification zone as well as an optical wave guide for the light field.

4 During epitaxial growth, thin monolayers of the (semiconducting) material are homogeneously deposited on a monocrystalline substrate from molecular beams.

**Figure 10.11** Standard package for laser diodes. The semiconductor device itself is hardly visible and has typical dimensions of 0.3 mm edge length. This type is called an edge emitter.



Owing to the construction, the laser field propagates along the surface of the crystals, and outcoupling takes place at the edge of a cleavage face. Therefore this type is called an “edge emitter,” in contrast to an alternative layout where light is emitted perpendicularly to the surface. This type will be introduced shortly in Section 10.5.2.

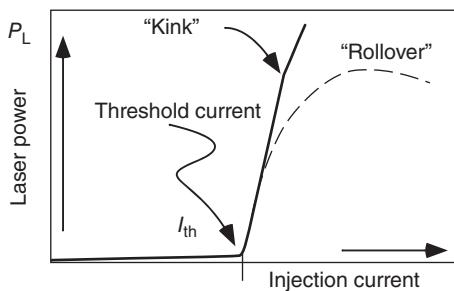
Laser crystals with a length of about 0.2–1 mm are produced by simply cleaving them from a larger epitaxially grown wafer. They can basically be inserted into a suitable package (Figure 10.11) without any further treatment and be contacted with standard techniques to facilitate handling. The transverse geometric properties of the laser field are determined by the shape of the amplification zone. In the far field an elliptical beam profile is generally observed, caused by the diffraction off the heterostructure and the transverse wave guiding. The light of edge-emitting laser diodes thus has to be collimated, which is quite costly for the purposes of application and is one reason for the development of surface emitters, which offer a circular beam profile from the beginning.

### 10.3.1.2 Laser Operation

In the most frequent and simplest case, the cleavage faces of the crystal already form a laser resonator. At a refractive index  $n = 3.5$ , the intrinsic reflectivity of a GaAs crystal is 30% and is often sufficient to support lasing due to the large gain coefficients of semiconductors. In other cases the reflectivity of the cleavage faces can be modified by suitable coatings. In Figure 10.12 the output power of a semiconductor laser as a function of the injection current is presented.

For many applications, for example, spectroscopy or optical communications, the use of single-mode lasers (both transverse and longitudinal) is important. The homogeneous gain profile of the laser diode offers excellent preconditions to implement single-mode operation even though the free spectral range of semiconductor lasers at  $\ell_{\text{typ}} = 0.3 \text{ mm}$  is, in spite of the substantial  $\Delta\nu_{\text{FSR}} = 150 \text{ GHz}$ , still very small compared to the gain bandwidth of 10 THz and more. In fact parasitic laser oscillations (“sidebands”) are very efficiently suppressed in many components.

The threshold currents of a laser diode vary depending on the layout, but the aim is always a laser threshold as small as possible. It has to be kept in mind that large current densities of  $100 \text{ kA cm}^{-2}$  and more occur, causing strong local heating and thus leading to damage of the heterostructures. For the same reason



**Figure 10.12** Current–power diagram of a laser diode. At large currents a “rollover” can occur due to heating of the pn junction.

the threshold current grows with temperature. In high-power lasers the so-called rollover occurs, for which increase of the injection current no longer leads to increase of the output power but on the contrary reduces it due to the heating of the pn junction! The relation between threshold current  $I_{\text{th}}$  and temperature follows an empirical law with a characteristic temperature  $T_0$  and current  $I_0$ :

$$I_{\text{th}} = I_0 \exp \left( \frac{T - T_0}{T_0} \right). \quad (10.10)$$

In conventional heterostructure lasers, the characteristic temperature has values of about  $T_0 = 60$  K, but in other layouts such as vertical cavity surface-emitting laser (VCSEL) or quantum film lasers (see Section 10.3.3), these values are increased in a favorable direction up to 200–400 K so that the temperature sensitivity of the components is significantly reduced. A qualitative semiconductor laser output power versus pump curve is presented in Figure 10.12 and reflects features of the idealized laser of Figure 9.1. From the slope of the power, the differential quantum efficiency can be obtained, which is typically 30% or more:

$$\text{Differential quantum efficiency} = \frac{e}{h\nu} \frac{dP}{dI}.$$

Sometimes there are so-called kinks in the power–current diagram. They are an indication of a modification of the laser mode, for example, caused by a charge carrier profile switching geometrically from one spatial mode to another at this particular current.

### 10.3.2 Spectral Properties

#### 10.3.2.1 Emission Wavelength and Mode Profile

The emission wavelength of a semiconductor laser is determined by the combined effect of gain profile and laser resonator as it is for other laser types. We first consider the wavelength selection of the “freely operating” laser diode without any additional optical elements.

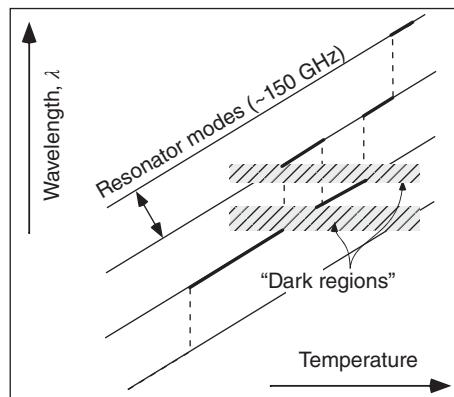
Single-mode operation occurs in many types of laser diodes. It is favored by a gain profile that is homogeneously broadened as a result of a large intraband relaxation rate. So at the gain maximum the mode starts lasing by itself. However, the detailed geometry of the often complex multilayer laser crystal can also allow multimode laser operation, and even in components explicitly called “single-mode laser” usually further parasitic modes may only be suppressed by a certain finite factor (typically 100×, or 20 dB).

Although the construction length of the resonator is generally very short ( $0.3\text{--}0.5\text{ mm}$ ,  $n \simeq 3.5$ ) and already for conventional components delivers a free spectral range of  $80\text{--}160\text{ GHz}$  (which can be much larger for VCSELs), there are still many resonator modes in the gain profile at a typical spectral width of some  $10\text{ nm}$  or some terahertz.

The refractive index determining the resonator frequency depends sensitively on the temperature as well as on the charge carrier density and the injection current, so that the exact laser frequency  $v_L$  can be tuned over considerable ranges by controlling these parameters:

- 1) Increasing the temperature of an external heat sink (e.g., a Peltier cooler) typically causes a redshift, that is, a frequency change with rate  $d\nu_L/dT = -30\text{ GHz K}^{-1}$ .
- 2) Variation of the injection current causes a shift  $d\nu_L/dI = \eta_{\text{th}} + \eta_n$ . The shift is due to temperature changes within the heterostructure ( $\eta_{\text{th}} \simeq -3\text{ GHz mA}^{-1}$ ) and also modifications of the charge carrier density ( $\eta_n \simeq 0.1\text{ GHz mA}^{-1}$ ). For slow current variations, the frequency change is dominated by the thermal redshift, but for modulation frequencies exceeding  $f_{\text{mod}} \geq 30\text{ kHz}$ , the influence of the charge carrier density dominates (see Section 10.4.1.2).

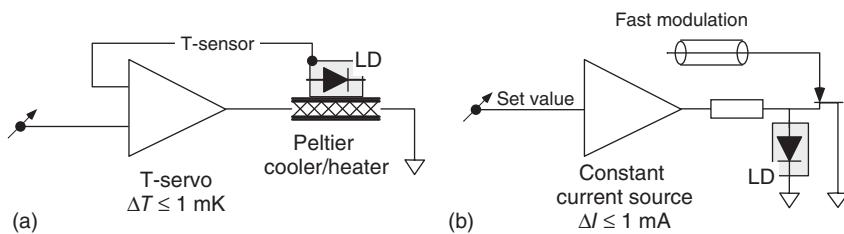
Unfortunately tuning by temperature and current at the pn junction only does not usually allow generation of every frequency within the gain profile. It is impaired by “dark” zones (Figure 10.13) since the gain profile and the mode structure of the resonator do not vary synchronously with each other. External optical elements can, however, also be used to access those forbidden domains (see Section 10.5.1).



**Figure 10.13** Mode jumps of diode lasers caused by temperature variations (“mode chart”).

### 10.3.2.2 Electronic Wavelength Control

When the exact frequency or wavelength of the laser radiation is important, for example, in spectroscopic applications, then the temperature at the laser diode junction and the injection current have to be controlled very precisely. The high sensitivity to temperature and current fluctuations sets high technical demands on the electronic control devices. If technically caused frequency fluctuations are to be kept lower than the typical  $5\text{ MHz}$  caused by intrinsic physical processes (see Section 10.4.2), then according to the variation rates given in the preceding section, obviously a temperature stability  $\delta T_{\text{rms}} \leq 1\text{ mK}$  and a current stability  $\delta I_{\text{rms}} \leq 1\text{ }\mu\text{A}$  have to be achieved with appropriate servo controllers.



**Figure 10.14** (a) Temperature control for laser diodes may use thermistors as temperature sensors. (b) Current control will typically inject a well-stabilized current that resists rapid variations. Fast modulation may then be realized by directly injecting additional small currents.

Considering it more exactly, the spectral properties of the servo controllers have to be investigated, but this would by far exceed the scope of this book. However, it is quite easy to see that the temperature control cannot have a large servo bandwidth due to its large thermal masses. The bandwidth of the current control is basically limited only by the capacitance of the laser diode itself, but it is advisable in terms of servo-control methods to limit the constant current source to a small internal bandwidth in order to reduce the current noise and instead of this to provide some additional fast high-impedance modulation inputs, for example, in Figure 10.14.

For wavelength stabilization the devices described here have a merely passive effect – they warrant tight control of operational parameters of the laser diode but do not interrogate the wavelength itself. For many applications, for example, optical wavelength standards, still better absolute stabilities are required, and deviations from a desirable wavelength must be directly sensed, for instance, through a spectroscopic signal, and corrected through suitable servo controls.

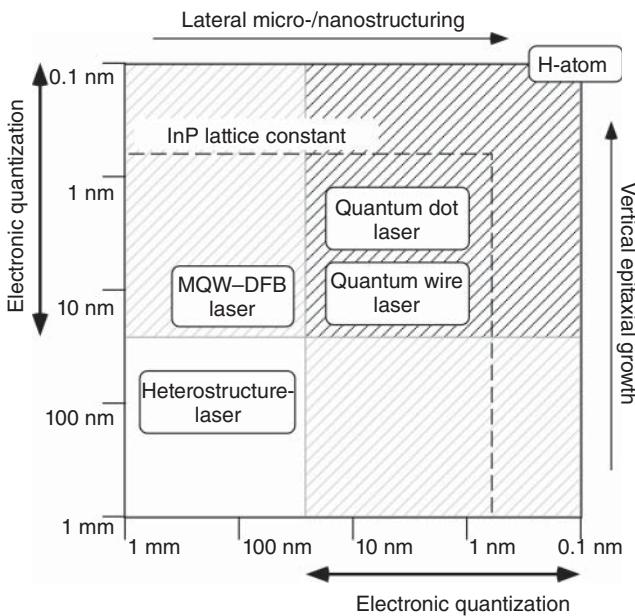
### 10.3.3 Quantum Films, Quantum Wires, and Quantum Dots

Conventional heterostructures serve to hinder the diffusion of electrons and holes and to concentrate the gain into a small zone. The charge carriers though move in a well with dimensions of about 100 nm like more or less classical point-like particles. By further miniaturization (see Figure 10.15), we reach the realm of quantized electronic motion in which the dynamics of the charge carriers in the vertical direction orthogonal to the layer system is characterized by discrete energy levels according to quantum mechanics.

Once miniaturization reaches the quantum border in one dimension, a “two-dimensional electron gas” is created, which we shall call a “quantum film” here. In the literature there are also other terms used, for example, quantum well (QW) lasers. Structures with reduced dimensionality offer lower threshold currents, larger gain, and lower temperature sensitivity than conventional DH lasers, advantages already essentially acknowledged since the early 1980s.

#### 10.3.3.1 Inversion in the Quantum Film

The two-dimensional character of the charge carrier gas causes a change in the DOS (see Section B.3), the fundamental origin of improved operation characteristics, for example, low threshold current and lower temperature sensitivity.



**Figure 10.15** Semiconductor miniaturization and semiconductor laser types with reduced dimensionality.

In addition to the kinetic energy of the transverse quantum state  $E_{Qi}$ , there are two continuous degrees of freedom with momentum components  $k_{\perp i}$ . For the electrons and holes in the  $i$ th subband of the quantum film, we find ( $k_{\perp i}^2 = k_{\perp ix}^2 + k_{\perp iy}^2$ )

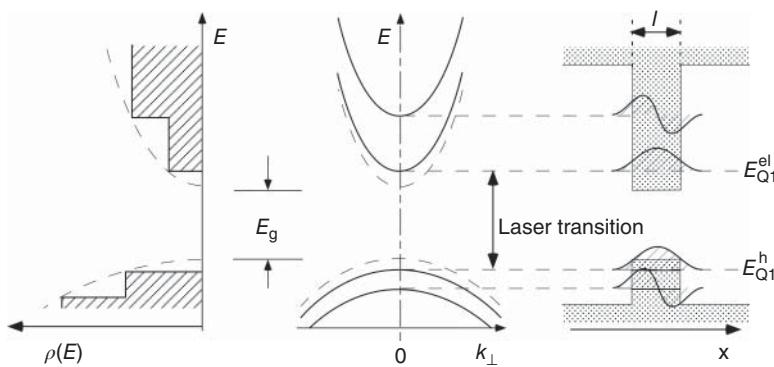
$$E_i = E_{VL} + E_{Qi} + \frac{\hbar^2 k_{\perp i}^2}{2m_{el,h}^*}.$$

Among the interesting properties of quantum film lasers is the possibility to control the transition wavelength by choosing the film thickness  $\ell$ , which determines the energy separation of the quantum states in the electronic and hole-like state. According to quantum mechanics we have  $E_{Q1} \simeq \hbar^2 / 2m_{el}^* \ell^2$ .

The DOS in the  $k$  plane is  $\rho_{el,h}(k) dk = k dk / 2\pi^2$  and can be converted into an energy density with  $dE = \hbar^2 k / m_{el,h}^* dk$ . In the transverse direction each quantum state (energy  $E_{Qi}^{el,h}$ , quantum number  $i$ ) contributes with the density  $\pi/\ell$ ,

$$\rho_{el,h}(E) dE = \sum_i \frac{m_{el,h}^{*i}}{\hbar^2 \ell} \Theta(E - E_{Qi}^{el,h}).$$

The theta function has the values  $\Theta(x) = 1$  for  $x > 0$  and  $\Theta(x) = 0$  for  $x \leq 0$ . Also the effective masses  $m_{el,h}^{*i}$  may depend on the quantum number. The DOS grows step-like (see Figure 10.16) in a quantum film every time the energy reaches a new transverse quantum state. There it has exactly the value corresponding to the volume material (dashed line in Figure 10.16).



**Figure 10.16** Band structure (middle) and density of states (left) in a quantum film. On the left the dashed curves show the corresponding density of states for the bulk material. The hatched curves on the right indicate the wave function of the stored electrons and holes in the 1D transverse potential.

The advantage of the QW laser becomes evident when we determine the dependence of the Fermi energy on the charge carrier concentration as we did on p. 326. With terms similar to the 3D case, for example,  $\alpha_{\text{el}}^i = E_L + E_{Q_i}^{\text{el}} - \epsilon_L$ , we obtain

$$n_{\text{el,h}} = \sum_i \frac{m_{\text{el,h}}^{*i} kT}{\hbar^2 \ell} e^{-\alpha_{\text{el,h}}^i} \int_0^\infty \frac{dx}{e^x + e^{-\alpha_{\text{el,h}}^i}}.$$

This integral can be analytically evaluated. Using the parameters of GaAs at  $T = 300$  K and for a quantum film with a thickness of  $\ell = 100$  Å, we find the relation

$$n_{\text{el}} = 3.3 \times 10^{15} \text{ cm}^{-3} \ln(1 + e^{-\alpha_{\text{el,h}}^i}),$$

from which the Fermi energy can be obtained. The value of the first factor is two orders of magnitude smaller than for the volume material (Eq. (10.5))! This indicates that in the QW laser, inversion can be expected already at considerably smaller charge carrier concentrations and thus smaller threshold current densities than in conventional DH lasers.

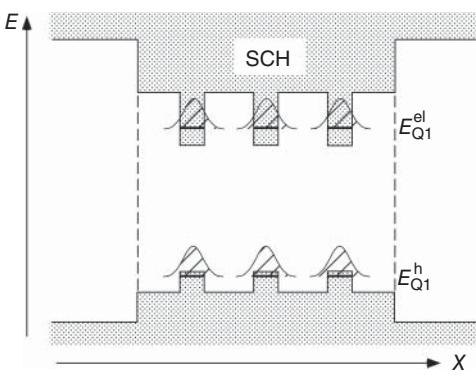
### 10.3.3.2 Multiple Quantum Well (MQW) Lasers

For a fair comparison with conventional DH lasers, it has to be taken into account that the total gain of a quantum film is smaller than that of a DH laser simply due to the smaller volume. This disadvantage can be largely compensated by introducing multiple identical quantum films in the volume of the laser light field.

In Figure 10.17 a multiple quantum well (MQW) structure is schematically presented. The charge carriers are to be “caught” in the potential wells, but the relaxation rate, for example, through collision with a phonon, is quite small due to the small film thickness. To increase the concentration of the charge carriers in the vicinity of the quantum films, an additional heterostructure is provided – the separate confinement heterostructure (SCH) shown in Figure 10.17. This structure also acts as a wave guide for the resonator field and focuses the light intensity

onto this zone, which is usually much smaller than an optical wavelength. Today MQW lasers have become a standard product of the optoelectronics industry.

Another interesting innovation has been introduced with the “strained quantum well.” They offer additional technical advantages since the effective masses are increased as a result of mechanical strain in the crystal lattices by a factor of 2. Therefore the DOS and the threshold current density decrease again.



**Figure 10.17** Potential landscape of a multiple quantum well (MQW) structure consisting of three quantum films. SCH: separate confinement heterostructure.

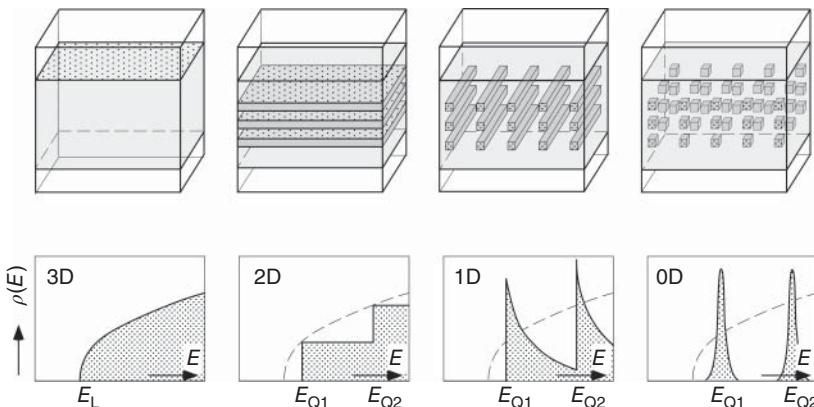
Let us once again summarize the advantages of quantum film lasers compared with conventional DH lasers:

- 1) The modified DOS causes lower threshold currents since fewer states per charge carrier are available, which can consequently be filled with lower currents. Typically threshold current densities of  $50\text{--}100 \text{ A cm}^{-2}$  are achieved. The lower threshold indirectly improves again the temperature sensitivity since there is less excess heat generated in the heterostructures.
- 2) The differential gain is larger than for the DH lasers since the electrically dissipated power growing with the current causes a lower reduction of the gain.
- 3) The threshold condition depends less strongly on the temperature. For conventional DH lasers the transparency threshold grows with  $T^{3/2}$ , and in quantum film lasers only in proportion to  $T$ . The characteristic temperatures according to Eq. (10.10) are about 200 K.

### 10.3.3.3 Quantum Wires and Quantum Dots

The reduced dimensionality of semiconductor structures can be extended by construction: two-dimensional quantum films become quasi-one-dimensional quantum wires and even zero-dimensional quantum dots when suitable methods of lateral microstructuring are chosen. In Figure 10.18 this evolution with its effect on the DOS is presented.

The properties of the DOS continue the tendency of the quantum film laser to realize overall gain already at small current densities. While the layer stack of the quantum film laser can simply be manufactured by controlling the growth processes (in Figure 10.18 in the vertical direction), the lateral properties have to be manufactured in general by a completely different process. On the one hand, there is no longer a big difference between manufacturing quantum wires and quantum dots from the technological point of view; on the other hand, the necessary lateral structural dimensions of 0.1–0.2 nm are not easily achievable with



**Figure 10.18** Evolution from the double-heterostructure laser over quantum films and wires to quantum dots.

standard methods of optical lithography. Also, the strictly periodic formation of quantum dots shown in Figure 10.18 has been difficult to realize up to now, but on the other hand, it is not necessary for the laser process either. Multiple quantum dots have been produced using self-organization of a heterogeneous growth process [154, 155].

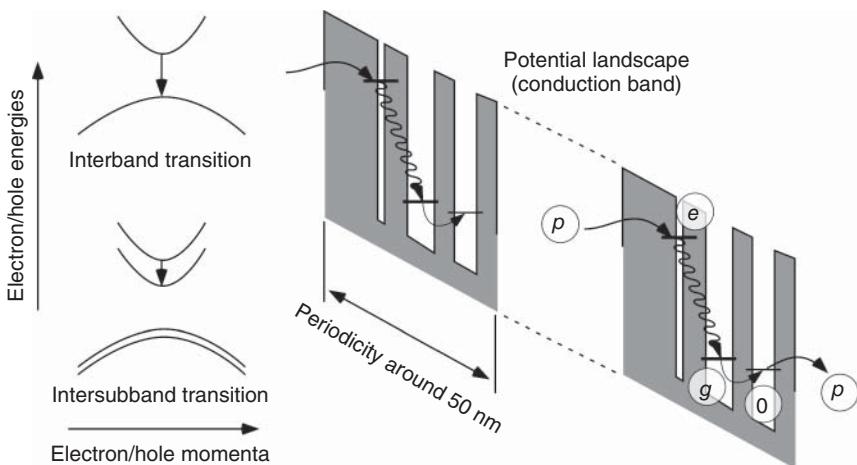
#### 10.3.4 Quantum Cascade Lasers

Heterostructures as described in the preceding sections are used to engineer the band structure of semiconductor compounds with respect to concentration of charge carriers for maintaining high gain density and bandgap control for wavelength variation. Stimulated emission is based on electron–hole recombination radiation that is on transitions between the conduction and the valence band named interband transition.

It was suggested early on [156] that band structure engineering could also be used to exploit transitions between different subbands of the conduction band to extract laser radiation from structured semiconductor compounds. Since the same kind of III–V materials with their naturally given bandgap energies are used, the energy spacings and hence intersubband transitions have much longer wavelengths in the infrared and terahertz domain.

In the quantum cascade laser (QCL), electrons move sequentially through a cascade of a group of QW (Figure 10.19). “Quantum confinement” causes energy levels in narrow wells to be at a higher energy than in wider wells. Transition frequencies between different subbands (in the transverse direction electrons can freely propagate) can be controlled by manufacturing quantum films of suitable width. Barrier widths between the wells determine the tunneling rates that may be assisted by non-radiative phonon scattering processes, for example, in order to rapidly empty the lower laser level  $g$  toward the ground state 0.

The QCL [157] can be understood as a four-level laser system consisting of a small group of QW (compare Figures 7.12 and 10.19 with identical level designations). Application of a voltage creates the staircase potential landscape across



**Figure 10.19** Band structure concept for quantum cascade lasers. Every electron propagates through the periodic structure and contributes with many photons to the laser field. The energy level structures corresponds to a four-level laser system; compare Figure 7.12.

the total structure. Under the action of the electric field, every electron then propagates through the structure and contributes with multiple photons to the laser field.

QCLs [158] are robust and compact devices offering a favorable “footprint” like all semiconductor lasers. They operate at room temperature (in stark contrast to their predecessors, lead salt diode lasers that needed a cryogenic environment), Watt level output powers are available, and the frequencies span the mid-infrared and terahertz region widely. This domain is full of molecular “fingerprint” spectra. Hence QCLs make ideal sources and are widely used for spectroscopic applications with respect to sensitive and selective molecule identification and detection.

## 10.4 Dynamic Properties of Semiconductor Lasers

Among the technically more attractive properties of the laser diode is the potential to modulate it directly by varying the injection current. For instance, the speed of switching the laser on and off determines the rate for generating digital signals and thus transmitting information. A rate equation approach makes an excellent approximation since the transverse relaxation is dominated by the fast rate of intraband scattering,  $\gamma'^{-1} = T_2 \simeq 1 \text{ ps}$ , and hence the polarization effects remain in equilibrium with the field amplitude (Section 9.2). To understand the dynamics of laser diodes, we use the amplitude equation (9.12) and the rate equation (9.20(ii)):

$$\begin{aligned}\dot{E}(t) &= \left( i \left[ \Omega - \omega - \frac{1}{2} \kappa \alpha n(t) \right] + \frac{1}{2} [\kappa n(t) - \gamma_c] \right) E(t), \\ \dot{n}(t) &= -\kappa n(t) n_{\text{ph}}(t) - \gamma n(t) + R.\end{aligned}\quad (10.11)$$

Here we use  $n(t)$  for the charge carrier density. The current density  $j$  “feeds” the dynamics with  $R = j/ed$ , we replace  $|E(t)|^2 \rightarrow n_{\text{ph}}(t)$ , and furthermore we now

use the photon lifetime  $\gamma_c \rightarrow 1/\tau_{ph}$  and recombination time  $\gamma \rightarrow 1/\tau_{rec}$  instead of the damping rates. At steady state ( $\dot{n} = \dot{n}_{ph} = 0$ ), we find

$$n^{st} = \frac{1}{\kappa\tau_{ph}} \quad \text{and} \quad \bar{n}_{ph} = \frac{1}{\kappa\tau_{rec}} \left( \frac{j}{j_{th}} - 1 \right),$$

where  $j_{th} = ed/\kappa\tau_{rec}\tau_{ph}$ . Mostly we are interested in small deviations from the stationary state. Then we can linearize

$$n(t) = n^{st} + \delta n(t) \quad \text{and} \quad n_{ph} = \bar{n}_{ph} + \delta n_{ph}$$

and find the equations of motion, in which we set  $j_0/j_{th} = I_0/I_{th}$ ,

$$\begin{aligned} \dot{\delta n}_{ph}(t) &= \frac{1}{\tau_{rec}} \left( \frac{I_0}{I_{th}} - 1 \right) \delta n(t), \\ \dot{\delta n}(t) &= \frac{j_{mod}}{ed} - \frac{1}{\tau_{rec}} \frac{I_0}{I_{th}} \delta n(t) - \frac{1}{\tau_{ph}} \delta n_{ph}. \end{aligned} \quad (10.12)$$

#### 10.4.1 Modulation Properties

We consider the effect of small harmonic modulations of the injection current  $j_{mod} = j_0 + j_m e^{-i\omega t}$  on the amplitude and the phase of the laser light field.

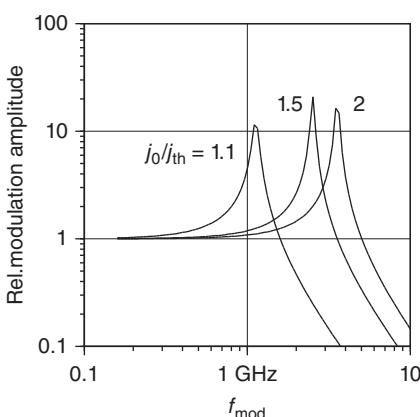
##### 10.4.1.1 Amplitude Modulation

The modulation of the number of photons is equivalent to the variation of the output power. Therefore we use  $\delta n_{ph}(t) = \delta n_{ph0} e^{-i\omega t}$  and  $\delta n(t) = \delta n_0 e^{-i\omega t}$ , and we can replace  $\delta n_{ph0} = -(I_0/I_{th} - 1)\delta n_0/i\omega\tau_{rec}$ . After a short calculation, we get

$$\delta n_{ph0} = -\frac{\tau_{ph} j_m}{ed} \frac{I_0/I_{th} - 1}{\omega^2 \tau_{rec} \tau_{ph} - (I_0/I_{th} - 1) + i(I_0/I_{th})\omega\tau_{ph}}, \quad (10.13)$$

resulting in an amplitude modulation coefficient proportional to

$$|\delta n_{ph0}| = \frac{\tau_{ph} j_m}{ed} \frac{I_0/I_{th} - 1}{\sqrt{[\omega^2 \tau_{rec} \tau_{ph} - (I_0/I_{th} - 1)]^2 + \omega^2 \tau_{ph}^2 (I_0/I_{th})^2}}. \quad (10.14)$$



**Figure 10.20** Amplitude modulation of a diode laser as a function of the modulation frequency.

In Figure 10.20 we present the response of a typical laser diode to a current modulation with frequency  $f_{mod} = \omega/2\pi$  calculated from Eq. (10.14). We have used a spontaneous recombination time  $\tau_{rec} = 2 \times 10^{-9}$  s and a photon lifetime of  $\tau_{ph} = 10^{-12}$  s. The frequency of the relaxation resonance grows with injection current as expected according to Eq. (9.25). Experimental data are well represented by this function.

For applications, for example, in optical communications, a large modulation bandwidth is important. In addition to this, the frequency response is to stay flat up

to frequencies as high as possible, and furthermore there should not be any major phase rotations ( $\delta n_{\text{ph}0}$  is a complex quantity!). Today in compact VCSEL components, modulation bandwidths of 40 GHz and more are achieved, and an end of this development is not yet in sight.

#### 10.4.1.2 Phase Modulation

Next we investigate the evolution of the phase  $\Phi(t)$  separating off the steady state in Eq. (10.11) with

$$E(t) \rightarrow \mathcal{E} \exp [i(\Omega - \omega - \alpha\gamma_c/2)] \exp [i\Phi(t)]$$

and find the coupling

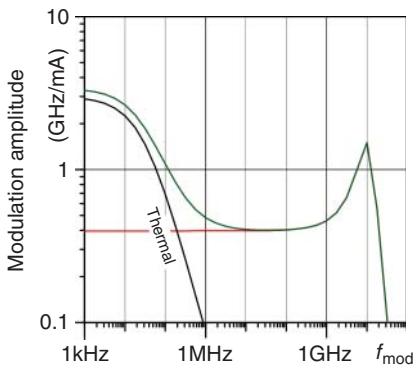
$$\dot{\Phi}(t) = \frac{1}{2}\alpha\kappa \delta n(t)$$

of charge carrier dynamics and phase evolution. We again expect a harmonic dependence  $\Phi(t) = \Phi_0 e^{-i\omega t}$ , which after a short calculation we can also express by the modulation amplitude of the photon number  $n_{\text{ph}}$  and so obtain the very transparent result where  $\alpha = (\omega - \omega_0)/\gamma$ , Eq. (9.8):

$$\Phi(t) = \Phi_0 e^{-i\omega t} = \frac{\alpha}{2} \frac{\delta n_{\text{ph}0}}{\bar{n}_{\text{ph}}} e^{-i\omega t}.$$

The result shows that the factor  $\alpha$  describes the coupling of the phase change to the amplitude change. In laser diodes it has typical values of 1.5–6. [It usually vanishes in gas lasers since those oscillate very close to the atomic or molecular resonance lines ( $\alpha \approx 0$ ).] It also plays a significant role for the linewidth of the laser diode, as we will see Section 10.4.2.

Up to now we understood the amplitude as well as the phase modulation only as a consequence of the dynamic charge carrier density. The modulation current moreover causes a periodic heating of the heterostructure, which modifies the optical length of the laser diode resonator as well and even dominates the modulation depth up to typical critical frequencies of some 10 kHz (Figure 10.21). Both the temperature and charge carrier density modulation, which we have already identified on p. 333 as the origin for the detuning of laser wavelength with injection current, contribute to the low-frequency limit of the phase modulation amplitude.



**Figure 10.21** Phase modulation of semiconductor lasers. The modulation index consists of thermal and charge carrier density contributions.

#### 10.4.2 Linewidth of the Semiconductor Laser

When the linewidth of a laser diode is calculated according to the Schawlow–Townes formula (Eq. (9.33)), a higher value than, for example, for the He–Ne laser is already expected from the beginning due to the large linewidth of the empty resonator  $\gamma_c \approx 10^{12}$ . In experiments, still larger linewidths of 10–100 MHz

are observed for a typical 1 mW laser diode and set in relation to the “pure” Schawlow–Townes limit  $\Delta\nu_{ST}$ . This broadening is described by the so-called  $\alpha$  parameter, which was already introduced in our simple laser theory describing the amplitude–phase coupling (Eq. (9.8)):

$$\Delta\nu'_{ST} = (1 + \alpha^2)\Delta\nu_{ST}.$$

With semiconductor lasers this is often called Henry’s  $\alpha$  parameter because C. Henry discovered that, albeit known from the early days of laser physics, it plays a much more significant role for diode lasers than for gas lasers [159].

The  $\alpha$  factor was initially introduced as an “abbreviation” for the normalized detuning in Eq. (9.8). A more detailed analysis shows that it gives the differential ratio of real and imaginary parts of the susceptibility or the refraction coefficient as well:

$$\alpha = \Delta n'/\Delta n''.$$

It can only be calculated with elaborate methods and detailed knowledge of the diode laser construction and is thus preferably obtained from experiment.

#### **Example: “Pure” Schawlow–Townes linewidth of a GaAs laser**

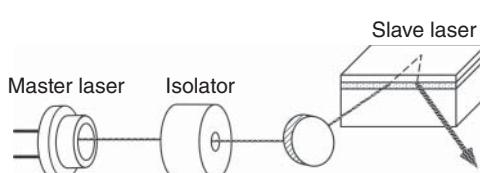
We determine the linewidth according to Eq. (9.33) of a GaAs laser for 1 mW output power and at a laser frequency of  $\nu_L = 350$  THz at 857 nm. The small Fabry–Pérot resonator with a length of 0.3 mm and a refraction coefficient 3.5 leads, for mirror reflectivities of  $R = 0.3$ , to a linewidth and to decay rates of  $\Delta\nu = \gamma_c/2\pi = 3 \times 10^{10}$ , which are much larger than for a typical GaAs laser and cause a very much larger Schawlow–Townes linewidth:

$$\Delta\nu_{ST} \approx \frac{\pi h \times 350 \text{ THz} (2\pi \times 50 \text{ GHz})^2}{1 \text{ mW}} = 1.5 \text{ MHz.}$$

In practice  $\Delta\nu'_{ST} = (1 + \alpha^2)\Delta\nu_{ST}$  with  $\alpha$  ranging from 1.5 to 6 is found for enhanced linewidths.

#### **10.4.3 Injection Locking**

In a conventional laser the oscillation of the light field starts by itself from the noise. We are going to study now how a laser oscillator reacts if it is already irradiated by an external monochromatic light field. The considerations are in principle valid for almost all types of lasers, but they are particularly important for the applications of laser diodes since in this way the preparation of a light field with good coherence length and high output power in functionally separated components (in a so-called “master–slave” arrangement) can be achieved.



**Figure 10.22** Injection locking. The coherent light field of the “master laser” is injected into a “slave laser” and leaves its coherence properties on it. The isolator serves to decouple the master laser from any radiation emitted by the slave laser.

In Figure 10.22 we have schematically presented a situation typical for laser diodes. In the “master laser” a laser light field with well-controlled coherence

properties is prepared. Its light is injected into a “slave laser” and determines the dynamical properties of the latter under conditions we are going to investigate here. The slave laser itself may generally have less advantageous coherence properties as long as it makes high output power available, for example, in Figure 10.22 from a broad stripe laser or a tapered amplifier.

Let us insert the coupling of the laser light field  $E$  in Eq. (10.11) to an external field  $E_{\text{ext}}$  in a heuristic way. The coupling term must have the same structure as the outcoupling term (i.e.,  $\dot{E} \propto (\gamma_{\text{ext}}/2)E_{\text{ext}}$  where  $\gamma_{\text{ext}}$  is the damping rate due to the outcoupling mirror), but the external field oscillates with its own frequency  $\omega_{\text{ext}}$ . We replace  $\kappa n \rightarrow G$  and write

$$\dot{E}(t) = \left[ i(\omega - \Omega - \frac{1}{2}\alpha G) + \frac{1}{2}(G - \gamma_c) \right] E(t) + \frac{1}{2}\gamma_{\text{ext}}E_{\text{ext}} e^{-i(\omega_{\text{ext}}-\omega)t+i\varphi}.$$

Then we find equations for the equilibrium that we separate into real and imaginary parts,

$$\begin{aligned} \text{(i)} \quad & \frac{1}{2}(G - \gamma_c) + \frac{\gamma_{\text{ext}}}{2} \frac{E_{\text{ext}}}{E} \cos \varphi = 0, \\ \text{(ii)} \quad & (\omega_{\text{ext}} - \Omega - \frac{1}{2}\alpha G) + \frac{\gamma_{\text{ext}}}{2} \frac{E_{\text{ext}}}{E} \sin \varphi = 0, \end{aligned} \quad (10.15)$$

which describe the amplitude (i) and the phase (ii), respectively. If we limit ourselves to the case of small coupling, the modifications of the field amplitude can be neglected. Then we can use the modified saturated gain,

$$G = \gamma_c - 2\Delta_M \cos \varphi,$$

from Eq. (10.15(i)) by introducing the frequency

$$\Delta_M := \frac{\gamma_{\text{ext}}}{2} \frac{E_{\text{ext}}}{E} = \frac{\gamma_{\text{ext}}}{2} \sqrt{\frac{I_{\text{ext}}}{I}}.$$

From Eq. (10.15(ii)) we obtain the relation

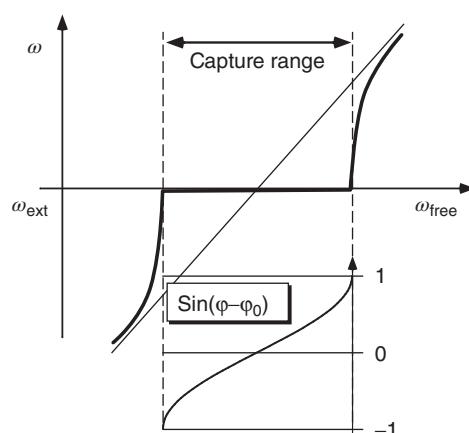
$$\omega_{\text{ext}} - (\Omega + \alpha\gamma_c/2) + \alpha\Delta_M \cos \varphi = \Delta_M \sin \varphi.$$

The result can be presented even more conveniently with  $\tan \varphi_0 = \alpha$  and  $\omega_{\text{free}} := \Omega + \alpha\gamma_c/2$ , which is the laser oscillation frequency in the absence of an injected field. For  $\alpha = 0$  it is known as the *Adler equation*:

$$\begin{aligned} \omega_{\text{ext}} - \omega_{\text{free}} &= \Delta_M \sqrt{1 + \alpha^2} \\ &\times \sin(\varphi - \varphi_0). \end{aligned} \quad (10.16)$$

Then we can derive immediately the limiting conditions for the so-called capture or “locking range”:

$$-1 \leq \frac{\omega_{\text{ext}} - \omega_{\text{free}}}{\Delta_M \sqrt{1 + \alpha^2}} \leq 1.$$



**Figure 10.23** Frequency characteristic and phasing of a slave laser on injection locking.

We find that the slave oscillator locks to the frequency of the external field, as shown in Figure 10.23. The locking range  $2\Delta_M$  is larger when more power is injected and when the coupling is stronger, that is, when the reflectivity of the resonator is lower. According to our analysis for a laser diode, which has typically a low reflectivity, the locking is furthermore supported by the phase–amplitude coupling described by the factor  $\sqrt{1 + \alpha^2}$ .

The phase condition shows that the locking is made possible through a suitable adjustment of the phase angle  $\varphi$  between master and slave oscillators. A more detailed analysis of the stability, which we skip here, shows that only one of the two adjustment solutions is stable according to Eq. (10.16).

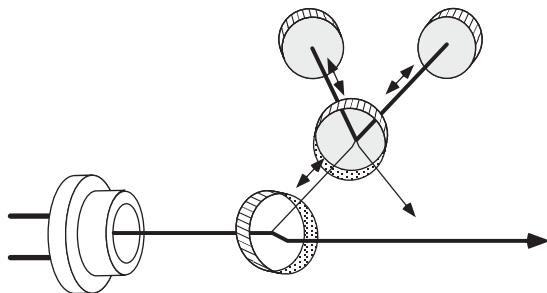
Outside the capture range, the locking condition cannot be fulfilled, but the external field there causes a phase modulation as well, which already leads to a frequency shift of the slave oscillator. The theoretical analysis is a bit more costly, but it shows, among other things, that close to the locking range additional sidebands are generated from the master and the slave light field due to nonlinear mixing processes.

### Example: Optical feedback and self-injection locking

The coherence properties of laser diodes are extraordinarily sensitive to backscattering from outside. Every randomly caused reflection can trigger considerable and uncontrollable frequency fluctuations. For critical applications, for example, in spectroscopy, therefore, optical isolators with a high extinction ratio have to ensure that the backscattering occurring at every optical element is suppressed.

The feedback from external components can be described as a form of “self-injection locking” in immediate analogy to the injection locking described in Section 10.4.3,

$$\dot{E}(t) = \left[ i \left( \omega - \Omega - \frac{1}{2} \alpha G \right) + \frac{1}{2} (G - \gamma_c) \right] E(t) + r(\omega) E(t) e^{-i\omega\tau},$$

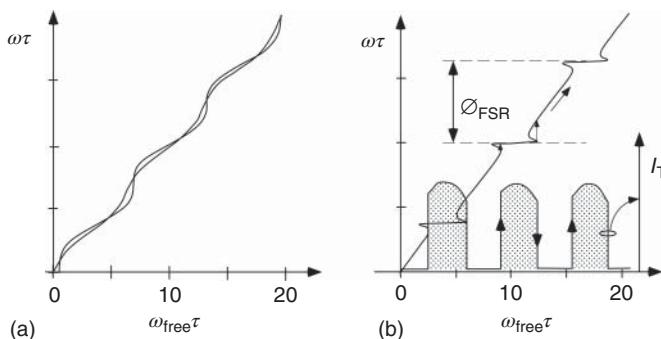


**Figure 10.24** Optical feedback from a folded resonator. Feedback can only take place in the case of resonance of laser frequency and resonator.

with  $\tau := 2\ell/c$  giving the delay time needed by the light to travel from the laser source to the scattering position at a distance  $\ell$  and back again. The reflection coefficient  $r(\omega)$  of the optical element may also depend on the frequency, for example, for the resonator in Figure 10.24 according to Eq. (6.14).

In direct analogy to the case of normal injection locking, the analysis leads again to a characteristic equation for the frequency, which now depends critically on the return phase  $\omega\tau$ :

$$\omega - \omega_{\text{free}} = r(\omega) = \sqrt{1 + \alpha^2} \sin[\omega(\tau - \tau_0)].$$



**Figure 10.25** The effect of feedback on the oscillator frequency of a laser diode. (a) A simple mirror with two different reflected waves. (b) Folded resonator according to Figure 10.24. The shadowed curve shows the expected transmission of the resonator at tuning of the laser frequency in the positive direction.

An overview can be most simply obtained graphically. In Figure 10.25 we present the situation for a simple mirror (a) and a Fabry-Pérot resonator (b). It is evident from Figure 10.25 that back-reflections from acoustically vibrating setups changing the return phase will always cause frequency fluctuations. A stable resonator, however, forces the laser frequency to oscillate at its eigenfrequency if the right conditions are chosen. Thus coupling to the external resonator results in improved coherence properties – the resonator works like a passive flywheel counteracting the phase fluctuations of the active oscillator [160].

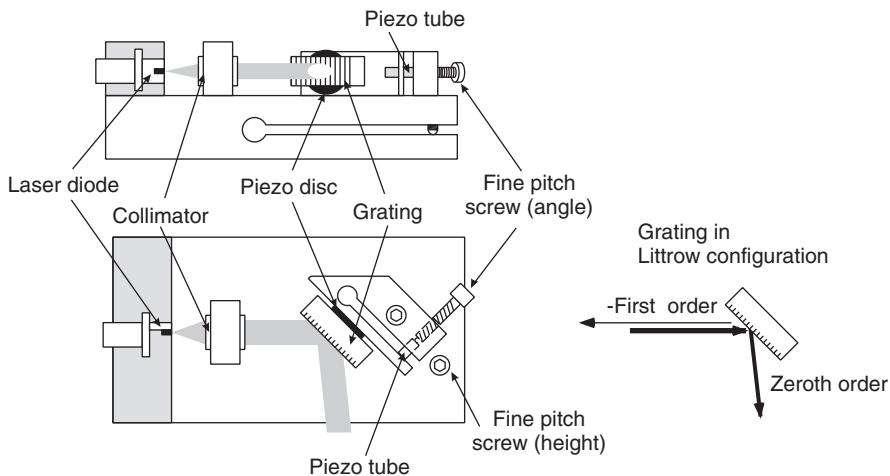
## 10.5 Laser Diodes, Diode Lasers, and Laser Systems

A laser diode emits coherent light as soon as the injection current exceeds the threshold current through the semiconductor diode. Specific applications, however, set different demands for the wavelength and the coherence properties of the laser radiation. In order to control these properties, the laser diode is used in different optical layouts and is integrated into “systems” that we are going to call “diode laser” to distinguish it from the optoelectronic “laser diode” component.

Owing to the microscopic dimensions of the laser crystal, additional devices such as filters may be immediately integrated during manufacture. Such concepts are realized with the so-called distributed feedback (DFB) and distributed Bragg reflector (DBR) lasers and VCSEL. Another possibility is to achieve frequency control by coupling the laser light back into the resonator as described in the previous section.

### 10.5.1 Tunable Diode Lasers (Grating Tuned Lasers)

Among the most unwanted properties of laser diodes are the mode hops that prevent continuous tuning along the entire gain profile as shown in Figure 10.13. This problem can be overcome by using an antireflection coating on the laser diode facets and inserting the chip as an amplifying medium in an outer resonator



**Figure 10.26** Construction of a diode laser system according to the Littrow principle. The – first order of the grating is retroreflected into the laser diode. The zeroth order makes light available for applications.

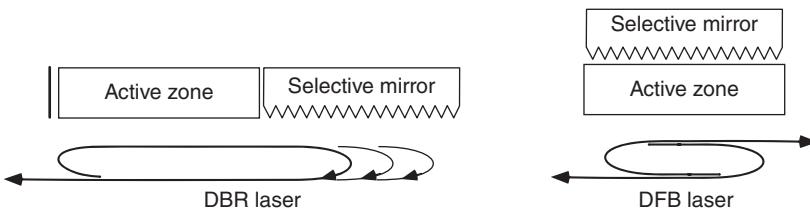
with suitable mirrors and filter elements. This “extended cavity” concept, though, gives up many advantages of the semiconductor laser, for example, the compact layout. Therefore the “external cavity” method is preferred. The external grating is mounted in the so-called Littrow arrangement for feedback (Figure 10.26). There the grating reflects about 5–15% of the power exactly back in – first order into the light source, while the rest is reflected away for applications. It thus causes frequency-selective feedback and a corresponding modulation of the gain profile. So by turning the grating, most laser diodes can now be tuned to almost every wavelength within their gain profile without any further modifications of their facet reflectivities.

### 10.5.2 DFB and DBR Lasers and VCSEL

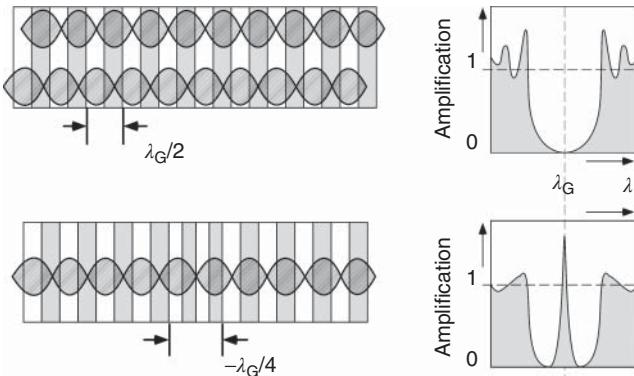
The integration of periodic elements for frequency selection has not only been studied with semiconductor lasers, but for them it is an attractive choice because the methods of microlithography are required for manufacturing anyway. The concepts of the DFB laser and DBR laser are implemented with lateral structures on a suitable substrate (edge emitters), while the VCSEL is realized by a vertical layer stack.

We already know the function of the integrated Bragg end mirrors from the fiber laser (Section 8.4.4), and both edge-emitting types differ only in the layout of the Bragg reflector. For the DBR laser it is set aside from the active zone as a selective mirror (and some parameters, such as the center wavelength, can possibly be controlled, e.g., by injecting a current for refractive index control). For the DFB laser the active zone and the Bragg grating (which is a phase grating in general) are integrated in one element (Figure 10.27).

Because of simpler and more reliable manufacturing methods, today the DFB laser is in more widespread use than the DBR variant among the edge emitters.



**Figure 10.27** Principal elements of DBR and DFB lasers.

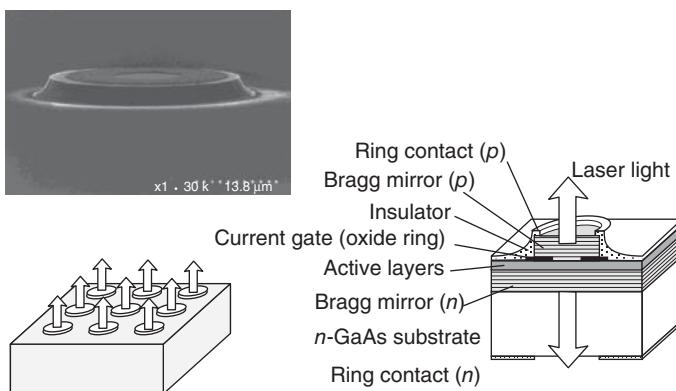


**Figure 10.28** Grating with uniform period and with  $\lambda/4$  shift in the DFB laser. The effective spectral gain profile is qualitatively drawn on the right-hand side.

Studying the spectral properties of the periodic DFB structure in more detail, one finds that light wave propagation is strongly suppressed in a region centered at the wavelength corresponding to the periodicity of the grating [137, 161]. The cause of this can be seen qualitatively in Figure 10.28. We can define two stationary waves, which experience a lower average refraction coefficient  $n_- = n - \delta n$  at one position and a higher one  $n_+ = n + \delta n$  at another position, so that for the same wavelength two frequencies  $v_{\pm} = n_{\pm}c/\lambda$  at the same separation from the center wavelength  $v_0 = nc/\lambda$  are allowed. A gain maximum is, however, generated exactly at this position if the so-called  $\lambda/4$  shift of the period is inserted at the center of the DFB structure.

A conceptual example of a VCSEL [162, 163] is presented in Figure 10.29. The layer structures are epitaxially grown. The active zone has a length of just one wavelength within the material, that is,  $\lambda/n \simeq 250$  nm at an emission wavelength of 850 nm. It is host to several closely adjacent quantum films with a typical thickness of 8 nm. Since the gain length is extremely short, the Bragg mirrors have to have a very high reflectivity of 99.5%. For this, typically 20–40  $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{Al}_y\text{Ga}_{1-y}\text{As}$  layer stacks are required with a refraction coefficient contrast as high as possible.

For the VCSELs the concentration of the injection current onto the desired cross-sectional area of the laser field is a huge technical challenge. In today's solutions, for example, the resistance of certain layers is strongly increased by proton bombardment, though this causes disadvantageous crystal damage in



**Figure 10.29** Concept of VCSELs. (The electron microscope photograph was provided by Dr. Michalzik, University of Ulm [163].)

adjacent material. Using another method, the upper Bragg stack is structured into round mesa-like mirrors, and finally a thin  $\text{Al}_{0.97}\text{Ga}_{0.03}\text{As}$  layer is chemically transformed into an insulating oxide, thus creating current apertures with an inner diameter of only a few micrometers.

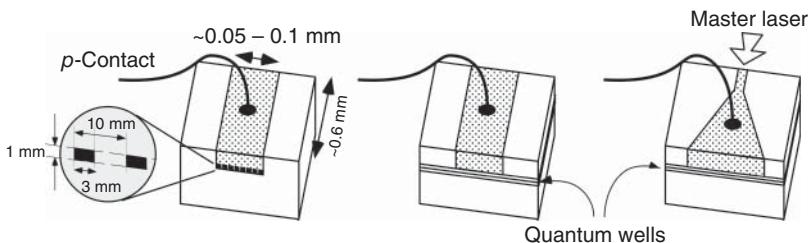
## 10.6 High-Power Laser Diodes

The direct conversion of electrical energy into coherent light with an excellent efficiency promises a wealth of applications. Coherent light provides this energy for, for example, cutting and welding in materials processing with a very high “quality,” so to speak, because its application can be controlled with a very good spatial and time resolution. So interest in increasing the output power of laser diodes up to the range of 1 kW and more was quite natural from the beginning.

The “quality” of a laser beam for machining applications depends on the total power available but at the same time depends crucially on its spatial properties, that is, the transverse coherence. For practical evaluation, it is customary to use the beam parameter product of beam waist  $w_0$  and divergence angle  $\theta_{\text{div}}$  (see p. 42). When this is normalized to the corresponding product for a perfect  $\text{TEM}_{00}$  Gaussian beam, it is called the  $M^2$  factor [161]:

$$M^2 = \frac{w_0 \theta_{\text{div}}(\text{measured})}{w_0 \theta_{\text{div}}(\text{perfect})}. \quad (10.17)$$

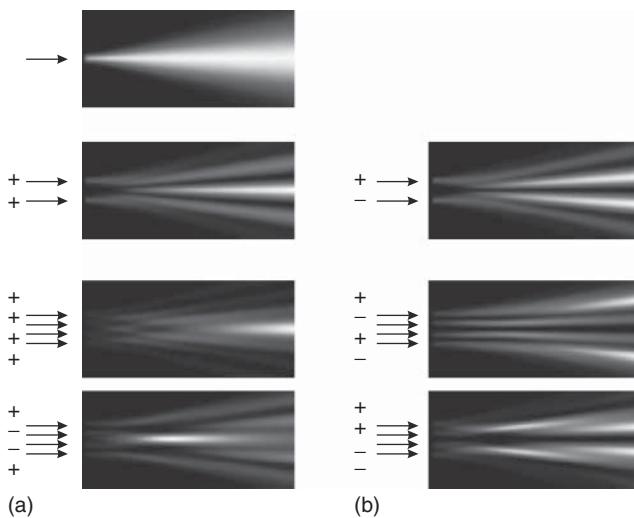
It is a measure of the performance of beam cross section and divergence and gives an estimate of the fraction of laser light propagating within the dominant Gaussian mode, for only this can be focused to the optimum, that is, limited by diffraction or transmitted through a spatial filter (see Figure 2.12). The  $M^2$  factor grows with decreasing beam quality and should differ as little as possible from unity.



**Figure 10.30** Concepts for high-power laser diodes: laser diode arrays, broad-area laser, and tapered amplifier.

As already indicated in Figure 10.12, the power increase just by increasing the injection current is seriously limited. On the one hand, owing to the excess heat, the “rollover” effect occurs; on the other hand, the light intensity becomes so high that the emitting facets suffer spontaneous damage, often leading to the total loss of the device. These problems are particularly severe for layers containing Al. For this reason, in high-power lasers, mostly Al-free quantum films are used at least for the gain zone. It is generally observed that the output power for a conventional single laser diode stripe with a facet of about  $1 \times 3 \mu\text{m}^2$  is limited to not more than some 100 mW. Therefore the output power of semiconductor devices can in principle only be increased by spreading the gain over facets as large as possible or over many facets and the volumes connected to them. Today the output power is increased by using laser arrays and broad-area and tapered amplifier lasers as schematically presented in Figure 10.30.

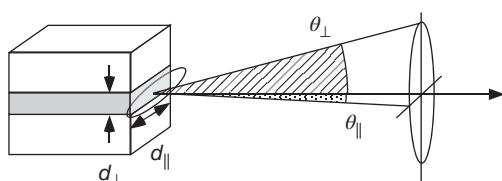
- 1) Several laser diode stripes can be placed on a single substrate without any problem. If the separations of the single stripes are not too big, the fields of adjacent modes overlap slightly and are coupled through their phase evolution, that is, the output power of all individual stripes is coherently coupled or “capable of interfering.” The far field of a laser array depends on the relative phase positions of the single stripes. In Figure 10.31 we show calculations of the idealized field distribution of two and four identical Gaussian emitters, considering all possible combinations of relative phase positions. A realistic laser array often shows a far field with two “ears.” Their origin becomes evident from this consideration.
- 2) In a broad-area laser diode, a wide diode volume is used for amplification, as the name indicates. However, the control of the transverse field distribution becomes more and more difficult with increasing power, so that broad-area lasers are limited to quite low powers (sub-watts).
- 3) Tapered amplifiers of trapezoidal shape are used to boost laser light from lower power to high power while transferring high spatial and longitudinal coherence. For this concept, the term master oscillator power amplifier (MOPA) has become established. The trapezoidal form has been chosen here to realize maximum gain, but to keep the power density low at the same time to avoid damage. With this concept  $M^2$  factors of 1.05 at output powers of some watts are achieved.



**Figure 10.31** Beam shapes of laser arrays: (a) the symmetric phase positions of the individual stripes; (b) the antisymmetric phase positions.

## Problems

- 10.1 Gain of the semiconductor laser** Consider a GaAs semiconductor laser at  $T = 0$  K. For the intrinsic charge carrier concentration, we take  $n = 1.8 \times 10^6$  cm $^{-3}$  for the recombination lifetime  $\tau = 50$  ns and for the bandgap  $E_g = 1.42$  eV. The effective mass of electrons and holes is  $0.07 m_{el}$  and  $0.5 m_{el}$ , respectively. Calculate the central emission wavelength, the bandwidth, and the maximal gain within the bandwidth for an amplifier with dimensions: length  $d = 200$   $\mu\text{m}$ , width  $w = 10$   $\mu\text{m}$ , and height  $h = 2$   $\mu\text{m}$ . The current injected into the laser pn junction is 1 mA.
- 10.2 Bit rate** Calculate the bit rate that can be handled by the amplifier of the previous problem. Compare with the bandwidth of an individual audio channel that requires a bit rate of 64 kbit s $^{-1}$ .
- 10.3 Beam profile of a broad stripe laser** The divergence of a Gaussian beam (see Section 2.3) with wavelength  $\lambda = 850$  nm is calculated to be  $\theta = \lambda/\pi w_0$  in the far field. All other beam geometries with larger divergence can be characterized with the  $M^2$  factor according to Eq. (10.17), or  $\theta = M^2 \lambda/\pi w_0$ .



**Figure 10.32** Components of a broad stripe laser diode.

A broad stripe laser shows an elliptic profile with a shorter and a longer axis (see Figure 10.32). Immediately at the diode facet beam, the near-field full width at half maximum (FWHM) in the narrow direction is  $2w_0 = d_{\perp} = 0.8 \mu\text{m}$ . The measured divergence is  $\theta_{\perp, \text{meas}} = 20^\circ$ . In the wider, horizontal direction, the corresponding values are  $2w_0 = d_{\parallel} = 100 \mu\text{m}$  and  $\theta_{\parallel, \text{meas}} = 10^\circ$ , respectively. Determine the  $M^2$  value of the laser beam. Note: The divergence angle is defined by the  $1/e^2$  points.

# 11

## Sensors for Light

The application of optical instruments depends critically on how sensitively light can be detected by means of suitable devices. Here we are rather blessed by the human eye, which is – despite all the weaknesses of its imaging optics – an enormously sensitive and versatile receiver.

From the historical point of view, above all, we find light-sensitive plates at the beginning of the development of optical sensors. Photographic emulsions in which light causes a permanent chemical change have been developed to high sensitivity, high resolution, and countless applications in more than a century of intensive work.

However, in a physical experiment or in a technical application, when the intensity of a light beam has to be detected and evaluated, then solid-state detectors (and among them especially semiconductor detectors) have outperformed films for quite a long time. They not only deliver an electrical signal that can be saved and recorded without a slow sequence of chemical processes but also are advantageous with regard to linearity.

Until recently, films used to be unbeatable for taking high-contrast pictures with high resolution. With the culture-driving development of semiconductor technology, and the opportunity of processing larger and larger (electronic) data streams faster and faster, that field of application runs into danger of being replaced by optoelectronic components as well. We report about this in the section about image sensors.

Optical sensors generally consist of physical materials that can be coarsely divided into two classes according to the effect of the incident light beam:

- 1) *Thermal detectors*: Ideal thermal detectors are *blackbodies* (see Section 7.3). This means that they absorb all incident light. The energy current of the incident light leads to a temperature increase compared to the environment, which is measured and converted into an electrical signal.  
Among the thermal detectors are thermopiles, bolometers, and pyroelectric detectors. The strengths of thermal detectors are their broad spectral sensitivity and their robust layout. Their most significant disadvantage is a slow rise time.
- 2) *Quantum sensors*: In a quantum detector, a light beam is converted into free charge carriers using the internal or external photo effect. The current or the charges are directly measured. The often used picture, according to which in a

photodiode photons are simply converted into electrons and counted, has to be taken with a pinch of salt. However, a more strict theoretical description of the photon counter is beyond the scope of this text [74, 86].

Among the quantum detectors are photomultipliers, on the one hand, and photoconductors and photodiodes, on the other hand. The historical development from electron tube to semiconductor technology can also be observed with these components. As suggested by the name, by means of quantum detectors, single electrons can be recorded. Their rise time is rarely more than 1  $\mu\text{s}$ , but often they have to be cooled and are subject to stronger spectral limitations than thermal detectors. In principle, also the emulsions of photographic films belong to the class of quantum sensors, since an individual photon is necessary to reduce each AgBr molecule and thus to cause blackening.

When an optical sensor has to be chosen for a certain application, from the physical point of view, it is of interest, for example, whether the detector has a sufficient sensitivity and a fairly short rise time to dynamically record the desired quantity. These properties can be found from the manufacturers' data sheets. For more insight we first have to strike out a bit further and talk about the noise properties of detector signals.

## 11.1 Characteristics of Optical Detectors

### 11.1.1 Sensitivity

In an optical sensor, light pulses are ultimately converted into electric signal voltages  $U(t)$  or signal currents  $I(t)$ . Since all electronic quantities are subject to similar procedures when measured, we use subscripts  $V_U$  and  $V_I$  for their identification. The *responsivity*  $\mathcal{R}$  describes the general response of the detector to the incident light power  $P_L$  without taking details such as wavelength, absorption probability, circuit wiring, and so on into account:

$$\text{Responsivity } \mathcal{R} = \frac{(V_U, V_I)}{P_L}. \quad (11.1)$$

The physical unit of responsivity is usually  $\text{V W}^{-1}$  (especially for thermal detectors) or  $\text{A W}^{-1}$ .

### 11.1.2 Quantum Efficiency

In a quantum detector photons are converted into electrons. Even single electrons may be amplified in such a way that their pulses can be registered and counted. Not every incident photon triggers an electron since the absorption probability is lower than unity or because other processes compete with the photo effect. The probability for registering an event for each incident photon is called the *quantum efficiency* (QE)  $\eta$ . The rate of arrival of photons  $r_{\text{ph}}$  at a detector with area  $A$  can easily be determined according to

$$r_{\text{ph}} = \frac{1}{h\nu} \int_A dx dy I(x, y). \quad (11.2)$$

If the entire radiation power is absorbed, Eq. (11.2) simplifies to  $r_{\text{ph}} = P_{\text{L}}/\hbar\nu$ . In an ideal quantum detector, this should become the photocurrent  $I = er_{\text{ph}}$ , but in physical reality there are competing processes reducing the QE. According to Eq. (11.1) the responsivity can be expressed in terms of these elementary quantities:

$$\mathcal{R} = \frac{r_{\text{el}}}{r_{\text{ph}}} \frac{e}{\hbar\nu} = \eta \frac{e}{\hbar\nu}. \quad (11.3)$$

A practical rule of thumb can be obtained by using the wavelength  $\lambda = c/v$  in  $\mu\text{m}$  instead of the frequency:

$$\mathcal{R} = \eta \frac{\lambda/\mu\text{m}}{1.24} \quad [\text{A W}^{-1}],$$

from which the QE can be determined for a known responsivity.

### 11.1.3 Signal-to-Noise Ratio

A quantity can only be recognized if it emerges “from the noise,” that is, if it is larger than the intrinsic noise of the detector. Formally the quantitative concept of “signal-to-noise ratio” (SNR) has been introduced:

$$\text{SNR} = \frac{\text{signal power}}{\text{noise power}}.$$

In this case we use a generalized concept of power  $\mathcal{P}_V(f)$  for an arbitrary physical quantity  $V(t) = \mathcal{V}(f) \cos(2\pi ft)$ . The average power is

$$\mathcal{P}_V(f) = \frac{1}{2}\mathcal{V}^2(f). \quad (11.4)$$

The physical unit of these powers is  $\text{A}^2$ ,  $\text{V}^2$ , ..., depending on the basic value. A fluctuating quantity such as noise current or voltage is determined not only by one amplitude at one frequency but also by contributions at many frequencies within the bandwidth  $\Delta f$  of the detector. The average power in a frequency interval  $\delta f$  can be measured with a filter of this bandwidth and with mid-frequency  $f$ . Therefore we define the power spectral density

$$v_n^2(f) = \frac{\delta V^2(f)}{\delta f},$$

so, for example,  $i_n^2(f)$  in  $\text{A}^2 \text{Hz}^{-1}$  for current noise and  $e_n^2(f)$  in  $\text{V}^2 \text{Hz}^{-1}$  for voltage noise. Since the contributions are not correlated, the square sum of the power contributions in small frequency intervals can be summed to give the average of the noise power (see Section A.1):

$$P_V = \int_{\Delta f} v_n^2(f) df. \quad (11.5)$$

If the noise in the bandwidth  $\Delta f$  is constant, the value of the noise power simplifies to  $P_V = v_n^2 \Delta f$ . For example, the r.m.s. value of the noise current  $I_{\text{rms}} = \sqrt{P_I}$  of a photodiode–amplifier combination reads  $I_{\text{rms}} = (i_n^2 \Delta f)^{1/2}$ , with  $i_n^2$  the constant value of the current noise spectral density.

Often the unphysical noise amplitude is given instead of the noise power,

$$\text{Noise amplitude} = (\text{noise power spectral density})^{1/2},$$

given in, for example,  $\text{A Hz}^{-1/2}$  or  $\text{V Hz}^{-1/2}$ . Very generally the noise contribution can be reduced by limiting the bandwidth of the detector. This advantage has to be traded in for reduced dynamic properties – faster signal variations can no longer be registered.

#### 11.1.4 Noise Equivalent Power (NEP)

The *noise equivalent power* (NEP) is the radiation power that is necessary to exactly compensate the noise power at the detector or to obtain an SNR of exactly unity. The lower the designed bandwidth of the detector, the lower is the minimum detectable power, but again at the expense of the bandwidth. The minimum detectable radiation power is therefore referred to 1 Hz bandwidth and is given by the unphysical noise amplitude density:

$$\text{NEP} = \frac{(\text{noise power spectral density})^{1/2}}{\text{responsivity}}.$$

Its physical unit is  $\text{W Hz}^{-1/2}$ . The manufacturer of a detector prefers to quote the spectral maximum of the responsivity, though it has to be taken into account that the value depends on the optical wavelength  $\lambda$  as well as on the electrical signal frequency  $f$ .

#### 11.1.5 Detectivity "D-Star"

For the sake of completeness, we mention the concept of “detectivity”  $D$  and  $D^*$  introduced to make different detector types comparable with each other. First, just the complement of the noise equivalent power  $D = \text{NEP}^{-1}$  was introduced as the detectivity. The variant of the “specific detectivity” called “D-star” ( $D^*$ ) has found widespread use since the responsivity of many detectors is proportional to the square root of the detector area  $A^{1/2}$ :

$$D^* = \frac{\sqrt{A}}{\text{NEP}}. \quad (11.6)$$

The reason for this is the limitation of the detection sensitivity by the thermal background radiation, especially for infrared detectors: the larger the detector area, the more blackbody radiation is absorbed. The physical unit of  $D^*$  is  $\text{cm}/(\text{W Hz}^{-1/2}) = 1 \text{ jones}$ , where the name of the inventor of the detectivity is used for abbreviation.  $D^*$  is a measure for the SNR in a bandwidth of 1 Hz when a detector with an area of diameter 1 cm is illuminated with a radiation power of 1 W.

#### 11.1.6 Rise Time

Often very fast events are to be recorded by means of optical detectors, which means that the detector has to react very rapidly to variations of the incident radiant flux. The “rise time”  $\tau$  is the time during which the current or voltage

change of the detector reaches  $(1 - 1/e)$  or 63% of the final value when the light source is switched on suddenly. In analogy to that, the “fall time” is also defined. They depend on the layout of the detector and can be influenced within the physical limits. Thermal detectors are inert and react with delay times of many milliseconds. The “rise time” of semiconductor detectors is generally limited by the capacitance of the pn junction and is only a few picoseconds in special cases. The finite response time of a detector can be taken into account by, for example, adding the time or frequency dependence to the responsivity of Eq. (11.3):

$$\mathcal{R}(f) = \frac{\mathcal{R}(0)}{1 + (2\pi f \tau)^2}.$$

The charge impulse of a photomultiplier can as well be shorter than 1 ns, though here the longer time of travel through the dynode layout has to be taken into account. The times of travel in cable connections also have to be taken into account in servo-control applications, since they limit their bandwidth.

### 11.1.7 Linearity and Dynamic Range

A linear relation of detector input power and output voltage or current provides optimal conditions for a critical analysis of the quantity to be measured. However, there is always an upper limit – ultimately due to the strong temperature load at high light power – at which deviations from linearity can be observed. The lower limit is mostly given by the NEP. A quantitative measure for the *dynamic range* can be given according to

$$\text{Dynamic range} = \frac{\text{saturation power}}{\text{NEP}}.$$

The dynamic range between these limits can be, for example, for photodiodes, an impressive six magnitudes or more.

## 11.2 Fluctuating Optoelectronic Quantities

In this section we collect physically different contributions of electrical noise generated in optoelectronic detectors. Besides the intrinsic contributions of the receiver–amplifier combination, such as dark current noise and amplifier noise, above all, there is the photon noise of the light source.

### 11.2.1 Dark Current Noise

A detector generates a fluctuating signal  $V_n(t)$  even when there is no incident light signal at all. In fact, the detector sensitivity is decreased not by the average of the background – this can be straightforwardly subtracted – but by its fluctuations. In a thermal detector, spontaneous temperature fluctuations cause the dark noise. In a quantum detector, generally charge carriers spontaneously generated, for example, by thermionic emission, are responsible for this. In the simplest case, the noise power density of the dark current  $I_D$  is  $i_D^2 = 2eI_D$  according to the Schottky formula (Eq. (A.13)). A proven but sometimes costly method for reducing the dark noise is to apply cryogenic cooling of the detector.

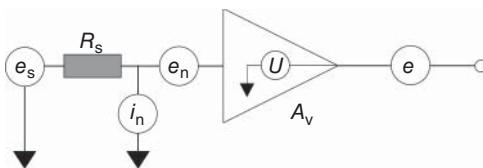
### 11.2.2 Intrinsic Amplifier Noise

Photomultipliers and avalanche photodiodes (APDs) have an internal amplification mechanism that multiplies the charge of a photoelectron by several orders of magnitude. The amplification factor  $G$  though is subject to fluctuations contributing to noise as well. The *excess noise factor*  $F_e$  is calculated according to

$$F_e^2 = \frac{\langle G^2 \rangle}{\langle G \rangle^2} = 1 + \frac{\sigma_G^2}{\langle G \rangle^2} \quad (11.7)$$

and can also be expressed through the variance of the amplification,  $\sigma_G^2 = \langle G^2 \rangle - \langle G \rangle^2$ . It affects dark and photocurrents in indistinguishable ways.

### 11.2.3 Measuring Amplifier Noise



**Figure 11.1** Noise sources of an idealized amplifier. Idealized voltage sources ( $e_i$ ) have negligible inner resistance, and current sources ( $i_n$ ) infinite inner resistance.  $R_S$  is the resistance of the detector, and  $A_V$  the amplification.

Depending on the application, a detector may operate as a voltage or a current source characterized by its internal resistance  $R_S$  (Figure 11.1). At the input of an idealized test amplifier, we find the voltage noise amplitude  $e_i$ , which consists of the uncorrelated contributions of the detector,  $e_S^2$ , with source resistance  $R_S$ , and the contributions of current and voltage

noise of the amplifier ( $i_n^2$  and  $e_n^2$ , respectively):

$$e_i^2 = e_S^2 + e_n^2 + i_n^2 R_S^2.$$

The noise voltage at the exit of the amplifier is then  $e = A_V e_i$ . The noise amplitude of the detector consists of the contributions of the dark current, the parallel resistance of detector and amplifier input, and the photon current  $i_{ph}^2$ :

$$e_S^2 = R_S^2 \left( i_{ph}^2 + i_D^2 + \frac{4kT}{R_S} \right).$$

The last contribution takes the thermal or *Johnson* noise of the detector resistance into account. For optical detection, the most desirable situation is obtained when the noise of the photoelectrons generated by the signal source ( $i_{ph}^2$ ) dominates all intrinsic amplifier contributions:

$$i_{ph}^2 > i_D^2 + \frac{4kT}{R_S} + \frac{e_n^2}{R_S^2} + i_n^2. \quad (11.8)$$

In practical applications it has to be taken into account that all the quantities mentioned previously depend on frequency. If possible, the frequency of the signal can be selected to minimize background noise. Advantageous conditions are generally found at high frequencies since all devices at frequencies below a certain *corner frequency*  $f_c$  show the so-called  $1/f$  or *flicker noise*, which approximately increases with  $1/f$  toward low frequencies. The typical spectral behavior of the amplifier noise is presented in Figure 11.2.

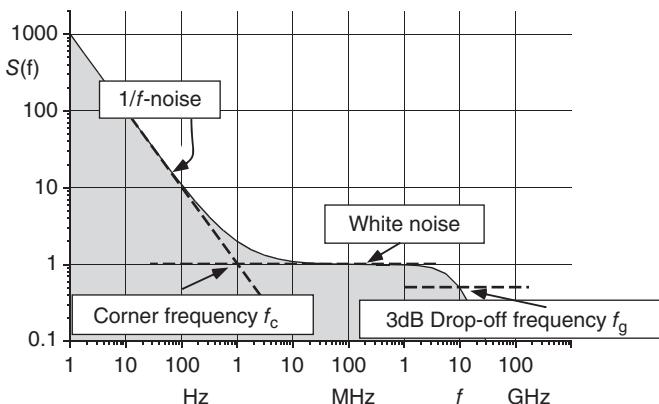


Figure 11.2 Spectral properties of typical amplifier noise, schematically.

### 11.3 Photon Noise and Detectivity Limits

By conversion of light into photoelectrons in an optoelectronic circuit, something like a copy of the photon current emerges. It is obvious that the fluctuations of the photon current are mapped onto the electron current as well. Now, for the rigorous description of the processes occurring during conversion of light into photoelectrons, a quantum theory of the electromagnetic field is required, but quantum electrodynamics does not offer an intuitive approach and is hence omitted here.

Instead, we assume that the probability of observing an event in a short time interval is proportional to  $\Delta t$  by taking into account the arrival rate of the photons (Eq. (11.2)) and the QE  $\eta$ :

$$p(1, \Delta t) = \eta r_{\text{ph}}(t) \Delta t. \quad (11.9)$$

Furthermore we assume that, for sufficiently small  $\Delta t$ , no double events occur and that the probabilities in sequential time intervals are statistically independent. The last assumption means that the photoemission process does not have any aftereffects in the detector; this is not necessarily the case anymore at high charge carrier density because they repel each other due to Coulomb forces.

These conditions lead to *Poisson statistics* of the counting events. The probability of finding  $K$  events in an arbitrary time interval  $\tau$  is

$$p_\tau(K) = p(K, t, t + \tau) = \frac{\bar{K}_\tau^K}{K!} e^{-\bar{K}_\tau}.$$

The average  $\bar{K}_\tau$  is according to Eq. (11.3)

$$\bar{K}_\tau = \eta r_{\text{ph}} \tau. \quad (11.10)$$

Random conversion of photons into photoelectrons leads to fluctuations of the photoelectron current. In addition, also the light intensity  $P_L(t)/A$  can vary. If it happens in a deterministic way, that is, predictably, we can define the power  $W_\tau$

integrated in the interval  $\tau$ ,

$$W_\tau(t) = \int_t^{t+\tau} P_L(t') dt',$$

and with the abbreviation  $\alpha = \eta/hv$ , we obtain the probability distribution

$$p_\tau(K) = \frac{(\alpha W_\tau)^K}{K!} e^{-\alpha W_\tau}. \quad (11.11)$$

The properties of the light source are reflected in the statistics of the photoelectrons, and therefore we consider the light field of a laser and a thermal light source as important examples.

### 11.3.1 Photon Statistics of Coherent Light Fields

The average power of a laser is constant; therefore the arrival rate of photons  $r_{\text{ph}}$  is constant as well, and we can directly take over the average from Eq. (11.10).

The statistical distribution is characterized by the variance

$$\sigma_{K_\tau}^2 = \overline{(K^2 - \bar{K}_\tau^2)},$$

which has the known value for Poisson statistics

$$\sigma_{K_\tau}^2 = \bar{K}_\tau. \quad (11.12)$$

From this relation it is also clear that relative fluctuations decrease with increasing number of events,

$$\frac{\sigma_{K_\tau}}{\bar{K}_\tau} = \frac{1}{\sqrt{\bar{K}_\tau}}, \quad (11.13)$$

and become very small for large  $\bar{K}_\tau$ . The noise caused by the grainy particle structure of the current is called *shot noise*. It also sounds very loud like the audible drubbing caused by raindrops falling on a tin roof.

We expect a random sequence of charge pulses for the photoelectron current. The spectrum of the current noise depends on the frequency and can be obtained directly from the Schottky formula (Eq. (A.13)):

$$i_{\text{coh}}^2 = 2e\bar{I}_{\text{ph}}. \quad (11.14)$$

This noise current also accounts for the contribution caused by the random conversion of photons into photoelectrons when the QE is lower than 100%. We can furthermore interpret the Schottky formula by identifying the r.m.s. value of the counting statistics  $\sigma_{K_\tau}^2$  in the time interval  $\tau$  with the variance of the number of charge carriers,  $\sigma_{K_\tau}^2 = \frac{1}{2}I_{\text{rms}}^2/e^2$ , which again leads to the result of Eq. (11.14) (the factor  $\frac{1}{2}$  occurs because the power spectral density is defined for positive Fourier frequencies only; see Section A.1).

A coherent light field generates the photocurrent with the lowest possible noise, and therefore it comes very close to our idea of a classical wave with constant amplitude and frequency. We may interpret the noise as a consequence of the “granularity” of the photocurrent and of its Poisson statistics, although it has to

be pointed out that we did not derive this result here but rather have put it in from the beginning.

### 11.3.2 Photon Statistics in Thermal Light Fields

A thermal light field generates an average photocurrent as well, though the intensity is not constant like in a coherent laser beam but subject to strong random fluctuations. Therefore, for the integrated power  $W_\tau$ , we can also give only probabilities  $p_{\tau,W}(W_\tau)$  with  $\int dW_\tau p_{\tau,W}(W_\tau) = 1$  here. The additional fluctuation of the amplitude results in *Mandel's formula*, which is formally similar to a Poisson transformation of the probability density  $p_{W_\tau}$  (see Eq. (11.11)):

$$p(K) = \int_0^\infty \frac{\alpha W_\tau}{K!} e^{-\alpha W_\tau} p_{\tau,W}(W_\tau) dW_\tau. \quad (11.15)$$

This contribution has double Poisson character, so to speak. It can be shown that the average of the counting events is

$$\bar{K}_\tau = \alpha \bar{W}_\tau$$

as before, and the variance is

$$\sigma_{K_\tau}^2 = \bar{K}_\tau + \alpha^2 \sigma_{W_\tau}^2. \quad (11.16)$$

Thus the variance of a fluctuating field, for example, blackbody radiation (see the excursion on p. 240), is in principle larger than for a coherent field. We will see, however, that detection of these strong fluctuations is possible at very short time scales only and hence beyond the dynamic properties of most photodetectors.

We can interpret the relation (11.16). The first term is caused by the random conversion of photons into photoelectrons and is a microscopic property of the light–matter interaction that cannot be removed. The second term represents the fluctuations of the recorded light field and also occurs without the randomness of the photoelectron generation process.

The calculation of  $\sigma_W^2$  in (11.16) is not a trivial problem at all. We consider the cases of extremely short and very long integration intervals  $\tau$ . A thermal light field is characterized by random amplitude fluctuations. For very short time intervals, even shorter than the very short remaining coherence time  $\tau_c$  of the light source of about 1 ps, we can assume a constant intensity so that  $W_\tau = P_L \tau$ . The intensity itself is randomly distributed and thus follows a negative exponential distribution:

$$p_{\tau,W} = e^{-W/\bar{W}_\tau} / \bar{W}_\tau.$$

By insertion into Eq. (11.15) and integration, the *Bose–Einstein distribution* of quantum statistics is obtained:

$$p_\tau(K) = \frac{1}{1 + \bar{K}_\tau} \left( \frac{\bar{K}_\tau}{1 + \bar{K}_\tau} \right)^K. \quad (11.17)$$

The variance of this field is

$$\sigma_K^2 = \bar{K}_\tau + \bar{K}_\tau^2,$$

and can be interpreted like (11.16) before. Its relative value always remains close to unity:

$$\frac{\sigma_K}{K_\tau} = \sqrt{\frac{\bar{K}_\tau}{1 + \bar{K}_\tau}}.$$

The distribution from Eq. (11.17) is well known for a light field when  $K$  is replaced by  $n$  and  $\bar{K}_\tau$  by the mean thermal number of photons:

$$\bar{n}_{\text{ph}} = \frac{1}{e^{h\nu/kT} - 1}. \quad (11.18)$$

The coherence time of a thermal light source though is so short that there exist hardly any detectors with appropriately short response and integration times. The more important limiting case for the thermal light field thus occurs for integration times  $\tau \gg \tau_c$ . For this case it can be shown [74] that the variance  $\sigma_K$  is well approximated by

$$\sigma_K^2 = \bar{K}_\tau \left( 1 + \frac{\bar{K}_\tau \tau_c}{\tau} \right). \quad (11.19)$$

Thus for most cases  $\sigma_K^2 \simeq \bar{K}_\tau$  as for the thermal light field (cf. Eq. (11.12)). Incidentally, these noise properties cannot tell us about the properties of the light field, coherent or thermal! The second term in Eq. (11.19) can be interpreted as the number of photons reaching the detector during a coherence interval. Only when this number becomes larger than unity can a significant increase of the fluctuations be expected.

Ambient radiation of a light source mostly corresponds to the spectrum of the blackbody radiation at 300 K. Its maximum lies at a wavelength of  $10 \mu\text{m}$  and decreases rapidly toward the visible spectral range. Unavoidably at least part of this radiation also enters the detector. Especially for infrared detectors the sensitivity is in general limited by the background radiation. For thermal radiation it is still valid that the coherence time is very short, so that the variance of the photoelectron noise of the thermal radiation can be calculated according to (11.19).

In order to determine the emission rate of photoelectrons  $r_{\text{el}}$ , we have to multiply the average photon number  $\bar{n}_{\text{ph}}$  of Eq. (11.18) with the density of oscillator modes  $\rho(\nu) = 8\pi\nu^2/c^3$  at frequency  $\nu$  to integrate over the detector area  $A$ . Accounting for the QE  $\eta(\nu)$  and in addition the radiative flux from half the solid angle  $2\pi$ , we arrive at

$$r_{\text{el}} = A \int_0^\infty d\nu \eta(\nu) \frac{2\pi\nu^2}{c^3} \frac{1}{e^{h\nu/kT} - 1}.$$

The spectrum of the charge carrier fluctuations is proportional to the variance of the arrival rate, which we can now calculate according to Eq. (11.19); as for the coherent light field, we obtain a white shot noise spectrum. Since the photoemission vanishes below a certain critical frequency  $\nu_g$  or a critical wavelength  $\lambda_g = c/\nu_g$ , the noise spectrum for a detector with the bandgap  $E_g = h\nu_g$  can be

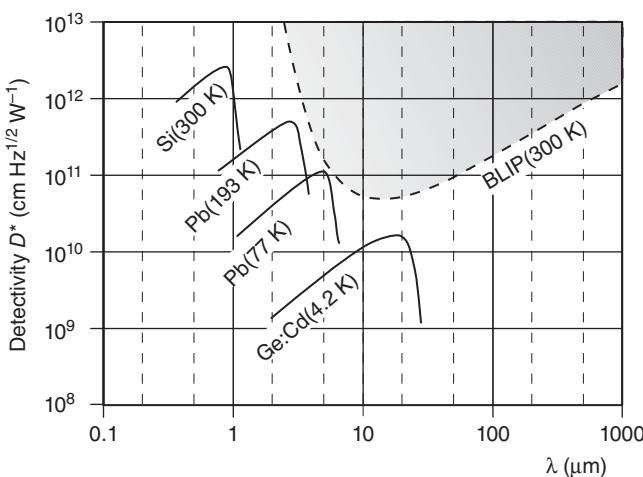


Figure 11.3 Specific detectivity for some important semiconductor detectors.

calculated according to

$$i_n^2 = 2e^2 r_{el} = 2e^2 A \int_{\nu_g}^{\infty} d\nu \eta(\nu) \frac{2\pi\nu^2}{c^2} \frac{1}{e^{h\nu/kT} - 1}.$$

When we additionally assume the QE has maximum value \$\eta(\nu) = 1\$ everywhere, then according to Eq. (11.6), we obtain the maximum specific detectivity \$D^\*(\lambda\_g, T)\$ of an ideal *background-limited photodetector* (BLIP detector), which depends on the environmental temperature \$T\$ and the critical wavelength \$\lambda\_g\$:

$$D^*(\lambda_g, T) = \frac{\lambda_g}{hc} \left( 2 \int_{c/\lambda_g}^{\infty} d\nu \frac{2\pi\nu^2}{c^2} \frac{1}{e^{h\nu/kT} - 1} \right)^{-1/2}.$$

This reaches a minimum at \$\lambda = 14 \mu m\$ (Figure 11.3). For large wavelengths \$D^\*\$ has to increase linearly since the thermal radiation power does not change anymore.

### 11.3.3 Shot Noise Limit and “Square-Law” Detectors

According to Eq. (11.14) the photoelectron noise generated by detection of a coherent laser beam is proportional to \$P\_L\$. This most favorable case is mainly realized with photodiodes. If the power is chosen large enough according to

$$P_L \geq \frac{h\nu}{\eta} \frac{1}{e^2} \left( 2e\bar{I}_D + \frac{4kT}{R_S} + \frac{e_n^2}{R_S^2} + i_n^2 \right) = \frac{h\nu}{\eta} r_{th}, \quad (11.20)$$

then the photon noise of the light beam dominates all other contributions in Eq. (11.8), which do not depend on the light power. This case is called the “shot-noise-limited” detection. Incidentally, the term within the large brackets in Eq. (11.20) can be interpreted as the rate \$r\_{th}\$ at which the detector–amplifier combination randomly generates charge carriers. Defining the minimum light

power by the value where the same number of charge carriers is generated (SNR  $\approx 1$ ), we find

$$P_{\min} = \frac{h\nu}{\eta} \sqrt{r_{\text{th}} \Delta f}$$

in a bandwidth  $\Delta f$ . For sufficiently long integration times (or correspondingly small bandwidth), in principle, arbitrarily small power may be registered. In practice, this potential though is impaired by the dynamics of the signal and slow drifts of the detector–amplifier properties.

Quantum detectors are also called “square-law” detectors since the trigger probability of a photoelectron is proportional to the square value of the field strength  $|E(t)|^2 = 2P_L(t)/ce_0A$  of the radiation field illuminating the detector area  $A$ . This is especially important for applying the so-called heterodyne detection. For this method the field of a local oscillator  $\mathcal{E}_{\text{LO}} e^{-i\omega t}$  (see Section 8.1.7.2) is superimposed with a signal field  $\mathcal{E}_S e^{-i(\omega+\omega_S)t}$  on the receiver. In general one chooses  $P_{\text{LO}} \gg P_S$ . The photocurrent will thus experience a variation in time

$$I_{\text{ph}} \simeq \frac{e\eta}{h\nu} (P_{\text{LO}} + 2\sqrt{P_S P_{\text{LO}}} \cos \omega_S t).$$

If LO and signal fields oscillate with the same frequency  $\omega$ , it is called a “heterodyne” detection; otherwise ( $\omega_S \neq 0$ ) it is a “heterodyne” detection. Superposition of optical fields on a square-law detector generates products of local oscillator and signal fields oscillating at difference frequencies, and thus it acts as an optical mixer.

The detection of a signal at a higher frequency is usually an advantage since it occurs at a lower noise power spectral density (Figure 11.2). When the LO power is increased until its shot noise density  $i_{\text{LO}}^2 = 2e^2\eta P_{\text{LO}}/\hbar\nu$  dominates all other contributions, the minimum detectable signal power no longer depends on the thermal noise properties of the detector. One has  $I_S = 2en\sqrt{P_{\min} P_{\text{LO}}}/\hbar\nu$ , and the minimum power  $I_S^2$  has to be larger than the noise power  $i_{\text{LO}}^2 \Delta f$  in the bandwidth of measurement,  $\Delta f$ :

$$P_{\min} = h\nu \Delta f / \eta.$$

In other words, within the time resolution  $\Delta f^{-1}$  of the detector, the signal light has to release at least one photoelectron to make detection possible.

## 11.4 Thermal Detectors

Thermal detectors (Figure 11.4) consist of a temperature sensor coated with an absorber material, for example, special metal oxides known from illumination technologies. Over wide wavelength ranges, they have very “flat” spectral dependences and are therefore very much sought after for calibration purposes.

In order to achieve a high sensitivity, that is, a large temperature increase  $\Delta T$ , the sensor should have a low heat capacity  $K$  as well as a low heat loss rate  $V$  to the environment caused by heat conduction due to the construction, convection, and

radiation. The temperature change of the probe follows the differential equation

$$\frac{d}{dt}\Delta T = \frac{P_L}{K} - \frac{V}{K}\Delta T, \quad (11.21)$$

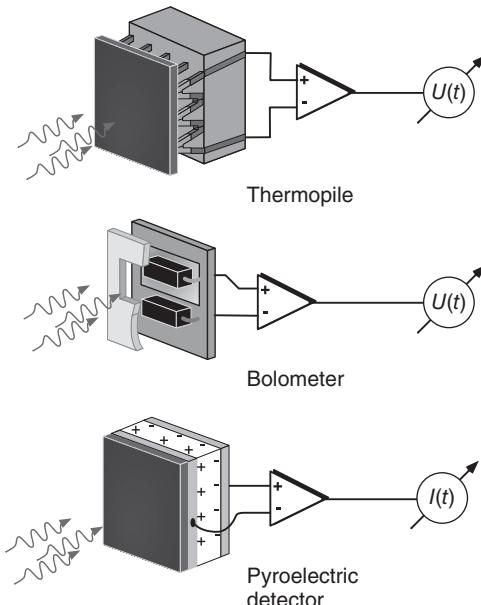
from which it can be seen immediately that a thermal detector integrates the incident light power for short times. In equilibrium the obtained temperature increase is  $\Delta T = P_L/V$ , from which the responsivity  $R_{\text{th}}$  is determined with the voltage–temperature coefficient of the thermal probe,  $C_{\text{TV}}$ :

$$R_{\text{th}} = C_{\text{TV}}/V.$$

However, compromises are necessary since the rise time is determined by the coefficient  $\tau = K/V$  according to Eq. (11.21). In the ideal case the minimum detectable power of a thermal detector is caused by unavoidable spontaneous temperature fluctuations, the spectral power density  $t^2 = 4k_B T^2 V/[V^2 + (2\pi Kf)^2]$  of which determines the theoretical responsivity limit ( $k_B$  = Boltzmann constant). For signal frequencies  $f$  far above the detector bandwidth  $\Delta f = 1/2\pi\tau$ , the idealized NEP can be given:

$$\text{NEP}_{\text{th}} = T\sqrt{2k_B V}.$$

Obviously it is profitable to lower the environmental temperature – a method that is used in particular for bolometer receivers.



**Figure 11.4** Thermal detectors.

#### 11.4.1 Thermopiles

In these, the light energy is absorbed by a thin blackened absorber plate in close thermal contact with a thin-layer pile of thermocouples made, for example, of copper–constantan. Since the voltage difference of a single element is very small, about 10–100 of them are connected in series with the “hot” ends receiving the radiation field to be detected and the “cold” ends kept at ambient temperature. The voltage of the thermopile is proportional to the temperature increase and thus to the power uptake of the absorber.

Thermopiles are mainly used in optics to determine the intensity of high-power light sources, especially laser beams. Owing to their integrating character, they are also capable of determining the average power of pulsed light sources.

### 11.4.2 Bolometers

The temperature increase by illumination can also be measured by means of a resistor with a large temperature coefficient. This is called a “bolometer.” For this application, especially semiconductor resistors called *thermistors* are of interest.

Bolometers are mainly used in a bridge circuit. Only one of two identical thermistors in the same environment is exposed to radiation so that fluctuations of the environmental temperature are already compensated. Very high sensitivities are obtained with bolometers operated at cryogenic temperatures when the heat capacity of the thermistor is very low.

### 11.4.3 Pyroelectric Detectors

In pyroelectric sensors a crystal is used with an electrical polarity that depends on temperature, for example, LiTaO<sub>3</sub>. The crystal is inserted into a capacitor. When the temperature changes, a charge is induced on the metallized faces, generating a transient current. The sensitivity for a crystal with pyroelectric coefficient  $p$ , heat capacity  $K$ , and distance  $d$  between the capacitor electrodes is

$$\mathcal{R} = p/Kd. \quad (11.22)$$

A pyroelectric detector registers only changes of the incident light power. According to Eq. (11.22) its sensitivity is significantly enhanced by thin-layer technology. Therefore the thickness of the crystal is only some 10 μm, which allows fast rise times. Wide spectral applicability of these detectors is achieved by using an appropriate broadband absorber.

Pyroelectric detectors are cheap and robust and are often used, for example, in the manufacture of motion sensors.

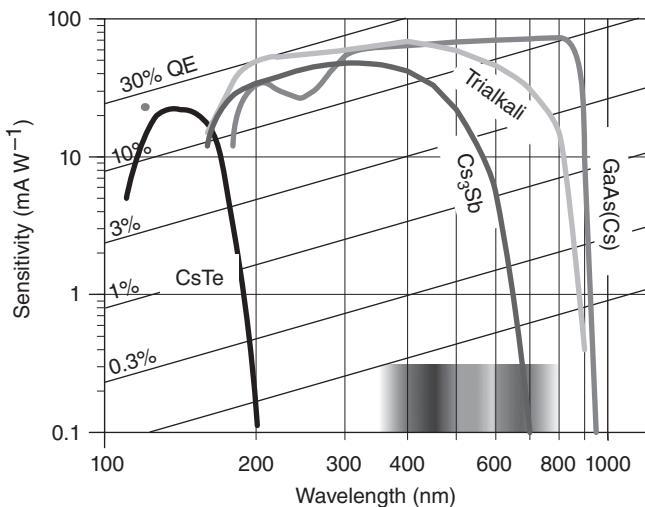
### 11.4.4 The Golay Cell

An unusual thermal detector is the radiation sensor called a *Golay cell* after its inventor. It is often used as a result of its high responsivity. The temperature increase by light absorption causes a pressure increase in a small container filled with xenon. On one side, the container is closed by a membrane that bulges due to the pressure increase. The small mechanical motion of the surface can be read out very sensitively by means of a “cat’s-eye technique.”

## 11.5 Quantum Sensors I: Photomultiplier Tubes

### 11.5.1 The Photoelectric Effect

It may be somewhat surprising that Albert Einstein got his Nobel Prize for physics in 1921 for his 1905 light quantum hypothesis of the photoelectric effect and not for any other of his scientific triumphs. He not only used Planck’s hypothesis that light energy could be absorbed only in light quanta with fixed value  $E_{\text{photon}} = h\nu$  but also expanded it by attributing quantum nature to the light field itself. According to Einstein’s simple concept, the maximum kinetic energy  $E_{\text{max}}$  of an electron



**Figure 11.5** Spectral responsivity of several important photocathodes.

that is emitted from the surface of a material with work function  $W$  is

$$E_{\max} = h\nu - W. \quad (11.23)$$

In general, though, only a few emitted electrons reach the maximum energy  $E_{\max}$ . More importantly, the photoelectric effect vanishes completely for frequencies  $\nu \leq W/h$ , the cutoff frequency or wavelength, which depends on the work function  $W$  of the used material.

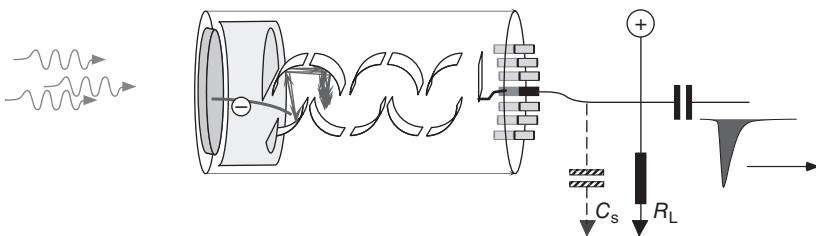
The photoelectric effect, or more shortly the photo effect, is the origin of all electric quantum sensors for light detection.

### 11.5.2 Photocathodes

Common metals mostly have very high values for the work function between 4 and 5 eV, corresponding to cutoff wavelengths between about 310 and 250 nm according to Einstein's equation (11.23). In vacuum it is also possible to use cesium, which immediately corrodes under atmospheric conditions but has the lowest work function of all metals with  $W_{Cs} = 1.92$  eV. By coating a dynode with cesium, a photocathode becomes sensitized for light frequencies extending nearly across the entire visible spectral range ( $\lambda < 647$  nm).

The probability of triggering a photoelectron by the absorption of a photon, the QE, is generally lower than unity (Figure 11.5). Owing to its high QE, approaching 30%, the semiconductor Cs<sub>3</sub>Sb is very often used for photocathode coating. It is inserted into vacuum tubes made from different glasses with differing transparencies. Such combinations have led to the classification of the spectral responsivity using the term *S-X cathode* ( $X = 1, 2, \dots$ ). The tri-alkali cathode S-20 ( $Na_2KC_2Sb$ ) has been used for a long time now, and 1% QE is achieved even at 850 nm.

Cs-activated GaAs offers 1% QE even further into the infrared at 910 nm wavelength. Still further in the infrared spectral range, the InGaAs photocathode



**Figure 11.6** Layout of a photomultiplier tube (PMT) with transparent dynode. The circuitry operates the PMT in the counting mode.

does not exceed 1% QE at any wavelength, but still has 0.1% QE at 1000 nm. In this spectral range, though, the internal photo effect in semiconductors has a very high QE; therefore here the photomultiplier tubes (PMTs) (Figure 11.6) compete with the APD discussed as follows, which can be considered as semiconductor-based PMTs.

Conversely there are situations in which a light detector ought to be sensitive only to UV wavelengths since then visible light such as daylight no longer contributes to the signal background and its noise. For this purpose the so-called *solar-blind* cathodes are used, which are made of, for example,  $\text{Cs}_2\text{Te}$  or  $\text{CsI}$ .

#### 11.5.2.1 Amplification

The success of PMT is not conceivable at all without the enormous amplification obtained with a secondary electron multiplier (SEM), which is connected to the photocathode. In an SEM, electrons are accelerated and cause multiple *secondary* electrons to be ejected from the anode. The multiplication factor for a layout with  $n$  dynodes at applied voltage  $U_{\text{PMT}}$  is  $\delta = c[U_{\text{PMT}}/(n + 1)]^\alpha$ . Up to 15 steps cause an avalanche-like amplification of the photocurrent  $I_{\text{ph}} = GI_{\text{el}}$ ,

$$G = \text{const} \times U_{\text{PMT}}^{\alpha n}, \quad (11.24)$$

with geometry and dynode material causing a slight attenuation of the theoretical amplification factor of a single step by a factor  $\alpha = 0.7\text{--}0.8$ . At the end of a cascade subjected to a total voltage of about 1–3 kV, a charge pulse with  $10^5\text{--}10^8$  electrons is available. The high intrinsic gain  $G$  leads to extreme sensitivity, which reaches values of  $R_{\text{PMT}} \simeq 10^4\text{--}10^7 \text{ A W}^{-1}$  according to

$$R_{\text{PMT}} = \frac{\eta Ge}{hv},$$

depending on layout and circuit wiring. Since the gain depends sensitively on the applied voltage due to (11.24), the voltage supply has to be stable and low noise.

#### 11.5.2.2 Counting Mode and Current Mode

The input channels of any electronic measuring amplifier usually expect a voltage at the input. Thus the current of the PMT has to be converted into a voltage by a load resistance  $R_L$ . Especially for low currents, the PMT works like an ideal current source, and thus  $R_L$  can be chosen arbitrarily large. In practice, however,

the rise time is limited by the load resistance and the stray capacitance of the anode to the layout

$$\tau = R_L C_S.$$

In addition, large load resistances cause the anode to discharge slowly. Thus the voltage of the last dynode stage is decreased and therefore also the efficiency of the anode for collecting secondary electrons: the characteristic curve becomes nonlinear and the PMT *saturates* for a certain light power.

For the circuitry wiring, the *counting mode* and the *current mode* are distinguished. The counting mode is suitable for very low light powers. The gain  $G$  is chosen very high and  $R_L$  so low that for a standard  $50\Omega$ , impedance typical voltage pulses of some  $10\text{ mV}$  and some nanoseconds width are observed. These pulses can be processed directly using commercial counting electronics. They cause the “clicks” of the *photon counter*. Because of the similarity to a *Geiger–Müller tube* for  $\alpha$  and  $\beta$  particles, this is also called the *Geiger mode*. Of course, a statistical distribution of impulses with different heights and widths is generated, from which signal photon pulses are selected by electronic discriminators.

The current mode is used for larger light intensities, with lower gain  $G$  and a load resistance adjusted to the desired bandwidth. The resistance should be chosen high enough to approach as close as possible an ideal current source.

### 11.5.2.3 Noise Properties of PMTs

A small current flows through a PMT even when the tube is operated in total darkness. It is called the “dark current”  $I_D$  and is mainly caused by thermionic emission of electrons from the photocathode, which are amplified indistinguishably from photoelectrons.

In the counting mode of the PMT, we can directly use the Schottky formula (Eq. (A.13)) if we insert the effective average charge  $\langle Ge \rangle$  of a single photoelectron to determine the power density of the shot noise of the dark counting rate  $R_D$ . In this case the NEP is calculated as

$$\text{NEP}_{\text{count}} = \frac{\sqrt{2R_D}}{\eta} h\nu, \quad (11.25)$$

where we have used the average gain  $\langle G \rangle$ .

If a PMT is used in the current mode, the fluctuations of the gain also cause noise: the noise power density of the current is then  $i_n^2 = \langle 2GeI_D \rangle = 2e\langle G^2 \rangle \langle I_D \rangle / \langle G \rangle$ , since the instantaneous gain is strictly related to the instantaneous current  $I_D$ . In current mode the result of (11.25) is increased by the *excess noise* factor  $F_e = \langle G^2 \rangle / \langle G \rangle^2$  of Eq. (11.7):

$$\text{NEP}_{\text{current}} = F_e \frac{\sqrt{2I_D / (eG)}}{\eta \langle G \rangle} h\nu.$$

Their enormous sensitivity has led to numerous applications for PMTs and in addition has caused the development of many specialized types. The most common models are the so-called *side-on* PMTs, in which the photoelectron is ejected from an opaque photocathode and first counter-propagates the light

beam. The *head-on* models are equipped with a transparent photocathode. From their rear the photoelectrons are sent into the secondary emission multiplier. They are advantageous when photocathodes with a large area are required, for example, in scintillation detectors. For applications in servo-control devices, though, PMTs have certain disadvantages when not only the rise time but also the delay time (caused by the travel time within the detector) plays a role.

#### 11.5.2.4 Microchannel Plates and Channeltrons

*Microchannel plates* (MCPs) are actually a variant of the SEM. A single microchannel consists of a glass capillary tube with a diameter of 6–20 µm. The wall is coated with a semiconducting material (e.g., NiCr) with relatively low conductivity. The ends of the tube are coated with a metal and operate as photocathode and anode, respectively; a high voltage drops along the walls and generates a “continuous dynode.” This type of secondary emission multiplier with a single channel is also known as a *channeltron*. Using an appropriate coating of the input facet, they can be converted into very compact photomultiplier devices. Their disadvantage is the saturation behavior, which generally begins at lower currents than in PMTs because of the high wall resistance.

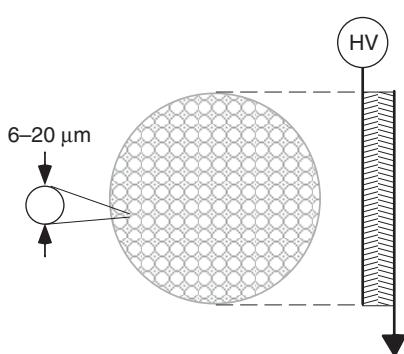


Figure 11.7 Microchannel plate (MCP), schematic.

An MCP consists of several thousand densely packed capillary tubes subjected in parallel to a high-voltage source and working like an array of SEM tubes (Figure 11.7). As *MCP-PMTs* they have advantages due to their excellent time resolution and their low sensitivity to magnetic fields (which influence the amplification behavior of every SEM). Moreover, they allow the detection of very low light intensities with spatial resolution and are therefore used to build the image intensifiers discussed in Section 11.7.3.

## 11.6 Quantum Sensors II: Semiconductor Sensors

In semiconductors, the photoelectrons do not have to be knocked out of the material but can internally generate free charge carriers. The *internal photo effect* is used in two different types of photodetectors: *photoconductors* and *photodiodes*. In photoconductors the photoelectric change of the conductivity is measured, while photodiodes are sources of photocurrent.

#### 11.6.1 Photoconductors

For the excitation of intrinsic photoelectrons, often a much lower energy is necessary than for the ejection of an electron out of a material. Photoconductors mostly manufactured using thin-layer technology therefore display their strength as infrared receivers.

In an *intrinsic* semiconductor, charge carriers can be generated by thermal motion or absorption of a photon. In this case the cutoff wavelength  $\lambda_g$  is determined by the energy of the bandgap according to Eq. (11.23) (Table 11.1). In Ge, for instance, it is 0.67 eV, corresponding to a cutoff wavelength of 1.85  $\mu\text{m}$ .

**Table 11.1** Bandgaps of selected semiconductors.

	Material	$E_g$ (eV) at 300 K	$\lambda_g$ ( $\mu\text{m}$ )
1	CdTe	1.60	0.78
2	GaAs	1.42	0.88
3	Si	1.12	1.11
4	Ge	0.67	1.85
5	InSb	0.16	7.77

**Table 11.2** Activation energies of dopants.

	Material	$E_A$ (eV) at 300 K	$\lambda_A$ ( $\mu\text{m}$ )
1	Ge:Hg	0.088	14
2	Si:B	0.044	28
3	Ge:Cu	0.041	30
4	Ge:Zn	0.033	38

The spectral sensitivity can be extended to even larger wavelengths by using extrinsic (doped) semiconductors (Table 11.2). The cutoff wavelength then increases with decreasing activation energy  $E_A$  of the donor atoms. Ge is used particularly often since its cutoff wavelength is extended, for example, by Hg dopants up to the 32  $\mu\text{m}$  limit.

### 11.6.1.1 Sensitivity

In a photoconductor the optically induced change of the conductivity is measured. Thus roles are played not only by the rate of charge carrier generation  $r_L$  – which behaves like the response of all quantum sensors – but also by the relaxation rate  $\tau_{\text{rec}}^{-1}$ , which ensures that the semiconductor returns to thermal equilibrium. For simplicity we assume the entire light power to be absorbed in the detector volume. Then the charge carrier density at constant light intensity is  $n_{\text{el,ph}} = \eta P_L \tau_{\text{rec}} / h\nu V_D$ .

However, the measured quantity is the conductivity  $\sigma$  and the current  $I = A\sigma U/\ell$ , respectively, flowing through a photoconductor of length  $\ell$  with effective diameter  $A$  when there is a voltage  $U$  across over it. It depends not only on the charge carrier densities  $n_{\text{el}}$  and  $p_h$  but also on the mobilities  $\mu_{\text{el}}$  and  $\mu_h$  of the electrons and holes, respectively. Owing to their low mobility, the holes contribute only negligibly to the conductivity, and hence

$$\sigma \simeq e n \mu_{\text{el}}. \quad (11.26)$$

By means of the photo effect, conductivity is generated within the photoconductor. It lasts until the electron–hole pair has recombined either still in the photoconductor itself or at the interfaces to the metallic connections. On the other hand, during the recombination time, a current flows that is determined by the mobility of the electrons. In the semiclassical Drude model, the drift velocity of the electrons can be connected with the applied voltage  $v_{\text{el}} = \mu_{\text{el}} U / \ell$  and also with the time  $\tau_d = \ell / v_{\text{el}}$  that it takes an electron to drift out of the photoconductor via the metallic leads. From  $I = A e n_{\text{el}} v_{\text{el}}$  the responsivity can be calculated as

$$\mathcal{R} = \frac{\eta e}{h\nu} \frac{\tau_{\text{rec}}}{\tau_d}.$$

Thus a photoconductor has an intrinsic gain  $G = \tau_{\text{rec}} / \tau_d$ , which can sometimes be smaller than unity. Moreover, the gain is obtained at the expense of a reduced detector bandwidth since the recombination rate  $\tau_{\text{rec}}$  determines the temporal behavior of the photocell as well.

#### 11.6.1.2 Noise Properties

The conductivity that is generated by thermal motion can be suppressed by routine cooling of the detector. So strictly speaking Eq. (11.26) has photoelectric and thermal parts:

$$\sigma = e(n_{\text{ph}} + n_{\text{th}})\mu_{\text{el}}.$$

The steady state of the conductivity in a photoconductor is determined by charge carrier generation and balanced by the recombination rate, which itself is a random mechanism. The shot noise of a photoconductor is called *generation–recombination noise* and is larger by a factor of 2 compared to the PMT or the photodiode:

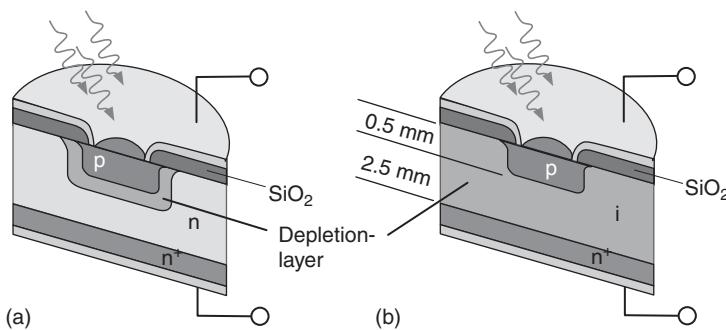
$$i_{\text{GR}}^2 = 4e\bar{I} \frac{\tau_{\text{rec}}}{\tau_d}.$$

At wavelengths of  $10 \mu\text{m}$  and beyond, the detectivity is generally limited by the thermal radiation background. Real detectors to a large extent operate at this limit.

#### 11.6.2 Photodiodes or Photovoltaic Detectors

Semiconductor photodiodes are among the most common optical detectors altogether because they are compact components and have many desirable physical properties, for example, high sensitivity, a fast rise time, and a large dynamical range. In addition, they come in numerous layouts and are straightforwardly interfaced with electronic semiconductor technology.

Their effect is based on the pn junction layer, which forms the so-called depletion layer where free charges are eliminated (see the following discussion). New electron–hole pairs are generated by absorption of light and accelerated by the internal electric field and thus cause a current flow in the test circuit. The *depletion region* acts as a nearly perfect current source, that is, with high internal impedance.



**Figure 11.8** Layouts of Si photodiodes. (a) Conventional design. (b) In the pin layout, the separation of the charge carriers is reached particularly fast.

### 11.6.2.1 pn and Pin Diodes

A depletion layer is formed close to the pn junction (Figure 11.8). Holes in the p-doped material and electrons in the n-doped material, respectively, diffuse to the opposite side and recombine there. The holes cause a positive space charge zone at the n side; since the electrons are in general more mobile than the holes, the corresponding negative zone is more extended on the p side. This process is finished when the electric field caused by the space charge prevents further diffusion of electrons and holes, respectively. A Si diode generates a known voltage drop of 0.7 V across the depletion layer.

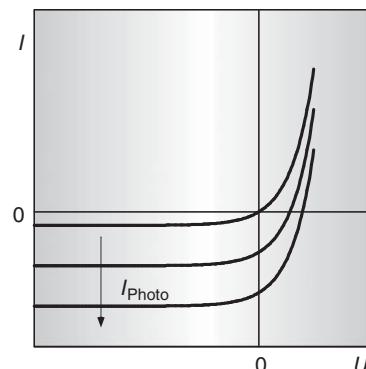
The construction of an efficient photodiode has the goal to absorb as much light as possible in the barrier layer so that the electric field, which can be increased still further by an external bias voltage, rapidly separates the electron–hole pairs. In contrast to a photoconductor, then recombination can no longer occur. This process can constructively be supported by inserting an insulating layer between the n and p layers, making the detector a pin photodiode. With this, the absorbing volume is increased, and additionally the capacitance of the barrier layer limiting the rise time is decreased.

### 11.6.2.2 Operating Modes

In Figure 11.9 a family of electrical characteristic curves of a photodiode is presented. The diagram results by adding the negative photocurrent  $-I_{\text{ph}}$  to the characteristic curve with  $I = I_s(e^{eV/kT} - 1)$  for a common diode.

There are three operating modes:

- *Photovoltaic mode*: When the photodiode is connected to an open circuit, then it is operated in the photovoltaic mode. Current flow is negligible ( $I = 0$ ), and the responsivity is given in  $\text{V W}^{-1}$ . This operating mode is also used in solar cells.



**Figure 11.9** Family of characteristics of a photodiode.

- *Short-circuit mode*: In the short-circuit mode, the current generated by the photoelectrons is measured and given in  $\text{A W}^{-1}$ .
- *Voltage bias operating mode*: In this most common operation mode, the barrier layer is further extended by a bias voltage so that a higher QE and shorter rise times are achieved.

### 11.6.3 Avalanche Photodiodes

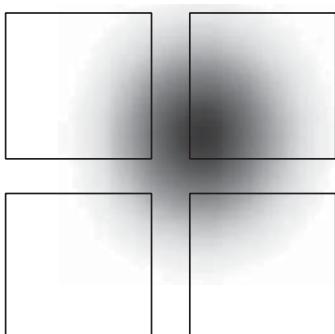
The principle of the APD has been known for a long time. However it was not possible to manufacture technically stable products until recently. In a way the APD realizes a photomultiplier based on semiconductor devices. If a very large bias voltage of several 100 V (in the reverse direction) is applied across the depletion region, then photoelectrons can be accelerated so strongly that they generate another electron–hole pair. Exactly as in the photomultiplier, a large amplification of the photoelectron can be achieved by a cascade of such ionization events. Therefore also the term “solid-state photomultiplier” is used occasionally.

The gain of APDs reaches 250 or more. Like in the usual pin Si photodiode, the photoelectrons are released in the depletion region with a correspondingly high QE. Therefore the responsivity of APDs can exceed  $100 \text{ A W}^{-1}$ .

For high light intensities, APDs are operated like PMTs in current mode. The gain, however, is sufficient to operate them in the Geiger mode for photon counting also. With ionization not only electrons but also holes are generated. If both charge carriers are generated with the same efficiency, then the detector is “ignited” by a first charge carrier pair and does not lose its conductivity since new electron–hole pairs are generated continuously.

In silicon the ionization coefficient for electrons is very much larger than that for holes. The current flow, however, cannot be stopped until all holes have left the depletion layer and only then can a new charge pulse be generated. In order to keep the resulting dead time as short as possible, the discharge can be passively quenched by a current-limiting resistor. Better conditions can be provided by interrupting the discharge current actively through suitable servo loops.

## 11.7 Position and Image Sensors



**Figure 11.10** Quadrant detectors for the localization of a laser beam.

The application of the highly integrated concepts of semiconductor technology to photodetectors, not only to Si but also to other materials, is quite obvious. Typically four photodiodes are combined on a Si substrate with relatively large area, forming a “quadrant detector” (Figure 11.10). This serves, for example, to determine the position of a light beam. By means of difference amplifiers, the detection of slight motions is possible with remarkable sensitivity. In another layout, photodiodes are used line-wise or column-wise with “diode arrays” in order, for example, to measure

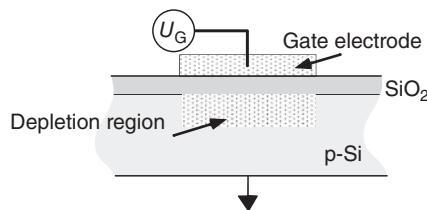
simultaneously the spectrum of a monochromator (see Figure 6.7) without mechanically actuating a grating. In a line camera, a movable mirror provides the line feed and thus allows recording of a full two-dimensional picture.

A two-dimensional array of photo-capacitors without movable parts can be applied for image formation. In such an array, the intensity distribution of a real image is stored as a two-dimensional charge distribution. The technical challenge is to “read out” the information saved in the capacitor charges on demand using electronic devices and at the same time to convert it into a time sequence of electrical impulses that are compatible with conventional video standards. For this purpose the concept of charge-coupled device (CCD) sensors developed in the 1970s based on metal–oxide–semiconductor (MOS) capacitors has gained wide acceptance, since such a sensor exhibits particularly low noise. Only in the infrared spectral range, when the sensors have to be cooled and the MOS capacitance decreases, do conventional pn capacitances equipped with MOS switches have advantages.

### 11.7.1 Photo-Capacitors

The charge generated by illumination in a common pn photodiode in the photovoltaic operating mode and with an open circuit does not drain but is stored in the capacitance of the space charge region. It operates as a potential well for the electrons released nearby, and we can call it a “photo-capacitor.” Such devices are of particular interest for image sensors since the image information can be first saved in the photo-capacitances and then be read out serially. Of course the charge will drain eventually by thermal motion, but the storage time is from several seconds up to minutes or hours depending on the system and temperature.

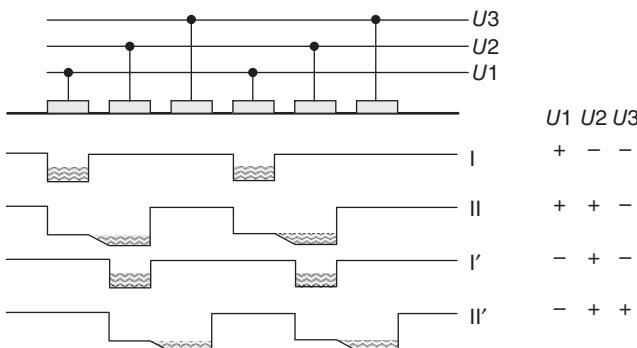
The MOS capacitors have proven themselves as photo-capacitors. At the MOS interface, which is also known as a Schottky contact, a potential is generated that serves to store photoelectrons. With MOS capacitors, large capacitance values are achieved. They prevent the stored charges from reducing the potential well, and thus the capacitor does not saturate with just a few photoelectrons or holes. A model of a MOS capacitor consisting of a metallic or polycrystalline Si gate, an  $\text{SiO}_2$  oxide layer, and p-Si is presented in Figure 11.11. For positive gate voltage  $U_G$ , a potential well for electrons is formed. Electrons ejected in the space charge region and stored in the potential well can later be released by decreasing the gate voltage. The storage time of photo-capacitors is limited by thermal relaxation and varies at room temperature from seconds up to several minutes.



**Figure 11.11** MOS photo-capacitor.  
Electrons optically generated are stored in the depletion region.

### 11.7.2 CCD Sensors

The heart of modern digital cameras is the CCD chip, which in its detector array generates a charge proportional to the intensity of the incident radiation and



**Figure 11.12** Three-phase operation of a CCD cell.

stores it in photo-capacitors until it gets read out by control electronics [164]. In comparison with the photographic plate, the CCD camera has the advantages of a large linear range, high QE of 50–80%, and direct generation of a voltage signal that can be digitized and processed by computer.

The key to the success of CCD sensors is the readout method, presented in Figure 11.12 using the example of a three-phase system. It is organized such that, by voltage control sequences of the gate electrodes, the charge stored in a sensor or *pixel* is transferred to the adjacent capacitor. The clock frequency of this periodic sequence can be more than 20 MHz. The average of the charge loss during the transmission is below  $10^{-6}$ . Therefore, even for many hundreds of transmission steps, generally more than 99.99% of the charge content of a pixel arrives at the readout amplifier.

An image sensor has to be read out line-wise. In order to prevent a long dead time being caused by this, and in addition to be able to accumulate more charges, the CCD sensors consist of an illuminated “image region” and a dark “storage region.” The formation of an image is finished by transferring all columns from the illuminated part in parallel and within 1 ms to the adjacent storage zone. While being transferred line-wise through a readout register step by step to the readout amplifier, already the next picture can be taken in the illuminated part.

The sensitivity of a CCD sensor is determined by the noise properties of each pixel, which on the one hand depend on the fluctuation of the thermally generated electrons and on the other hand are mostly dominated by the so-called readout noise. This is added to the charge content of a pixel by the readout amplifier. Since this noise contribution occurs only once per readout process, it is often favorable to accumulate charges generated photoelectronically on the sensor for as long as possible. For this, though, only slow image sequences can be achieved. The noise properties of a CCD sensor are often given in the unit “electrons/pixel” indicating the r.m.s. width of the dark current amplitude distribution.

The spatial resolution of a CCD sensor is determined by the size of the pixels, whose edge length today is typically 1–25  $\mu\text{m}$ . The resolution of course cannot be better than the optical image system, that is, the camera lens. The determination of the positions of small objects is, however, sometimes possible with sub-pixel resolution. If the point spread function of the optical imaging system is known, it can be fitted to the distribution extending over several pixels. The center value can then be evaluated with sub-pixel resolution.

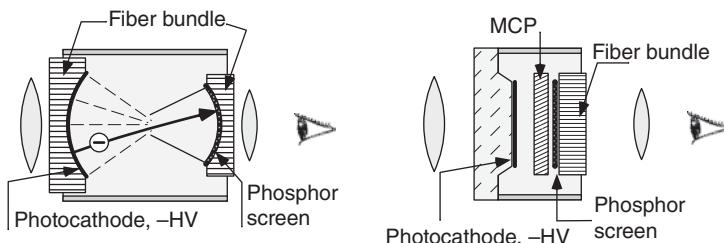


Figure 11.13 Concepts for image amplifiers of first and second generation.

### 11.7.3 Image Intensifiers

For image amplifiers, the extremely sensitive properties of a photomultiplier based on the conversion of light into electrons are used in detectors with spatial resolution. The potential for applications of image amplifier tubes and their variants is quite high since they allow more than just the taking of pictures of extremely faint objects. The concept can be transferred to many kinds of radiation, for example, infrared radiation or X-rays, which are not visible to the human eye and common cameras at all, but can cause localized ejection of electrons. Such devices are also called *image converters*.

In Figure 11.13 we present two widely used concepts for optical image intensifiers. On the left is the first-generation concept, in which a picture is guided through fiber optics to a photocathode. The electrons emitted there are accelerated by electro-optics and projected onto a luminescent screen. Its luminescence can be observed by eye or by camera. There can be up to  $150 \text{ lm} \text{ lm}^{-1}$  of image intensification.<sup>1</sup>

On the right, a model of the so-called second generation is shown. Here by means of a channel plate (MCP; see p. 370), a gain of  $10^4$  and more can be achieved. The spatial resolution of the incident optical image is slightly decreased by the spread of the electron bunches emitted from the MCP. The image of a single fluorescing atom in Figure 14.16 was taken with such an ICCD camera.

Image intensifiers not only allow the observation of very faint signals. The high voltage necessary at the channel plate for amplification can be switched on and off on a nanosecond scale, and this makes it possible to realize cameras with extremely high shutter speed.

## Problems

- 11.1 Thermal detectors** Consider the differential equation for the increase in temperature  $\Delta T$  of an ideally black detector with heat capacity  $K$  and total heat loss rate  $V$ . What determines the rise time  $\tau$ ? Derive the sensitivity  $\mathcal{R}$  of a thermopile with the gross Seebeck coefficient  $C_{TU}$ .

<sup>1</sup> Here the SI unit lumen (lm) is used, which measures the light current emitted by a point source with 1 candela (cd) light intensity into a solid angle of 1 steradian (sr):  $1 \text{ lm} = 1 \text{ cd sr}^{-1}$ . Light intensity is measured in the SI unit candela (cd). At 555 nm wavelength its value is  $1 \text{ cd} = (1/683) \text{ W sr}^{-1}$ ; at other wavelengths, it is referred to the spectrum of a blackbody radiator operated at the melting point of platinum.

The power density  $t^2$  of the spontaneous temperature fluctuations at frequency  $f$  and the absolute temperature  $T$  is  $t^2 = 4k_B T^2 V / (V^2 + (2\pi K f)^2)$ . Show that the noise equivalent power far below the maximal bandwidth, that is, for  $2\pi f \tau \ll 1$ , is  $\text{NEP} = T \sqrt{2k_B V}$ .

- 11.2 **Photocell** Design a simple electric circuit that transforms the change of conductivity in a photoconductor into a linear voltage change.
- 11.3 **Photovoltaic detector (I)** Study the  $I$ - $V$  characteristic curves of the photodiode (Figure 11.9). Identify the positions where (a) the photovoltaic mode, (b) short-circuit mode, and (c) the bias mode are located.
- 11.4 **Photovoltaic detector (II)** What laser power incident on a Si photodiode is required to operate the detector with shot-noise-limited conditions? The sensitivity is  $\mathcal{R} = 0.55 \text{ A W}^{-1}$  at 850 nm, the impedance of the photodiode is  $100 \text{ M}\Omega$ , and the dark current is  $I_D = 100 \text{ pA}$ . Take an amplifier with noise figures  $e_n = 10 \text{ nV Hz}^{-1/2}$  and  $i_n = 1 \text{ pA Hz}^{-1/2}$ .
- 11.5 **Photomultiplier** What is the minimal power detectable for a photomultiplier operated in current mode in a bandwidth of  $\Delta f = 1 \text{ Hz}$  with the following specifications: quantum efficiency  $\eta = 10\%$ , laser wavelength  $\nu_L = 600 \text{ nm}$ , and dark current  $I_D = 1 \text{ fA}$ . What is the rate of clicks indicating the arrival of photons?

## 12

# Laser Spectroscopy and Laser Cooling

In Chapter 7 on light and matter, we theoretically investigated the occupation number and the polarizability of an ensemble of atomic or other microscopic particles. Here we turn to the question of how these quantities are observed and used in experiments. They are not observed directly but through their effect on certain physical properties of a sample, for example, on the fluorescence of an excited sample or the absorption and dispersion of a probe beam. There are also numerous alternative methods of detection, for example, through the effect on acoustic or electrical properties of the sample.

For a wider overview over the extended field of laser spectroscopy, we refer the reader to [165], for example.

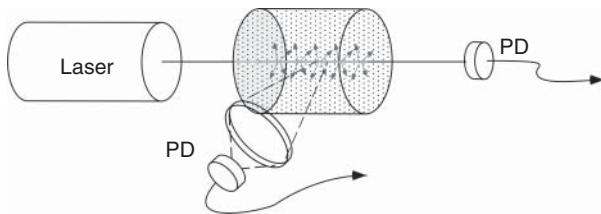
## 12.1 Laser-Induced Fluorescence (LIF)

Fluorescence is caused by spontaneous emission. We observe it, for example, when a laser beam passes through a gas cell (Figure 12.1). It corresponds to radiation damping and can only occur when an atom is in an excited state. In the Bloch equations (7.36) we have taken the fluorescence into account phenomenologically with the decay rates  $\gamma$  and  $\gamma'$ . A single particle in the excited state emits energy  $\hbar\omega$  during its lifetime  $\gamma$ , and an intense resonant laser beam keeps half of all particles excited on average. We can also express the saturated fluorescence power  $P_{\text{sat}}$  by the saturation intensity  $I_0$  according to Eq. (7.42):

$$P_{\text{sat}} = \hbar\omega \frac{\gamma}{2} = \frac{\gamma/2}{\gamma'} \sigma_Q I_0.$$

More generally, the intensity of the fluorescence to be observed is proportional to the excitation probability  $(w + 1)/2$ , where  $w$  is the  $z$ -component of the Bloch vector (see Section 7.2.3), and to the particle density  $N/V$ . In addition, we have to take the experimental setup (losses, solid angle of observation, etc.) into account with a geometry factor  $G$ . The observed fluorescence intensity may then be defined by  $I_{\text{fl}} = GP_{\text{sat}}$ :

$$I_{\text{fl}} = G \frac{N}{V} \hbar\omega \frac{1}{2} (1 + w) = G \frac{N}{V} P_{\text{sat}} \frac{s}{1 + s}.$$



**Figure 12.1** Laser spectroscopy. The spectral properties of a sample can be detected by laser-induced fluorescence (LIF) or by absorption. For the detection of dispersive properties, interferometric experiments are necessary.

The saturation parameter  $s$  is proportional to the intensity of the exciting laser field  $I$  according to Eq. (7.40). For large excitation field intensity ( $s \gg 1$ ), it can be found immediately that

$$I_{\text{fl}} \simeq G \frac{N}{V} P_{\text{sat}}.$$

In the limiting case of low excitation ( $s \ll 1$ ), laser-induced fluorescence (LIF) allows linear mapping of selected properties of a sample such as particle density, damping rates  $\gamma, \gamma'$ , and so on. The spectral dependence of the low-intensity resonance line at  $\omega_0$  is Lorentz shaped in the stationary case,

$$I_{\text{fl}}(\omega) \simeq G \frac{N}{V} P_{\text{sat}} s = G \frac{N}{V} \frac{\gamma \gamma'}{2} \frac{I}{(\omega - \omega_0)^2 + \gamma'^2},$$

and a fluorescence profile like the one shown in Figure 7.2 is obtained. With LIF, for example, spatially resolved density measurements of known atomic or molecular gases can be carried out.

## 12.2 Absorption and Dispersion

Like fluorescence, linear absorption and dispersion of the driving laser field occur at low saturation intensities only. Therefore we determine the more general absorption coefficient and the real part of the refractive index according to Eq. (7.23):

$$\begin{aligned} \alpha(\omega) &= -\frac{\omega}{2I(z)} \Im \{ \mathcal{E}(z) \mathcal{P}^*(z) \} = \frac{N}{V} \frac{\omega}{2I(z)} d_{\text{eg}} \mathcal{E}_0 v_{\text{st}}, \\ n'(\omega) - 1 &= \frac{c}{2I(z)} \Re \{ \mathcal{E}(z) \mathcal{P}^*(z) \} = \frac{N}{V} \frac{c}{2I(z)} d_{\text{eg}} \mathcal{E}_0 u_{\text{st}}. \end{aligned} \quad (12.1)$$

When we insert the dipole components ( $d = d_{\text{eg}}(u_{\text{st}} + i v_{\text{st}})$ ) from Eq. (7.44), we again obtain the relations

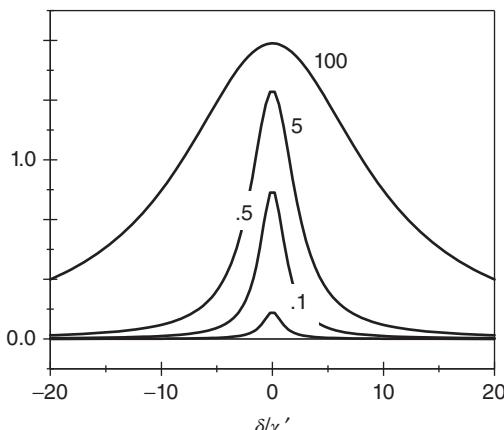
$$\begin{aligned} \alpha(\omega) &= -\frac{N}{V} \frac{\gamma}{2\gamma'} \frac{w_0 \sigma_Q}{1 + I/I_0 + [(\omega - \omega_0)/\gamma']^2}, \\ n'(\omega) - 1 &= -\frac{N}{V} \frac{\gamma}{2\gamma'} \frac{\lambda}{2\pi} \frac{w_0 \sigma_Q (\omega - \omega_0)/\gamma'}{1 + I/I_0 + [(\omega - \omega_0)/\gamma']^2}, \end{aligned} \quad (12.2)$$

by also accounting for Eqs. (7.21) and (7.41). From these relations the limiting case of small intensities can be reduced again to the classical case (7.20) without any further difficulties ( $I/I_0 \ll 1$  and  $\omega_0 = -1$ ). There, the absorption coefficient and refraction coefficient depend only on atomic properties (decay rates  $\gamma, \gamma'$ , detuning  $\delta = \omega - \omega_0$ , particle density  $N/V$ ) and not on the incident intensity. Conversely, these physical quantities can be determined using absorption spectroscopy. Since the determination of the refractive index generally requires an interferometric method, and thus considerable instrumental effort, the absorption measurement is the preferred method.

### 12.2.1 Saturated Absorption

For increasing intensity ( $I/I_0 \simeq 1$ ), the “saturation” of a resonance plays a more and more important role since the absorption coefficient becomes nonlinear: it itself depends on the intensity (Figure 12.2). For the sake of clarity, we introduce the resonant unsaturated absorption coefficient  $\alpha_0 = -\sigma_Q \omega_0 (N/V) \times (\gamma/2\gamma')$  (=  $\sigma_Q (N/V) \times (\gamma/2\gamma')$  for optical frequencies), and with the new linewidth  $\Delta\omega = 2\gamma_{\text{sat}}$  according to Eq. (12.4), we write

$$\begin{aligned}\alpha(\omega) &= \alpha_0 \frac{\gamma'^2}{(\omega - \omega_0)^2 + \gamma'^2(1 + I/I_0)} \\ &= \alpha_0 \frac{\gamma'^2}{(\omega - \omega_0)^2 + \gamma_{\text{sat}}^2}.\end{aligned}\quad (12.3)$$



**Figure 12.2** Saturation of resonance lines.

Normalized fluorescence intensity as a function of normalized detuning  $\delta/\gamma'$ . The parameter gives the incident laser power normalized to the saturation intensity,  $I/I_0$ . The maximum fluorescence intensity occurs at even occupation of the atomic levels.

So, in spite of the saturation at large intensity  $I \geq I_0$ , the Lorentz shape of the resonance line is preserved, though it becomes wider. It is also straightforward to show that on resonance ( $\omega = \omega_0$ ), the intensity no longer decreases exponentially following Beer's law but decreases linearly for large  $I/I_0$  according to

$$\frac{dI}{dz} = -\alpha(I)I \simeq -\alpha_0 I_0.$$

Here it is worth introducing another new parameter,

$$\gamma_{\text{sat}} = \gamma' \sqrt{1 + I/I_0}. \quad (12.4)$$

Nonlinear saturated absorption plays an interesting role in overcoming the Rayleigh–Abbe limit of resolution in microscopy into the nanometer domain; see Section 5.5.3.

### 12.3 The Width of Spectral Lines

The observation of fluorescence and absorption spectra is among the simplest and thus most common method of spectroscopy. Physical information is contained in the center frequency value of a line as well as in its shape and width. As a measure for the width (Figure 12.3), usually the full width at half maximum is used, that is, the full frequency width between the values for which the resonance line reaches the half maximum value.<sup>1</sup> For intensities below the saturation value,  $I/I_0 \ll 1$ , the transverse relaxation rate  $\gamma'$  can be inferred from Eqs. (7.40) and (12.2):

$$\text{FWHM} \quad \Delta\omega = 2\pi\Delta\nu = 2\gamma'.$$

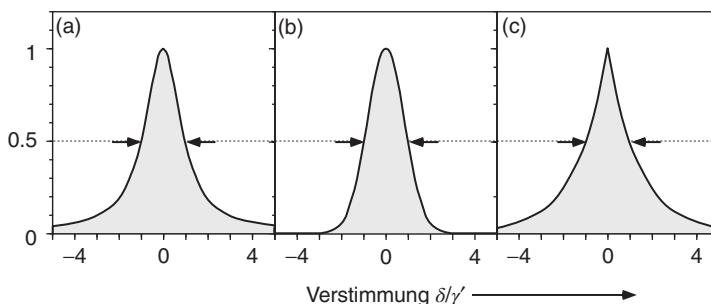
For a free atom, which can release its energy only by radiative decay, we have  $\gamma' = \gamma/2$  and thus

$$\Delta\omega = 2\pi\Delta\nu = \gamma. \quad (12.5)$$

In dilute atomic or molecular samples, the  $Q$  value or factor of the resonance, that is, the ratio between resonance frequency and FWHM, can easily assume very large values of  $10^6$  and more for optical frequencies of  $10^{14}$ – $10^{15}$  Hz:

$$Q = \nu/\Delta\nu.$$

It is obvious that, for decreasing linewidth  $\Delta\nu$  of a spectral line, the  $Q$  value and thus the “definition” of the center wavelength or frequency of a resonance line increase. The precise experimental preparation and measurement of such “sharp” resonances are a primary goal for spectroscopists. This goal requires a deep understanding of the physical mechanisms determining the position of a line, its width, and shape. Usually the natural linewidth caused by the spontaneous decay of excited states is considered to be the lower limit. It has been known, though, for a long time that this decay rate can be modified by the properties of the environment. For instance, the results of measurements are systematically influenced in the vicinity of conductive or reflecting walls (see Section 14.3.3) [166].



**Figure 12.3** Important spectral line shapes: (a) Lorentz line, (b) Gaussian profile, and (c) profile of time-of-flight broadening, presented for identical half-widths.

<sup>1</sup> Often the short forms FWHM and HWHM for full width and half-width at half maximum, respectively, are used.

Let us now present the most important limiting cases only; an extensive microscopic theory would go far beyond the scope of this chapter. Also the interaction of different broadening mechanisms is often complex, has to be described by mathematically elaborate convolutions, and so is neglected here.

### 12.3.1 Natural Width and Homogeneous Linewidth

The dream of the precision spectroscopist is a motionless particle in free space [167] whose resonance linewidth is limited only by the finite lifetime  $\tau$  of an excited state according to Eq. (12.5). It is called the “natural linewidth”  $\Delta\nu = \Delta\omega/2\pi = \gamma_{\text{nat}}/2\pi$  and is identical with the Einstein  $A$  coefficient of the spontaneous decay rate:

$$\gamma_{\text{nat}} = A_{\text{Einstein}} = \frac{1}{\tau}.$$

For an estimate of the natural width of typical atomic resonance lines, a characteristic dipole can be estimated with the Bohr radius,  $d_{\text{eg}} = er_{\text{eg}} = ea_0$ . For a red atomic resonance line ( $\lambda = 600 \text{ nm}$ ), we find from Eq. (7.49)

$$A_{\text{Einstein}} \simeq 10^8 \text{ s}^{-1}.$$

The resonance frequency of a free undisturbed particle is still shifted by the Doppler effect ( $\Delta\omega = kv = 2\pi v/\lambda$ ; see the following), which we discuss in the next section. However, for a long time, it has been possible to prepare almost motionless atoms and ions routinely in atom and ion traps using the method of *laser cooling*; see Section 12.5 [168]. Since the motion-induced frequency shift is caused only by the component of motion in the direction of the exciting or emitting light, the natural linewidth of an atomic or molecular resonance can be observed also with atomic beams.

The natural linewidth is identical for all particles of an ensemble. Such line broadening is called “homogeneous.”

### 12.3.2 Doppler Broadening and Inhomogeneous Linewidth

During the emission of a photon, not only the energy difference between the internal excitation states of the atom is carried away but also the momentum  $\hbar\mathbf{k}$ . For low velocities ( $v/c \ll 1$ ), we can take the difference between the resonance frequency in the laboratory frame ( $\omega_{\text{lab}}$ ) and in the rest frame ( $\omega_{\text{rest}} = (E - E')/\hbar$ ) from momentum and energy conservation:

$$\begin{aligned} m\mathbf{v}' + \hbar\mathbf{k} &= m\mathbf{v}, \\ E' + \frac{1}{2}mv'^2 + \hbar\omega_{\text{lab}} &= E + \frac{1}{2}mv^2. \end{aligned}$$

In nearly all cases the atomic momentum is much larger than the recoil experienced by the emission process,  $\hbar k/mv \ll 1$ . Thus we can neglect the term  $\hbar^2 k^2/2m$  in

$$\frac{m}{2}v^2 = \frac{m}{2}v'^2 + \frac{\hbar^2 k^2}{2m} + \hbar\mathbf{k}\mathbf{v} \simeq \frac{m}{2}v'^2 + \hbar\mathbf{k}\mathbf{v},$$

and arrive at the linear Doppler shift:

$$\omega_{\text{lab}} = \omega_{\text{rest}} + \mathbf{k}\mathbf{v}. \quad (12.6)$$

The direction within the laboratory frame ( $\mathbf{k}$ ) is determined either by the observer (in emission) or by the exciting laser beam (in absorption). The radiation frequency of a source appears to be blueshifted toward shorter wavelengths if it travels toward the observer and redshifted toward longer wavelengths if it moves away.

In a gas the molecular velocities are distributed according to the Maxwell–Boltzmann law. The probability  $f(v_z)$  of finding a particle at temperature  $T$  with velocity component  $v$  in an interval  $dv_z$  is

$$f_D(v_z) dv_z = \frac{1}{\sqrt{\pi} v_{\text{mp}}} e^{-(v_z/v_{\text{mp}})^2} dv_z, \quad (12.7)$$

where  $\int_{-\infty}^{\infty} dv_z f_D(v_z) = 1$ . The most probable velocity  $v_{\text{mp}}$  is ( $k_B$  is Boltzmann constant, and  $T$  is absolute temperature)

$$v_{\text{mp}} = \sqrt{2k_B T/m}.$$

For common temperatures the velocities of the molecular parts of a gas generally are between 100 and 1000 m s<sup>-1</sup> so that typical shifts of  $kv/\omega = v/c \simeq 10^{-6}$ – $10^{-5}$  or some 100–1000 MHz are expected. The natural linewidth of atomic or molecular resonance transitions is in general much smaller and therefore masked by the Doppler shift. For this reason the methods of *Doppler-free spectroscopy* (Section 12.4) have been an important topic of research for many years.

If the emission of the particles is otherwise undisturbed, the spectral line shape and width of the absorption line of the gas can be obtained from the superposition of all contributing undisturbed absorption profiles according to Eq. (12.3):

$$\alpha_D(\omega) = \int_{-\infty}^{\infty} dv_z f_D(v_z) \alpha(\omega + kv_z).$$

If  $\alpha(\omega)$  has Lorentz shape, the line profile  $\alpha_D$  described by this mathematical convolution is called a *Gauss–Voigt profile*. At room temperature, in many gases, the decay rate  $\gamma$  of an optical transition is much smaller than the typical Doppler shift  $kv_{\text{mp}}$ . Then the distribution function  $f_D(v_z)$  virtually does not change in that range where  $\alpha(\omega + kv_z)$  differs significantly from zero. It can be replaced by its value at  $v_z = (\omega - \omega_0)/k$  and pulled out of the integral. The integration over the remaining Lorentz profile results in a constant factor,

$$\alpha_D(\omega) = \alpha_0 f_D\left(\frac{\omega - \omega_0}{k}\right) \frac{\pi\gamma'}{k\sqrt{1+I/I_0}},$$

and with  $\sqrt{\pi \ln 2} = 2.18$ , we arrive eventually at the Gaussian profile

$$\alpha_D(\omega) = \frac{2.18\alpha_0}{\sqrt{1+I/I_0}} \frac{\gamma'}{\Delta\omega_D} \exp\left[-\ln 2\left(\frac{\omega - \omega_0}{\Delta\omega_D/2}\right)\right]. \quad (12.8)$$

Here we have already introduced the Doppler FWHM or Doppler width

$$\Delta\omega_D = \omega_0 \sqrt{\frac{8k_B T \ln 2}{mc^2}}.$$

The absorption coefficient is reduced by approximately the factor  $\gamma'/\Delta\omega_D$  since the line intensity is now spread over a very much bigger spectral range. It is useful to express the Doppler width in units of the dimensionless atomic mass number  $M$  and the absolute temperature  $T$  in *kelvin*,

$$\Delta\nu_D = \Delta\omega_D/2\pi = 7.16 \times 10^{-7} \sqrt{T/M} \nu_{\text{rest}}, \quad (12.9)$$

where  $\nu_{\text{rest}}$  is the resonance frequency of the particle at rest.

Doppler broadening is an example of an “inhomogeneous” linewidth. In contrast to the homogeneous line, each particle contributes to the absorption of the fluorescence line with a different spectrum depending on its velocity.

### 12.3.3 Pressure Broadening

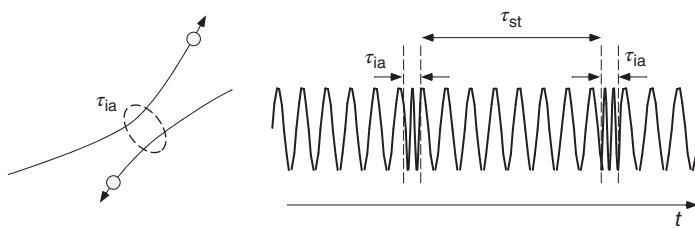
In a gas mixture atoms and molecules continuously experience collisions with neighboring particles that disturb the motion of the orbital electrons for a short time. During the collision the frequency of the emission is slightly changed compared to the undisturbed case. For neutral atoms or molecules, the interaction can be described, for example, by a van der Waals interaction, causing a mutual polarization of the collision partners. In a plasma the interaction of the charged particles is much stronger.

It is useful to consider first the relevant time scales that determine collisional processes and are compiled in Table 12.1. The interaction between neutral particles is generally *short range*, that is, it is significant only over a short distance of the order of the diameter of the atom or the molecule. The duration of the *interaction time*  $\tau_{\text{ia}}$  can therefore be estimated from the typical transit time across an atomic diameter. For thermal velocities, according to this, some 10–1000 oscillation cycles occur during the collision. The mean time interval between collisions (or the inverse collision rate)  $\tau_{\text{coll}}$  can be determined from the collision cross section  $\sigma_A$  and the mean velocity  $v$  following the known formula  $\tau_{\text{coll}} = n\sigma_A v$ . It is much larger than the interaction time  $\tau_{\text{ia}}$  even under atmospheric conditions, and thus electronic motion is rarely disturbed by the collisions. In a simple model

**Table 12.1** Relevant times for collisional broadening.

Process	Formula <sup>a)</sup>	Conditions	Duration
Optical cycle	$\tau_{\text{opt}} = 1/\nu_{\text{opt}}$		$10^{-14}\text{--}10^{-15} \text{ s}$
Interaction time	$\tau_{\text{ia}} = d_{\text{atom}}/v_{\text{therm}}$	$T = 300 \text{ K}$	$10^{-12}\text{--}10^{-13} \text{ s}$
Time between collisions	$\tau_{\text{coll}} = n\sigma_A v_{\text{therm}}$	$T = 300 \text{ K},$ $n = 10^{19} \text{ cm}^{-3}$	$10^{-7}\text{--}10^{-9} \text{ s}$
Natural lifetime	$\tau = A_{\text{Einstein}}^{-1}$		$10^{-8} \text{ s}$

a)  $d_{\text{atom}} = 2 \text{ \AA}$ ,  $\sigma_A = \pi d_{\text{atom}}^2/4$ .



**Figure 12.4** Disturbance of radiative processes by collisions in a neutral gas. The duration of the collisions is very short compared to the collision rate ( $\tau_{\text{coll}}^{-1}$ ) and to the lifetime of the excited state. The influence of collisions can be modeled through random phase jumps of an otherwise undisturbed wave.

all details of the molecular interaction are therefore negligible, and the effect of the collision can be reduced to an effective random phase shift of the otherwise undisturbed optical oscillation (Figure 12.4).

Let us first consider the intensity spectrum  $\delta I(\omega)$  of a damped harmonic wave train that starts at  $t_0$  and is simply aborted after a randomly chosen time  $\tau$ :

$$\delta I = I_0 \left| \int_t^{t_0 + \tau} e^{[-i(\omega_0 - \omega) - \gamma']t} dt \right|^2 = I_0 e^{-2\gamma' t_0} \left| \frac{e^{[i(\omega_0 - \omega) - \gamma']\tau} - 1}{i(\omega_0 - \omega) - \gamma'} \right|^2.$$

The dependence on the start time  $t_0$  can be eliminated immediately by integration,  $I(\omega) = 2\gamma' \int \delta I(\omega, t_0) dt_0$ . The phase jumps (and thus the periods of the undisturbed radiation times) are distributed randomly and occur with a mean rate  $\gamma_{\text{coll}} = \tau_{\text{coll}}^{-1}$ . Then we can calculate the shape of the collision-broadened spectral line with the probability distribution  $p(\tau) = e^{-\tau/\tau_{\text{coll}}} / \tau_{\text{coll}}$ :

$$I(\omega) = I_0 \int_0^\infty \left| \frac{e^{[i(\omega_0 - \omega) - \gamma']\tau} - 1}{i(\omega_0 - \omega) - \gamma'} \right|^2 \frac{e^{-\tau/\tau_{\text{coll}}}}{\tau_{\text{coll}}} d\tau.$$

The result is

$$I(\omega) = \frac{I_0}{\pi} \frac{\gamma' + \gamma_{\text{coll}}}{(\omega_0 - \omega)^2 + (\gamma' + \gamma_{\text{coll}})^2}.$$

The Lorentzian line shape is maintained; the effective collisional broadening rate  $\gamma_{\text{coll}}$  though has to be added to the transverse relaxation rate  $\gamma'$ . Since all particles of an ensemble are subject to the same distribution of collisions, this line broadening is homogeneous like the natural line shape.

Spectral lines are affected not only by pressure broadening but also by a pressure shift of the center of mass of a line. With increasing pressure, the number of collisions between the particles of a gas increases. Naively we can imagine that the volume available for the binding orbital electrons is reduced, and in quantum mechanical systems, volume reduction is always associated with an increase of the binding energy. The pressure shift therefore generally causes a shift to blue frequencies.

#### 12.3.4 Time-of-Flight (TOF) Broadening

Light-matter interaction of atoms and molecules in a gas or in an atomic beam is mostly restricted to a finite period. For example, for  $v = 500 \text{ m s}^{-1}$ , an atom

needs  $\tau_{tr} = 2 \mu\text{s}$  to pass a beam of diameter  $d = 1 \text{ mm}$ . However, the relaxation of many optical transitions occurs on the nanosecond scale, during which time an atom travels a few micrometers at most. The stationary solutions for (7.36) are a good approximation in these cases. However, in focused laser beams or for slowly decaying transitions, equilibrium is never reached, and the line shape is dominated by transient interaction corresponding to the finite time of flight  $\tau_{tr}$ . Slow or long-lived transitions are of particular interest since the corresponding very sharp resonance lines are excellent objects for precision measurements at low intensities. The two-photon spectroscopy of the hydrogen atom (see the example on p. 392) is an exceptionally beautiful example of this.

Slow transitions (with small transition dipole moment) that are only briefly subjected to a weak light field have  $\Omega_R < \gamma', \tau_{tr}^{-1}$ , and we can assume that the population of the ground state is virtually unchanged ( $w(t) \approx w(t=0) = -1$ ). Let us consider atoms or molecules crossing a laser beam that is assumed to have Gaussian envelope with  $1/e^2$  radius  $w_0$  (Figure 12.5). The Rabi frequency  $\Omega_R(z) = (d_{eg}\mathcal{E}_0/\hbar) \exp[-(z/w_0)^2]/\sqrt{\pi}$  is now a function of position, and we have to solve the first optical Bloch equation of (7.38):

$$\frac{d}{dt}\rho_{eg} = v \frac{d}{dz}\rho_{eg} = -(\gamma' + i\delta)\rho_{eg} + i\Omega_R(z).$$

We calculate the mean absorption coefficient of a single dipole with velocity  $v$  from Eq. (7.23):

$$\langle \alpha(v) \rangle = \frac{\omega}{2I} \frac{1}{w_0} \int_{-\infty}^{\infty} dz \operatorname{Im} \{ d_{eg} \mathcal{E}(z) \rho_{eg}(z, v) \}. \quad (12.10)$$

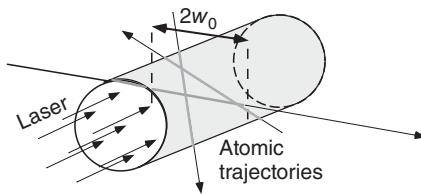
Before the particle enters the light field, there is no dipole moment. Thus we have  $\rho_{eg}(z = -\infty) = 0$ , and the general solution can be given by

$$\rho_{eg}(z, v) = i \frac{d_{eg}\mathcal{E}_0}{\hbar} e^{-(\gamma' + i\delta)z/v} \int_{-\infty}^z \frac{dz'}{v} e^{(\gamma' + i\delta)z'/v} e^{-(z'/w_0)^2}.$$

If the typical time of flight is small compared to the typical decay time,  $\gamma' \ll \tau_{tr}^{-1}$ , we can neglect  $\gamma'$ . By inserting  $\rho_{eg}(z, v)$ , then the integral in Eq. (12.10) can be evaluated analytically:

$$\langle \alpha(v, \delta) \rangle = \frac{\omega}{2I} \frac{|d_{eg}\mathcal{E}_0|^2}{\hbar} \frac{w_0}{v} e^{-(\delta w_0/2v)^2}.$$

In order to determine the absorption coefficient of a gaseous sample with a cylindrical laser beam, we would still have to average over all possible trajectories, but this results only in a modified or effective beam cross section whose details we skip over here.



**Figure 12.5** Atomic trajectories crossing a laser beam with Gaussian profile.

The summation of the velocity distribution in a two-dimensional gas ( $f(v) dv = (v/\bar{v}^2) \exp[-(v/\bar{v})^2] dv$ , as the velocity component along the direction of the laser beam does not play any role here) results in

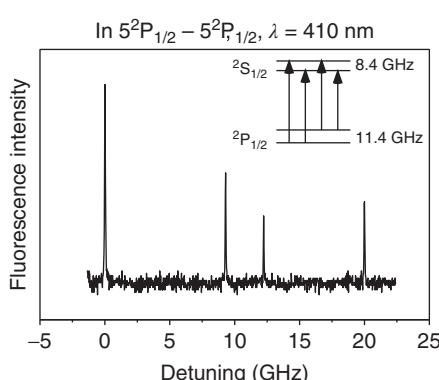
$$\alpha(\delta) = \int_0^\infty dv f(v) \langle \alpha(v) \rangle = \alpha_0 e^{-|\delta w_0/\bar{v}|} = \alpha_0 e^{-|\delta \tau_{\text{tr}}|},$$

whose form has already been presented in Figure 12.3c. The effective width of this line is determined by  $\tau_{\text{tr}} = w_0/\bar{v}$ .

## 12.4 Doppler-Free Spectroscopy

The linewidth of atomic and molecular resonances at room temperature is usually dominated by the Doppler effect. The intrinsic and physically attractive properties of an isolated particle are revealed only at velocity  $v = 0$ . Laser spectroscopy offers several nonlinear methods where light–matter interaction is effective for selected velocity classes only. The result is called “Doppler-free” spectroscopy.

### 12.4.1 Spectroscopy with Molecular Beams



**Figure 12.6** Fluorescence spectrum of an indium atomic beam obtained with a blue diode laser.

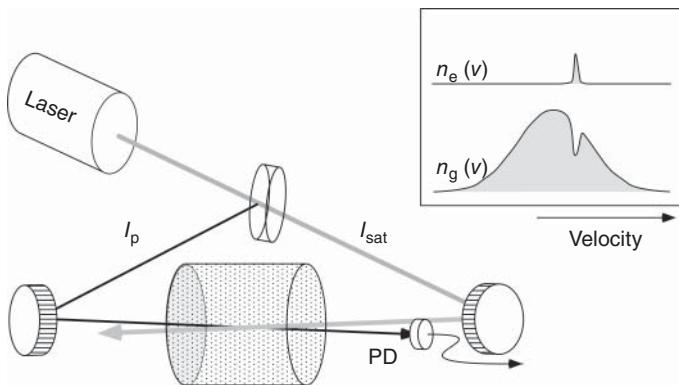
As soon as tunable lasers became available in the 1970s, high-resolution optical spectra were obtained with molecular or atomic beams. In such an apparatus, where the transverse velocities of molecules are reduced to near zero by geometric collimation, resolutions of  $\Delta v/v \simeq 10^8$  and better are routinely achieved.

The example in Figure 12.6 was recorded with an indium atomic beam. The transverse velocities were limited to  $v \leq 5 \text{ m s}^{-1}$  by appropriate apertures so that the residual Doppler effect  $kv \leq 10 \text{ MHz}$  was significantly smaller than the natural linewidth of 25 MHz.

### 12.4.2 Saturation Spectroscopy

By a resonant laser light field, atoms are promoted to the excited state, and as a result the occupation number difference is modified. In an inhomogeneously broadened spectral line profile such as the Doppler profile, then for not too large intensities a spectral hole is “burnt” into the velocity distribution, which is qualitatively presented in Figure 12.7.

In a gaseous sample the laser absorption spectrum reflects the velocity distribution of the atom, as described earlier. Atoms resonant with the laser are



**Figure 12.7** Principle of saturation spectroscopy. Inset: a laser beam with frequency  $\omega$  is used to “burn” a spectral hole into the ground-state velocity distribution at  $kv = \omega_0 - \omega$ ; simultaneously, an excited-state population with a narrow velocity distribution is generated.

excited and thus change the absorption properties of this velocity class. This modification can be probed with a second auxiliary laser. In the simplest arrangement, the absorption spectrum of one of two counter-propagating laser beams is measured.

Figure 12.7 shows the basic arrangement for the so-called saturation spectroscopy. In order to simplify the theoretical description, we assume that the intensities of saturation ( $I_{\text{sat}}$ ) and probe beam ( $I_p$ ) are small in comparison with the saturation intensity (Eq. (7.42)),  $I_{\text{sat,p}}/I_0 \ll 1$ , and do not directly influence each other. Let us calculate the absorption coefficient according to Eq. (12.2) by again using the Maxwell–Gauss velocity distribution  $f_D(v)$  from Eq. (12.7) and carrying out the Doppler integration:

$$\alpha_p(\delta) = \frac{\omega}{2I} \int_{-\infty}^{\infty} dv f_D(v) d_{eg} \mathcal{E} v_{st}^+(\delta, v).$$

We now distinguish the forward (“+”) and the backward (“−”) traveling laser beams, and from Eq. (7.44) we use

$$v_{st}^+(\delta, v) = -\gamma' d_{eg} \mathcal{E} w_{st}^- / \{1 + [(\delta - kv)/\gamma']^2\},$$

but following (7.39) we insert  $w_{st}^- = -1/(1 + s^-) \simeq -(1 - s^-)$  in order to account for the modification of the occupation number by the second counter-propagating laser beam, which has the saturation parameter

$$s^- = (I_{\text{sat}}/I_0) / \{1 + [(\delta + kv)/\gamma']^2\}.$$

Since the Doppler profile varies only slowly compared to the narrow Lorentzian contributions of each velocity class, at the detuning  $\delta = \omega_0 - kv$ , we can again pull  $f_D(v)$  out of the integral:

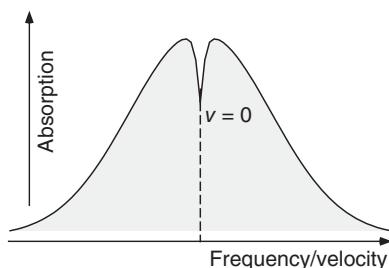
$$\alpha_p(\delta) = \alpha_0 f_D(\delta/k) \left( 1 - \frac{1}{\pi} \frac{I_{\text{sat}}}{I_0} \int_{-\infty}^{\infty} dv \frac{\gamma'^2}{\gamma'^2 + (kv - \delta)^2} \frac{\gamma'^2}{\gamma'^2 + (kv + \delta)^2} \right).$$

The evaluation of the integral [169] again results in a Lorentzian curve, which, due to our assumption of a very low saturation ( $s^\pm \ll 1$ ), has the natural linewidth  $2\gamma'$ :

$$\alpha_p(\delta) = \alpha_0 f_D(\delta/k) \left( 1 - \frac{I_{\text{sat}}}{I_0} \frac{\gamma'^2}{\gamma'^2 + \delta^2} \right).$$

The saturation resonance occurs exactly at the velocity class with  $v = 0$  (Figure 12.8). A more complete calculation shows the width corresponding to the width saturated by both laser fields according to Eq. (12.4) [169]:

$$\gamma_{\text{sat}} = \gamma' [1 + (I_{\text{sat}} + I_p)/I_0].$$



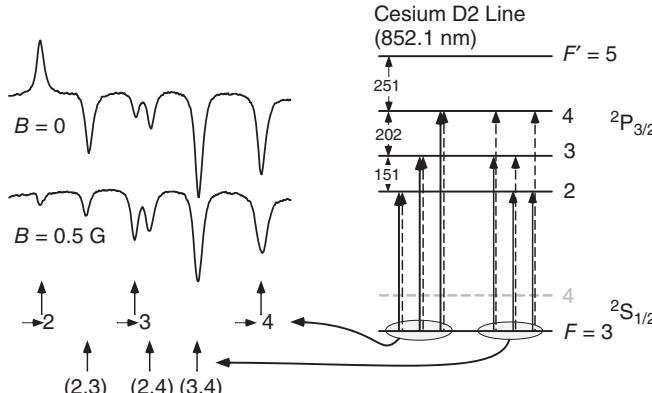
**Figure 12.8** Doppler profile with Lorentz-shaped saturation resonance. The Doppler-free line leads to an increased transparency.

vapor cell. Their vapor pressure at room temperature already leads to absorption lengths of only a few centimeters. In Figure 12.9 characteristic absorption lines are presented together with an energy diagram of the cesium D2 line at 852.1 nm.

From the  ${}^2S_{1/2}$ ,  $F = 3$  hyperfine state, three transitions with different frequencies to  ${}^2P_{3/2}$ ,  $F' = 2, 3, 4$  are available. From our simple analysis, we

The concept of saturation spectroscopy explains the occurrence of spectral holes in the Doppler profile (or in other inhomogeneously broadened spectral lines). In realistic experiments, though, it is influenced by further phenomena, for example, optical pumping or magneto-optical effects, all of which are collected a bit less precisely under the term “saturation spectroscopy” (Figure 12.8).

A simple experiment, though complex in its interpretation, can be carried out with diode lasers and a cesium or rubidium



**Figure 12.9** Saturation spectrum of a cesium vapor cell at the 852.1 nm D2 line. Here the  $F = 3 \rightarrow F = 2, 3, 4$  lines of the D2 line are presented. The second hyperfine transition from the ground state ( $F = 3 \rightarrow F = 3, 4, 5$ ) is at a distance of 9.2 GHz and cannot be seen here. The transition  $F = 3 \rightarrow F = 5$  is forbidden according to dipole selection rules ( $\Delta F = 0, \pm 1$ ). The separations of the hyperfine structure levels in the excited state are given in megahertz.

thus expect three line-shaped incursions in the absorption, but we observe six instead! And not only this. If the magnetic field is manipulated – in the upper spectra the geomagnetic field of 0.5 G is reduced to below 0.01 G by means of compensating coils – then even a reversal of selected lines can be observed. The reasons for this complex behavior are explained in detail in [170] and can only be sketched here.

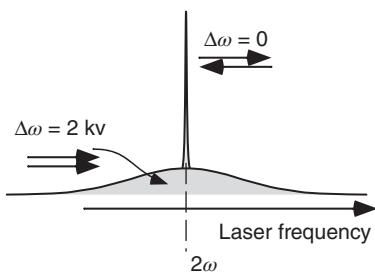
- 1) *Number of lines:* For velocities  $v \neq 0$  two different excited states can be coupled at the same time if the frequency differences are compensated by the Doppler effect. They also cause velocity-dependent population redistribution and lead to additional resonances called *crossover* lines. In Figure 12.9 three of those cases are presented, for example, at  $\omega = (\omega_{F=3 \rightarrow F=4} + \omega_{F=3 \rightarrow F=3})/2$ . Here they are particularly prominent since one of the laser fields can empty one of the two lower hyperfine levels ( $F = 3$ ) in favor of the other one ( $F = 4$ ) by optical “depopulation pumping,” which effectively removes these absorbers from the other light beam.
- 2) *Line reversal:* In simple laboratory setups, there is no care taken about compensating the geomagnetic field of 0.5 G. Then the lower form of the spectrum in Figure 12.9 is observed. The geomagnetic field, which does not have any well-defined direction relative to the laser polarization, is too small to split the lines visibly. But atoms are microscopic gyromagnets, and they can rapidly change their orientation by precession, and therefore all of them without exception can be excited by the light field. Effectively, the  $m$  quantum number is not a “good” quantum number in the geomagnetic field.

If this precession is suppressed, atoms can be trapped in “dark states” due to optical pumping and thus no longer participate in the absorption process and increase the transparency. But also the opposite effect occurs if they are pumped back to absorbing substates with the right choice of frequencies or polarizations (in Figure 12.9: orthogonal linear polarizations of pump and probe beam induce repopulation pumping) and by this increase the absorption. A detailed understanding here requires detailed knowledge of the level structure.

### 12.4.3 Two-Photon Spectroscopy

In the interaction of light and matter, usually electric dipole transitions are of interest because their relative strength dominates all other types. We understand these processes as absorption or emission of a photon, without actually defining the term “photon” [171] more precisely here. In Chapter 14 we discuss the “photonic” or quantum nature of light fields in more detail.

Besides the dipole interaction, higher multipole transitions or multiphoton processes occur. The latter are nonlinear in the intensities of the participating light fields. A simple and illustrative example is two-photon spectroscopy. For this in an atom or a molecule, a polarization  $P_{2\text{ph}} \propto E_1(\omega_1)E_2(\omega_2)$  is induced, causing absorption of radiation. Two-photon transitions follow different selection rules regarding the participating initial and final states – for example,  $\Delta\ell = 0, \pm 2$  has to be fulfilled for the angular momentum quantum number. Furthermore the calculation of the transition probabilities may raise problems



**Figure 12.10** Two-photon spectroscopy: Doppler background and Doppler-free resonance line.

where we expect *ad hoc* from second-order perturbation theory of quantum physics that matrix elements have to have the form [172]

$$M_{if} = \sum_s \left( \frac{\langle i|dE_1|s\rangle\langle s|dE_2|f\rangle}{E_i - E_s - \hbar\omega_1} + \frac{\langle i|dE_2|s\rangle\langle s|dE_1|f\rangle}{E_i - E_s - \hbar\omega_2} \right).$$

Transition rates are proportional to  $|M_{if}|^2$ , and the square value will also be proportional to the product  $I_1 I_2$  of both participating fields. A more detailed calculation shows that, as in the one-photon process, a Lorentz line with width  $2\gamma' = 2/T_2$  is obtained, which in the case of free atoms is identical with the natural linewidth. A simplified model for anharmonic oscillators as well conveys an impression about the origin of the two-photon absorption (Section 15.1).

Like in saturation spectroscopy, two-photon spectroscopy allows the nonlinear generation of signals at velocity  $v = 0$ . Here the absorption has to occur from two exactly counter-propagating laser beams with identical frequency since in that way the linear Doppler shift is just compensated (Figure 12.10):

$$\begin{aligned} (E_1 - E_2)/\hbar &= \omega_1 + \mathbf{k}\mathbf{v} + \omega_2 - \mathbf{k}\mathbf{v} \\ &= \omega_1 + \omega_2. \end{aligned}$$

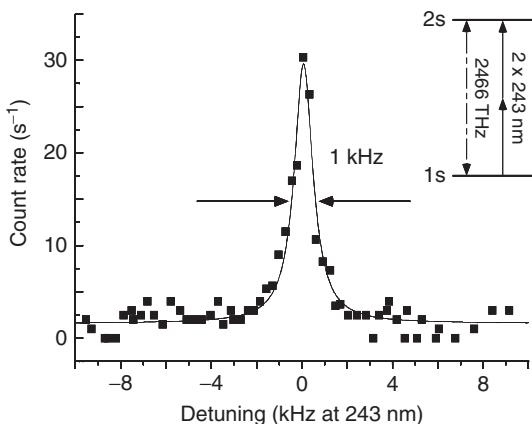
As a result Doppler-free spectra are obtained whose linewidths are limited by the natural lifetime or the time of flight for very long-lived states (see Section 12.3.4). In contrast to saturation spectroscopy, though, not only one selected velocity class at  $v = 0$  contributes to the signal with width  $\Delta v = \gamma/k$  but all velocity classes! The total strength of the Doppler-free resonance therefore is as large as that of the Doppler-broadened one and can be very easily separated from it (Figure 12.8).

### Example: The mother of all atoms: Two-photon spectroscopy of the hydrogen atom

The hydrogen atom is an atom of outstanding interest for spectroscopists. In contrast to all other systems, it is a two-body system and allows direct comparison with theoretical predictions, especially of quantum electrodynamics.<sup>2</sup> Its energy levels are principally determined just by the Rydberg constant, which as a result of two-photon spectroscopy is today the most exactly measured physical constant of all.

<sup>2</sup> This assertion though is challenged since at present the physical significance of the extremely precise measurement is limited by the relatively insufficient knowledge of the structure of the proton, which consists of several particles and is in fact not point-like as assumed by Dirac theory.

**Figure 12.11** Two-photon resonance of the 1s–2s transition of atomic hydrogen. See text. By permission of Hänsch and Walther [173].



The most interesting transition wavelength for precision measurements is the 1s–2s transition driven by  $\times 243$  nm. This wavelength can be generated experimentally in a much more convenient way than the 121.7 nm of the directly adjacent 1s–2p Lyman  $\alpha$  line. Furthermore, and in contrast to the adjacent 2p state (lifetime 0.1 ns), the decay rate of this metastable level is only about  $7\text{ s}^{-1}$  and promises a very unusually narrow linewidth of just 1 Hz!

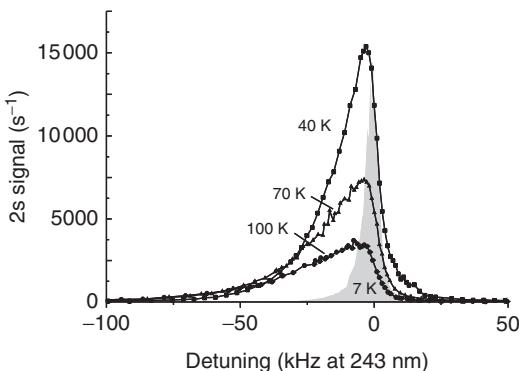
For many years T. W. Hänsch (born 1941, Nobel Prize 2005; see Section 9.5.7) and his coworkers have studied more and more exactly the 1s–2s transition of atomic hydrogen and are steadily approaching this ultimate goal of spectroscopy. At present their best published value is about  $\Delta\nu \simeq 1\text{ kHz}$  at 243 nm [173], that is, for a transition frequency of  $v_{1s2s} = 2466\text{ THz}$  already a  $Q$  value of more than  $10^{12}!$  (Figure 12.11). By a phase coherent comparison of the optical transition frequency with the time standard of the cesium atomic clock, the 1s–2s transition frequency has meanwhile become the best-known optical frequency of all (and thus wavelength as well; see p. 34) [174]:

$$f_{1s2s} = 2\ 466\ 061\ 413\ 187.103(46)\text{ kHz.}$$

During detection of the 1s–2s spectrum, another interesting spectroscopic effect occurs: the observed lines are asymmetric and slightly shifted to red frequencies with increasing velocity of the atoms. The reason for this is the Doppler effect of second order, which is not suppressed in two-photon spectroscopy. For the hydrogen atom it plays an important role due to its low mass and therefore high velocity. Only for very low velocities is the relatively symmetric signal of Figure 12.12 observed.

The line shift caused by the second-order Doppler effect is proportional to  $\Delta\nu_{2o} = \omega(v/c)^2/2$  and can be explained by the time dilation known from the special theory of relativity. In a moving atomic inertial reference frame, time seems to run more slowly than for an observer at rest in the laboratory frame.

In the experiment the observation has been made that the different line shapes of Figure 12.12 are a function of the temperature of the nozzle from which the hydrogen atoms are ejected into the evacuated spectrometer. They have a velocity distribution corresponding to the nozzle temperature, and they travel through



**Figure 12.12** Second-order Doppler effect of the two-photon spectrum of atomic hydrogen.

the exciting UV laser beam for a length of about 30 cm. The linewidth is determined by the time of flight.

## 12.5 Light Forces

When light–matter interaction of atoms and molecule is analyzed, usually the influence on the internal dynamics is in the foreground. In Section 2.5 we have studied optomechanical effects of light at the macroscopic level already. Absorption and emission of light also change the external mechanical state of motion of microscopic or atomic particles and are widely used to control their motion.

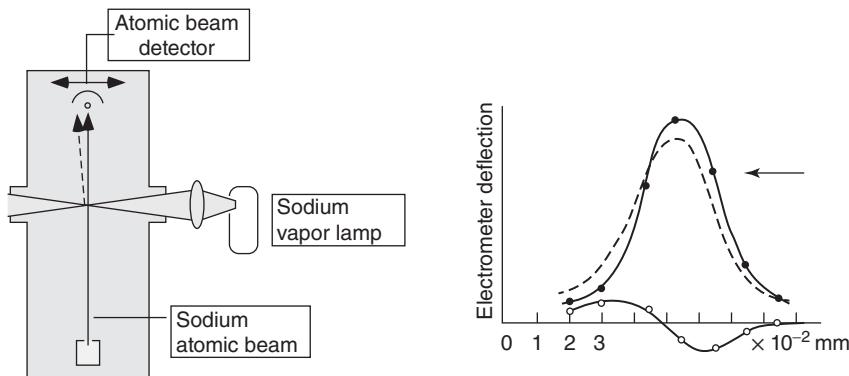
Photons have momentum  $\hbar\mathbf{k}$ , and during absorption and emission, this momentum has to be transferred to the absorber as a result of momentum conservation. For these processes we expect recoil effects, and the corresponding forces are called *light forces*<sup>3</sup> (Figure 12.13). Though the photon picture derived from quantum mechanics is very useful, light forces are known from classical light–matter interaction in an analogous way – for example, the Poynting vector describes the momentum density of the propagating electromagnetic field. So let us begin with a study of the mechanical effect of a planar electromagnetic wave on a classical Lorentz oscillator.

An inhomogeneous electric field exerts a force on a particle carrying a dipole, whether induced or permanent, which we may describe component-wise,  $\mathbf{d} = (d_x, d_y, d_z)$ . For an oscillating dipole, we furthermore have to average over an oscillation period  $T = 2\pi/\omega$  of the field,  $\langle F \rangle = T^{-1} \int_0^T F(t) dt$ :

$$F_i^{\text{el}} = \left\langle \sum_j d_j(t) \frac{\partial}{\partial X_j} E_i(t) \right\rangle \quad \text{or} \quad \mathbf{F}^{\text{el}} = \langle (\mathbf{d}(t) \cdot \nabla) \mathbf{E}(t) \rangle. \quad (12.11)$$

While this analysis seems straightforward, there *is* a problem. In a planar wave traveling in free space, the electromagnetic field is transverse, and thus in the linear Lorentz model, the induced dipole has to be transverse. The electric field of a planar wave, on the other hand, can change only in the propagation direction  $\mathbf{k}$ ;

<sup>3</sup> The reader is referred to [168] for more details on this topic.



**Figure 12.13** First observation by Frisch [175] of the deflection of an atomic beam by light forces. The data are taken from the original publication. The solid line shows the atomic beam profile without the light beam, and the dashed line that with the light beam. The difference is given in the lower part.

hence  $\mathbf{d} \perp \nabla$  and one should not expect any electrical force at all from Eq. (12.11). In a realistic light beam, though, with, for example, a Gaussian-shaped envelope, of course transverse electric dipole forces do occur, which we illustrate for the case of a standing-wave field in Section 12.5.4.

We must not forget, however, that there are also magnetic forces in general acting on neutral polarizable atoms. They are caused by the Lorentz force on the electric current in the atom, which is given by the time derivative of the dipole moment,

$$\mathbf{F}^{\text{mag}} = \langle \dot{\mathbf{d}} \times \mathbf{B} \rangle = \frac{1}{c} \langle \dot{\mathbf{d}} \times (\mathbf{e}_k \times \mathbf{E}) \rangle, \quad (12.12)$$

and these magnetic forces exert a net force on the entire atom.

### 12.5.1 Radiation Pressure in a Propagating Wave

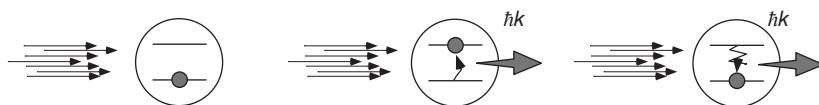
Let us calculate this latter force for a linear electronic Lorentz oscillator with eigenfrequency  $\omega_0$  subject to a planar transverse wave  $\mathbf{E} \perp \mathbf{k}$ . By using the complex polarizability  $\alpha = \alpha' + i\alpha''$  (see Section 7.1), we find

$$\dot{\mathbf{d}}(t) = -i\omega\alpha(\delta)\mathbf{E}(t) \quad \text{with} \quad \alpha(\delta) = \frac{q^2/2m\omega_0}{\delta - i\gamma/2},$$

and the average over an electromagnetic cycle is evaluated by means of the Poynting theorem (Section A.2), which picks out the imaginary or absorptive part of the polarizability. We finally arrive at

$$\mathbf{F}^{\text{mag}} = \mathbf{k}\alpha''(\delta)|\mathbf{E}|^2 = \mathbf{k}\alpha''(\delta)I/c\epsilon_0. \quad (12.13)$$

The force derived here predicts a light force that is parallel to the wave vector  $\mathbf{k}$  of propagation and the intensity  $I$  of the light field. It is called “radiation pressure” or “spontaneous force” since it depends on the absorption and spontaneous reemission of photons (Figure 12.14).



**Figure 12.14** Absorption–emission cycle and momentum transfer of the spontaneous force. During absorption, momentum is always transferred in the direction of the laser beam. The recoil of spontaneous emission is exerted in random directions. Thus on average, for many cycles, there is no momentum transferred on emission.

We expect that the classical treatment is a good approximation for low intensities ( $I/I_0 \ll 1$ , where  $I_0$  is the saturation intensity from Eq. (7.41)). For larger intensities, we have to treat the internal atomic dynamics according to the Bloch equations. We may seek a shortened path to the results of the semiclassical treatment by replacing the classical Lorentz oscillator in an *ad hoc* way by the Bloch oscillator through  $\mathbf{d} \cdot \mathbf{E} = \alpha \mathbf{E} \cdot \mathbf{E} \rightarrow (u + iv)\hbar\Omega_R$ . Using the normalized intensity  $s_0$  from Eq. (7.41), we now obtain

$$\mathbf{F}^{\text{mag}} = M\mathbf{a} = \hbar\mathbf{k} \frac{\gamma}{2} \frac{s_0}{1 + s_0 + (2\delta/\gamma)^2} \quad \text{with} \quad s_0 = I/I_0. \quad (12.14)$$

The force grows linearly with “photon momentum”  $\mathbf{k}$  and (for small  $s_0$ ) with the intensity  $I$  of the light field. It is proportional to the absorptive component with the characteristic Lorentz line shape. The force is also a result of spontaneous emission and proportional to the corresponding rate  $\gamma$ . The radiation pressure force is therefore interpreted as a *spontaneous force* as well. For large intensities ( $s \gg 1$ ), it saturates at the value  $\mathbf{F}^{\text{sp}} \rightarrow \hbar\gamma\mathbf{k}/2$  while exerting maximum acceleration

$$a_{\max} = \hbar k \gamma / 2M. \quad (12.15)$$

On average, a strongly driven atom is excited with a probability of 50% and can take on the momentum  $\hbar\mathbf{k}$  with each emission cycle. In Table 12.2 we have collected the relevant physical parameters for the mechanical effect of light on some important atoms. For their “cooling transitions” with wavelength  $\lambda$  and decay rate of spontaneous emission  $\gamma$ , we list the following:  $v_{\text{th}}$ , initial thermal velocity of the atomic beam;  $a/g$ , maximum acceleration  $a$  caused by radiation pressure (light

**Table 12.2** Overview of mechanical parameters for important atoms subject to light forces (see text for details).

Atom	$\lambda$ (nm)	$\gamma$ ( $10^6 \text{ s}^{-1}$ )	$v_{\text{th}}$ ( $\text{m s}^{-1}$ )	$a/g$	$\tau$ (ms)	$\ell$ (cm)	$N$
${}^1\text{H}$	121	600	3 000	$1.0 \times 10^8$	0.003	4.5	1 800
${}^7\text{Li}$	671	37	1 800	$1.6 \times 10^5$	1.2	112	22 000
${}^{23}\text{Na}$	589	60	900	$0.9 \times 10^5$	0.97	42	30 000
${}^{133}\text{Cs}$	852	31	320	$0.6 \times 10^4$	5.9	94	91 000
${}^{40}\text{Ca}$	423	220	800	$2.6 \times 10^5$	0.31	13	34 000

force), normalized to the gravitational acceleration  $g = 9.81 \text{ m s}^{-2}$ ;  $\tau$ , stopping or deceleration time for thermal atoms;  $\ell$ , stopping or deceleration distance for thermal atoms; and  $N$ , number of photons scattered during this stopping time.

### Excuse: Zeeman slowing

The spontaneous force is perfectly suited to slow down atoms from large thermal velocities (several  $100 \text{ m s}^{-1}$ ) to extremely low ones (some  $\text{mm s}^{-1}$  or  $\text{cm s}^{-1}$ ), provided we can exploit the maximum acceleration. In the laboratory system, however, the atomic resonance frequency  $\omega_0$  is shifted by the Doppler effect,  $\omega_{\text{lab}} = \omega_0 + \mathbf{k}\mathbf{v}$ , and an atom would lose resonance with the slowing laser – which is effective only within a natural linewidth – after only a few cycles.

This problem can be overcome by either tuning the laser synchronously with the slowing-down process ("chirp slowing") or by compensating the Doppler shift by means of a spatially variable magnetic field (Figure 12.15) in which the Zeeman effect<sup>4</sup> ( $\delta_{\text{Zee}} = \mu B / \hbar$ , with effective magnetic moment  $\mu$ , and typically  $\mu / \hbar = 2\pi \times 14 \text{ MHz mT}^{-1}$ ) compensates the change by the Doppler shift ("Zeeman slowing"):

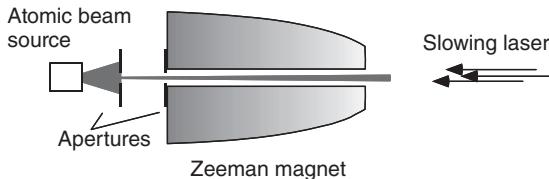
$$\delta = \omega_L - (\omega_0 + kv - \mu B / \hbar).$$

For Zeeman slowing along the atomic trajectory, a constant acceleration  $v = -a_{\text{sp}}t$  as large as possible is desirable. The compensation field is thus formed according to

$$B(z) = B_0 \sqrt{1 - z/z_0}.$$

The construction length  $z_0$  cannot be changed in general. It determines that only velocities with  $v \leq v_0 = (2a_{\text{sp}}z_0)^{1/2}$  can be slowed down, as found by a short calculation. Moreover, it also sets limits for the magnetic field strength,  $B_0 \leq \hbar v_0 / \mu$ .

In Figure 12.16 the effect of laser slowing on the initial thermal distribution is shown. At the end of the Zeeman slower, a narrow distribution is generated whose mean velocity is tunable by the laser frequency and the magnetic field. Its width is limited by the so-called Doppler temperature (see Eq. (12.19)). The Zeeman slower is well suited to prepare "cold" atomic beams with large intensities [176]. A "cold" atomic beam not only exhibits a low mean velocity but also has much smaller velocity spread than the initial thermal beam.

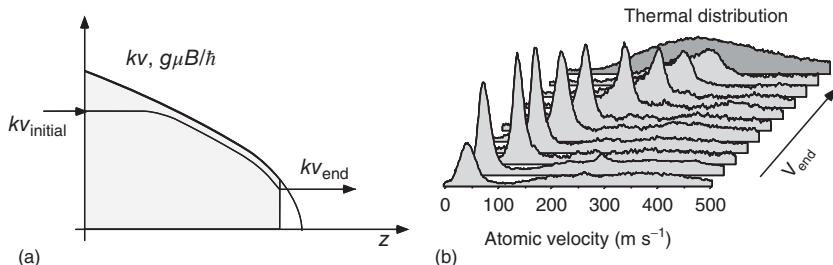


**Figure 12.15** A Zeeman slower for the deceleration of atomic beams. The magnetic field compensates the variation of the Doppler shift caused by deceleration.

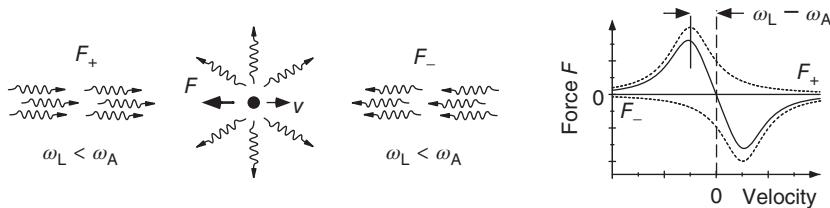
### 12.5.2 Damping Forces

Let us consider the effect of the spontaneous light force exerted by two counter-propagating laser beams with identical frequency. For this effectively one-dimensional situation, we assume that the forces according to Eq. (12.14)

<sup>4</sup> More precisely the Zeeman shift depends on the magnetic quantum number  $m$  and the Landé  $g$ -factors of the excited and ground states:  $\mu = \mu_B(m_g g_e - m_g g_g)$  (here  $\mu_B$  = Bohr magneton). Optical pumping with circularly polarized light leaves only the highest  $m$  value with  $mg \approx 1$  of significance.



**Figure 12.16** (a) Evolution of the velocity in the Zeeman slower. Velocity and magnetic field are given in terms of their corresponding Doppler and Zeeman shifts. (b) Experimental velocity profiles of an atomic beam at the exit of the Zeeman slower.



**Figure 12.17** Light forces for counter-propagating laser fields depending on the velocity or detuning.

simply add up, that is, we neglect interference effects:

$$F = F_+ + F_- = \frac{\hbar k \gamma}{2} \left( \frac{s_0}{1 + s_0 + (2\delta_+/ \gamma)^2} - \frac{s_0}{1 + s_0 + (2\delta_- / \gamma)^2} \right). \quad (12.16)$$

The Doppler shift  $\delta_\pm = \delta_0 \pm kv$  now depends on the direction of the light wave (Figure 12.17). For a red-detuned laser field ( $\delta_0 < 0$ ), the atomic Doppler-shifted resonance frequency lies always closer to the laser frequency of the counter-propagating beam. In this case the atom is always slowed down by radiation pressure – its motion is damped by a damping force. Very low velocities with  $kv/\delta_0 \ll 1$  are of special interest in order to estimate the ultimate limit for the reduction of velocities. We use an expansion of the force equation (Eq. (12.16)) in terms of the velocity  $v$  and find

$$\frac{dp}{dt} = F \simeq -\frac{8\hbar k^2 \delta_0}{\gamma} \frac{s_0}{[1 + s_0 + (2\delta_0 / \gamma)^2]^2} v = -\alpha m v. \quad (12.17)$$

For  $\delta_0 < 0$  we find a viscous damping force with coefficient  $\alpha$ . While the radiation pressure causes only retardation or acceleration, true laser cooling relies on such damping forces.

The one-dimensional concept of laser cooling can be extended to three dimensions by exposing an atom to counter-propagating laser beams in all directions of space. For this at least four tetrahedrally arranged laser beams have to be used. This situation corresponds to the strongly damped motion in a highly viscous liquid and is called “optical honey” or “optical molasses.”

### 12.5.3 Heating Forces, Doppler Limit

The spontaneous light force causes not only an acceleration in the propagation direction of the beam (which can be combined with cooling) but also a fluctuating force leading to heating of an atomic ensemble.<sup>5</sup> In a simple model we can consider the heating effect by the stochastic effect of the photon recoil  $\hbar k$  caused by spontaneous emission in analogy with the Brownian motion or diffusion of molecules. If  $N$  photons are randomly scattered, then for the average  $\overline{p}_N$  and the variance  $(\Delta^2 p)_N = \overline{p_N^2} - \overline{p_N}^2$  of the atomic momentum change by the isotropic emission, we have

$$\overline{p_N} = 0 \quad \text{and} \quad (\Delta^2 p)_N = \overline{p_N^2} = N\hbar^2 k^2.$$

The heating force or power can now be estimated from the scattering rate for photons,  $dN/dt = (\gamma/2)s_0/[1 + s_0 + (2\delta/\gamma)^2]$ , so

$$\left( \frac{d}{dt} \overline{p^2} \right)_{\text{heat}} = \frac{\hbar^2 k^2 \gamma}{2} \frac{s_0}{1 + s_0 + (2\delta/\gamma)^2} = 2D, \quad (12.18)$$

where the relation with the diffusion constant  $D$  is taken from the theory of Brownian motion. In equilibrium, we expect the heating and the cooling or damping power to exactly compensate each other:

$$(d \overline{p^2}/dt)_{\text{heat}} + (d \overline{p^2}/dt)_{\text{cool}} = 0.$$

For the cooling power we use relation (12.17)

$$\frac{d}{dt} p = -\alpha p \quad \text{and} \quad \left( \frac{d}{dt} p^2 \right)_{\text{cool}} = -2\alpha p^2,$$

and for the stationary state, we thus obtain

$$p^2 = D/\alpha = M k_B T.$$

It is associated with a characteristic temperature that is obtained by explicit insertion of  $D$  and  $\alpha$  from (12.17) and (12.18):

$$k_B T_{\text{Dopp}} = -\frac{\hbar\gamma}{2} \frac{1 + (2\delta/\gamma)^2}{4\delta/\gamma}. \quad (12.19)$$

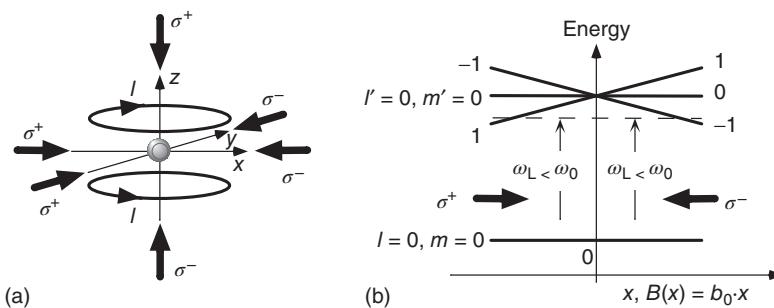
When this temperature reaches its lowest value at  $2\delta/\gamma = -1$  with  $k_B T_{\text{Dopp}} = \hbar\gamma/2$ , it is called the *Doppler temperature*.

The Doppler limit has played an important role for many years since it was considered the fundamental limit of laser cooling. It was therefore a big surprise when in experiments considerably lower, so-called sub-Doppler, temperatures were observed. Optical pumping processes are the origin of sub-Doppler laser cooling, and thus they occur only in atoms with a complex magnetic fine structure.

<sup>5</sup> This fact reflects the very fundamental law that dissipative processes (here, damping forces) are always associated with fluctuations and thus heating processes.

### Excuse: Magneto-optical trap (MOT)

In optical molasses, atomic gases are cooled down to the millikelvin (mK) range and lower by very efficient laser cooling. However, atoms cannot be stored in the intersection region of four or more laser beams only by radiation pressure, because they diffuse out of the overlap volume – the dissipative forces do not define a binding center. This problem has been solved by the invention of the *magneto-optical trap* (MOT), in which the radiation pressure is spatially modified by a quadrupole field.



**Figure 12.18** Magneto-optical trap. (a) Schematic setup with three pairs of counter-propagating laser beams. A set of two coils carries opposing currents, which generate a quadrupolar magnetic field with a zero at the center. (b) Spatial dependence of energy levels for a model  $J = 0$  to  $J = 1$  atomic transition. In the given direction, the magnetic field increases linearly from the center.

In one dimension, the magneto-optical trap (MOT) can be explained using the simplified example of an atom with a  $J = 0 \rightarrow J = 1$  transition (Figure 12.18b), which is exposed to a pair of counter-propagating light beams with circular polarization of opposite handedness in a linearly increasing magnetic field ( $\sigma^+ \sigma^-$  configuration). Away from the center and for red detuning ( $\omega_L < \omega_0$ ), a sufficiently slow atom is always much more in resonance with the laser beam whose radiation pressure is directed to the center of the quadrupole field. Thus the atom will experience a force directed toward this center.

In three dimensions, a spherical quadrupole field has to be used. It is generated by two parallel coils with currents flowing in opposite directions ("anti-Helmholtz coils"). The handedness of the circular polarizations has to be chosen in accordance with the magnetic field (Figure 12.18a). The simple one-dimensional concept has proven successful in three dimensions as well. The realization of the MOT with simple vapor cells has significantly contributed to its widespread use. The MOT is used in numerous laboratories in experiments with laser-cooled samples of neutral atoms. In the MOT, an equilibrium between loading rate (by capture of atoms from the slow part of the thermal distribution) and loss rate (by collisions with "hot" atoms) is built up, which typically contains some  $10^8$  atoms and has a volume of 0.1 mm diameter. The residual pressure of the cell must not be too high, because the loading of atoms into the MOT must not be interrupted by collisions with fast atoms during the capture process (which takes some milliseconds).

#### 12.5.4 Dipole Forces in a Standing Wave

Let us now evaluate the magnetic force (Eq. (12.12)) for the case of a standing-wave field generated from counter-propagating plane waves. In this standing wave the  $B$ -field is shifted by  $90^\circ$  with respect to the electric field,

$$\mathbf{E}(z) = 2\mathbf{E}(t) \cos(kz) \quad \text{and} \quad \mathbf{B}(z) = (2i/c)\mathbf{e}_k \times \mathbf{E}(t) \sin(kz),$$

and the time average picks out the real part in this case (compare Eq. (12.13)):

$$\mathbf{F}^{\text{mag}} = k\alpha'(\delta) \sin(2kz) |\mathbf{E}|^2.$$

This force is called the *dipole force* and can be derived from a potential

$$U_{\text{dip}} = \alpha'(\delta) I(z) / 2c\epsilon_0.$$

The interpretation is obvious as well: the force shows a dispersive frequency characteristic, that is, it changes sign with detuning from the resonance frequency. An interesting application of dipole forces in a standing wave is realized with “atom lithography” and is described in the excursion earlier.

In order to proceed to the semiclassical description, we may again use the trick from the previous section (see Eq. (12.14)), which yields here

$$U_{\text{dip}} = \frac{1}{2} \hbar \delta \ln(1+s).$$

Dipole forces derive from a conservative potential and thus should be disturbed by spontaneous events as little as possible. Therefore, in applications, a large detuning  $\delta \gg \gamma'$  is chosen, and correspondingly small saturation parameters  $s \simeq (I/I_0)/(\delta/\gamma')^2$  (Eq. (7.40)) are obtained, so that to a good approximation the dipole potential results in

$$U_{\text{dip}}(\mathbf{r}) \simeq \frac{I}{I_0} \frac{\hbar \gamma'^2}{2\delta}.$$

Dipole forces, though, only exist if the intensity of the electromagnetic field depends on position, for example, in the standing wave mentioned previously. Also a Gaussian beam profile makes an inhomogeneous light field and indeed provides an optical dipole trap for atoms and molecules [177], closely resembling the macroscopic optical tweezers of Section 12.5.6. Dipole forces always occur when coherent fields are superimposed. The details can be complicated because of the three-dimensional vector nature of the fields and can cause the appearance of “optical lattices” [178]. These are standing-wave fields with periodicity in one to three dimensions in which laser-cooled atoms move like in a crystal lattice.

#### Excuse: Atom lithography

By means of standing-wave fields, apparently strong forces can be exerted on the motion of atoms. Direct experimental proof is not very simple since the atomic trajectories are manipulated on a microscopically small scale. A transparent example for the application, though, is the so-called atom lithography, which is an example of atom nanofabrication

(Continued)

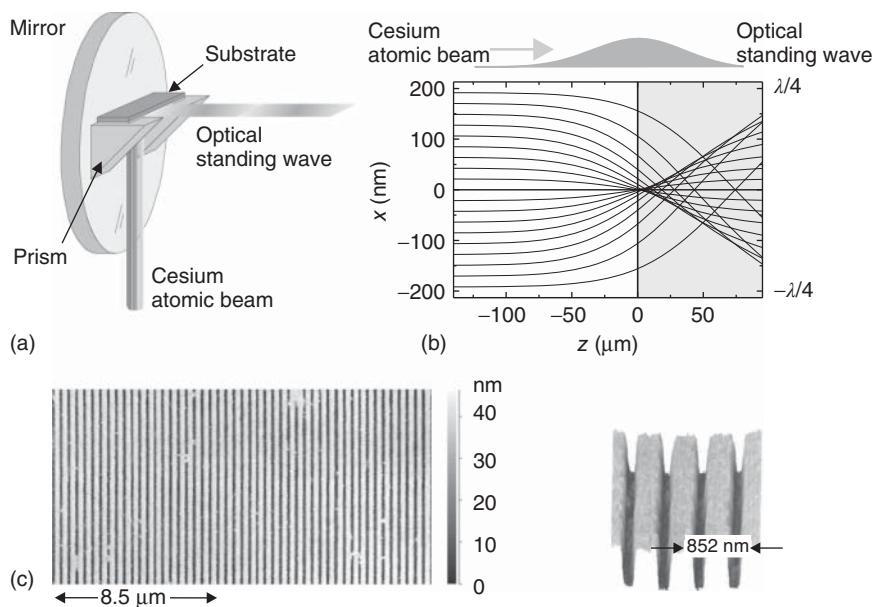


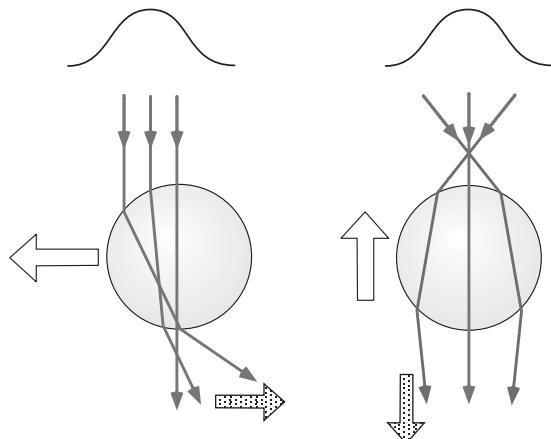
Figure 12.19 Atom lithography – see text for details.

#### Excuse: (Continued)

[179]. With this method the dipole forces of a standing wave serve to periodically modulate the intensity of an atomic beam when it is deposited onto a surface. The surface is physically or chemically modified only where the atoms hit. The experimental concept is presented in Figure 12.19 and is as outlined as follows:

- A substrate is exposed to an atomic beam that has just passed a standing wave generated by a mirror arranged behind the substrate.
- The simulation of atomic trajectories in a half-wave shows that the atoms are focused onto the surface in close analogy to an optical lens. Even spherical aberrations are visible. Here the standing wave creates an array of cylindrical lenses with width  $\lambda/2$  each. The periodic field of micro-lenses generates changes on the substrate by either growing layers (“direct deposition”) or causing chemical reactions (“neutral atom lithography”) with dimensions considerably below optical wavelengths. Therefore atom lithography is included in the class of methods allowing structuring at nanometer scales.
- An example is nanoscale grooves manufactured with this method. A Cs atomic beam is sent through a near-resonant standing-wave light field at  $\lambda = 852 \text{ nm}$ . The atoms modify the reactivity of a so-called self-assembled monolayer of organic thiol molecules covering the surface and corresponding to the resist in conventional optical lithography. In the final steps the chemical pattern is transformed into an array of grooves by wet etching processes.

**Figure 12.20** The effect of optical tweezers in terms of ray optics. White arrows: forces on the glass sphere. Dotted arrows: momentum change of the light rays. The transversal profile of a laser beam is indicated above the spheres.



### 12.5.5 Generalization

Let us briefly discuss the relation of electric and magnetic forces again that we started at the beginning of this section. We can also express the magnetic force (Eq. (12.12)) according to

$$\mathbf{F}^{\text{mag}} = \left\langle \frac{d}{dt}(\mathbf{d} \times \mathbf{B}) \right\rangle - \langle \mathbf{d} \times \dot{\mathbf{B}} \rangle.$$

The first term vanishes when averaged over a period. If the particle velocity is small,  $\dot{\mathbf{R}} \ll c$ , furthermore,  $d\mathbf{B}/dt \simeq d\mathbf{B}/dt = \nabla \times \mathbf{E}$  can be replaced, yielding

$$\mathbf{F}^{\text{mag}} = \langle -\mathbf{d} \times \nabla \times \mathbf{E} \rangle,$$

or component-wise

$$F_i^{\text{mag}} = \left\langle \sum_j d_j \left[ \frac{\partial}{\partial X_i} E_j - \frac{\partial}{\partial X_j} E_i \right] \right\rangle.$$

Comparison with Eq. (12.11) shows that the total force can generally be determined from

$$\mathbf{F} = \mathbf{F}^{\text{el}} + \mathbf{F}^{\text{mag}} = \left\langle \sum_j d_j \nabla E_j \right\rangle.$$

### 12.5.6 Optical Tweezers

In the last section we investigated the mechanical effect of light beams on microscopic particles such as atoms. Especially for dipole forces, we can give a macroscopic analogue that is used more and more widely, the so-called optical tweezers [180].

The dispersive properties of an atom are in fact similar in many ways to those of a transparent dielectric glass sphere for which we can describe the effect of macroscopic light forces qualitatively and in terms of ray optics.

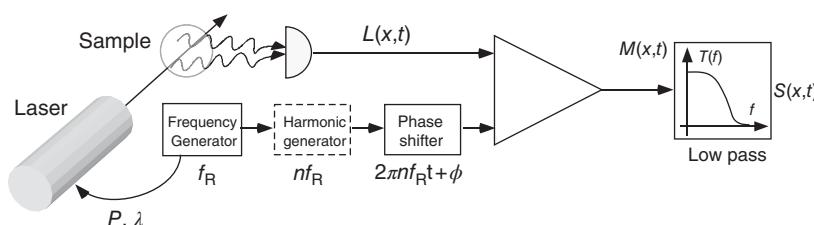
In Figure 12.20 the position of a glass sphere is either transversely shifted away from the axis of a Gaussian laser beam (left) or longitudinally shifted away from

the focus of a focused beam. Taking into account that the refraction of light beams causes transfer of momentum just as for atoms, we can infer from the directional changes of the beams that there is a mechanical force exerted on the glass sphere.

Optical tweezers are useful as nonmaterial micromanipulators in microscopy; for instance, bacteria in liquids can be caught, trapped, and moved.

## Problems

- 12.1 Lock-in amplifier** Consider the experimental setup from (Figure 12.21): a sample is irradiated with laser light. The frequency of the laser beam is modulated at  $\omega_L(t) = \omega_L^0 + \delta\omega_{\text{mod}} \cos 2\pi f_R t$ . The fluorescence of a Lorentzian resonance line is observed with a photodiode, yielding the spectral line shape  $L(x) = 1/(1+x^2)$  with  $x := (\omega_L - \omega_A)/\gamma$ . Here  $\omega_A$  marks the resonance frequency, and  $\gamma$  the resonance linewidth.
- The signal  $L(x, t)$  delivered by the photodiode is *mixed* with the reference signal,  $R(t) = R_0 \sin(2\pi f_R t + \phi)$ . Mixing is equivalent to electronic multiplication and generates an intermediate signal  $M(x, t)$ . Calculate  $M(x, t)$  and show that a low pass with cutoff frequency  $f_G \ll f_R$  generates an output signal  $S(x)$ , which is proportional to the derivative  $dL(x, t)/dx$  of the original line shape. How does the signal depend on the modulation amplitude  $\delta\omega_{\text{mod}}$ ?
  - Sometimes it is useful to replace the fundamental modulation frequency  $f_R$  by its multiples  $2f_R$  or  $3f_R$ . Calculate the output signals for this case. (c) What limits the bandwidth of the lock-in amplifier? What is the price to be paid for increased bandwidth?
- 12.2 Ramsey spectroscopy and spin echo** The *method of separated oscillating fields*, for which N. Ramsey (born 1915) was awarded the Nobel Prize in 1989, was originally invented for microwave spectroscopy of the hyperfine states of atoms traveling in an atomic beam. In that case, the undamped Bloch equations (7.31) can be used as a very good approximation. Today, Ramsey's method is in widespread use also in optical spectroscopy because for selected quantum states laser light transitions with Rabi frequencies exceeding the decay rates of these states are straightforward to obtain. The time sequence of operations is



**Figure 12.21** Schematic diagram of a lock-in amplifier. The phase-sensitive rectifier is the most important component.

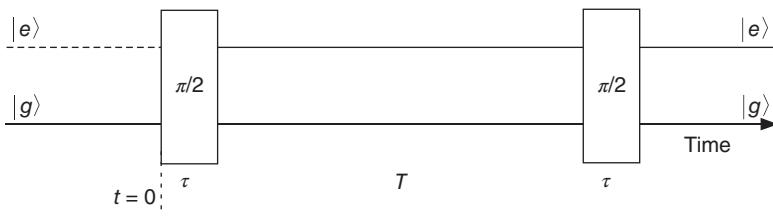


Figure 12.22 Sequential operations of a Ramsey experiment.

symbolically shown in Figure 12.22. Begin by considering the undamped Bloch equations. For an arbitrary state vector in a two-level system, we have  $|\psi\rangle = c_g|g\rangle + c_e|e\rangle$ , and the Bloch vector components are  $\mathbf{u} = \{u = c_e^*c_g + c_e c_g^*, v = -ic_e^*c_g - c_e c_g^*, w = |c_e|^2 + |c_g|^2\}$ .

- Show that a resonant microwave or light pulse (Rabi frequency  $\Omega_R$ , duration  $\tau$ , detuning  $\delta$ ) rotates the Bloch vector  $\mathbf{u}(\tau)$  about the  $u$  axis into the state  $\mathbf{u}(\tau) = \Theta(\tau)\mathbf{u}(0)$ . Give the rotation matrix explicitly. Take the initial value  $\mathbf{u} = (0, 0, -1)$ . Graphically sketch the probability  $P_g(\tau) = |c_g(\tau)|^2$  to find the atom after a pulse with duration  $\tau$  in state  $|g\rangle$ .
- Show that the free precession of the Bloch vector in the absence of any driving field describes a rotation about the  $w$  axis depending on detuning and give the rotation matrix. How does a Bloch vector  $\mathbf{u} = (0, 0, -1)$  evolve after excitation with a resonant  $\pi/2$  pulse (i.e., with  $\Omega\tau = \pi/2$ )?
- In Ramsey's method, two resonant  $\pi/2$  pulses (duration  $\tau$ , delay time  $T$ ) are applied onto the system. In the measurement, the population of the excited ( $P_e$ ) or the lower ( $P_g$ ) quantum state is recorded. Write a matrix formalism for the sequence of operations in the Ramsey method. Calculate and sketch the time evolution of the  $w$ -component of the Bloch vector for initial states  $\mathbf{u}(0) = (0, 0, -1)$  and for all atoms with identical interaction time. Show  $P_e(\tau)$  as a function of  $T$ . What is the origin of the interferences?
- In experiments, the Ramsey interferences or oscillations are frequently subject to a damping that can be interpreted as dephasing in an inhomogeneous ensemble, that is, an ensemble with a distribution  $p(\delta)$  of small detunings  $\delta$ . Consider a model where the detunings are characterized by Gaussian distribution,  $p(\delta) = \exp[-(\delta - \bar{\delta})^2/(2\sigma^2)]/\sqrt{\pi}\sigma$ . Determine the influence of this distribution onto the Ramsey signal.
- In a modified experiment, apply first a  $\pi$  pulse after time  $T$  and the second  $\pi/2$  pulse only with delay time  $2T$ . Study the distribution of populations around time  $2T$ . Show that the so-called *refocusing pulse* causes a revival of the population oscillations after time  $2T$ , also called *echo pulse*. Interpret the action of the  $\pi$  pulse. What is the phasing of the echo signal?

**12.3 Doppler effect and two-photon spectroscopy** In an atom or molecule traveling with velocity  $\mathbf{v}$ , an excitation is induced from an energy state  $E_a$  to  $E_b$  by absorption of a photon with wave vector  $\mathbf{k}$  and frequency  $\nu_{\text{lab}}$ . According to the special theory of relativity, the frequency of the moving atom in its rest frame is shifted in the frame of the laboratory by the Doppler effect,  $\nu_{\text{lab}} = \nu_{\text{rest}}(1 - (v/c)^2)^{1/2}/(1 - v_{\parallel}/c)$  with  $v_{\parallel} = \mathbf{k} \cdot \mathbf{v}/k$ .

(a) Expand the frequency shift in terms of  $v/c$ . (b) A two-photon transition may be induced by absorption of one photon from two counter-propagating laser beams each. Show that for this two-photon transition the first-order Doppler shift is suppressed. Furthermore show that the now dominating second-order Doppler effect causes a shift  $\nu_{\text{lab}} = (\nu_{\text{rest}}/2) \times (1 - (v/c)^2/2)$ . (c) Calculate the line shape in a thermal gas as a function of the temperature  $T$ . Neglect the natural line width and use the thermal distribution  $p(E) = 2\sqrt{E} \exp(-E/k_B T)/[(k_B T)^{3/2} \sqrt{\pi}]$  in a gas.

## 13

### Coherent Light–Matter Interaction

In Chapters 7 and 12 we were mostly concerned with the question of how the populations of quantum states of matter evolve when they interact with light fields. Spectroscopic application of intense and narrowband laser radiation opened the path to precision measurement, efficient detection of the physical properties of, for example, atmospheric gases in remote sensing, and more.

Laser spectroscopy, however, offers still more: the regime of the so-called coherent phenomena and coherent control. As outlined in Section 7.2.6 the damped optical Bloch equations (7.36) are essentially governed by two kinds of time constants, or rates, respectively: the Rabi frequency  $\Omega_R = d \cdot E/\hbar$  describes the (coherent) coupling of light and matter, and the relaxation rates  $\{\gamma, \gamma'\}$  take radiative damping phenomenologically into account. (In fact, quantum electrodynamics does not give new insight into the consequences of damping, but it provides a more rigorous justification of radiation damping.)

It is thus useful to distinguish two limiting cases of radiation–matter interaction in accordance with the ratio of Rabi frequency and damping rates. The case of *strong coupling* is of particular interest since it produces transient oscillatory phenomena that can only be realized with intense laser fields, not with classical light sources. Strong coupling is associated with superposition of quantum states and is hence the origin of many phenomena that do not exist in classical radiation–matter interaction.

#### 13.1 Weak Coupling and Strong Coupling

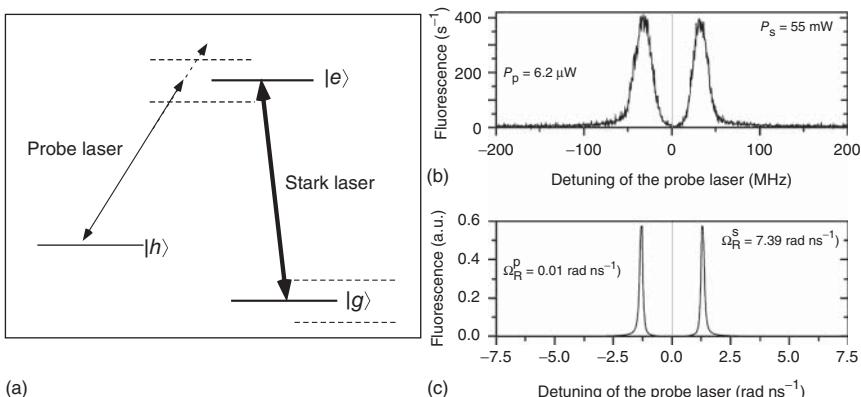
- *Weak coupling*  $\Omega_R \ll \gamma, \gamma'$ : Optical excitation of, for example, atoms or molecules with thermal light sources or with low-intensity laser beams is typically a case of weak coupling. In this case only a few particles are excited, and the steady-state values of the occupation probability are nearly unchanged. The  $w$ -component of the Bloch vector (see Section 7.2.3) remains near its initial value  $w = -1$  to a good approximation. Rate equations are usually sufficient to describe the dynamics of the system. The energy absorbed from the exciting light field is irreversibly transferred to the damping reservoir.

- *Strong coupling*  $\Omega_R \gg \gamma, \gamma'$ : Once the intensity of a driving laser field exceeds the saturation intensity,  $I/I_0 > 1$  (Eqs. (7.40) and (7.42)), the coupling of light field and polarizable matter is stronger than damping induced through coupling to the bath, usually the electromagnetic vacuum. For time scales that are short compared to the damping time  $1/\gamma$ , then transient phenomena occur, for instance, oscillations of the occupation numbers (see Figure 7.6), where the energy is periodically exchanged between the driving light field and the absorbing medium. In this case, the spectral lines are broadened (saturation broadening; see Section 12.2.1), or they show the AC Stark effect (see next section). Under strong coupling conditions, the driving field-matter system undergoes coherent evolution, that is, with fixed phase relation, for time scales  $t < \gamma^{-1}$ .

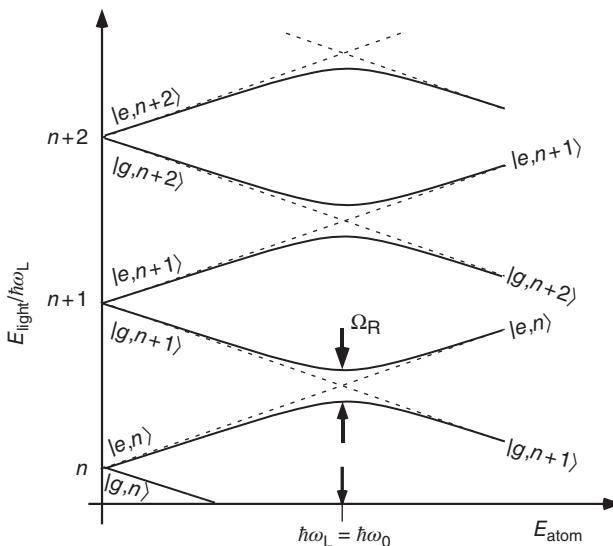
### 13.1.1 AC Stark Effect and Dressed-Atom Model

The spectrum of a resonance line that is excited by a very strong laser, that is, with multiple saturation intensity  $I/I_0 \gg 1$  (Eq. (7.42)), causes saturation broadening of the resonance line as discussed in Section 12.2.1. Let us modify the experiment by resonantly exciting the  $e-g$  transition with a strong laser field (Figure 13.1) and probing the system with an auxiliary transition  $h-e$  and a weak second laser. Instead of a single resonance line, the spectrum of the probe laser exhibits a doublet, which is called *Autler-Townes doublet* or *AC Stark splitting*.

The splitting of the atomic level can be understood with the so-called *dressed-atom* model [181]. We have used product states from atomic and field quantum states already in Section 14.3.2 in order to simplify the notation of Weisskopf-Wigner theory. Here we extend this method. We consider products of the atom quantum states  $\{|g\rangle, |e\rangle\}$  with frequency difference  $\omega_0 = (E_e - E_g)/\hbar$



**Figure 13.1** AC Stark or Autler-Townes splitting of a two-level system of the neon atom. The intensity of the Stark laser that is tuned into exact resonance,  $\omega_{ge} = \omega_{\text{Stark}}$ , is 10 times the saturation intensity. (a) Energy level diagram. (b) Experimental spectrum recorded with a weak probe laser ( $P_s, P_p$ : powers of Stark and probe laser, respectively). (c) Calculated spectrum ( $\Omega_R^S, \Omega_R^P$ : Rabi frequencies relevant for the Stark and the probe laser transition, respectively). (Adapted from [16].)



**Figure 13.2** Energy diagram of the *dressed-atom* model as a function of atomic excitation energy  $\hbar\omega_0$  and for different photon numbers  $n$ . Dashed lines: unperturbed states. Solid lines: dressed states with dipole interaction. (Adapted from [181].)

and a strong light field with near-resonant frequency  $\omega_L$  and number of quantum states  $|n\rangle$  with a large photon number  $n \gg 1$  as well. Near resonance ( $\omega_0 \simeq \omega_L$ ), the

$$\text{unperturbed dressed states } \{|g, n+1\rangle, |e, n\rangle\}$$

have nearly identical (degenerate) energy. The light field corresponds precisely to a single mode, which is, for instance, realized by a Gaussian beam traveling in free space. More rigorously we should use the so-called coherent states from Section 14.5.2.2, which are constructed from a superposition of field states with different photon numbers and provide a more realistic description of a strong laser field. The result is not significantly modified by this simplification, however.

In Figure 13.2 the energy values of the unperturbed dressed states are given with dashed lines as a function of atomic resonance frequency corresponding to a ground-state energy of  $-\hbar\omega_0/2$  and  $+\hbar\omega_0/2$  for the excited state. The energy expectation value of the unperturbed Hamiltonian  $\hat{H} = \hbar\omega_L \hat{a}^\dagger \hat{a} + \hbar\omega_0 \hat{\sigma}^\dagger \hat{\sigma}$  is then  $E_{g,n+1} = (n+1)\hbar\omega_L - \hbar\omega_0/2$  and  $E_{e,n} = n\hbar\omega_L + \hbar\omega_0/2$ , respectively. Introducing the detuning  $\delta = \omega_L - \omega_0$ , we can rewrite the energies  $E_{g,n+1} = (n+1/2)\hbar\omega_L - \hbar\delta/2$  and  $E_{e,n} = (n+1/2)\hbar\omega_L + \hbar\delta/2$ , respectively. At resonance  $\omega_L = \omega_0$  the unperturbed quantum states  $\{|g, n+1\rangle, |e, n\rangle\}$  are perfectly degenerate for any photon number  $n$ .

The degeneracy at the resonance  $\hbar\omega_L = \hbar\omega_0$  is lifted due to light-matter interaction. Formally this is described by the total Hamiltonian that includes the Jaynes-Cummings operator (14.9):

$$\hat{H} = \hbar\omega_L \hat{a}^\dagger \hat{a} + \hbar\omega_0 \hat{\sigma}^\dagger \hat{\sigma} + \hbar g (\hat{a}^\dagger \hat{\sigma} + \hat{\sigma}^\dagger \hat{a}).$$

The energies of the new eigenstates can be calculated from the  $\{|g, n+1\rangle, |e, n\rangle\}$  base states by conventional diagonalization. We use

$$\begin{pmatrix} H_{ee} & H_{eg} \\ H_{ge} & H_{gg} \end{pmatrix} = \begin{pmatrix} -\hbar\delta/2 & \hbar g\sqrt{n+1} \\ \hbar g^* \sqrt{n+1} & \hbar\delta/2 \end{pmatrix},$$

where we have omitted the constant  $\hbar\omega_L(n+1/2)$  term. The matrix eigenvalues  $\Lambda_{\pm}$  are calculated with standard methods, and one finds that they agree with half the Rabi frequency  $\Omega_R$  (Eq. (7.33)):

$$\Lambda_{\pm} = \pm((\hbar\delta/2)^2 + (\hbar g\sqrt{n+1})^2)^{1/2} = \hbar\Omega_R/2. \quad (13.1)$$

In Figure 13.2 the new eigenvalues  $|\pm, n\rangle$  are shown with solid lines. They have the general form

$$\begin{aligned} |+, n\rangle &= \cos\theta |e, n\rangle + \sin\theta |g, n+1\rangle, \\ |-, n\rangle &= \sin\theta |e, n\rangle - \cos\theta |g, n+1\rangle, \end{aligned}$$

with

$$\begin{aligned} \cos\theta &= \frac{(\Omega + \delta)/2}{((\Omega + \delta)^2/4 + g^2(n+1))^{1/2}}, \\ \sin\theta &= \frac{g\sqrt{n+1}}{((\Omega + \delta)^2/4 + g^2(n+1))^{1/2}}. \end{aligned}$$

For an illustration let us study two important limiting cases:

- *AC Stark splitting*,  $\delta = \omega_L - \omega_0 = 0$ : Here we have  $\Omega = \Omega_R = 2g\sqrt{n+1}$ , and the mixing angles are exactly equal,  $\theta = \pi/4$ ,  $\cos\theta = \sin\theta = 1/\sqrt{2}$ . The width of the splitting is  $\Lambda_+ - \Lambda_- = 2g\sqrt{n+1} = \Omega_R$ , which for large photon numbers varies only slowly with  $\sqrt{n}$ . It is also called Autler–Townes or Rabi splitting.
- *AC Stark shift*,  $|\delta| \gg g$ : Away from the resonance we have  $\cos\theta \approx 1$ ,  $\sin\theta \approx 0$ , and thus the original states are only slightly modified. By Taylor expansion of Eq. (13.1), we find

$$\Lambda_+ - \Lambda_- = \hbar\delta \left( 1 + \frac{g^2(n+1)}{\delta^2/4} \right)^{1/2} \approx \hbar\delta + \frac{\hbar g^2(n+1)}{\delta/2}.$$

Thus at large detunings, the separation of the atomic energy levels experiences a small shift proportional to the intensity,  $I = (ce_0/2)E_\omega^2(n+1)$ . The alternative calculation of this *light shift* or *AC Stark shift* by second-order perturbation theory is a standard textbook problem in quantum mechanics (see, e.g., [94]).

## 13.2 Transient Phenomena

So far we have considered the interaction between a light field and matter particles by means of the optical Bloch equations (Eq. (7.36)) and mostly concentrating on stationary solutions. In the last section, though, we had to

investigate the dynamic behavior in order to describe time-of-flight broadening of long-lived states. It is in general always necessary to take the dynamic properties into account whenever the interaction time scale is short compared to the relevant damping times  $T_{1,2}$ .

Let us study important special cases as examples of dynamic light–matter interaction:  $\pi$  pulses, rapid turn-on processes, and the effect of a sequence of short light pulses.

### 13.2.1 $\pi$ Pulses

First, we consider once again the undamped case of the optical Bloch equations (7.31). For the frequent case of an atom initially in the ground state ( $w(t=0) = -1$ ), for  $\delta = 0$  the resonant solution,

$$(u, v, w)(t) = (0, \sin \theta(t), \cos \theta(t)),$$

can be easily found. The light field simply causes rotation of the Bloch vector in the  $vw$  plane. The rotation angle  $\theta(t)$  is determined by the pulse area,

$$\theta(t) = \int_{-\infty}^t \Omega_R(t') dt' = -\frac{d_{eg}}{\hbar} \int_{-\infty}^t \mathcal{E}_0(t') dt', \quad (13.2)$$

where  $\Omega_R(t) \propto \sqrt{I(t)}$ . If the rotation angle assumes the value  $\theta = \pi$ , then the atom is promoted exactly from the ground state to the excited state. If the value is  $2\pi$ , then the atom finishes the interaction again in the ground state.

Let us estimate what kind of light pulse is required to drive a  $\pi$  pulse for an atomic resonance line that has dipole moment  $d_{eg} \simeq ea_0 = 0.85 \times 10^{-29}$  C m. It is clear that the total rotation angle  $\theta(t \rightarrow \infty)$  depends on the so-called pulse area only. For a light pulse with constant intensity and period  $T$ , the necessary intensity and pulse duration can be determined according to (7.30),  $\pi = (ea_0/\hbar)\mathcal{E}_0 T$ . The numerical value for the corresponding intensity seems to be enormously high at first:

$$I_0 \simeq 120 \text{ kW mm}^{-2} (T/\text{ps})^{-2}.$$

But it has to be taken into consideration that the pulses are very short, so that the average power of a picosecond laser does not need to be very high. Standard pulsed lasers of the mode-locked type (see Section 9.5.2) operate at a pulse rate of 80 MHz, and for an area of  $1 \text{ mm}^2$ , an average total power of  $\langle P \rangle = 80 \text{ MHz} \times T \times P_0 \simeq 10 \text{ W} \times (T/\text{ps})^{-1}$  is necessary. Commercial laser systems offer average output powers exceeding 1 W, which is quite sufficient if the pulse lengths are slightly increased to 10 ps. Even then the excitation time is only about 1/1000 of the lifetime of an excited atomic state.

### 13.2.2 Free Induction Decay

At both the beginning and the end of an interaction period in light–matter coupling, transient oscillations can occur like for the classical damped oscillator of Section 7.1.1. While stationary behavior is characterized by an oscillation at the driving frequency  $\omega$ , immediately after turn-on (or turn-off), we also

expect dynamic evolution at the eigenfrequency  $\omega_0$  of the system, which though is damped out very rapidly (with time constant  $\gamma^{-1}$ ). General time-dependent solutions of the (optical) Bloch equations have already been given by Torrey (discussed in [97]) in 1950. However, they are transparent and easily understood only for special cases, for example, at exact resonance ( $\delta = 0$ ). The dynamic phenomena are also known as *optical nutation*.

An interesting case occurs for the so-called free induction decay (FID). It describes especially the decay of the macroscopic polarization of a sample in the absence of laser light, for example, after the application of a very short laser pulse with large intensity. The polarization of an individual particle may live for a much longer time than the macroscopic polarization of an ensemble, which is affected by “dephasing” of the individual particles.

The evolution of the Bloch vector components depends of course on the detuning,  $\mathbf{u} = \mathbf{u}(t, \delta)$ . We conveniently use Eq. (7.38) for analysis. For very large intensity ( $\Omega_R \gg \delta$ ) and very short time, we can neglect the detuning at first, since, during the coupling period, the Bloch vector does not have any time to precess by a significant angle. Thus a short and strong initial pulse rotates the dipole  $\rho_{eg}$  to the angle given by Eq. (13.2), and we arrive at

$$\rho_{eg}(0, \delta) = u(0, \delta) + iv(0, \delta) = i \sin \theta, \quad (13.3)$$

where we fully neglect the time elapsed. Once the light pulse is turned off, free precession occurs according to

$$\rho_{eg}(t, \delta) = i \sin \theta e^{-(\gamma' + i\delta)t}. \quad (13.4)$$

In a large sample, there often exists an inhomogeneous distribution  $f(\omega_0)$  of eigenfrequencies of the individual particles and thus of the detunings. In a gas cell this distribution is determined, for example, by the Doppler shift, with  $\delta_D = \Delta\omega_D / 2\sqrt{\ln 2}$ :

$$f(\delta) = \frac{1}{\sqrt{\pi}} e^{-(\delta/\delta_D)^2}.$$

Following an excitation with a  $\pi/2$  pulse with  $\sin \theta = 1$ , the free evolution of the macroscopic polarization is calculated from

$$P(t) = \frac{N_{at}}{V} d_{eg} e^{-i\omega_0 t} \int_{-\infty}^{\infty} f(\delta) e^{-(\delta/\delta_D)^2} e^{-(\gamma' + i\delta)t} d\delta. \quad (13.5)$$

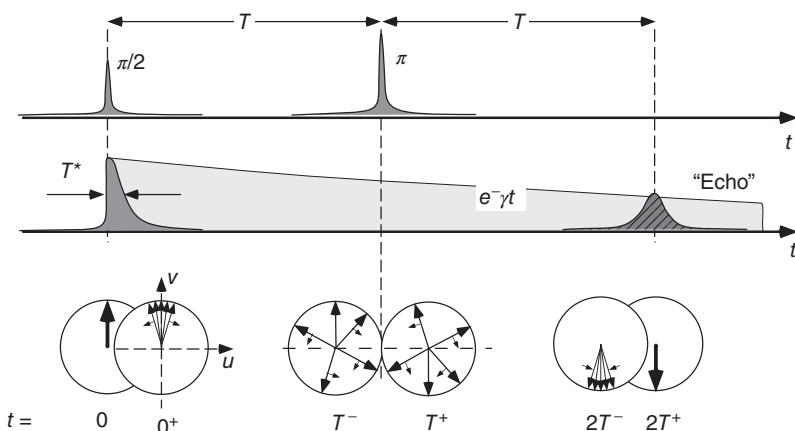
If  $\gamma' \ll \delta_D$ , we can neglect the slow decay, and the integral straightforwardly yields

$$P(t) = \frac{N_{at}}{V} d_{eg} e^{-i\omega_0 t} e^{-(\delta_D t/2)^2}.$$

The macroscopic polarization thus decays with the relaxation time  $T_2^*$ ,

$$T_2^* = \frac{4 \ln 2}{\Delta\omega_D} \ll \gamma'^{-1} = T_2,$$

much faster than the microscopic polarization whose relaxation determines the fastest time scale for individual particles. This rapid decay is a consequence of



**Figure 13.3** Free induction decay (FID) and photon echo. In this example a sample is excited by a  $\pi/2$  pulse and after a time  $T \ll \gamma^{-1}$  is subject to another  $\pi$  pulse that causes inversion of precession (top row). After the excitation light pulse, free induction decay is observed at first, which emerges from the initially cooperative emission of all excited atoms and is emitted into the direction of the excitation laser (middle row). Afterward the polarization of microscopic particles continues to decay by spontaneous emission. After time  $2T$  an echo pulse is observed in the direction of the excitation pulses. The precession of the Bloch vector components in the  $uv$  plane is marked in the bottom row.

the dephasing of the precession angles of the microscopic dipoles. Experimental observation for typical atomic resonance transitions in a gas cell must have resolution better than 1 ns and requires considerable effort; thus slower and weaker transitions are more appropriate to observe this phenomenon.

In the middle row in Figure 13.3 the time evolution of the radiation field is presented, which is caused by the macroscopic polarization and contains the cooperative radiation field of all excited microscopic dipoles of the sample. At the beginning, constructive interference of the microscopic dipole fields generates a radiation field propagating exactly in the direction of the exciting laser beam. For a perfectly synchronized phase evolution, a well-directed, accelerated, and exhaustive emission of the excitation energy would be observed due to the so-called superradiance. In an inhomogeneous sample, however, this emission ceases very rapidly as a result of the destruction of the phase synchronization ("dephasing"); the stored excitation energy is then released only by common spontaneous emission with a lower rate and isotropically.

### 13.2.3 Photon Echo

The method of "photon echoes" for inhomogeneously broadened lines – like many other optical phenomena – has been stimulated by the "spin echo" method at radio frequencies, which was discovered by I. Hahn for nuclear magnetic resonance. If a sample is excited by two or more short light pulses ( $T \ll \gamma^{-1}$ ), under certain conditions, it emits an "echo pulse" that follows the excitation pulses in their direction and seems to appear from nowhere. This contradiction is again due to the different evolutions of the microscopic and macroscopic

polarizations in a large sample of atoms, molecules, or other microscopic objects, which we have already just met in FID.

The photon echoes can of course be observed only within the natural lifetime of the microscopic polarization. Let us consider the evolution of an individual single dipole with detuning  $\delta$  under the effect of two resonant light pulses. After time  $T$  the dipole has reached the value

$$\rho_{eg}(t, \delta) = i \sin \theta e^{-(\gamma' + i\delta)T}$$

according to Eqs. (13.3) and (13.4). The application of a  $\pi$  pulse now generates an inversion of the  $(v, w)$  components (“phase reversal”). Formally this situation is identical with an inversion of the detuning, that is, after the  $\pi$  pulse, we have

$$\rho_{eg}(t, \delta) = i \sin \theta e^{-(\gamma' - i\delta)T} e^{-(\gamma' + i\delta)(t-T)}.$$

The development of the macroscopic polarization can now be given again by Eq. (13.5):

$$P(t) = \frac{N_{at}}{V} d_{eg} e^{-i\omega_0 t} e^{-[\delta_0(t-2T)/2]^2} e^{-\gamma' t}.$$

After time  $t = 2T$  the precession phase angle of each of the microscopic dipoles coincides again; macroscopic polarization is thus restored and once more causes cooperative emission of a macroscopic and coherent radiation field in the direction of the exciting light beam. This pulse is called a “photon echo.”

### 13.2.4 Quantum Beats

Simultaneous excitation of two or more electronic states by a short light pulse causes observation of a damped oscillation in the fluorescence. These oscillations are usually called *quantum beats*.

In order to realize coherent superposition of several adjacent quantum states, the inverse period of the light pulse  $T^{-1}$  (or in other words its “bandwidth”  $\Delta\nu = 1/T$ ) has to be larger than the frequency separation of the states from each other. Thus the spectral structure of the system is in fact not resolved by the exciting light pulse!

A simple quantum mechanical description assumes that the coherent superposition of two excited states decays freely and spontaneously after the excitation. For a single decaying channel, it can be shown that one can describe the time evolution of the excited state with the wave function  $|\Psi(t)\rangle = e^{-\gamma t} e^{-i\omega t} |e\rangle$ . Furthermore the observed fluorescence intensity is proportional to the square of the induced dipole moment  $|\langle g| \hat{d}_{eg} |e(t)\rangle|^2$ , and the following can be easily calculated.<sup>1</sup>

$$I_{fl} = I(0) e^{-2\gamma t}.$$

---

<sup>1</sup> For a rigorous theoretical treatment, one has to consider quantum states  $|e\rangle|0\rangle$ , the product states for the excited atom and the electromagnetic vacuum field, and all states  $|g\rangle|1_k\rangle$  for the ground state and field modes with wave vector  $\mathbf{k}$ . Here we restrict ourselves to an ad hoc treatment of the time evolution of the excited state only.

If two states  $|e_{1,2}\rangle$  with excitation frequencies  $\omega_{1,2}$  are prepared in a coherent superposition  $|\Psi(t=0)\rangle = |e_1\rangle + |e_2\rangle$ , then one has

$$|\Psi(t)\rangle = |e_1\rangle e^{-i\omega_1 t} e^{-\gamma_1 t} + |e_2\rangle e^{-i\omega_2 t} e^{-\gamma_2 t},$$

and the emitted field contains also the beat frequency  $\Delta\omega = \omega_1 - \omega_2$ . For the special case  $\gamma_1 = \gamma_2$ , one can calculate

$$I_{\text{fl}} = I(0) e^{-\gamma' t} (A + B \cos \Delta\omega t).$$

The quantum-beat method has proven to be very useful, for example, to investigate the fine structures of excited atomic or molecular states with broadband pulsed laser light, which themselves do not provide the necessary spectral resolution. For a systematic experiment, it is, though, necessary to use laser pulses of good quality, that is, precisely known shape (so-called transform-limited pulses) to guarantee coherence conditions.

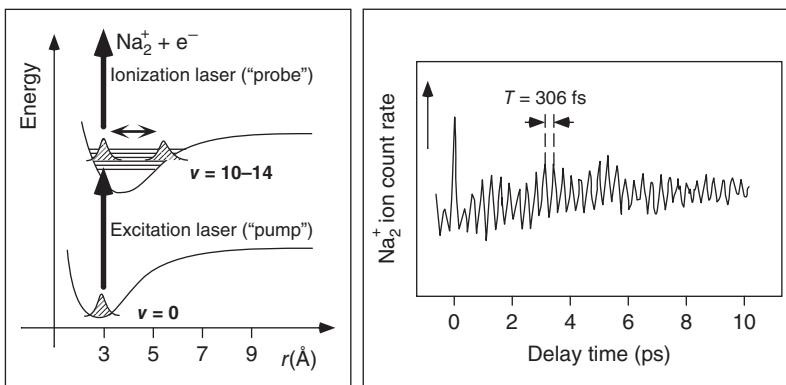
### 13.2.5 Wave Packets

A natural extension of quantum beats is offered by wave packets in microscopic systems, which are generated by coherent superposition of many quantum states. With extremely short laser pulses (10 fs corresponds to a bandwidth of 17 THz!), for example, in a molecule, numerous vibrational states can be superimposed coherently [182]. A large density of electronic states is also offered from Rydberg states in atomic systems. Rydberg atomic states have very large principal quantum numbers  $n > 10$  [183]. Neither the atomic Rydberg states nor the molecular vibrational states are usually very strongly radiating states, and therefore it is quite difficult to detect them with common fluorescence detectors. In vacuum, though, the weakly bound Rydberg states can be detected by field ionization and the molecular states by multiphoton ionization. These charged products can be detected with such high sensitivity and selectivity that only a few excited particles are required for such experiments.

With this evolution of the old quantum-beat method, it is conceivable that a wave packet is prepared from excited quantum states by a light pulse and subsequently propagates freely, that is, undisturbed by further light interaction. As long as we use a perfect harmonic oscillator, the wave packet will even propagate dispersion-free and return periodically to the origin.

Real molecules though have a strong anharmonicity, which leads to the loss of phase coherence of the atomic wave function like for the FID. The total wave function is then more or less spread over the energetically allowed space. For many systems however – in this case without application of an external pulse – the wave packet reoccurs. The phenomenon of collapse and revival of an oscillation was already predicted by Poincaré for classical oscillators. It occurs always when a finite number of oscillations are superimposed; the larger the number, the more time this return takes (Figure 13.4).

The dynamic evolution of a wave packet in molecules or Rydberg atoms can be investigated experimentally by so-called “pump–probe” experiments (Figure 13.4). With the first pulse a physical excitation is generated; with the



**Figure 13.4** Two-photon stepwise ionization of  $\text{Na}_2$  molecules as a function of the delay time between the two pulses. The ion current is plotted as a function of the delay of the ionization or “probe” pulse from the excitation or “pump” pulse. The duration of the laser pulses was 70 ps. The oscillation shows a beat signal that originates from the superposition of two contributions with periods 306 and 363 fs [182].

second one the dynamic evolution is probed after a variable time delay. We introduce a transparent example, multiphoton ionization of the model system of  $\text{Na}_2$  molecules, in a qualitative way.

A molecular beam with  $\text{Na}_2$  molecules is excited by a sequence of laser pulses (pulse period 70 fs,  $\lambda = 627 \text{ nm}$ ). The first laser pulse transfers molecules from the ground state ( $v = 0$ ) to an excited state in which several oscillation states ( $v \approx 10-14$ ) are superimposed. A further laser pulse generated by the same laser in this experiment generates  $\text{Na}_2^+$  molecules by two-photon ionization. These ions can be detected by a secondary electron multiplier, for example, a channeltron with a probability approaching 100%. In the experiment more filters such as mass spectrometers are used to separate the  $\text{Na}_2^+$  signal from the background. If the ionization pulse is delayed, an oscillation of the ion current can be observed as a function of the delay time. Since a beat is observed, the spectrum has to consist of two oscillation frequencies. The first one at 306 fs is caused by the oscillation of the wave packet in the molecular potential, and the second one at 363 fs by the interaction of the detection laser with yet another higher-lying molecular potential.

Using laser pulses of an extremely short period, it has become possible to resolve the dynamics of molecular wave packets directly on the femtosecond time scale. These and other methods are used more and more in the so-called femtochemistry.

## 14

# Photons: An Introduction to Quantum Optics

The question whether the quantum properties of the electromagnetic field (and not only of material particles) play a significant role for light–matter interactions is the central topic of *quantum optics*. Here we have to explore in more detail concepts such as “photon,” “stimulated emission,” and “spontaneous emission” beyond the interpretation we have already given in Chapter 7. Spontaneous emission is one of the most fundamental processes of light–matter interaction at the microscopic quantum level – a single atom and a single photon. Before we can treat this seminal problem that contributed in an outstanding way to the emergence of quantum electrodynamics (QED), we have to introduce some formal methods for the description of quantized light fields.

## 14.1 Does Light Exhibit Quantum Character?

It seems obvious today that the physical properties of matter at microscopic scales must be described by quantum theories. But the connection of the microscopic and the macroscopic points of view is not always so clear, a discrepancy contributing to the hierarchy of theories given in Table 7.1 as well. Namely, a large number of phenomena of light–matter interaction can straightforwardly be explained with semiclassical theories (the realm of quantum electronics in Table 7.1), that is, a quantum theory of matter interacting with a classical electric field characterized by amplitude and phase. Semiclassical treatment is sufficient for the majority of processes discussed in the chapters on laser spectroscopy (Chapter 12) and nonlinear optics (Chapters 15 and 16).

The answer to the question as to which phenomena cannot be physically understood without the quantum properties of the electromagnetic field is indeed not so easy to find. For instance, the widespread view that the photoelectric effect constitutes a proof for the quantum nature of light turns out to be an elegant way of describing a resonance phenomenon induced by the quantum structure of a conducting material such as a metal but not the quantum nature of the electromagnetic fields [86]. On the other hand, the quantum nature of light is of unrivaled and well-recognized importance for the birth of quantum theory. In 1900, Max Planck introduced the quantum hypothesis to account for experimental contradictions regarding the spectrum of blackbodies

(see Section 7.3). Albert Einstein received the Nobel Prize in 1921 “for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect,” dating back to his famous *annus mirabilis*, 1905. The term *photon* was not used before Lewis [171] wrote (1926) in a correspondence:

I therefore take the liberty of proposing for this hypothetical new atom, which is not light but plays an essential part in every process of radiation, the name *photon*.

Practically speaking, the quantum nature of light becomes visible if, for example, individual atoms only interact with a light field or if the intensity of a light field is so low that sensitive photon counters must be used for detection. It is generally accepted that the following phenomena and their theoretical description have experimentally established the need for a quantum theory of light, which includes the field of *quantum optics*:

- *Lamb shift.* The relativistically correct theory of the hydrogen atom by P. Dirac predicts that the electronic  $^2S_{1/2}$  and  $^2P_{1/2}$  states with main quantum number  $n = 2$  are perfectly degenerate. In 1947, however, W. Lamb discovered a small splitting. It has since been called the *Lamb shift* and is explained through the so-called vacuum fluctuations of the electromagnetic field.
- *Spontaneous emission.* The first successful calculation of the rate of spontaneous emission (perhaps the most fundamental of all light–matter interaction processes) was carried out in 1930 by Weisskopf and Wigner [88] on the basis of the quantization of the electromagnetic field introduced by P. Dirac.
- *The spectrum of resonance fluorescence.* The fluorescence spectrum of a driven atom, one of the simplest possible quantum oscillators, is different from the spectrum of a classical oscillator.
- *Photon correlations, “bunching,” and “anti-bunching”.* When instead of the spectrum the complementary temporal dynamics of a light source such as a single atomic oscillator is recorded by a photon counter, the fluctuations – quantitatively determined by temporal correlation measurements of the intensity – have so-called nonclassical character.

In this chapter we will treat these phenomena with the exception of the Lamb shift, which is a standard topic for many textbooks on QED (see, e.g., [184]), supplemented with more recent developments. Before we can describe these phenomena theoretically, it is necessary to prepare theoretical tools for the description of the electromagnetic field in quantum theory.

## 14.2 Quantization of the Electromagnetic Field

A rigorous foundation of the quantization of the electromagnetic field requires the application of the Lagrange formalism, which exceeds the scope of this book, and so the reader is referred to standard textbooks on advanced QED, for example, [184, 185]. Here we restrict ourselves to a heuristic approach

taking advantage of the analogy of the Hamilton formalism in classical and quantum physics.

Let us consider a volume  $V$  with perfectly conducting walls and decompose its electromagnetic field into its eigenmodes. From the beginning we normalize the amplitude  $E_{\mathbf{k},\epsilon}(t) = E_\omega \alpha_{\mathbf{k},\epsilon}(t)$  of each eigenmode in such a way that for  $|\alpha_{\mathbf{k},\epsilon}(t)|^2 = 1$  the volume  $V$  stores the energy of exactly one “photon” (see Section 2.1.8) with energy  $\hbar\omega$ :

$$E_\omega = \left( \frac{\hbar\omega}{2\epsilon_0 V} \right)^{1/2} \quad \text{with} \quad c\epsilon_0 \int_V |E_\omega|^2 dV = \hbar\omega. \quad (14.1)$$

Each eigenmode has an index  $\mathbf{k}$  for the wave vector and  $\epsilon$  for the polarization state. These numbers will later be equivalent to the quantum numbers of the quantum field. The electromagnetic field constructed from these quantities is

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \mathbf{Y}(\mathbf{r}, t) = \sum_{\mathbf{k},\epsilon} |\alpha_{\mathbf{k},\epsilon}| E_\omega \sin(\omega t - \mathbf{k}\mathbf{r}) \\ &= -i \sum_{\mathbf{k},\epsilon} E_\omega \epsilon [\alpha_{\mathbf{k},\epsilon}(t) e^{i\mathbf{k}\mathbf{r}} - \alpha_{\mathbf{k},\epsilon}^*(t) e^{-i\mathbf{k}\mathbf{r}}] = -i(\mathbf{E}^{(+)} - \mathbf{E}^{(-)}). \end{aligned}$$

We could have used just as well the definition  $\mathbf{X} = \mathbf{E}^{(+)} + \mathbf{E}^{(-)}$ , which differs from  $\mathbf{Y}$  only in the phase, which may be chosen freely at this point.  $\mathbf{X}$  and  $\mathbf{Y}$  are called *quadratures* of the electromagnetic field. In interferometric experiments the phase can be defined by a strong (i.e., classical) local oscillator field, and the two quadratures can be measured by interference with the in-phase and out-of-phase components of the local oscillator.

According to Maxwell's equations the (normalized) amplitudes  $\alpha_{\mathbf{k},\epsilon}(t)$  depend on the current density, which has Fourier components

$$j_{\mathbf{k},\epsilon} = \frac{1}{V} \int d^3 r \epsilon \mathbf{j}(\mathbf{r}) e^{-i\mathbf{k}\mathbf{r}}.$$

The amplitude  $\alpha_{\mathbf{k},\epsilon}(t)$  obeys the equation of motion

$$\dot{\alpha}_{\mathbf{k},\epsilon}(t) + i\omega \alpha_{\mathbf{k},\epsilon}(t) = \frac{i}{\sqrt{2\epsilon_0 \hbar \omega_{\mathbf{k}}}} j_{\mathbf{k},\epsilon}(t).$$

In free space we have  $j_{\mathbf{k},\epsilon}(t) = 0$  and thus  $\alpha_{\mathbf{k},\epsilon}(t) = \alpha_{\mathbf{k},\epsilon}(0) e^{-i\omega t}$ . For accelerated charges we have Fourier components  $j_{\mathbf{k},\epsilon}(t) \neq 0$ , and the field is now driven, for instance, by the oscillating current of an atom.

We now take the next step toward a quantum theoretical description of electromagnetic fields by the direct analogy of classical field amplitudes and quantum field operators, that is, we simply promote the normalized amplitudes of the classical electromagnetic field to *field operators*:

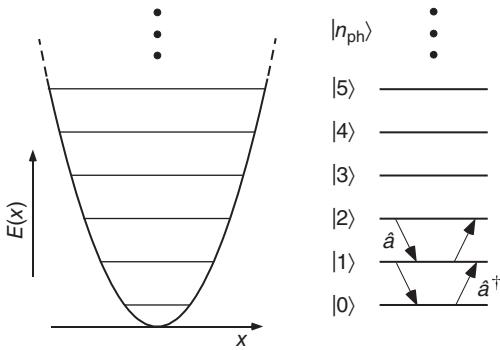
$$\alpha_{\mathbf{k},\epsilon}(t) \rightarrow \hat{a}_{\mathbf{k},\epsilon}(t) \quad \text{and} \quad \alpha_{\mathbf{k},\epsilon}^*(t) \rightarrow \hat{a}_{\mathbf{k},\epsilon}^\dagger(t).$$

As a consequence, the operator of the electromagnetic field now reads

$$\hat{\mathbf{E}} = -i(\hat{\mathbf{E}}^{(+)} - \hat{\mathbf{E}}^{(-)}) = -i \sum_{\mathbf{k},\epsilon} E_\omega \epsilon [\hat{a}_{\mathbf{k},\epsilon}(t) e^{i\mathbf{k}\mathbf{r}} - \hat{a}_{\mathbf{k},\epsilon}^\dagger(t) e^{-i\mathbf{k}\mathbf{r}}], \quad (14.2)$$

while the Hamilton operator takes the well-known form

$$\hat{H}_{\text{field}} = \sum_{\mathbf{k}, \epsilon} \hbar \omega_{\mathbf{k}} (\hat{a}_{\mathbf{k}, \epsilon}^{\dagger} \hat{a}_{\mathbf{k}, \epsilon} + 1/2). \quad (14.3)$$



**Figure 14.1** The quantum state ladder of the harmonic oscillator (potential and quantum states on the left) is perfectly analogous to the quantum states of an electromagnetic field mode. Electromagnetic field operators are constructed from raising ( $\hat{a}^\dagger$ ) and lowering operators ( $\hat{a}$ ).

interaction takes place in an intense laser field. In many cases then all modes of the electromagnetic field except the driving mode of the laser field can be neglected.

The Hamilton operator of an individual electromagnetic state in (14.3) is formally equivalent to the Hamilton operator of a material particle with mass  $m$  confined by harmonic potential  $V(x) = m\omega^2x^2/2$ . With spatial and momentum operators,  $\{\hat{x}, \hat{p}\}$ , we have (see, e.g., [184, 185])

$$\hat{H} = \hbar\omega(a a^\dagger + 1/2) = \frac{\hat{p}^2}{2m} + \frac{\omega^2 \hat{x}^2}{2m}.$$

For the harmonic oscillator we thus find

$$\hat{x} = \sqrt{2m\hbar/2\omega} (a + a^\dagger) \quad \text{and} \quad \hat{p} = -i\sqrt{2m\hbar\omega/2} (a - a^\dagger).$$

Because of the formal analogy with the quadratures of the electric field operator, we expect that  $\hat{X} = (\hat{\mathbf{E}}^{(+)} + \hat{\mathbf{E}}^{(-)})$  and  $\hat{Y} = -i(\hat{\mathbf{E}}^{(+)} - \hat{\mathbf{E}}^{(-)})$  obey an uncertainty relation like  $\hat{x}$  and  $\hat{p}$ .

Let us now use the Dirac notation to collect important relations for the field operators, the annihilation ( $\hat{a}$ ), the creation ( $\hat{a}^\dagger$ ), and the number operator  $\hat{n} = \hat{a}^\dagger \hat{a}$ :

$$\begin{aligned} \hat{a}_{\mathbf{k}, \epsilon}^{\dagger} |n\rangle_{\mathbf{k}, \epsilon} &= (n+1)^{1/2} |n+1\rangle_{\mathbf{k}, \epsilon}, \\ \hat{a}_{\mathbf{k}, \epsilon} |n\rangle_{\mathbf{k}, \epsilon} &= n^{1/2} |n-1\rangle_{\mathbf{k}, \epsilon}, \\ \hat{a}_{\mathbf{k}, \epsilon}^{\dagger} \hat{a}_{\mathbf{k}, \epsilon} |n\rangle_{\mathbf{k}, \epsilon} &= \hat{n} |n\rangle_{\mathbf{k}, \epsilon} = n |n\rangle_{\mathbf{k}, \epsilon}. \end{aligned} \quad (14.4)$$

Quantum states of the electromagnetic field are fully classified by quantum numbers  $(\mathbf{k}, \epsilon)$  and the photon occupation number  $n_{\mathbf{k}}$  (see Figure 14.1). The sum of all electromagnetic field states from Eq. (14.3) with respect to quantum numbers corresponds to the continuous spectrum of free space. Light-matter interaction in realistic physical situations such as spontaneous emission or the Lamb shift is dominated by this wide spectrum.

Another limiting case is realized when light-matter

Every number state  $|n\rangle$  can be generated from the vacuum state  $|0\rangle$  through  $n$ -fold application of the creation operator:

$$|n\rangle = \frac{1}{\sqrt{n!}} (\hat{a}^\dagger)^n |0\rangle \quad \text{and} \quad \langle n| = \frac{1}{\sqrt{n!}} \langle 0| (\hat{a})^n. \quad (14.5)$$

The state of the electromagnetic vacuum – that is, the field without any photons – is described by a product state, for instance,

$$|\text{Vac}\rangle = |0000 \dots 0000\rangle$$

where every digit represents the occupation number of some specific mode of the field. We can use this state to show immediately that the expectation value of an electromagnetic field vanishes but not the variance, which is a measure for the intensity of the so-called vacuum fluctuations:

$$\langle \text{Vac} | \hat{\mathbf{E}} | \text{Vac} \rangle = 0 \quad \text{and} \quad \langle \text{Vac} | \hat{\mathbf{E}} \hat{\mathbf{E}}^\dagger | \text{Vac} \rangle > 0 \quad (14.6)$$

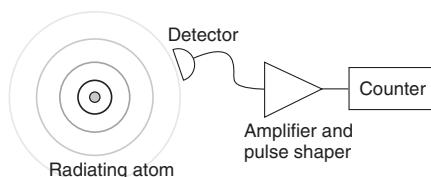
The variance is furthermore proportional to the energy density  $U = \epsilon_0 \langle \mathbf{E} \mathbf{E}^* \rangle$ . This quantity diverges in any volume having an infinite number of states as well as in every single point. This observation hints at the limited validity of this type of QED but is beyond the scope of this text [184].

### 14.3 Spontaneous Emission

One of the simplest, perhaps even *the* simplest of all, processes of radiation–matter interaction is constituted by the radiative decay of an initially excited atom in free space. This event, where a single atom makes the light source and emits precisely one photon into its empty environment (the *vacuum*), is called *spontaneous emission*. Here again the use of the word *photon* offers a convenient manner of speaking.

One should be aware of possible misinterpretation caused by the photon picture. When only microscopic objects are participating in the processes of light–matter interaction, then the role of the detector can no longer be neglected.

A consistent view regards a *photon* as that incident that occurs with the “click”<sup>1</sup> of the detector, which is always an object of our macroscopic world. Here, an individual atom emits a radiation field in close analogy to a classical microscopic dipole antenna. The total energy emitted corresponds to the energy  $\hbar\omega$  of a photon. In the language of the quantum theory of measurements, the detector causes *reduction*



**Figure 14.2** For light–matter interactions at the microscopic level, the role of the light source, the radiation field, and detector must all be taken into account.

<sup>1</sup> The electric pulse from photon detectors such as photon multiplier can always be converted electronically into an acoustic “click.”

(or *projection*) of the spreading wave function of the electromagnetic field onto the electron generated at the detector by the photoelectric effect. This electron is amplified by, for example, the photomultiplier (see Section 11.5), yielding an electric pulse that can be recorded and counted by conventional electronics (Figure 14.2).

### 14.3.1 Vacuum Fluctuations Perturb Excited Atoms

In the realm of classical electrodynamics, the treatment of radiation fields emitted by dipole oscillators and their back-action onto the motion of the charged particles (“radiation reaction”) leads into contradictions that cannot be resolved within the theories of Maxwell and Newton (see Section 7.1.1, [87]) Neither does quantum mechanics alone – the treatment of electronic motion in atoms by the rules of quantum mechanics – deliver a conclusive concept. As is well known the dipole moment of an atom vanishes in every pure quantum state – so how can a radiation field arise? Indeed, in Section 7.2.6, we have used a purely phenomenological approach by introducing a damping term in the semiclassical theory of light–matter interaction, for instance, in the optical Bloch equations presented.

Only when a full quantum theory of the electromagnetic field is applied does a consistent theory of radiative damping become possible. Let us consider the radiative decay of an atom or a molecule prepared in an initial excited state  $|i\rangle$  into a final state  $|f\rangle$ . In a perturbative approach, the transition rate can be calculated by Fermi’s golden rule,

$$W_{i \rightarrow f} = \frac{2\pi}{\hbar} |M_{if}|^2 \delta(E_f - E_i), \quad (14.7)$$

where  $M_{if}$  denotes the matrix element (having the dimension of an energy) corresponding to the transition. For the case of an excited atom radiating spontaneously into its vacuum environment, we can write the quantum states as product states (later we will call them *dressed states*) of atomic states  $|e\rangle, |g\rangle$  and field states with indices  $\alpha = (\mathbf{k}, \epsilon)$ . At the beginning ( $t = 0$ ) the field is in the vacuum state, but in the end there are many different modes that can be excited. Thus we have initial and final states

$$\begin{aligned} |i\rangle &= |e\rangle|000\dots000\rangle, \\ |f\rangle_\alpha &= |g\rangle|000..1_\alpha..000\rangle. \end{aligned} \quad (14.8)$$

For the dipole operator we use like in Section 7.2.2, the atomic lowering and raising operators  $\hat{\sigma} = |e\rangle\langle g|$  and  $\hat{\sigma}^\dagger = |g\rangle\langle e|$ , yielding  $\hat{\mathbf{d}} = -e\hat{\mathbf{r}} = -e\mathbf{r}_{eg}(\hat{\sigma} + \hat{\sigma}^\dagger)$ . The field operator is now expressed with operators (14.2):

$$\hat{\mathbf{d}} \cdot \hat{\mathbf{E}} = -e\hat{\mathbf{r}} \cdot \epsilon(\hat{a} + \hat{a}^\dagger)E_\omega = -e(\mathbf{r}_{eg} \cdot \epsilon)E_\omega(\hat{a} + \hat{a}^\dagger)(\hat{\sigma} + \hat{\sigma}^\dagger).$$

Near resonance we can again apply the rotating wave approximation from Section 7.2.3, that is, neglect terms  $\hat{a}\hat{\sigma}$  and  $\hat{a}^\dagger\hat{\sigma}^\dagger$ , which oscillate rapidly with  $2\omega$ . One obtains an effective operator for the dipole interaction:

$$\hat{V}_{\text{dip}} = -\hat{\mathbf{d}} \cdot \hat{\mathbf{E}} \simeq e(\mathbf{r}_{eg} \cdot \epsilon)E_\omega(\hat{a}\hat{\sigma}^\dagger + \hat{a}^\dagger\hat{\sigma}) = \hbar g (\hat{a}\hat{\sigma}^\dagger + \hat{a}^\dagger\hat{\sigma}). \quad (14.9)$$

This is called the *Jaynes–Cummings model* and has gained enormous relevance since it allows one to describe the quantum nature of light–matter interaction in

an exactly solvable model. The coupling constant

$$g = er_{\text{eg}}E_\omega/\hbar = er_{\text{eg}} \left( \frac{\omega}{2\hbar\epsilon_0 V} \right)^{1/2}$$

is called the *vacuum Rabi frequency*. It describes the coupling strength of an atomic dipole to an electromagnetic field in the absence of any excitation. The amplitude  $\mathcal{E}_\omega = (\hbar\omega/2\epsilon_0 V)^{1/2}$  (Eq. (14.1)) may be interpreted as the average electric field strength of a single photon with energy  $\hbar\omega$  stored in a volume  $V$ .

In order to calculate the total transition rate according to Eq. (14.7), we now have to take all possible final states  $\alpha$  into account. Replacing the energies by transition frequencies and using  $\delta(E) = \hbar\delta(\omega)$  in (14.7), we find

$$W_{i \rightarrow f} = 2\pi \sum_{\alpha=k,e} |g_\alpha|^2 \delta(\omega - \omega_{if}).$$

Next we replace the summation by integration in  $\mathbf{k}$ -space over the volume  $V_k$ ,  $\sum_{\mathbf{k}} \rightarrow \int_{V_k} d^3k$ , and with density  $\rho_{\mathbf{k}}(\mathbf{k}) = 1/(2\pi)^3$  (see Section B.3). The density of states in free three-dimensional (3D) space is derived from

$$\rho_{\text{free}}(\omega) = 2 \int_{V_k} d^3k \rho_{\mathbf{k}}(\mathbf{k}) \delta(\omega - |\mathbf{ck}|) = \frac{\omega^2}{\pi^2 c^3}.$$

The factor 2 accounts for the two orthogonal polarizations allowed for each field mode. In the final evaluation, the scalar product  $\mathbf{r}_{\text{eg}} \cdot \mathbf{\epsilon}$  of the dipole operator (Eq. (14.9)) is averaged in 3D space and causes another factor of 1/3:

$$W_{i \rightarrow f} = \frac{1}{3} \frac{2\pi}{\hbar} e^2 r_{if}^2 \frac{\hbar\omega_{if}}{2\epsilon_0} \frac{\omega_{if}^2}{\pi^2 c^3} = \frac{e^2 r_{if}^2 \omega_{if}^3}{3\hbar\epsilon_0 \pi c^3}.$$

This result is identical with the Einstein  $A$  coefficient for spontaneous emission and also the result by Wigner and Weisskopf, which is discussed in the next section. This agreement is by no means obvious since the golden rule applied here is valid only for short times, when the initial state has not yet significantly changed (Figure 14.3).

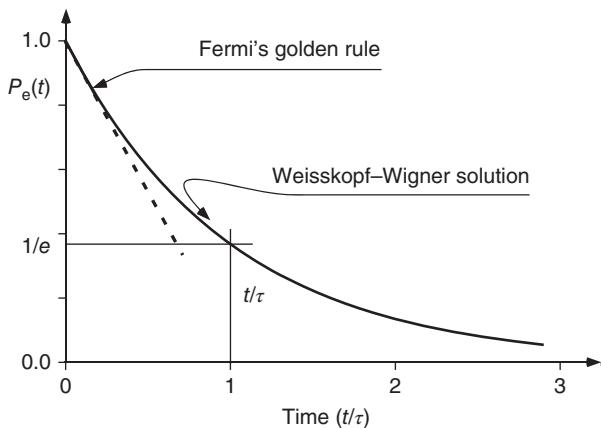
The decay rate  $W_{i \rightarrow f} = A_{i \rightarrow f} = \gamma = 1/\tau$  determines the natural line width  $\gamma = \Delta\omega = 2\pi\Delta\nu$  in spectroscopic observations of an electromagnetic transition; see Section 12.3.

### 14.3.2 Weisskopf and Wigner Theory of Spontaneous Emission

The theoretical problem of spontaneous emission was solved for the first time in 1930 by Weisskopf (1908–2002) and Wigner (1902–1995) [88], who applied the ideas of QED developed by Dirac (1902–1984) [186]. Following their treatment, let us consider an atomic superposition state with eigenstates from (14.8)

$$|\Psi(t)\rangle = C_i(t)e^{-i\omega_i t}|i\rangle + \sum_{\alpha} C_{f\alpha} e^{-i(\omega_f + \omega)t}|f\rangle_{\alpha}.$$

The interaction picture of quantum mechanics gives the equation of motion  $i\hbar|\dot{\Psi}(t)\rangle = \hat{V}_{\text{dip}}|\Psi(t)\rangle$ . Application of the dipole operator (14.9) and inspection



**Figure 14.3** Exponential decay of the excitation probability according to the Weisskopf and Wigner theory. Fermi's golden rule perfectly agrees for short times.

of the coefficients yields the system of equations

$$\begin{aligned}\dot{C}_i(t) &= -i \sum_{\alpha} g_{\alpha} e^{-i(\omega-\omega_{if})t} C_{f\alpha}(t), \\ \dot{C}_{f\alpha}(t) &= -ig_{\alpha}^* e^{i(\omega-\omega_{if})t} C_i(t).\end{aligned}$$

Since infinitely many states  $|f\rangle_{\alpha}$  and coefficients  $C_{f\alpha}$  contribute to the evolution, the system of equations is infinitely large as well! Let us integrate the second equation formally:

$$\dot{C}_i(t) = - \sum_{\alpha} |g_{\alpha}|^2 \int_0^t e^{-i(\omega-\omega_{if})(t-t')} C_i(t') .$$

We now apply the so-called coarse-grained solution where we assume that  $C_i(t') \approx C_i(t)$  and hence can be pulled out of the integral. Finally we apply the well-known result from complex analysis ( $\mathcal{P}$  denotes the principal value integral)

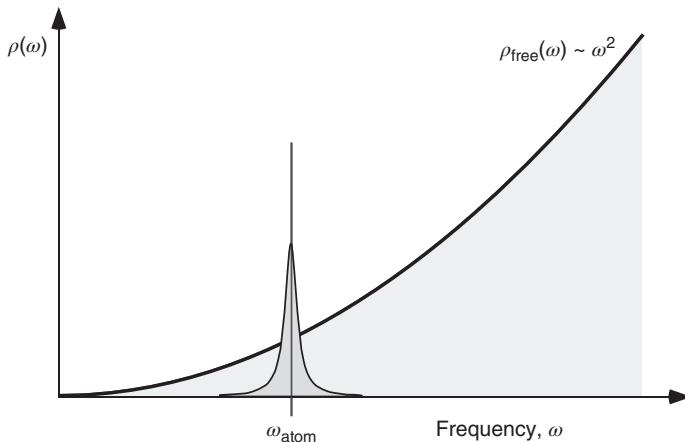
$$\lim_{t \rightarrow \infty} \int_0^t dt' e^{-i(\omega-\omega_{if})(t-t')} = \pi \delta(\omega - \omega_{if}) - \mathcal{P} \frac{i}{\omega - \omega_{if}}. \quad (14.10)$$

The imaginary part in Eq. (14.10) causes a very small frequency shift, in analogy with a classical damped oscillator. It corresponds to the famous *Lamb shift*, which is however measurable for very few atomic resonance lines only.<sup>2</sup> Here we assume that this contribution is already included into the resonance frequency  $\omega_{if}$ . We finally obtain

$$\dot{C}_i(t) = -\frac{\gamma}{2} C_i(t) \quad \text{with} \quad \gamma = 2\pi \sum_{\alpha} |g_{\alpha}|^2.$$

The sum over quantum numbers  $\alpha$  is calculated as in the preceding section and yields the same coefficient. In this case we have derived an exponential law

<sup>2</sup> For hydrogen, the degeneracy of the  $S_{1/2}$  and  $P_{1/2}$  states facilitates this observation of the Lamb shift. In all other atoms quantum states are generally isolated, and a shift can only be detected if a theoretical model gives the unperturbed position of the atomic state with sufficient accuracy.



**Figure 14.4** Interaction of an excited atomic state (sharp resonance line at  $\omega = \omega_{\text{atom}}$ ) with the electromagnetic vacuum, here represented by its density of states  $\rho_{\text{free}}(\omega)$ . In this picture, the natural linewidth of the excited state is a result of the mixing of a discrete atomic state with the continuous spectrum of the electromagnetic vacuum.

of decay valid at all times, in excess of the perturbative approach by Fermi's golden rule.

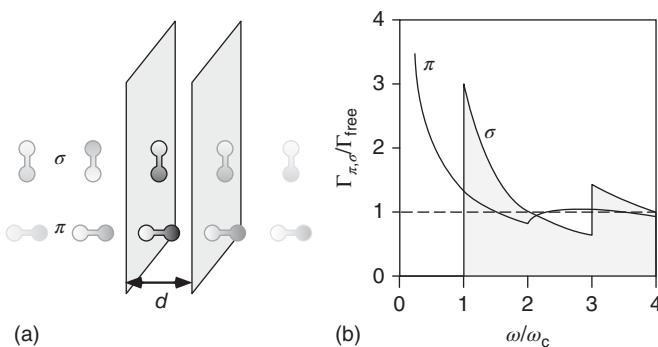
While in conventional quantum mechanics all atomic eigenstates have exact, *sharp* eigenvalues and are stable in time, the interaction with the electromagnetic vacuum induces decay of excited states. Long-lived states occur if the coupling with the electromagnetic field is weak, for example, in the metastable  ${}^3\text{S}$  state of the helium atom (Figure 14.4).

### 14.3.3 Suppression of Spontaneous Emission

The natural decay of an excited atomic or molecular quantum state seems to have inevitable fundamental character. In an environment with conducting surfaces, the decay process can, however, be modified or even turned off. As an example we consider an atom that for the sake of transparency we model as a microscopic dipole antenna positioned between two large mirror planes with separation  $d$ . The radiation field of the atom is reflected from the walls and interacts again with the radiating dipole. Depending on the phase, the reflected field is reabsorbed and inhibits decay or constructively interferes with the emitted field, enhancing the decay. One can construct the reflected radiation field from the method of image charges (Figure 14.5) and then calculate the modified decay rate of the original atom in analogy to an antenna array of image dipoles [166]. It is even simpler to determine which waves are allowed to propagate between the mirrors. The metallic walls form a simple wave guide with a cutoff frequency and wavelength for electric fields polarized orthogonal to the normal to the walls. The cutoff frequency is

$$\omega_c = \pi c/d.$$

Thus the atomic radiation field cannot propagate if the separation of the mirrors is smaller than half the resonance wavelength,  $d < \pi c/\omega_{\text{atom}} = \lambda_{\text{atom}}/2$ , and



**Figure 14.5** Suppression of spontaneous emission between plane mirrors. (a) Image charge model. At small separation  $d < \lambda/2$  the interference field of the dipole and its image dipoles causes suppression of the radiation field for the  $\sigma$  and enhancement for the  $\pi$ -orientation. (b) Modified decay rate for  $\sigma$ - and  $\pi$ -orientation as a function of mirror separation, normalized to the decay rate in free space  $\Gamma_{\text{free}}$ .

spontaneous emission is suppressed. Since optical wavelengths of atomic transitions have micrometer scales, an experiment was carried out with atoms traveling between a pair of 10 mm long mirrors separated by a very tiny gap of  $d = 1.5 \mu\text{m}$  only. It was found that atoms entering the gap in an excited state with lifetime  $1.6 \mu\text{s}$  were still excited when leaving the gap after  $20 \mu\text{s}$  [166].

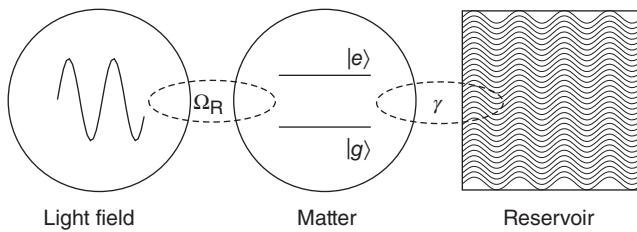
#### 14.3.4 Interpretation of Spontaneous Emission

The example of suppression of spontaneous emission shows impressively that the radiation properties of a microscopic particle are influenced by the environment. Spontaneous emission in particular is not an inevitable law of nature, as was pointed out by D. Kleppner with the phrase *turning off the vacuum* [187]. This interpretation draws on a picture where spontaneous emission is induced, stimulated by the fluctuating field of the electromagnetic vacuum. While this picture is close to our intuition, one has to keep in mind that from the theoretical point of view, there is no compelling argument for interpretation – one could alternatively also regard the fluctuations of electric motion in the atom as the first cause initiating spontaneous emission.

The modification is strongest in highly reflecting cavities. This field of *cavity QED* has been studied for many years now [188]. The experimental demonstration of cavity QED phenomena has not only delivered many proofs and illustrations [189, 190] for the relevance of QED. It has also opened the view that quantum properties of matter can be controlled and made useful for applications.

#### 14.3.5 Open Quantum Systems and Reservoirs

In quantum theory, microscopic physical systems are treated with the Hamilton operator: “Give me your Hamiltonian, and I predict the properties of your system.” A shortcoming of this approach, though, is the closed nature of Hamiltonian systems. In general they have only few degrees of freedom, and damping is not of relevance. All real systems are, however, coupled to the environment



**Figure 14.6** The coupling of a two-level system (representing “matter”) with an intense coherent light field is described by the Rabi frequency  $\Omega_R$ . At the same time, matter is always coupled to a bath, which causes damping. Depending on which of the two rates is larger, the case of weak or strong coupling is realized (Section 13.1).

with a continuous spectrum of degrees of freedom. These environments are also called *bath* or *reservoir* (Figure 14.6). Examples include the electromagnetic vacuum, blackbody radiation, or the lattice vibrations (*phonons*) of solid materials, which can be characterized by a temperature. While energy coupled to a true bath never returns to its original system (for instance, in spontaneous emission), it was known to Poincaré already at the end of the nineteenth century that in closed (sub)systems of Hamiltonian character, there is a reoccurrence of the excitation of the initial state always.

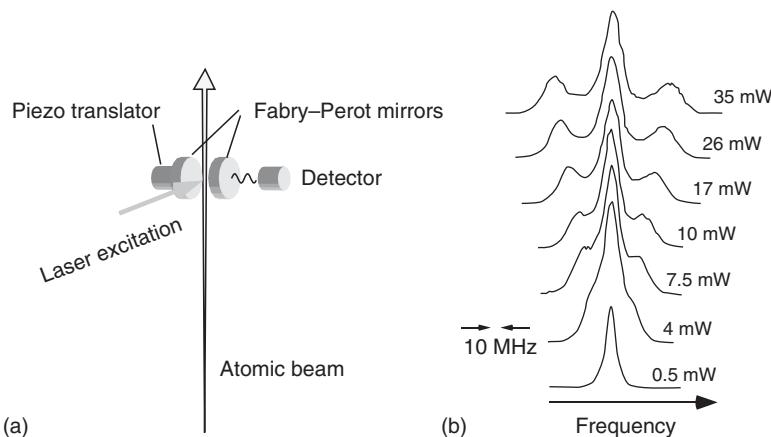
The most important bath in optics is the electromagnetic vacuum, which is the subject of QED. It is subject to fluctuations that are, for instance, manifest in its spectrum [136]. Spontaneous emission is a fundamental example of a simple Hamiltonian system with only two degrees of freedom coupled to a reservoir with a very large number of states. The treatment by Weisskopf and Wigner allows one to calculate the strength of the coupling of an atom and its bath, the electromagnetic vacuum. For a detailed analysis we refer to [135, 136].

## 14.4 Resonance Fluorescence

The term *resonance fluorescence* denotes the process where an individual atom absorbs or scatters radiation from a light field, typically a laser field, by stimulated or spontaneous emission. Stimulated emission causes reinforcement of the driving light field, while spontaneous emission radiates the absorbed energy into all directions. Resonance fluorescence has played a central role in the history of quantum optics, since the only difference with spontaneous emission itself is the addition of a driving light field. The dynamics of this process is governed by quantum fluctuations, which become apparent in the spectral properties of the combined physical system of light and matter. A complementary situation occurs in the temporal evolution of the radiation field of the atom. For instance, the so-called *anti-bunching* phenomenon (see Section 14.5.4.2) carries information about the dynamical evolution of the atomic radiation field.

### 14.4.1 The Spectrum of Resonance Fluorescence

Today, trapped individual ions and atoms make ideal objects to study resonance fluorescence in detail, but it was studied before with dilute atomic beams as



**Figure 14.7** Experimental analysis of resonance fluorescence from a sodium atomic beam irradiated with resonant light at  $\lambda = 589$  nm. (a) Schematic experimental apparatus. (b) Spectrum as a function of laser intensity. Two-level systems were obtained by the method described in Section 7.2.1. (Adapted from [191] and with kind permission by H. Walther.)

well [191]. The arrangement of such an experiment and the spectrum recorded with the apparatus are shown in Figure 14.7. With increasing intensity one observes a splitting of the initially solitary resonance line into the so-called *Mollow triplet*, which is named after the author of the first theoretical calculation of this spectrum [192]. Such experiments, albeit they are conceptually straightforward, were made possible only by the advent of widely tunable laser sources in the 1970s.

#### 14.4.2 Spectra and Correlation Functions

Optical spectra are counted among the most valuable experimental quantities in many areas of physics, yielding information about the energetic structure and – in a complementary way – temporal dynamics. As was shown previously (Section 14.3.2), the electromagnetic vacuum and quantum fluctuations play an important role in the discussion of the physical description of the spectrum of spontaneous emission. Theoretical prediction of the spectrum of a light field can be obtained by deriving the time-dependent *correlation function* based on the quantum properties of light and matter. The complementary spectrum is then calculated by Fourier transform.

The spectrum of a dynamical physical quantity such as the electric field strength  $E(t)$  is straightforwardly measured by monitoring the power  $I(\omega) \propto \{ |E(t)|^2 \}_\omega = \{ E(t) \cdot E^*(t) \}_\omega$  transmitted through a spectral filter at frequency  $\omega$  and with bandwidth  $\Delta\omega$ . The power spectral density is then obtained from  $i(\omega) = I(\omega)/\Delta\omega$ . This concept can be generalized for arbitrary physical quantities  $A$  where  $|A|^2$  is the generalized power. For calculating the spectral properties, we take advantage of the Fourier transform where products such as  $E(t) \cdot E^*(t')$  as well as the correlation functions already introduced in Section 6.2.1 play an important role. The correlation functions can be obtained from the equations of motion of the system and allow the derivation of the spectral properties of the light fields involved.

In Section 14.5.1 we will directly explore the complementary temporal dynamics associated with the correlation functions, yielding advanced information on the coherence properties of light fields.

The light field radiated by an excited atom is proportional to its dipole moment, or in quantum physics to the expectation value of the dipole operator,  $\hat{E}^+ \propto er_{eg}\hat{\sigma}^\dagger$ ,  $\hat{E}^- \propto er_{eg}\hat{\sigma}$ , and so on. Using the dipole operators from Eq. (7.26), the intensity  $I_{fl}(t)$  incident on a detector can be written as

$$I_{fl}(t) = \frac{c\epsilon_0}{2} \langle \hat{E}^+(t)\hat{E}^-(t) \rangle = \beta I_0 \langle \sigma^\dagger(t)\sigma(t) \rangle = \beta \frac{I_0}{2} (\langle \sigma_z \rangle + 1). \quad (14.11)$$

Here we have normalized the total fluorescence to its maximum value given by the saturation intensity  $I_0 = \pi hc\gamma/\lambda^3$  (Eq. (7.42)). The factor  $\beta \leq 1$  describes the geometrical fraction of the total fluorescence that reaches the detector. Formally we have again the structure of the pseudo-spin system of Eq. (7.28). The fluorescence intensity is proportional to the occupation number of the upper state because of  $\sigma_z + 1 = |e\rangle\langle e|$ . In a more precise description, the field at the detector is proportional to the retarded value of the atomic dipole radiator,  $\hat{E}^+(t) \propto \hat{\sigma}^\dagger$  ( $\tilde{t} = t - |\mathbf{r}|/c$ ). For stationary problems this retardation has no consequences, however, and can be neglected.

A classical physical quantity such as the electric field strength  $E(t)$  is connected with its Fourier components  $\mathcal{E}(\omega)$  by

$$\begin{aligned} E(t) &= \int_0^\infty \Re(\mathcal{E}(\omega)e^{i\omega t}) d\omega \\ &= \frac{1}{2} \int_0^\infty \{ \mathcal{E}(\omega)e^{i\omega t} + \mathcal{E}^*(\omega)e^{-i\omega t} \} d\omega. \end{aligned}$$

Here  $E(t)$  is a real number; hence  $\mathcal{E}(\omega) = \mathcal{E}^*(-\omega)$  and

$$\mathcal{E}(\omega) = \frac{1}{\pi} \int_{-\infty}^\infty E(t)e^{-i\omega t} dt.$$

In a measurement the spectral power density  $S_E(\omega)$  of this quantity is determined with a filter of bandwidth  $\Delta\omega$ :

$$\begin{aligned} S_E(\omega) &= \frac{c\epsilon_0}{2} |\mathcal{E}(\omega)|^2 \Delta\omega \\ &= \frac{c\epsilon_0}{2} \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^2 T} \int_{-T/2}^{T/2} E^*(t')e^{-i\omega t'} dt' \int_{-T/2}^{T/2} E(t)e^{i\omega t} dt. \end{aligned}$$

If the spectrum is not explicitly dependent on time, we can substitute  $t' - t \rightarrow \tau$ , yielding

$$S_E(\omega) = \lim_{T \rightarrow \infty} \frac{c\epsilon_0/2}{(2\pi)^2 T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} E^*(t')E(t' + \tau)e^{i\omega\tau} dt' d\tau.$$

For the stationary case (which is characterized by dynamic fluctuations, of course!), we let the integration time  $T$  grow very large. Introducing the time-averaged ( $\{\dots\}_t$ ) correlation function, we have

$$G_{EE}(\tau) = \{E^*(t)E(t + \tau)\}_t = \frac{1}{T} \int_{-T/2}^{T/2} E^*(t')E(t' + \tau) dt'. \quad (14.12)$$

Thus the spectrum  $S_E(\omega)$  can theoretically be obtained by calculating the Fourier transform of the *autocorrelation function*  $G_{EE}(\tau)$ , which is no longer explicitly dependent on time:

$$S_E(\omega) = \frac{c\epsilon_0/2}{(2\pi)^2} \int_0^\infty G_{EE}(\tau) e^{i\omega\tau} d\tau. \quad (14.13)$$

This relation is also known as the Wiener–Khinchin theorem (see Section A.1). For  $\tau = 0$  the correlation function is simply proportional to the intensity,  $G_{EE}(\tau) = \langle E^*(t)E(t) \rangle_t = 2I/c\epsilon_0$ . Using this result we introduce the normalized *coherence function* of the first order,

$$g^{(1)}(\tau) = \frac{\langle E^*(t)E(t+\tau) \rangle_t}{\langle E^*(t)E(t) \rangle_t} = \frac{c\epsilon_0}{2} \frac{G_{EE}(\tau)}{I}, \quad (14.14)$$

which is defined in straight analogy with the visibility of a classical interferometer (Eq. (6.5)). Thus the spectrum of such a classical system can be calculated once the time evolution of  $E(t)$  is known.

### Example: Spectrum of a driven classical oscillator

Classical charged oscillators are driven by an electromagnetic field  $Ee^{-i\omega_L t}$  as described in Section 7.1. The harmonic oscillator with  $\omega_{\text{osc}} \simeq \omega_L$ , mass  $m$ , and charge  $-e$  obeys the equation of motion (7.1),

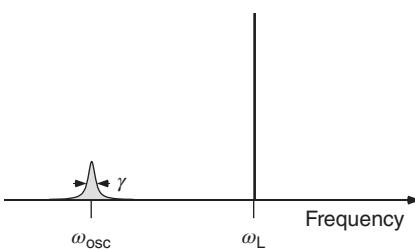
$$\ddot{x} + \gamma\dot{x} + \omega_{\text{osc}}^2 x = (-e/m)Ee^{-i\omega_L t},$$

which we have reduced for the sake of simplicity to a single coordinate again.

The field emitted is proportional to the dipole moment,  $d(t) \propto E_{\text{dip}}(t)$ . In addition to the stationary solutions for  $d(t) = -ex(t)$  treated in Section 7.1.1, we have to consider the transient solutions now as well.

For an oscillator initially at rest, we have  $x(t=0) = 0$  and  $d(t) = d_0(e^{-i\omega_L t} - e^{-i\omega_{\text{osc}} t}e^{-\gamma t/2})$ . The correlation function is now evaluated from the stationary and the transient solution using Eq. (14.12). We find contributions at frequencies  $\omega_L$ ,  $\omega_{\text{osc}}$ , and  $|\omega_L - \omega_{\text{osc}}|$ . Neglecting the latter one that occurs at radio frequencies, we find

$$G_{EE}(\tau) = \frac{c\epsilon_0}{2} |E_0|^2 \{ e^{-i\omega_L \tau} - e^{-i\omega_{\text{osc}} \tau} e^{-\gamma \tau/2} / (\gamma T) \}.$$



**Figure 14.8** Spectrum of a driven classical oscillator. Driving frequency  $\omega_L$ , oscillator frequency  $\omega_{\text{osc}}$ .

Finally, the Fourier transformation according to (14.13) yields the delta function in Figure 14.8 at  $\omega_L$  and the resonance at  $\omega_{\text{osc}}$ . This Lorentzian-shaped spectral contribution is only relevant for the initial transient dynamics and thus depends on the averaging time  $T$ . The contribution to the total spectrum is of order  $1/\gamma T$ , which becomes very small, even negligible, at long observation times. It is hence of technical and not of fundamental physical nature. The delta-like line represents the light field

scattered by the dipole oscillator driven by an external monochromatic field. This process is called *Rayleigh scattering* and occurs for all types of polarizable matter at low light field intensities, that is, in the classical limit.

#### 14.4.3 Spectra and Quantum Fluctuations

In classical physics, that is, for classical fields, correlation functions, products such as  $E(t)E(t')$ , can be calculated once the time evolution of, for example, the field  $E(t)$  is known. In quantum mechanics, the correlation function is calculated from the expectation value of the corresponding product of field operators,  $\{E(t)E^*(t')\}_t \rightarrow \{\langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t') \rangle\}_t$ . Since these operators do not necessarily commute, the expectation value of the operator product is in general not identical with the product of its individual expectation values:

$$G_{\hat{E}\hat{E}} = \langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t+\tau) \rangle \neq \langle \hat{E}^{(-)}(t) \rangle \langle \hat{E}^{(+)}(t+\tau) \rangle.$$

We can define an analogue of the classically expected correlation function through

$$G_{EE}^{\text{cl}}(\tau) = \langle \hat{E}^{(-)}(\tau) \rangle \langle \hat{E}^{(+)}(0) \rangle, \quad (14.15)$$

and we use this definition to extract contributions that have genuine quantum character by

$$G_{\hat{E}\hat{E}}(\tau) - G_{EE}^{\text{cl}} = \langle \hat{E}^{(-)}(\tau)\hat{E}^{(+)}(0) \rangle - \langle \hat{E}^{(-)}(\tau) \rangle \langle \hat{E}^{(+)}(0) \rangle.$$

This difference of the full quantum correlation function and the classical analogue is useful for physical interpretations of radiation–matter processes.

Explicit calculation of expectation values like  $\langle \hat{E}^{(-)}(\tau)\hat{E}^{(+)}(0) \rangle$  is beyond the scope of this book. Here we restrict ourselves to demonstrate application of the *Onsager–Lax* or *quantum regression theorem* [136]. For so-called Markov processes<sup>3</sup> the operator products obey a set of equations of motion identical with those of the operators themselves. For a set of operators  $\hat{O}_i(t)$  with linear equations of motion,

$$\frac{\partial}{\partial t} \langle \hat{O}_i(t) \rangle = \sum_j G_{ij}(t) \langle \hat{O}_j(t) \rangle, \quad (14.16)$$

we use the corresponding system of equations for the expectation values of the correlation functions:

$$\frac{\partial}{\partial \tau} \langle \hat{O}_i(\tau)\hat{O}_k(0) \rangle = \sum_j G_{ij}(\tau) \langle \hat{O}_j(\tau)\hat{O}_k(0) \rangle. \quad (14.17)$$

If the solutions for the equations of motion (14.16) are known, the solutions for Eqs. (14.17) are known as well. Spectral properties of the system are then calculated using the Wiener–Khinchin theorem (Eq. (14.13)). The spectral properties of interacting light–matter systems can be treated by means of the optical Bloch equations (7.36) from Section 7.2.6, which are an example for the system of operator equations (14.16). It is remarkable that for the description of the system,

<sup>3</sup> Markov processes have no *memory* (they are *delta correlated*).

stationary solutions are no longer sufficient however. Transient solutions play a central role, reflecting the fact that quantum fluctuations continuously cause small perturbations of the system with subsequent relaxation back toward the equilibrium state.

#### 14.4.4 Coherent and Incoherent Contributions of Resonance Fluorescence

Let us try to distinguish classical and nonclassical contributions in the fluorescence spectrum of, for example, atoms. Let us consider first the classical analogue by evaluating Eq. (14.11) using the expectation values of the field amplitudes  $\hat{E} \propto \hat{\sigma}$ :

$$I_{\text{coh}} = \beta I_0 \langle \sigma^\dagger(t) \rangle \langle \sigma(t) \rangle.$$

The solutions for the stationary amplitude are straightforwardly taken from  $\langle \sigma(t) \rangle = \langle \sigma^\dagger(t) \rangle^* = \frac{1}{2}(u_{\text{st}} + iv_{\text{st}}) e^{-i\omega_L t}$  using the solutions  $\{u_{\text{st}}, v_{\text{st}}\}$  of the optical Bloch equations for the two-level atom, Eqs. (7.44) and (7.45). For the special case of perfect resonance ( $\delta = \omega - \omega_0 = 0$ ), we find for an atom in free space ( $\gamma' = \gamma/2$ )

$$\langle \sigma(t) \rangle = \langle \sigma^\dagger(t) \rangle^* = -i \frac{v_{\text{st}}}{2} e^{-i\omega_L t} = \frac{-i}{\sqrt{2}} \frac{\sqrt{I/I_0}}{1+I/I_0} e^{-i\omega_L t}$$

and

$$I_{\text{coh}} = \beta \frac{I_0}{2} \frac{I/I_0}{(1+I/I_0)^2}, \quad (14.18)$$

respectively. The correlation function  $(c\epsilon_0/2)G_{EE}^{\text{cl}}(\tau) = I_0 \langle \sigma^\dagger(\tau) \rangle \langle \sigma(0) \rangle$  can be determined from these solutions as well, and the spectrum is calculated by Fourier transform (Eq. (14.13)), yielding

$$S_E^{\text{coh}}(\omega) = \frac{I_0}{2} \frac{I/I_0}{(1+I/I_0)^2} \delta(\omega - \omega_L) \xrightarrow{I/I_0 \gg 1} 0. \quad (14.19)$$

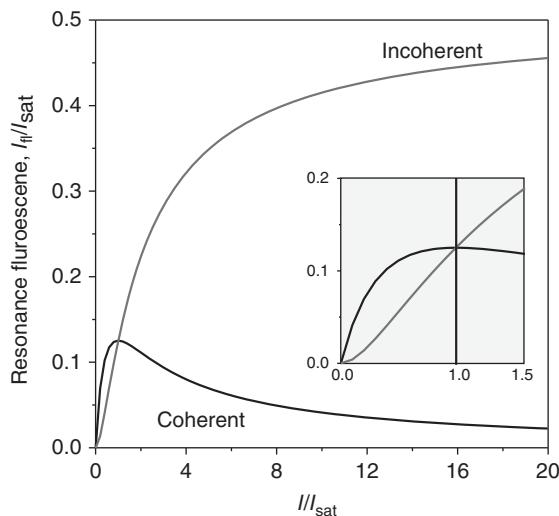
It turns out that this spectrum is delta shaped in close analogy with the Rayleigh scattering line of the fictitious classical oscillator from the example on p. 430. Note that the intensity of this contribution vanishes with increasing intensity of the driving field! The total fluorescence intensity is calculated from Eq. (14.11), and with  $\langle \sigma_z \rangle = w_{\text{st}} = -(1+I/I_0)^{-1}$ , we find

$$I_{\text{fl}}(t) = \beta \frac{I_0}{2} (\langle \sigma_z \rangle + 1) = \beta \frac{I_0}{2} \frac{I/I_0}{1+I/I_0}.$$

Now we can extract the contribution caused by quantum fluctuations by calculating the so-called *incoherent* part  $I_{\text{inc}}$  from

$$\begin{aligned} I_{\text{fl}} &= \beta I_0 [\langle \sigma^\dagger \rangle \langle \sigma \rangle + (\langle \sigma^\dagger \sigma \rangle - \langle \sigma^\dagger \rangle \langle \sigma \rangle)] \\ &= I_{\text{coh}} + I_{\text{inc}} = \beta \frac{I_0}{2} \left[ \frac{I/I_0}{(1+I/I_0)^2} + \frac{(I/I_0)^2}{(1+I/I_0)^2} \right]. \end{aligned}$$

**Figure 14.9** Coherent and incoherent contributions in resonance fluorescence of a two-level atom.



In Figure 14.9 both the coherent and the incoherent fractions are shown as a function of the normalized intensity  $I/I_0$ . At large intensities the incoherent part survives alone:

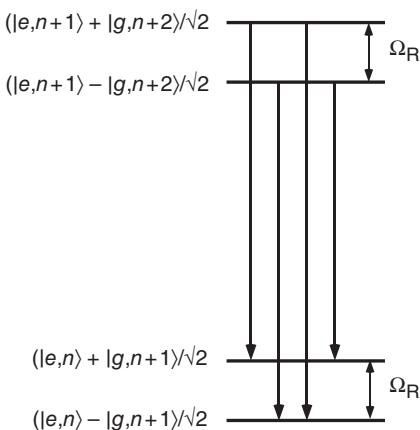
$$I_{\text{inc}} = \frac{I_0}{2} \frac{(I/I_0)^2}{(1 + I/I_0)^2} \xrightarrow{I/I_0 \gg 1} \frac{1}{2} I_0.$$

The distribution of coherent and incoherent contributions marks the difference of *polarization* and *excitation* or *occupation*. At low intensities an atom couples vacuum field states with the driving light field in a fully coherent way, like a microscopic beam splitter (BS). Strong light fields, on the other hand, generate occupation of the excited state. Spontaneous emission from the excited state generates a light field, which is uncorrelated in phase (incoherent) with the driving field.

#### 14.4.4.1 The Mollow Triplet

The dressed-atom model (Section 13.1.1) offers a convenient approach to understand the shape of the spectrum of resonance fluorescence. This model fully describes the properties of an atom interacting with a strong light field above the saturation intensity  $I > I_0$ . While all processes of stimulated emission are implicitly accounted for by this model, spontaneous emission causes scattering of light fields out of the intense laser beam, that is, transitions  $|i, n\rangle \rightarrow |f, n-1\rangle$ . Let us restrict our discussion to the most prominent case of exact resonance. There, each state  $|\pm\rangle$  contains even amounts of the unperturbed atomic ground and excited states  $\{|g\rangle, |e\rangle\}$ . Therefore, spontaneous emission is allowed for all transitions  $|\pm, n\rangle \rightarrow |\pm, n-1\rangle$  and with equal strength, as shown in Figure 14.10.

We can use the reduced energy level scheme of Figure 14.10 to find the lines of the resonance fluorescence spectrum. For intense laser beams (large  $n$ ), the splitting of neighboring doublets ( $\Omega_R = 2g\sqrt{n+1}$ ) is equal for slightly different photon numbers  $n$  to a very good approximation. Thus we find two transitions with identical frequency  $\omega_L = \omega_0$  forming the central line. Furthermore there



**Figure 14.10** Reduced energy level diagram of the dressed-atom model with allowed transitions. Compare this with Figure 13.2.

the correlation function from Eq. (14.12),  $G_{\sigma\sigma}(\tau) = (2I_0/c\epsilon_0)\langle\sigma^\dagger(\tau)\sigma(0)\rangle$ . Subsequent Fourier transform following (14.13) yields

$$I_{\text{fl}}(\omega) = I_0 \int_0^{\infty} \langle\sigma^\dagger(\tau)\sigma(0)\rangle e^{i\omega\tau} d\tau. \quad (14.20)$$

Here the term  $\langle\sigma^\dagger(\tau)\sigma(0)\rangle$  is time dependent. It resembles the transient phenomena occurring in the usual optical Bloch equations (Section 7.2.3). For instance, the complete solution for the occupation number with initial value  $\langle\sigma_z(t=0)\rangle = 0$  is

$$\langle\sigma_z(t)\rangle = \frac{1}{1+I/I_0} \left[ 1 - e^{-3\gamma t/4} \left( \cosh(\kappa t) + \frac{3\gamma}{4\kappa} \sinh(\kappa t) \right) \right]$$

with  $\kappa = [(\gamma/4)^2 - \Omega_R^2]^{1/2}$ .

For large Rabi frequencies  $\Omega_R \gg \gamma/4$ , we have  $\kappa \simeq \pm i\Omega_R$ , that is, the transient corresponds to a damped oscillation. The initial condition for the correlation function is  $\langle\sigma^\dagger(0)\sigma(0)\rangle = (\langle\sigma_z\rangle + 1)/2$ .

We give the result for the spectrum in the case of perfect resonance and for strong fields,  $I/I_0 \gg 1$ :

$$S_E(\omega) = \frac{I_0}{2\pi} \frac{I/I_0}{1+I/I_0} \left( \frac{\delta(\omega - \omega_0)}{1+I/I_0} + \frac{\gamma/4}{(\gamma/2)^2 + (\omega - \omega_0)^2} \right. \\ \left. + \frac{3\gamma/16}{(3\gamma/4)^2 + (\omega - (\omega_0 + \Omega_R))^2} + \frac{3\gamma/16}{(3\gamma/4)^2 + (\omega - (\omega_0 - \Omega_R))^2} \right).$$

The sidebands of the spectrum of resonance fluorescence appear exactly at the frequencies of the transient phenomena; see Figures 14.7 and 14.10. One may interpret this observation with continuous relaxation of small perturbations induced by quantum fluctuations.

are two sidebands at  $\omega_L = \omega_0 + \Omega_R$  and  $\omega_L = \omega_0 - \Omega_R$ , respectively. The spectral triplet is called *Mollow triplet* [192] and agrees very well with the measurements for large intensities ( $I/I_0 \gg 1$ ) shown in Figure 14.7.

We have so far found the position of the spectral lines but not yet obtained information about their line shapes. As outlined in Section 14.4.2, the theoretical calculation of the line shapes is based on the Onsager–Lax theorem but exceeds the scope of the present book. Instead, we only quote the result following the discussion by Walls and Milburn [136] and add some physical interpretation.

The spectrum of resonance fluorescence  $I_{\text{fl}}(\omega)$  is calculated by taking

## 14.5 Light Fields in Quantum Optics

Up to now our characterization of light fields has been dominated by their spectral properties. A *thermal field* (also *chaotic field*) has a broad spectrum and shows strong amplitude fluctuations, while laser light exhibits high spectral purity and small amplitude variations, as described in Section 9.4.1. The noise properties of light fields were of obvious importance already in the sections on detectors, Sects. 11.3.1 and 11.3.2. Here we want to extend our discussion to the quantum nature of those types of light fields that are of fundamental importance in experiments. For this purpose we heuristically extend the formal tools for the description of quantum fields. For more rigorous theories that require, for example, a quantum theory of the generation of photoelectrons in photomultipliers, we refer to presentations like [85, 136, 193]. An extensive overview of experimental work is found in [194].

### 14.5.1 Fluctuating Light Fields

An idealized classical light field has fixed amplitude and phase. All real light fields are subject to fluctuations, however, in both amplitude and phase. These variations are easily caused by disturbances of the light source originating in environmental conditions such as acoustic vibrations or temperature modifications. The technical noise can be eliminated by suitable servo-control measures, albeit the implementation may be very involved.

Beyond the technical noise all light fields exhibit intrinsic fluctuations caused by the quantum nature of the field. They are responsible for the physical limits of the ultimate physical coherence properties of a light source. For their characterization we have already introduced correlation functions in Section 6.2. For instance, the longitudinal or temporal coherence is measured with a Michelson interferometer (Section 6.4). It is quantitatively determined by measuring the interference contrast – the visibility (Eq. (6.5)) – as a function of the difference in the length of the two arms. The temporal first-order correlation function  $G_{EE}(\tau)$  (Eq. (14.12)) has already played a central role in the theoretical treatment of the spectrum of resonance fluorescence. It was shown that classical electrodynamics and QED can lead to different predictions for the spectral properties, documenting the quantum nature of light. In this section we extend the concepts characterizing fluctuations by introducing the second-order coherence, which allows to unambiguously distinguish light fields of classical and nonclassical character.

#### 14.5.1.1 First-Order Coherence

Consider an electromagnetic field  $\hat{E}(t)$  superposed with itself at some detector with a time delay  $\tau$ , measuring  $\hat{E}(t)^*\hat{E}(t + \tau)$ . This can be achieved, for instance, with a Michelson interferometer (Section 6.4) with path length difference  $d = c\tau$ . In analogy with classical electrodynamics (Eq. (14.14)), the photodetector records the (normalized) signal

$$g^{(1)}(\tau) = \frac{\langle \hat{E}^*(t)\hat{E}(t + \tau) \rangle}{\langle \hat{E}^*(t)\hat{E}(t) \rangle} = \frac{\langle \hat{a}^\dagger(t)\hat{a}(t + \tau) \rangle}{\langle \hat{a}^\dagger(t)\hat{a}(t) \rangle}. \quad (14.21)$$

(Here  $\langle \cdot \cdot \cdot \rangle$  symbolizes both the calculation of expectation values and temporal averaging.) The first-order coherence function is evaluated exactly like the visibility (Eq. (6.5)) defined in Section 6.2.1 by analyzing the interference contrast. The first-order coherence can be different for quantum fields and their classical counterparts. It is difficult though to experimentally identify the difference because in both cases we have

$$0 \leq |g^{(1)}(\tau)| \leq 1.$$

There is thus no unambiguous signature in first-order coherence for a typical quantum field. Again in analogy with the definition of visibility for conventional interferometer, fields with  $|g^{(1)}| = 1$  are called coherent or more precisely *first-order coherent*, with  $|g^{(1)}| = 0$  called *incoherent*.

#### 14.5.1.2 Second-Order Coherence

Very prominent differences with unique signatures for nonclassical field states are found for the so-called coherence of the second order. It is straightforwardly defined for classical as well as for quantum fields by extending the definition of first-order coherence. For simplicity we concentrate on the case of a signal generated at a single location. For classical fields, the correlation function of the second-order corresponds to the intensity–intensity correlation function. It is measured by comparing the intensity recorded with a detector at different times  $\{t, t + \tau\}$ :

$$g_{\text{cl}}^{(2)}(\tau) = \frac{\{E^*(t)E^*(t + \tau)E(t + \tau)E(t)\}_t}{\{E^*(t)E(t)\}_t} = \frac{\{I(t + \tau)I(t)\}}{\{I(t)\}^2}. \quad (14.22)$$

Writing the instantaneous intensity  $I(t) = \{I\} + \delta I(t)$  with average value  $\{\delta I(t)\} = 0$ , one finds

$$\begin{aligned} \{I(t + \tau)I(t)\} &= \{(\{I\} + \delta I(t + \tau))(\{I\} + \delta I(t))\} \\ &= \{I\}^2 + \{\delta I(t + \tau)\delta I(t)\} \\ &\rightarrow \{I\}^2 + \{\delta I(t)^2\} \quad \text{for } \tau \rightarrow 0. \end{aligned}$$

For  $\tau = 0$  the quantity  $\{\delta I(t)^2\} > 0$  gives precisely the variance of the intensity fluctuations. Thus we have for classical fields, e.g. thermal fields (Figure 14.11),

$$1 \leq g_{\text{cl}}^{(2)}(\tau = 0) \leq \infty. \quad (14.23)$$

Very generally we also expect that for large times  $\tau$  any correlation is lost, that is,  $\{\delta I(t + \tau)\delta I(t)\} = 0$  and

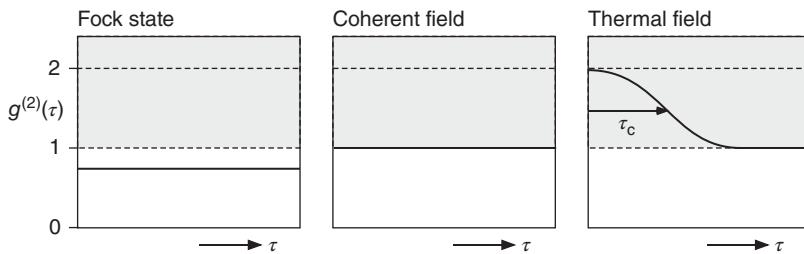
$$g^{(2)}(\tau) \rightarrow 1 \quad \text{for } \tau \rightarrow \infty.$$

It is worth noting that the result in Eq. (14.23) holds for  $\tau = 0$  only; otherwise we can only infer from the positive definiteness of the intensity and Eq. (14.22):

$$0 \leq g_{\text{cl}}^{(2)}(\tau = 0) \leq \infty \quad \text{for } \tau \neq 0$$

In quantum optics, the normalized second-order correlation function has a form that is conveniently written with annihilation and creation operators:

$$g^{(2)}(\tau) = \frac{\langle \hat{a}^\dagger(t)\hat{a}^\dagger(t + \tau)\hat{a}(t + \tau)\hat{a}(t) \rangle}{\langle \hat{a}^\dagger(t)\hat{a}(t) \rangle^2}. \quad (14.24)$$



**Figure 14.11** Theoretical prediction for the second-order correlation function  $g^{(2)}(\tau)$  for different types of light fields. The region  $0 \leq g^{(2)}(\tau) < 1$  is only accessible with nonclassical light fields. Details of the different kinds of light fields are explained in Section 14.5.2.

For  $\tau = 0$  we use the number operator  $\hat{n} = \hat{a}^\dagger \hat{a} = \hat{a}\hat{a}^\dagger + 1$  and find

$$g^{(2)}(\tau = 0) = \frac{\langle \hat{n}(\hat{n} - 1) \rangle}{\langle \hat{n} \rangle^2}.$$

Here, all expectation values are calculated from products of operators with their Hermitian conjugates. They are hence positive definite. In contrast to the case of the second-order correlation function for classical fields, however, no further general properties are found, as a consequence of, for instance, the non-commutativity of the operators. We thus have

$$0 \leq g^{(2)}(\tau = 0) \leq \infty. \quad (14.25)$$

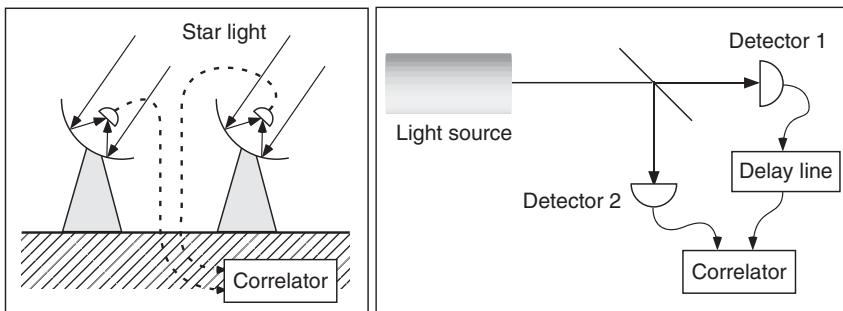
By comparison with Eq. (14.23), we can now give a criterion by which we can unambiguously distinguish fields with predominant classical and quantum character:

$$0 \leq g_{\text{non-class.}}^{(2)}(\tau = 0) < 1.$$

Light fields that fulfill this condition are called *nonclassical light fields*.

#### 14.5.1.3 Hanbury Brown and Twiss Experiment

An important experimental concept for the observation and analysis of fluctuating light fields was proposed in 1956 by the Australian astronomers Hanbury Brown and Twiss [195]. Their setup was stimulated by the intention to determine the diameter of stars by analyzing intensity correlations in the light fields received from the same star but with two separate telescopes. The idea was to overcome the resolution limit of conventional telescopes, which is insufficient to directly observe the size of a star. As explained in Section 5.6.1, the light of a star is equivalent to a point-like source, that is, a source with perfect transverse coherence. When the light fields received from a single star by two independent telescopes are superposed, interference should take place in analogy with Young's double-slit experiment. Also, the interference contrast is expected to vanish once the separation of the two telescopes exceeds  $d > \lambda z_s / 2\pi D$  (see Section 6.3.1) where  $\lambda$  denotes the wavelength of observation,  $z_s$  the distance, and  $D$  the diameter of the star. For a star with known separation, then the diameter could be inferred from a measurement of the interference contrast.



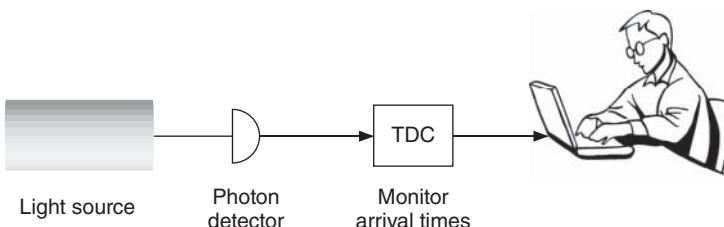
**Figure 14.12** Schematic of the Hanbury Brown and Twiss experiment. On the left we show the original idea of the experiment, intended to determine the diameter of a star by means of intensity–intensity correlations from two separate telescopes.

In the initial experiments it was indeed attempted to directly measure the first-order coherence function by means of a Michelson interferometer (Section 6.4), that is, by directly observing the interference of the fields received by the telescopes. This method was, however, severely impaired by wave front disturbances caused by atmospheric turbulence relevant already at a few meters of separation of the telescopes. The setup by Hanbury Brown and Twiss overcomes this problem since the measured intensities are not sensitive to phase disturbances and can even be compared *a posteriori* as a function of delay time  $\tau$ . The classic experimental scheme is shown in Figure 14.12. The field radiated by some light source is split, and two detectors record the intensity as a function of time. An electronic correlator (e.g., an electronic multiplier) then calculates  $g^{(2)}(\tau)$ .

A more modern variant of the Hanbury Brown and Twiss setup is shown in Figure 14.13. In this case, the arrival times of photons are recorded. From this record one determines the conditional probability to detect a second photon if a first one was already observed as a function of delay time (compare Section 14.5.4.2 and Figure 14.16).

#### 14.5.2 Quantum Properties of Important Light Fields

Let us apply the concepts that we have developed for the description of the quantum properties of optical fields to different types of light fields. In the following,



**Figure 14.13** Hanbury Brown and Twiss experiment with modern equipment. A time-to-digits converter (TDC) records the arrival times of photons. The analysis of the correlations is carried out *a posteriori* by a computer program.

we will take isolated resonator modes or Gaussian beams with a pure transverse  $\text{TE}_{mn}$  mode as a realization of quantum states of an optical light field.

#### 14.5.2.1 Fock States or Number States

The eigenstates  $|n\rangle_{\mathbf{k},\epsilon}$  of the number operator,  $\hat{n}_{\mathbf{k},\epsilon} = \hat{a}_{\mathbf{k},\epsilon}^\dagger \hat{a}_{\mathbf{k},\epsilon}$ , are called Fock states with

$$\hat{n}_{\mathbf{k},\epsilon}|n\rangle_{\mathbf{k},\epsilon} = n|n\rangle_{\mathbf{k},\epsilon}.$$

We have shown in Eq. (14.6) that Fock or number states<sup>4</sup>  $|n\rangle$  do not have a definite amplitude,  $\langle n|\hat{E}|n\rangle = 0$ . The average photon number, on the other hand, is uniquely defined,  $\bar{n} = n$ . This is confirmed by the vanishing variance  $\Delta n^2 = \langle n|\hat{n}^2|n\rangle - \langle n|\hat{n}|n\rangle^2 = 0$ .

For the coherence properties of the first and second orders, we calculate straightforwardly from Eqs. (14.21) and (14.24)

$$|g^{(1)}(\tau)| = 1$$

$$g^{(2)}(\tau) = 1 - 1/n < 1.$$

The result shows that the Fock state exhibits first-order coherence. Furthermore we have strictly nonclassical character because of  $g^{(2)}(\tau) < 1$ , that is, there is no classical analogue of this field state. Although Fock states seem to be the natural quantum states associated with the field and number operators of the electromagnetic field, their experimental realization is not so simple and an active field of research.

#### 14.5.2.2 Coherent Light Fields and Laser Light

The classical concept of an electromagnetic wave with amplitude and phase has proven overwhelmingly successful within the wave theory of light. As we have explained already in Section 14.2, there is a close formal analogy of electromagnetic oscillations and harmonically bound particles with mass. In 1926 E. Schrödinger discovered the so-called coherent states (also *Glauber states*, named after R. Glauber (born 1925, Nobel Prize 2005)). With coherent states an excellent approximation of a classical harmonic oscillator described with amplitude and phase is obtained. This concept was applied by R. Glauber to electromagnetic field states in the early 1960s [193].

The coherent state is constructed to be an eigenstate of the non-Hermitian annihilation operator  $\hat{a}$  (Eq. (14.4)):

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle. \quad (14.26)$$

Seeking an expansion of this state in Fock states  $|n\rangle$ ,

$$|\alpha\rangle = c \sum_n \langle n|\alpha\rangle|n\rangle,$$

we use Eq. (14.5),

$$\langle n|\alpha\rangle = \frac{1}{\sqrt{n!}} \langle 0|\hat{a}^n|\alpha\rangle = \frac{\alpha^n}{\sqrt{n!}} \langle 0|\alpha\rangle,$$

<sup>4</sup> In most cases we will now drop the indices  $\{\mathbf{k}, \epsilon\}$ , which identify a specific mode. It is in general clear which mode is meant.

and find

$$|\alpha\rangle = c \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle.$$

From the normalization condition  $\langle\alpha|\alpha\rangle = 1$  and the Taylor series  $\sum_n |\alpha|^{2n}/n! = \exp(-|\alpha|^2)$ , we can then directly give the expansion of the coherent state in terms of Fock states:

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (14.27)$$

It is not difficult to calculate the coherence properties of the coherent state:

$$|g^{(1)}(\tau)| = 1,$$

$$g^{(2)}(\tau) = 1.$$

It is obvious to use coherent states as the appropriate quantum states for the description of laser light. Because of their enormous significance, let us collect some important properties:

- *Average photon number.* The average photon number is directly calculated from the definition (Eq. (14.26))

$$\bar{n} = \langle\alpha|\hat{n}|\alpha\rangle = \langle\alpha|\hat{a}^\dagger\hat{a}|\alpha\rangle = |\alpha|^2 \quad \text{and} \quad |\alpha| = \sqrt{\bar{n}}.$$

- *Variance of the photon number.* We have  $\langle\alpha|\hat{n}^2|\alpha\rangle = |\alpha|^4 + |\alpha|^2$ , and thus one finds

$$\Delta n^2 = \langle\alpha|\hat{n}^2|\alpha\rangle - \langle\alpha|\hat{n}|\alpha\rangle^2 = |\alpha|^2 \quad \text{or} \quad \Delta n = \sqrt{\bar{n}}.$$

- *Coherent states are almost orthogonal.* It is not surprising that the coherent states are not orthogonal because they are eigenstates to a non-Hermitian operator. However, for reasonably large  $\alpha$  and  $\beta$ , they are almost orthogonal; the factor

$$\langle\alpha|\beta\rangle = \exp\left(-\frac{1}{2}|\alpha|^2 - \frac{1}{2}|\beta|^2 + \alpha^*\beta\right)$$

vanishes rapidly if  $\alpha$  and  $\beta$  are just slightly different.

- *Minimum uncertainty states – phase and amplitude.* One can show that the variances of the quadratures of the electromagnetic field,  $\hat{X}$  and  $\hat{Y}$  (see Section 14.2) [136], are independent of the amplitude of the coherent state,  $\alpha = |\alpha|e^{i\phi}$ . Thus the coherent state should be considered the quantum analogue of the classical field with amplitude and phase.
- *Photon number distribution.* The probability to find  $n$  photons in a given mode of the radiation field (i.e., to record a certain number of counts in a certain time interval) is

$$p_n = \Delta n^2 = \langle n|\alpha\rangle = \exp(-|\alpha|^2) \frac{|\alpha|^{2n}}{n!} = \exp(-\bar{n}) \frac{\bar{n}^n}{n!}.$$

This distribution corresponds precisely to the Poisson distribution.

### 14.5.2.3 Thermal Light Fields

A thermal light field – the type of light field for more or less all natural light sources, also called chaotic light field – may be represented by its photon number distribution,

$$p_n = \frac{\bar{n}^n}{(1 + \bar{n})^{1+n}}, \quad (14.28)$$

where  $\bar{n}$  is the average photon number. Experimental observation with an incandescent lamp is quite difficult, though. The typical time scale of fluctuation is characterized by the coherence time  $\tau_c$  of the light source, which is roughly inversely proportional to the bandwidth of the light source. For a white spectrum lamp, this would require detectors to operate at the scale of a few optical cycles in the femtosecond domain. At large averaging times, on the other hand, when rapid fluctuations are averaged out, the distribution (14.28) approaches the Poisson distribution valid for coherent states only.

The average photon number has been fixed by definition of (14.28), and the variance of the field is calculated from this distribution:

$$\langle \hat{n} \rangle = \bar{n} \quad \text{and} \quad \Delta n^2 = \bar{n}^2 + \bar{n}.$$

One can furthermore show [85] that the first-order coherence functions for the two most important spectral line shapes, Gaussian and Lorentzian, with bandwidths  $\Delta\omega \simeq \tau_c^{-1}$  are given by

$$\begin{aligned} g^{(1)}(\tau) &= \exp(-i\omega\tau - (\tau/\tau_c)^2/2) && \text{Gaussian shape,} \\ g^{(1)}(\tau) &= \exp(-i\omega\tau - |\tau/\tau_c|) && \text{Lorentzian shape.} \end{aligned}$$

The first-order coherence function is indeed identical for the classical and the quantum description of thermal fields.

An interesting relation can be shown for the second-order coherence function of a fluctuating chaotic light field [85]:

$$g^{(2)}(\tau) = 1 + |g^{(1)}(\tau)|^2.$$

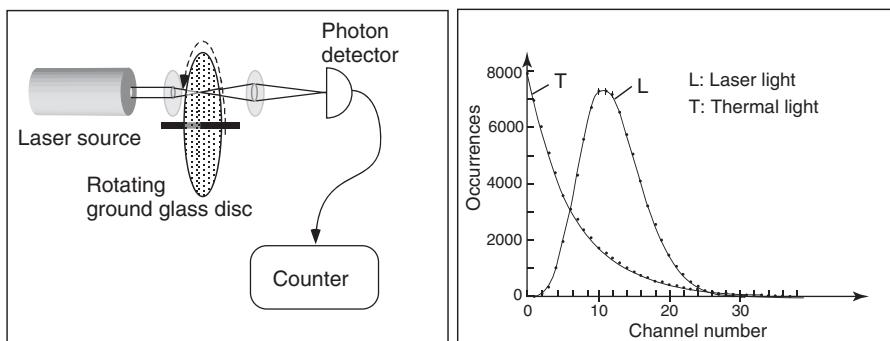
The special case  $\tau = 0$  yields for a light field showing perfect first-order coherence for short times

$$g^{(2)}(\tau = 0) = 2.$$

This situation is called *bunching*. It was demonstrated with a so-called pseudo-thermal light field [196]. It is discussed in Section 14.5.4.1 and shown in Figure 14.15.

### 14.5.3 Photon Number Distribution

In an experiment, the photon number distribution is measured by recording the number of photons, or more precisely the number of photoelectrons, generated in a detector as a function of the time interval  $T$ . In the experiment it is furthermore important to select a single spatial mode of the radiation field as well.



**Figure 14.14** Photon number distribution for a (pseudo-)thermal and for a coherent laser beam. In this experiment from 1965, a HeNe laser was used as the light source. The pseudo-thermal light was generated by focusing the HeNe laser beam onto a  $20\text{ }\mu\text{m}$  diameter spot on a rotating ground glass disc with typical length scale  $3\text{ }\mu\text{m}$  of the random surface structure. (After [84].)

For a laser beam this is simply the Gaussian beam, for example, a  $\text{TEM}_{00}$  mode. From an extended thermal source, one has to equivalently select a sufficiently small (point-like) surface area by apertures to warrant transverse coherence (see Section 6.3.1), resulting in low intensity. From repeated measurements the distribution of occurrences, that is, the histogram in Figure 14.14, can be extracted.

As mentioned earlier, the experimental observation of the distinctions between different types of light fields is challenging due to the fluctuations that are inversely proportional to the bandwidth of the light source. Hence they occur at very short time scales in particular for thermal sources. The interesting part of the information is then easily averaged out for longer integration times  $T \gg \tau$  of the detector. In Sections 11.3.1 and 11.3.2 we have already noted that laser light and thermal light have no measurable difference with conventional setups and detectors.

Already in the early days of the laser, the so-called *pseudo-thermal* light sources were used to simulate the fluctuations of a thermal light beam by manipulating a coherent laser beam. In the experiments by Arecchi and coworkers [84], laser light was focused for this purpose onto a rotating ground glass disc. At rest, diffraction and scattering from the rough surface of the disc cause the speckle pattern that was described in Section 6.9. For a rotating disc, the intensity at the detector shown in Figure 14.14 fluctuates at an ever higher rate with the rotation speed of the glass disc. Controlling the rotation speed, and for the conditions described in Figure 14.14, allows one to set the effective coherence time of the transmitted light to some  $50\text{--}1000\text{ }\mu\text{s}$ , which is a convenient domain for experiments.

Incidentally, Purcell [197] had predicted before that these noise properties should also be compatible with the fluctuations of classical thermal and even laser light sources. The experiments by Arecchi, however, showed that the statistical distribution of photoelectrons is correctly described by the theory of QED, which was applied to this case by Glauber [193].

### 14.5.4 Bunching and Anti-bunching

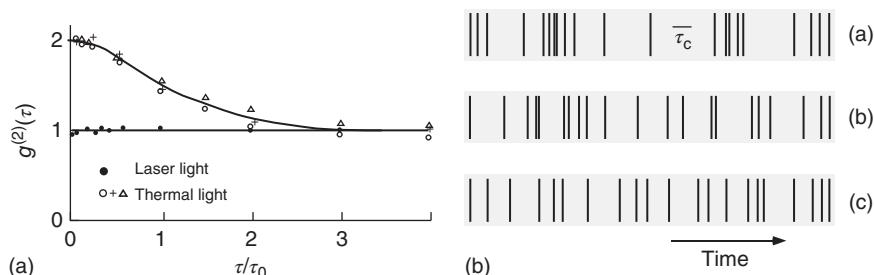
#### 14.5.4.1 Bunching

The pseudo-thermal light source that was used to measure the photon number distribution of Figure 14.14 can also be used to experimentally determine the coherence or correlation function of the second order,  $g^{(2)}(\tau)$  from Eq. (14.22). The result is shown in Figure 14.15 for the case of laser light and again for pseudo-thermal light. Here, the delay time  $\tau$  was normalized to the effective coherence time  $\tau_0$ , which was controlled by the rotation speed of the ground glass disc.

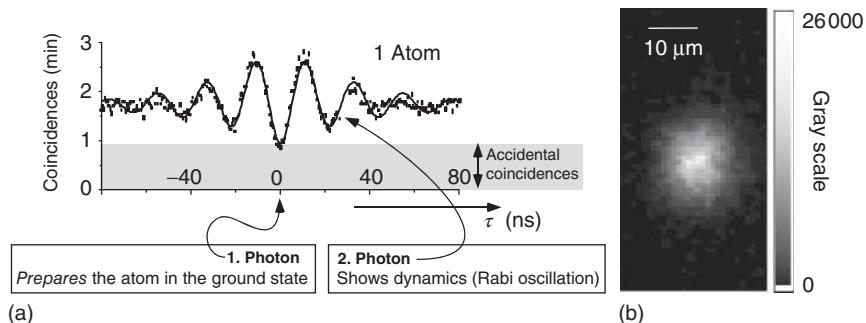
To the right of the experimental result, we show examples of series “clicks” of a photon counter, which are interpreted as photoelectrons generated by photons recorded with a detector. Photons from a coherent (laser) light field are recorded at fully random times, that is, there is no correlation, no “memory” of the arrival of a previous photon. One may imagine that the photons arrive at the detector like raindrops on a tin roof, which also arrive randomly. A thermal light field, however, shows the *bunching* phenomenon. Once a photon has been recorded, it is more likely to immediately receive a second photon than at later times. The phenomenon of *anti-bunching*, in contrast, describes special light sources where the photons seem to “repel” each other. In this case, the arrival times of photons are more evenly spread out, more regular than for the fully random case.

#### 14.5.4.2 Anti-bunching

A single atom is a very special, simple microscopic light source. In Section 14.4.1 we have treated its spectral properties. The so-called *anti-bunching* phenomenon occurs in resonance fluorescence and shows a clear signature for a nonclassical light field, that is,  $g^{(2)}(\tau) < 1$ . Neglecting the background the arrival of a photon in this experiment carries the information that an atom was excited and is now in the ground state. The experimental data shown in Figure 14.16 are thus a measure of the probability of finding an atom in the excited state again if a photon was



**Figure 14.15** (A) Measurement of the normalized second-order coherence function for coherent (black dots) and thermal light sources. Pseudo-thermal light was generated by the method shown in Figure 14.14. The different symbols are associated with the rotation speed controlling the effective coherence time  $\tau_0$ . (Arecchi *et al.* 1966 [196]. Reproduced with permission of Elsevier.) (B) Examples of photoelectron series (“clicks”) for different types of photon statistics: (a) thermal light source, (b) laser light source, and (c) light source with anti-bunching effect. Compare with Figure 14.11. (After [85].)



**Figure 14.16** (a) Raw data (dots) of the rate of photon coincidences showing *anti-bunching* in resonance fluorescence of a single neutral cesium atom stored in a magneto-optical trap (see excursion on p. 400) The oscillations can be interpreted in terms of Rabi oscillations. (b) Image of a fluorescing single Cs atom (wavelength 852 nm) taken with an ICCD camera. Exposure time 1 s. (After [198].)

recorded at time  $\tau$  before the current event. There are two properties that are particularly conspicuous.

After subtraction of the background of random coincidences, no further photons are recorded for delay time  $\tau = 0$ . This observation can be interpreted with the excitation dynamics of the atom. It must be re-excited before another photon can be emitted. In the language of the quantum theory of measurement, one can also say that recording of a fluorescent photon projects or prepares the atom in its ground state.

With increasing delay time  $\tau$ , the conditional probability to observe a second photon shows Rabi oscillations that are damped within some 30 ns – the lifetime of the excited state of the cesium atoms used in this experiment – to the equilibrium value. The atom is strongly saturated in this case. The corresponding spectrum of resonance fluorescence (see Section 14.4.1) would hence show sidebands corresponding to the oscillations observed here in the time domain; compare Figure 14.7.

The first anti-bunching experiment was carried out with an extremely diluted atomic beam [199]. Today this experiment can be carried out with single trapped atoms (Figure 14.16) [198], ions [200], or also solid-state sources [201, 202], which do not need extensive equipment for storage of single microscopic particles. For the observation of anti-bunching, it is mandatory to prepare an individual microscopic system such as a single atom or ion. In an ensemble of  $N$  fluorescing atoms, the rate of random coincidences grows like  $N(N - 1)$ ; hence the anti-bunching signal vanishes rapidly within the background.

## 14.6 Two-Photon Optics

In previous sections we have studied photon correlation phenomena that uncover information about the fluctuation properties of light fields. A typical experiment measures photon coincidences as a function of delay time, and the typical separation of two individual photons must be large enough (at least tens of picoseconds)

to be measurable with the experimental equipment. Although there exist correlations between photons, it remains impossible to predict the *arrival time* of an individual photon at the detector.

Deterministic experimental control of single photons requires exactly that: a light source that allows the generation of a single photon states at a well-defined instant in time and propagation of this photon state from one point to another. Such deterministic *single-photon sources* are very desirable devices and an active field of laboratory research, which in the future may offer technically robust solutions. Since about 1995 efficient two-photon sources have been available offering devices where loosely speaking pairs of photons are simultaneously generated but propagate in spatially well-separated directions. The photon pairs – more precisely one should speak of a two-photon quantum state – are still generated at random. Since they travel on different paths, however, detection of one of the photons may serve as a flag to indicate the presence of the other one.

The experimental realization of two-photon sources has made numerous experiments possible where the quantum properties of the two-photon state can be used very efficiently today. One example is the field of *quantum cryptography*, which allows fundamentally secure transmission of messages. The development of two-photon light sources relying on spontaneous parametric down-conversion (SPDC) was one of the achievements of L. Mandel (1927–2001), one of the pioneers of quantum optics.

#### 14.6.1 Spontaneous Parametric Fluorescence, SPDC Sources

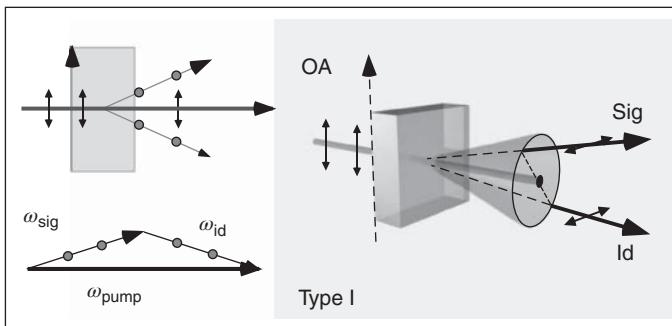
Spontaneous parametric fluorescence or SPDC may be considered the elementary process of the parametric oscillator, which is discussed in detail in Section 15.6. Here we restrict ourselves to a grossly simplified description. In a nonlinear material driven by a strong monochromatic laser field (frequency  $\omega_0$ ), a polarization is induced, which causes fluorescence of photon pairs with frequencies ( $\omega_1, \omega_2$ ). Of course, the material must be transparent for all relevant wavelengths. In a simplified interpretation the conditions for photon pair production to be fulfilled are conservation of energy ( $\hbar\omega_0 = \hbar\omega_1 + \hbar\omega_2$ ) and momentum (“phase matching”) in the crystal  $\hbar\mathbf{k}_0 = \hbar\mathbf{k}_1 + \hbar\mathbf{k}_2$ :

$$\omega_0 = \omega_1 + \omega_2,$$

$$(n_0\omega_0/c) \mathbf{e}_0 = (n_1\omega_1/c) \mathbf{e}_1 + (n_2\omega_2/c) \mathbf{e}_2.$$

In isotropic media it is impossible to meet all conditions simultaneously because of dispersion. For normal dispersion we have typically  $2n_0 > n_1 + n_2$ . A solution is offered by birefringent materials (for details see Section 15.4.3). For type I configuration both fluorescence photons have orthogonal polarization with respect to the driving light field. For type II their polarizations are mutually parallel and orthogonal. Momentum conservation requires a small angle between the directions of emission, which is exaggerated in Figure 14.17.

The geometry of phase matching for SPDC fluorescence is schematically shown in Figure 14.17. The nonlinear crystal is driven with a focused beam of short-wavelength laser light. For symmetry reasons the allowed directions of the fluorescence photons form a conical surface. For the type I configuration,



**Figure 14.17** Spontaneous parametric down-conversion for type I phase matching. For historical reasons the two fluorescence photons are labeled *signal* (*sig*) and *idler* (*id*) photon. OA: optical axis.

the indices of refraction of the two colors  $\omega_1$  and  $\omega_2$  differ only slightly due to dispersion, and the cones are perfectly superposed for the degenerate case  $\omega_1 = \omega_2$ . In order to filter individual colors out of the rainbow-like spectrum, one can use interference filters or apertures. A case of particular interest occurs if two photons with identical color are emitted. Such photons that can be prepared by an interference filter at twice the wavelength of the pump laser can interfere with each other.

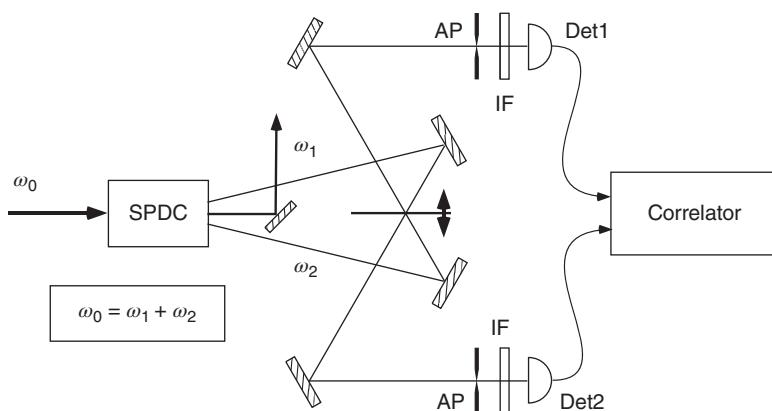
With this two-photon SPDC source, numerous experiments were carried out, which occasionally stretch our imagination [203]. Here we focus our attention on the experiment by Hong *et al.* [204], which was the first one to demonstrate the interference properties of the two photons. Later on we will introduce the extended two-photon source invented by Kwiat and coworkers [205]. Within only 10 years, it has become a standard photon light source in quantum optics. It is more and more simply called an *SPDC source*.

#### 14.6.2 Hong–Ou–Mandel Interferometer

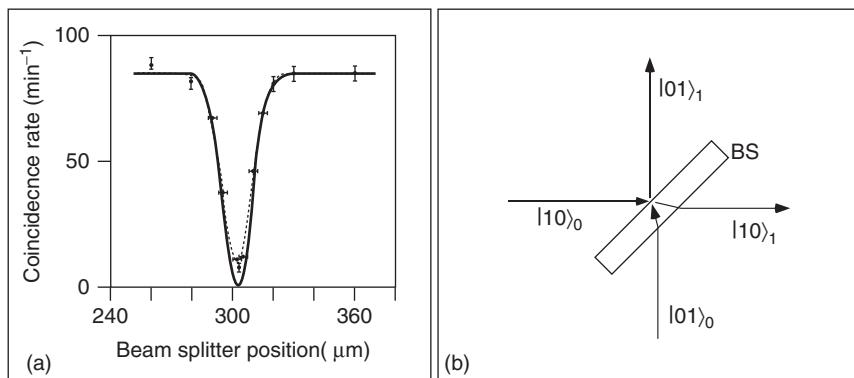
The interferometer setup conceived by L. Mandel and his coworkers is shown in Figure 14.18. The two photons generated by the SPDC source are superposed on the BS in analogy with a Michelson interferometer. A detector in each exit records the arrival of photons. A valid coincidence is recorded only if a valid event has taken place simultaneously in both arms, that is, if and only if the two photons are distributed onto the two detectors.

The rate of coincidences is measured as a function of the difference in path length taken by the photons on their route to the BS. Here, the beam splitter itself (BS in Figure 14.18) is slightly displaced for this modification. The apertures and the interference filters serve to single out indistinguishable fluorescence photons with identical color. The result – the rate of coincidences as a function of the path difference – is shown in Figure 14.19.

The rate of coincidences shows a sharp drop when the path length for the two photons is identical, that is, at zero path difference. In this case the two photons arrive simultaneously at the BS. Since no coincidences are found, they must have traveled together in either of the two arms, activating only one of the detectors.



**Figure 14.18** Two-photon interferometer after Hong, Ou, and Mandel. SPDC: spontaneous down-conversion source for the generation of photon pairs (see Figure 14.17); BS: beam splitter; AP: apertures; IF: interference filter; Det: detectors.



**Figure 14.19** (a) Coincidence rate as a function of the beam splitter position (BS in Figure 14.18) in the Hong–Ou–Mandel interferometer. (After [204].) (b) Designations of the quantum field states in Figure 14.18.

If the photons arrive separately at the detector, they are transmitted with 50% probability into either of the two arms. Thus we expect to find coincidences in 50% of all possible cases.

Let us consider a simple model for the interference effect of Figure 14.19. On arrival of a photon wave packet in state  $|01\rangle_0$ , a departing wave packet  $(|01\rangle_1 + |10\rangle_1)/\sqrt{2}$  is generated at the BS. A wave packet arriving in state  $|10\rangle_0$  is transformed into  $(|01\rangle_1 - |10\rangle_1)/\sqrt{2}$ , since the reflection at the BS takes place at the thicker and the thinner medium, respectively. With field operators

$$\hat{a}_1^\dagger |00\rangle_1 = |10\rangle_1 \quad \text{and} \quad \hat{a}_2^\dagger |00\rangle_1 = |01\rangle_1,$$

we can describe the transformation of the incoming single-photon states by the 50 : 50 BS into the outgoing photon states by

$$\hat{s}_1 = (\hat{a}_1^\dagger + \hat{a}_2^\dagger)/\sqrt{2} \quad \text{and} \quad \hat{s}_2 = (\hat{a}_1^\dagger - \hat{a}_2^\dagger)/\sqrt{2}.$$

Two-photon states arriving simultaneously at the detector can then be described by the product operator  $\hat{s}_1\hat{s}_2$ , and we find the new two-photon state for perfect interference:

$$\hat{s}_1\hat{s}_2|00\rangle_1 = \{(\hat{a}_1^\dagger)^2 - (\hat{a}_2^\dagger)^2\}|00\rangle_1/2 = (|20\rangle_1 - |02\rangle_1)/2.$$

Quantum interference obviously causes *both* photons to propagate either in one or in the other arm, but definitely not distributed into the two arms. This interpretation precisely agrees with observation, the now so-called Hong–Ou–Mandel dip shown in Figure 14.19. The interference is only perfect for exact temporal coincidence of the photons at the BS. If the wave packets do not overlap, the photons do not meet at the BS. Then, each photon is transmitted with 50% probability into the two interferometer arms, yielding a coincidence measurement in 50% of all possible cases.

Finally, we have to explain the effective length of the photon wave packets. The interference contrast in Figure 14.19 vanishes for a path length difference of  $\Delta x/2 = 16 \mu\text{m}$  (the factor of 2 takes into account that the displacement acts on both arms) corresponding to a delay time  $\Delta\tau = \Delta x/c \simeq 100 \text{ fs}$ . This time corresponds to the (inverse) spectral width of the interference filters, which were used for the preparation of photons of identical color.

## 14.7 Entangled Photons

### 14.7.1 Entangled States According to Einstein–Podolsky–Rosen

One of the best-known – and very often misunderstood – paradoxes in physics is the Einstein–Podolsky–Rosen paradox (*EPR paradox*), which was named after the authors of a famous publication from 1935: Einstein *et al.* [206]. Until the 1990s, this topic was preferentially considered a curiosity rather than an important discovery in physics. This situation prevailed even though theoretical work by Bell [207] had pointed out that the predictions by Einstein and his colleagues, motivated by classical interpretations of physics, could be cast into a theory with quantitative predictions distinguishing the quantum and the classical world. A first experimental test was realized with the work by Aspect and his coworkers in 1981 [208], which confirmed the prediction of quantum theory and further shattered classical interpretations.

Let us now introduce the EPR paradox itself and the role of entanglement. Its significance for optical experiments was significantly boosted in 1995 when the already mentioned extended parametric two-photon source [205] was announced. It allows efficient generation of entangled photon pairs, that is, two-photon states exhibiting pure quantum correlations. Groundbreaking experiments including quantum communication and quantum teleportation have become possible with this source.

#### 14.7.1.1 The Einstein–Podolsky–Rosen (EPR) Paradox

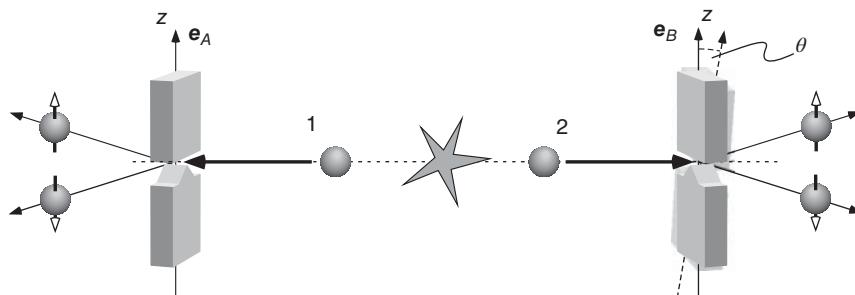
Einstein never questioned the overwhelming success of quantum theory, nor its potential to correctly predict the outcome of physical experiments. However,

he was never satisfied with the wave function description and its probabilistic interpretation. Apparently he was especially irritated by the uncertainty relation. This states that the expectation values of quantities that are represented by noncommuting operators such as position and momentum,  $(\hat{x}, \hat{p})$ , or the components of a spin,  $\hat{\sigma} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$ , can never be predicted with exact certainty. Measurement of one of the quantities seems to always act back on the other one, lending some probabilistic character to that measurement. Einstein considered quantum physics *incomplete*. He argued that there should be some *super theory* reproducing the results of quantum mechanics but in a fully deterministic way. As a consequence he created the following requirements that a physical theory should obey:

If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity. [206])

With a simple two-particle system, Einstein and his coauthors constructed a highly interesting physical situation with the only intention to point out the contradictory nature of quantum theory. With correlated two-particle systems, it seemed that the limitations of the uncertainty principle imposed on the expectation values of conjugated physical quantities could be lifted. A simplified physical example that is today the conceptual basis for most experiments (Figure 14.20) was introduced by Bohm and Aharonov [209]. It is based on a molecule such as  $Hg_2$ , which consists of two spin-1/2 atoms (two-level atoms with the only internal degree of freedom described by the spin-1/2 operators  $\sigma_{1,2}$ ) for atoms 1 and 2. The molecule has total spin 0. Thus the total spin wave function of the (singlet) system can be written as

$$\Psi_{EPR} = \frac{1}{\sqrt{2}}(\psi_+(1)\psi_-(2) - \psi_-(1)\psi_+(2)), \quad (14.29)$$



**Figure 14.20** Gedanken experiment by Bohm and Aharonov [209] illustrating the idea of the EPR paradox. Two inhomogeneous magnetic fields are used to determine the value of the spin components by means of a Stern–Gerlach experiment. In quantum optics, this experiment is realized with entangled photons. Then, the Stern–Gerlach magnets are replaced with polarizers.

where  $\psi_+(1)$  defines the “spin-up” quantum state of atom 1 with spin  $+\hbar/2$  and with  $\psi_-(2)$  for atom 2 correspondingly. When the system is dissociated, the particles separate from each other, but the total angular momentum and hence the quantum state described by (14.29) are preserved. In Figure 14.20 atom 1 moves toward detector  $A$ , and atom 2 toward  $B$ . Once the atoms have traveled some distance, no direct interaction is possible.

If we analyze the spin components of the system with two analyzers, for instance, with the Stern–Gerlach magnets ( $A$  and  $B$  in Figure 14.20), with coordinate systems  $\{\mathbf{e}_A\}$ ,  $\{\mathbf{e}_B\}$  both parallel to the  $z$  direction, we expect perfect correlation: if a measurement for atom 1 in the  $\{\mathbf{e}_A\}$  direction results in “+” (“−”), the result for atom 2 is predicted to be “−” (“+”), without any uncertainty and exactly as demanded by Einstein. Such correlations are actually well known from everyday life: consider two little balls instead of the two-spin system, a white and a black one. Put one of them into your left trouser pocket and the other one into your right pocket. After a while you have completely forgotten which ball resides in which pocket. But taking out one of the balls instantaneously also defines the result for the other pocket – without any mutual disturbance of these “measurements.”

The problem becomes interesting when the analyzers, the measurement apparatus, no longer employ identical basis systems. For instance, one could use for  $\{\mathbf{e}_B\}$  the  $x$  axis, which is orthogonal to  $\{\mathbf{e}_A\}$ . According to Einstein’s argument a measurement at  $B$  would then instantaneously also imply the value for the  $x$ -component of atom 1 spin at  $A$  – in striking contrast to the prediction of quantum mechanics that inhibits this measurement because of the noncommutativity of the spin component operators. Einstein’s assumption of the lack of completeness in quantum mechanics initiated Bohm to develop a theory of so-called hidden variables[209, 210] with the idea of lending deterministic character to quantum theory.

The gedankenexperiment of Bohm does not rely on the application of a molecular or atomic system. The concept can perfectly be realized with any quantum two-level system. In quantum optics, for instance, photon pairs with entangled polarization states are well suited to realize a quantum state consistent with Eq. (14.29). A source of entangled photon pairs now employed in numerous applications is presented in Section 14.7.4.

### 14.7.2 Bell’s Inequality

For more than 30 years, the EPR paradox was considered a mere curiosity. Only when J. Bell in 1964 published the inequality now bearing his name [207] was the route opened to create experimental situations that could quantitatively, by laboratory measurements, distinguish quantum theory from deterministic theories with hidden variables.

The spin components  $\hat{\sigma}_{1,2}$  are measured with analyzers  $A$  and  $B$ , that is, by the scalar products  $\hat{\sigma}_1 \cdot \mathbf{e}_A$  and  $\hat{\sigma}_2 \cdot \mathbf{e}_B$ , respectively. In addition we assume that there is an additional parameter or set of parameters that we call  $\lambda$ . For a two-level system, the measurement can only have the results

$$A(\mathbf{e}_A, \lambda) = \pm 1, \quad B(\mathbf{e}_B, \lambda) = \pm 1. \quad (14.30)$$

If we call  $\rho(\lambda)$  the probability distribution of  $\lambda$ , we can calculate the expectation values of the measurements  $E(\mathbf{e}_A, \mathbf{e}_B)$  from

$$E(\mathbf{e}_A, \mathbf{e}_B) = \int d\lambda \rho(\lambda) A(\mathbf{e}_A, \lambda) B(\mathbf{e}_B, \lambda), \quad (14.31)$$

provided that  $B$  cannot depend on  $\mathbf{e}_A$  and  $A$  not on  $\mathbf{e}_B$ . Furthermore we can make use of the strict anticorrelation for parallel analyzers,  $A(\mathbf{e}_A, \lambda) = -B(\mathbf{e}_A, \lambda)$ . Thus we also have

$$E(\mathbf{e}_A, \mathbf{e}_B) = - \int d\lambda \rho(\lambda) A(\mathbf{e}_A, \lambda) A(\mathbf{e}_B, \lambda).$$

If we introduce yet another base unit vector  $\mathbf{e}_C$ , we can write

$$\begin{aligned} E(\mathbf{e}_A, \mathbf{e}_B) - E(\mathbf{e}_A, \mathbf{e}_C) \\ = - \int d\lambda \rho(\lambda) (A(\mathbf{e}_A, \lambda) A(\mathbf{e}_B, \lambda) - A(\mathbf{e}_A, \lambda) A(\mathbf{e}_C, \lambda)) \\ = \int d\lambda \rho(\lambda) A(\mathbf{e}_A, \lambda) A(\mathbf{e}_B, \lambda) (A(\mathbf{e}_B, \lambda) A(\mathbf{e}_C, \lambda) - 1), \end{aligned}$$

where he have inserted  $A(\mathbf{e}_B, \lambda) A(\mathbf{e}_B, \lambda) = 1$  because of (14.30). The result of the measurement  $A(\mathbf{e}_A, \lambda) A(\mathbf{e}_B, \lambda)$  cannot be smaller than  $-1$ . Thus the inequality

$$|E(\mathbf{e}_A, \mathbf{e}_B) - E(\mathbf{e}_A, \mathbf{e}_C)| \leq \int d\lambda \rho(\lambda) (A(\mathbf{e}_B, \lambda) A(\mathbf{e}_C, \lambda) - 1)$$

must hold. With the definition (14.31) we finally obtain Bell's inequality:

$$1 + E(\mathbf{e}_B, \mathbf{e}_C) \geq |E(\mathbf{e}_A, \mathbf{e}_B) - E(\mathbf{e}_A, \mathbf{e}_C)|. \quad (14.32)$$

The quantum mechanical expectation value  $E_Q$  can be calculated explicitly for the singlet state (14.29), yielding

$$E_Q(\mathbf{e}_A, \mathbf{e}_B) = \int dV \Psi^*(\hat{\sigma}_1 \cdot \mathbf{e}_A)(\hat{\sigma}_2 \cdot \mathbf{e}_B)\Psi = -\mathbf{e}_A \cdot \mathbf{e}_B.$$

We consider the special case  $\mathbf{e}_A \cdot \mathbf{e}_B = 0$ ,  $\mathbf{e}_A \cdot \mathbf{e}_C = \mathbf{e}_B \cdot \mathbf{e}_C = 2^{-1/2}$ , and by insertion into (14.32), we find

$$1 - 2^{-1/2} = 0.29 \geq |0 + 2^{-1/2}| = 0.71.$$

The result is manifestly different for quantum theory and hidden variables theories!

### 14.7.3 Bell's Inequality and Quantum Optics

The orthogonal polarization states of photons had already caught D. Bohm's attention as a good candidate for two-state systems applicable in a test of Bell's inequality. The first optical experiments [208] were carried out in 1981 with pairs of photons that were generated in a cascading decay of two subsequent atomic transitions. Since then experiments demonstrating the violation of Bell's inequality have been made more and more stringent. One reason for this effort is the existence of so-called loopholes, which reduce the significance of an experiment, for instance, the finite detection probability of the detectors. This

discussion has not been completed and exceeds the scope of the present text. Here we restrict ourselves to the presentation of modern experimental concepts.

From the very beginning of experimental studies of Bell's inequalities in most experiments, a variant was analyzed, which had been proposed in 1969 by Clauser and coworkers [211]. They introduced a version that is tolerant with respect to imperfections of the detector. Thus the condition Eq. (14.30) is expressed less rigorously:

$$|A(\mathbf{e}_A, \lambda)| \leq 1, \quad |B(\mathbf{e}_B, \lambda)| \leq 1. \quad (14.33)$$

One can show (see Problem 14.4 and [211–213]) that a theory with hidden variables obeying these conditions must fulfill the Clauser–Horne–Shimony–Holt (CHSH) inequality

$$-2 \leq S(\mathbf{e}_A, \mathbf{e}_{A'}, \mathbf{e}_B, \mathbf{e}_{B'}) \leq 2 \quad (14.34)$$

with the definition

$$S(\mathbf{e}_A, \mathbf{e}_{A'}, \mathbf{e}_B, \mathbf{e}_{B'}) = E(\mathbf{e}_A, \mathbf{e}_B) - E(\mathbf{e}_A, \mathbf{e}_{B'}) + E(\mathbf{e}_{A'}, \mathbf{e}_B) + E(\mathbf{e}_{A'}, \mathbf{e}_{B'})$$

following Eq. (14.31). In an experiment, the expectation values are determined from coincidence measurements occurring with rates  $R_{++}(\mathbf{e}_A, \mathbf{e}_B)$ , and so on:

$$\begin{aligned} &E(\mathbf{e}_A, \mathbf{e}_B) \\ &= \frac{R_{++}(\mathbf{e}_A, \mathbf{e}_B) - R_{-+}(\mathbf{e}_A, \mathbf{e}_B) - R_{+-}(\mathbf{e}_A, \mathbf{e}_B) + R_{--}(\mathbf{e}_A, \mathbf{e}_B)}{R_{++}(\mathbf{e}_A, \mathbf{e}_B) + R_{-+}(\mathbf{e}_A, \mathbf{e}_B) + R_{+-}(\mathbf{e}_A, \mathbf{e}_B) + R_{--}(\mathbf{e}_A, \mathbf{e}_B)}. \end{aligned} \quad (14.35)$$

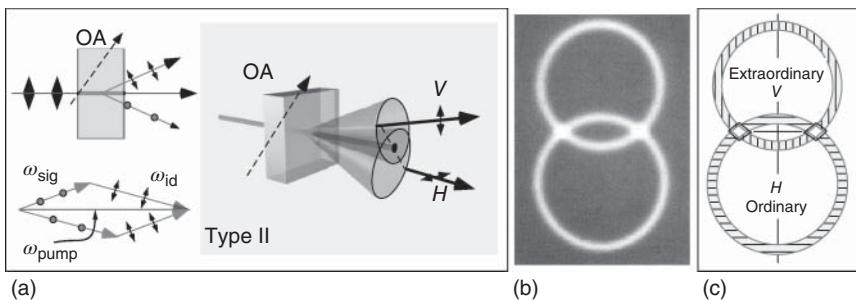
Quite often the experimental setup helps to simplify this relatively complex form. For instance, the results do not depend on a specific orientation  $\mathbf{e}_{A,B}$ , but only on the relative angle  $\mathbf{e}_A \cdot \mathbf{e}_B = \cos \alpha$ .

#### 14.7.4 Polarization-Entangled Photon Pairs

It is not so straightforward to produce two microscopic particles that show entanglement, that is, a non-factorizable quantum state like Eq. (14.29)<sup>5</sup> with nonlocal character. In 1995 Kwiat and his colleagues [205] created an extended concept of the SPDC sources described in Section 14.6.1, which serves now as a robust source of entangled photon pairs.

The fundamental concept is shown in Figure 14.21. In contrast to the source from Figure 14.17, in this case, type II phase matching is employed, which produces photon pairs with orthogonal polarization states. The optical axis is no longer orthogonal to the direction of propagation. It is tilted by an angle that is close to the angle required for collinear phase matching (see details in Section 15.4.3.1). Here one can obtain a situation where the ordinary and the extraordinary fluorescence light is emitted into two cones with diverging axes. The cones thus intersect at exactly two points where both polarizations exist. If one furthermore uses an interference filter to prepare photon pairs with identical

<sup>5</sup> There exist numerous natural entangled states in nature, for example, the singlet ground state of the He atom corresponds to an entangled state of its two electrons. However, this entanglement is not useful for controlled applications since it is impossible to individually manipulate the electrons.



**Figure 14.21** Generation of polarization-entangled photon pairs by spontaneous parametric down-conversion; compare Figure 14.17. (a) Geometry of phase matching: the upper cone has extraordinary (V) and the lower one ordinary (H) polarization. (b) Photograph of spontaneous fluorescence looking straight into the crystal. OA: optical axis. (After [205] and with permission by A. Zeilinger.) (c) Photon pairs emitted into the lined cross sections of the extraordianrily and the ordinarily polarized cones exhibit entanglement.

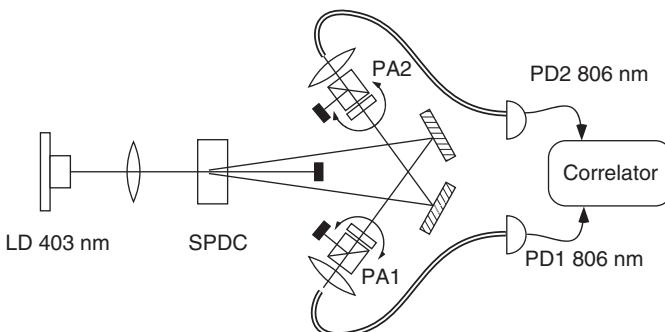
color, the emitted photons are fully indistinguishable. They form a quantum state

$$|\Psi_{\text{EPR}}\rangle = \frac{1}{2^{1/2}} (|H\rangle_1|V\rangle_2 + e^{i\phi}|V\rangle_1|H\rangle_2),$$

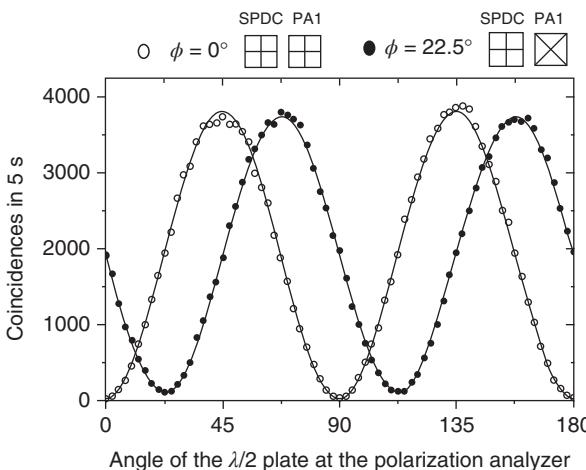
which corresponds precisely to the EPR state suggested by D. Bohm. Since the photons propagate into different directions, one can use retarder plates (e.g.,  $\lambda/2$ ; see Section 3.4.3) to generate further EPR states, for instance,  $(|H\rangle_1|H\rangle_2 + e^{i\phi}|V\rangle_1|V\rangle_2)/\sqrt{2}$ . With the present source one can generate entangled photon pairs at a rate of several 100 kHz and straightforwardly take advantage of the quantum nature of the photon pairs with so-called Bell experiments.

#### 14.7.5 A Simple Bell Experiment

The simple setup shown in Figure 14.22 agrees very well with the concept of Bohm presented in Figure 14.20. Today, blue laser diodes (Section 10.2.1)



**Figure 14.22** Experimental setup [214] for the measurement of polarization correlations of entangled photon pairs. LD: laser diode; SPDC: SPDC source of photon pairs; M: mirrors; PA: polarization analyzer, consisting of a rotatable  $\lambda/2$  plate and a polarizing beam splitter; PD: photodetector operated in counting mode.



**Figure 14.23** Polarization correlations of photon pairs that are measured in two different sets of bases. The axes of the bases for the SPDC source and analyzer 1 (PA1) are indicated above the graphs; see Figure 14.22. (After [214].)

offering some 10 mW output power are sufficient to generate a comfortable rate of two-photon pairs from an SPDC source and realize a Bell experiment with relatively simple means.

The result and shape of the correlation or coincidence measurements are shown in Figure 14.23. The SPDC source emits photon pairs with entangled polarizations with orientations defined by the axes of the SPDC nonlinear crystal. The first analyzer is oriented in a fixed position with respect to these axes. At the 0° position of the  $\lambda/2$  plate, photons of the first light ray are detected in a (0°, 90°) base, at 22.5°, however, with (45°, -45°) base axes.

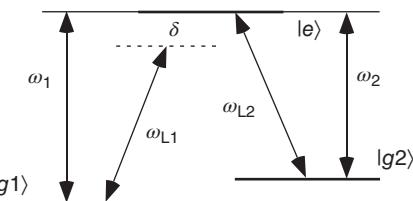
If we rotate the  $\lambda/2$  retarder plate for the second light ray by an angle  $\alpha$ , one expects a variation of the coincidence rate  $\propto \sin^2 \alpha$ . For identical orientation of the base sets, no coincidences can be found at all because of the anticorrelation of horizontal and vertical polarization. At  $\alpha = 45^\circ$  we find the maximal possible rate of coincidences. In Figure 14.23 the  $\sin^2$  variation with  $\alpha$  is shown. The same dependence is also found for the case  $\phi = 22.5^\circ$ , which is an immediate consequence of entanglement already.

A central result of the CHSH inequality (Eq. (14.34)) – the physical interpretation of the measured quantity corresponds to an attempt to quantitatively grasp a measure of entanglement – can be obtained from measurements for special orientations of the polarizer. A strong violation of the inequality is expected for base orientations  $(a, a', b, b') = (45^\circ, 0^\circ, 22.5^\circ, -22.5^\circ)$ . For this situation one calculates for the expectation value from quantum theory a result that maximally violates the inequality:

$$S_{\text{QM}} = 2\sqrt{2} = 2.82 \geq 2.$$

The result of the analysis of the experimental data from [214] yielded  $S_{\text{exp}} = 2.732 \pm 0.017$ . This value violates any theory with hidden parameters by 40 standard deviations. It indicates the validity of quantum physics and renders nonlocality an implicit element of physical reality.

**Figure 14.24** Three-level system with two driving light fields.



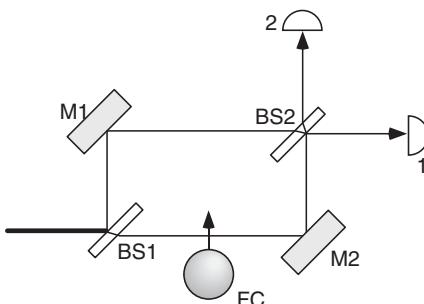
## Problems

- 14.1 “Dressed states” and three-level atom** In Figure 14.24 a three-level system is shown, which couples the two ground states  $|g1\rangle, |g2\rangle$  with two laser light fields  $\mathcal{E}_{1,2}$  to the excited level  $|e\rangle$ . Use the Rabi frequencies  $\Omega_{1,2} = d_{1,2}\mathcal{E}_{1,2}/\hbar$  as measures for the coupling strength. Laser 1 will be tuned across the resonance (laser frequency  $\omega_{L1} = \omega_1 + \delta$ ), and the frequency of laser 2 is firmly set at exact resonance,  $\omega_{L2} = \omega_2$ . Determine the position of the energy levels using the dressed-atom model (Section 13.1.1) as a function of detuning  $\delta$ . For this purpose the dressed-atom model is extended to three levels. Show that for perfect resonance ( $\delta = 0$ ) a so-called dark state occurs, which no longer couples to the driving light fields.

- 14.2 Light-sensitive fireworks** A manufacturer takes a new product to the market: firecrackers that are ignited on exposure to light. The fuses are so sensitive that a single photon suffices for ignition. Unfortunately an entire shipment of functioning firecrackers is mixed up with a shipment of malfunctioning samples.

In order to save at least part of the shipment, a quantum optician proposes the following test: the firecrackers are placed into one arm of a Mach–Zehnder interferometer (MZI) (Figure 14.25) where they are exposed to a very dilute light beam corresponding to isolated photons. The empty MZI is symmetrically aligned, such that by interference all light is detected in exit 1. The malfunctioning fuse will not cause any modification of the propagating light beam. A functioning fuse, however, absorbs the light beam and fires the cracker. Which events are registered

**Figure 14.25** Mach–Zehnder interferometer for Single-photon detection. M: mirrors; FC: firecracker; BS: beam splitters.



by the detector? Show that 50% of the functioning firecrackers are ignited, 25% are falsely identified as malfunctioning, and 25% are correctly sorted out as functioning devices.

- 14.3 Transformation of Bell's states** An SPDC source generates pairs of polarization-entangled photons (Section 14.7.4) that are emitted in the singlet state  $\Psi_{\text{EPR}} = (|H\rangle_1|V\rangle_2 + e^{i\phi}|V\rangle_1|H\rangle_2)/\sqrt{2}$ . Which optical elements can be used to transform them to other entangled (non-factorizable) quantum states?
- 14.4 Clauser–Horne–Shimony–Holt (CHSH) inequality** Suppose the expectation value  $s$  of a correlation measurement of the components of two spin-1/2 particles by two different *Stern–Gerlach* analyzers  $A$  and  $B$  with orientations  $\{\mathbf{e}_A, \mathbf{e}_B\}$  depends on a hidden parameter  $\lambda$  with

$$\begin{aligned} s(\mathbf{e}_A, \mathbf{e}_{A'}, \mathbf{e}_B, \mathbf{e}_{B'}) \\ = A(\mathbf{e}_A \lambda)B(\mathbf{e}_B, \lambda) - A(\mathbf{e}_A \lambda)B(\mathbf{e}_{B'}, \lambda) + A(\mathbf{e}_{A'} \lambda)B(\mathbf{e}_B, \lambda) + A(\mathbf{e}_{A'} \lambda)B(\mathbf{e}_{B'}, \lambda). \end{aligned}$$

Show that for  $\{A(\mathbf{e}_A \lambda), B(\mathbf{e}_B, \lambda)\} = \pm 1$ , only the values  $s = \pm 2$  are possible. Define the ensemble average using Eq. (14.31) through

$$S(\mathbf{e}_A, \mathbf{e}_{A'}, \mathbf{e}_B, \mathbf{e}_{B'}) = E(\mathbf{e}_A, \mathbf{e}_B) - E(\mathbf{e}_A, \mathbf{e}_{B'}) + E(\mathbf{e}_{A'}, \mathbf{e}_B) - E(\mathbf{e}_{A'}, \mathbf{e}_{B'}). \quad (14.31)$$

Show that the condition

$$-2 \leq S(\mathbf{e}_A, \mathbf{e}_{A'}, \mathbf{e}_B, \mathbf{e}_{B'}) \leq 2$$

must hold for theories with hidden variables  $\lambda$ . Furthermore, calculate the expectation value according to quantum mechanics for the singlet state (14.29); see Section 14.7.2 as well.

# 15

## Nonlinear Optics I: Optical Mixing Processes

Up to now we have mostly considered polarizations connected linearly with the driving field. The theory of the *linear response* was completely sufficient as long as only classical light sources were available. Since the invention of the laser, we have been able to drive matter so strongly that, besides linear contributions to the polarization (like in Eq. (7.13)), nonlinear ones also become noticeable (Figure 15.1).

### 15.1 Charged Anharmonic Oscillators

We can modify the classical model of Section 7.1.1 to obtain a simplified microscopic model of the properties of nonlinear interactions of light and matter. For this purpose, we add a weak anharmonic force  $\max^2$  to the equation of motion of the linear oscillator. This model reflects, for example, the situation of the potential of a charge in a crystal with a lack of inversion symmetry. At the same time we neglect the linear damping by absorption and scattering, which are undesired for the application and the study of nonlinear processes and, as we will see, make the formal treatment even more complex. So we consider the undamped equation (Figure 15.2)

$$\ddot{x} + \omega_0^2 x + \alpha x^2 = \frac{q}{m} \mathcal{E} \cos(\omega t).$$

We now seek a solution  $x(t) = x^{(1)}(t) + x^{(2)}(t)$  with  $x^{(1)}$  as the already known linear part,  $x^{(1)}(t) = x_L \cos(\omega t)$ .<sup>1</sup> The amplitude is  $x_L = q\mathcal{E}/m(\omega_0^2 - \omega^2)$ , and the small nonlinear perturbation ( $|x^{(2)}| \ll |x^{(1)}|$ ) approximately fulfills the equation

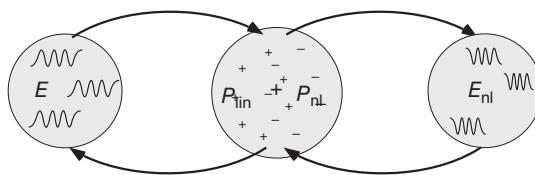
$$\ddot{x}^{(2)} + \omega_0^2 x^{(2)} = -\alpha[(x^{(1)})^2 + 2x^{(1)}x^{(2)} + \dots] \simeq -\alpha x_L^2 \cos^2(\omega t).$$

We now split the nonlinear polarization into a constant and a term oscillating at twice the frequency of the driving field  $2\omega$ :

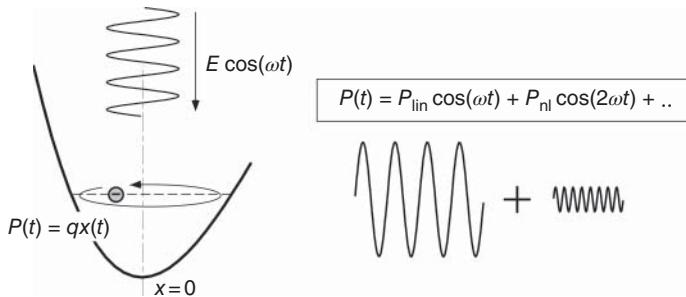
$$x^{(2)} = x_{DC}^{(2)} + x_{2\omega}^{(2)}.$$

---

<sup>1</sup> In transparent materials the electronic resonances are far away, and we can neglect the absorptive contribution ( $\propto \sin \omega t$ ) to a good approximation.



**Figure 15.1** A nonlinear polarization  $P_{\text{nl}}$  is generated in a dielectric material at large intensities, in addition to linear interactions. It acts as the source of a new electromagnetic field  $E_{\text{nl}}$ , which acts back onto the polarization.



**Figure 15.2** Charged oscillator in an anharmonic potential. By the anharmonic motion, the harmonics of the driving frequency  $\omega$  are excited. In a real crystal,  $x(t)$  has to be replaced by an appropriate normal coordinate.

We find the solution

$$\begin{aligned} x_{\text{DC}}^{(2)} &= -\frac{\alpha x_{\text{L}}^2}{2\omega_0^2}, \\ x_{2\omega}^{(2)} &= -\frac{\alpha x_{\text{L}}^2}{2(\omega_0^2 - 4\omega^2)} \cos(2\omega t). \end{aligned}$$

The first term describes the shift of the mean position of the charge caused by the asymmetry of the potential. Thus the optical wave causes a constant, macroscopic polarization of the sample, which we can as well interpret as “optical rectification” or as “inverse Kerr effect” (see Section 3.5.1).

The second term describes the first harmonic of the charge at frequency  $2\omega$ . For suitable conditions, which will be discussed in more detail in Section 15.4 on frequency doubling, the sample emits a coherent electric field at this frequency!

In analogy with the linear case, we can introduce a nonlinear susceptibility describing nonlinear light–matter interaction. It causes a harmonic wave at frequency  $2\omega$  and is connected with a new polarization at this frequency:

$$P_{2\omega}(t) = -\frac{\alpha(q/m)^2}{2(\omega_0^2 - \omega^2)^2(\omega_0^2 - 4\omega^2)} \mathcal{E}^2 \cos(2\omega t).$$

We thus obtain a nonlinear susceptibility

$$\chi(2\omega) = -\frac{1}{\epsilon_0} \frac{\alpha(q/m)^2}{2(\omega_0^2 - \omega^2)^2(\omega_0^2 - 4\omega^2)}. \quad (15.1)$$

It is interesting to note that it shows a resonance at  $\omega_0 = 2\omega$ , which may be interpreted through two-photon absorption.

## 15.2 Second-Order Nonlinear Susceptibility

We can generally describe the response of a sample to one or more optical waves by means of nonlinear susceptibilities. In the following we only consider monochromatic electric fields, which we split into positive and negative frequency parts using the complex notation,

$$\begin{aligned}\mathbf{E}(\mathbf{r}, t) &= (\mathbf{E}^{(+)} + \mathbf{E}^{(-)})/2, \\ \mathbf{E}^{(+)}(\mathbf{r}, t) &= \epsilon e^{-i(\omega t - \mathbf{k}\mathbf{r})}, \\ \mathbf{E}^{(-)}(\mathbf{r}, t) &= (\mathbf{E}^{(+)}(\mathbf{r}, t))^*,\end{aligned}$$

and correspondingly for the dielectric polarization  $\mathbf{P}^{(\pm)}$ . If the field is linearly polarized, the amplitude is calculated from

$$|\mathcal{E}| = \sqrt{\frac{I}{n c \epsilon_0}},$$

because, in this definition,  $|\mathbf{E}|^2 = \mathbf{E}^{(+)} \mathbf{E}^{(-)}$ . The linear relation of field strength and polarization is already known from Eq. (7.15). In order to avoid the elaborate presentation using the convolution integral, we symbolize it here by the  $\odot$  sign:

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \chi^{(1)} \odot \mathbf{E}(\mathbf{r}, t).$$

Furthermore we use an additional superscript index “(1)” to identify the linear or first-order contribution. In the most important case of monochromatic fields, a simple product is recovered from the temporal convolution (Eq. (7.15)).

At high field intensities the nonlinear contributions of the polarization also lead to perceptible effects:

$$\begin{aligned}\mathbf{P}(\mathbf{r}, t) &= \mathbf{P}^{\text{lin}}(\mathbf{r}, t) + \mathbf{P}^{\text{NL}}(\mathbf{r}, t) \\ &= \epsilon_0 [\chi^{(1)} \odot \mathbf{E}(\mathbf{r}, t) + \chi^{(2)} \odot \mathbf{E}(\mathbf{r}, t) \odot \mathbf{E}(\mathbf{r}, t) + \dots].\end{aligned}$$

Terms of second order and higher are the topics of nonlinear optics; they are also called “nonlinear products.” In general, the interaction is anisotropic ( $\chi^{(1)} = \chi_{ij}^{(1)}, \chi^{(2)} = \chi_{ijk}^{(2)}$ , etc.,  $\{i, j, \dots\} = \{x, y, z\}$ ) and depends on the individual vector components; thus nonlinear products for all the relevant field components can occur ( $(\mathbf{E} \odot \mathbf{E})_{ij} = E_i \odot E_j$ , etc.).

### 15.2.1 Mixing Optical Fields: Three-Wave Mixing

For each order of  $\chi^{(n)}$ , a series of new frequencies is generated through the “mixing” products, the  $e^{-i\omega_i t} e^{-i\omega_j t} \dots e^{-i\omega_n t}$  terms. It is thus much simpler for nonlinear optics to sort the contributions to the polarization by their frequency components  $\omega = \omega_i \pm \omega_j \pm \dots \pm \omega_n$ . A general term for the polarization of second order can, for instance, be given component-wise by

$$P_i(\omega) = \sum_{jk} \sum_{mn} \chi_{ijk}^{(2)}(\omega; \omega_m \omega_n) E_j(\omega_m) E_k(\omega_n). \quad (15.2)$$

Since in nonlinear optics a new wave is created as a result of the nonlinear polarization, we speak of *three-wave mixing* processes.

For a simplified one-dimensional and isotropic case ( $j = k$ ), we can extract all the frequency components of nonlinear polarization from

$$\begin{aligned} E(\mathbf{r}, t)^2 &= \left[ \sum_m (E_m^{(+)} + E_m^{(-)}) \right]^2 \\ &= \sum_i \left[ (E_m^{(+)})^2 + E_m^{(+)} E_m^{(-)} + 2 \sum_{n \neq m} (E_m^{(+)} E_n^{(+)} + E_m^{(+)} E_n^{(-)}) \right] + \text{c.c.} \end{aligned}$$

Already by irradiation with just two optical waves ( $m = 1, 2$ ) of different frequencies ( $\omega_{1,2}$  in Eq. (15.2)), nonlinear polarizations at five different sums and difference frequencies are produced, which act as the driving force for generation of a new wave at the mixing frequency (Figure 15.3):

$\mathbf{P}(2\omega_1)$	second harmonic frequency (1)	(SHG),
$\mathbf{P}(2\omega_2)$	second harmonic frequency (2)	(SHG),
$\mathbf{P}(\omega_1 + \omega_2)$	sum frequency	(SUM),
$\mathbf{P}(\omega_1 - \omega_2)$	difference frequency	(DIF),
$\mathbf{P}(\omega = 0)$	optical rectification	(OR).

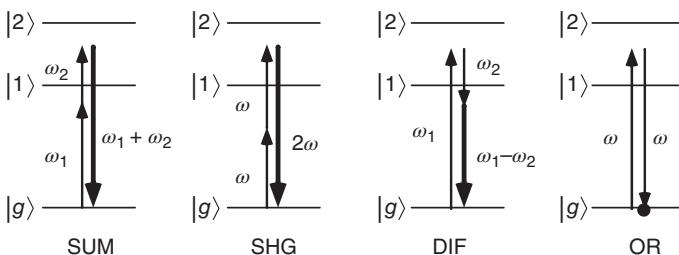
Two field components each with frequencies  $\omega_1, \omega_2$  generate a polarization at frequency  $\omega$ . The corresponding susceptibility is characterized by the notation

$$\chi_{ijk}^{(2)}(\omega; \omega_1, \omega_2), \omega = \omega_1 + \omega_2.$$

The indices “ $ijk$ ” can represent every Cartesian coordinate from ( $x, y, z$ ) and take the tensorial character of the susceptibility into account. Therefore for each frequency combination in principle, there are 27 tensor elements in second order.

Neglecting the Cartesian dependence for the moment, the following relations can be found by splitting the polarization into Fourier components  $P_i(\mathbf{r}, t) = (P_i^{(+)} + P_i^{(-)})/2$  and by comparison with Eq. (15.3):

$$\begin{aligned} P^{(+)}(\omega = 2\omega_1) &= \epsilon_0 \chi^{(2)}(\omega, \omega_1, \omega_1) (E_1^{(+)})^2, \\ P^{(+)}(\omega = 2\omega_2) &= \epsilon_0 \chi^{(2)}(\omega, \omega_2, \omega_2) (E_2^{(+)})^2, \\ P^{(+)}(\omega = \omega_1 + \omega_2) &= 2\epsilon_0 \chi^{(2)}(\omega, \omega_1, \omega_2) E_1^{(+)} E_2^{(+)}, \\ P^{(+)}(\omega = \omega_1 - \omega_2) &= 2\epsilon_0 \chi^{(2)}(\omega, \omega_1, -\omega_2) E_1^{(+)} E_2^{(-)}, \end{aligned}$$



**Figure 15.3** Passive  $\chi^{(2)}$  processes. SUM = sum frequency generation; SHG = second harmonic generation; DIF = difference frequency generation; OR = optical rectification.

$$\begin{aligned} P^{(+)}(\omega = 0) = 2\epsilon_0 & [\chi^{(2)}(0; \omega_1, -\omega_1)E_1^{(+)}E_1^{(-)} \\ & + \chi^{(2)}(0; \omega_2, -\omega_2)E_2^{(+)}E_2^{(-)}]. \end{aligned}$$

### 15.2.2 Symmetry Properties of Susceptibility

The search for crystals with large nonlinear coefficients is a matter of continued scientific research. The symmetry properties of real crystals play an important role [215] and will be here subject to a short consideration with regard to nonlinear optics. For the sake of simplicity, we restrict this discussion to the second-order effects.

#### 15.2.2.1 Intrinsic Permutation Symmetry

Using two fundamental waves and one polarization wave, six different mixing products can be generated if we additionally require  $\omega = \omega_1 + \omega_2$ :

$$\begin{array}{lll} \chi_{ijk}^{(2)}(\omega; \omega_1, \omega_2); & \chi_{ijk}^{(2)}(\omega_1; -\omega_2, \omega); & \chi_{ijk}^{(2)}(\omega_2; \omega, -\omega_1); \\ \chi_{ijk}^{(2)}(\omega; \omega_2, \omega_1); & \chi_{ijk}^{(2)}(\omega_1; \omega, -\omega_2); & \chi_{ijk}^{(2)}(\omega_2; -\omega_1, \omega). \end{array}$$

The upper row is identical to the lower one when coordinates  $(i, j)$  are permuted along with the corresponding frequencies:

$$\chi_{ijk}^{(2)}(\omega; \omega_1, \omega_2) = \chi_{ikj}^{(2)}(\omega; \omega_2, \omega_1).$$

#### 15.2.2.2 Real Electromagnetic Fields

Since the harmonic time dependence of  $P^{(-)}$  is connected with  $P^{(+)}$  by replacement of  $\omega_i \rightarrow -\omega_i$ , the following has to be valid:

$$\chi_{ijk}^{(2)}(\omega_i; \omega_j, \omega_k) = \chi_{ijk}^{(2)}(-\omega_i; -\omega_k, -\omega_j)^*.$$

#### 15.2.2.3 Loss-Free Media

In loss-free media the susceptibility is real. Then we have

$$\chi_{ijk}^{(2)}(\omega_i; \omega_j, \omega_k) = \chi_{ijk}^{(2)}(-\omega_i; -\omega_k, -\omega_j). \quad (15.4)$$

In addition, “complete permutation symmetry” holds, that is, all frequencies can be permuted if the corresponding Cartesian indices are permuted at the same time. For this it has to be taken into account that the sign of the commuted frequencies has to change in order to meet the condition  $\omega = \omega_1 + \omega_2$ :

$$\chi_{ijk}^{(2)}(\omega; \omega_1, \omega_2) = \chi_{jik}^{(2)}(-\omega_1; -\omega, \omega_2) = \chi_{jik}^{(2)}(\omega_1; \omega, -\omega_2).$$

In the last step we used Eq. (15.4). A proof of this symmetry can be based on the quantum mechanical calculation of  $\chi$  or the energy density in a nonlinear medium.

### 15.2.3 Two-Wave Polarization

In the previous section we have seen that one or more new polarization waves result as a mixing product of two input fields:

$$\begin{aligned} P^{(+)} &= \epsilon_0 \chi^{(2)}(\omega; \omega_1, \omega_2) E_1^{(+)} E_2^{(+)}, \\ P_1^{(+)} &= \epsilon_0 \chi^{(2)}(\omega_1; -\omega_2, \omega) E_2^{(-)} E_1^{(+)}, \\ P_2^{(+)} &= \epsilon_0 \chi^{(2)}(\omega_2; \omega, -\omega_1) E_1^{(+)} E_2^{(-)}. \end{aligned} \quad (15.5)$$

At the same time a new field at the frequency of the polarization wave has to emerge, which by nonlinear interaction now itself contributes to the polarization at the already existing frequencies. This nonlinear coupling describes the back-action of the nonlinear polarization onto the fundamental waves, for example, the exchange of energy. With the symmetry rules of Section 15.2.2, we can confirm that for the approximation of loss-free media, the  $\chi^{(2)}$  coefficients in Eq. (15.5) are identical! In Section 15.3.1 we shall go further to investigate the coupling of *three* waves.

#### 15.2.3.1 Contracted Notation

In nonlinear optics the “contracted notation” is used very often, which at first is defined by the tensor

$$d_{ijk} = \frac{1}{2} \chi_{ijk}^{(2)}.$$

The notation is now simplified, and the number of possible elements is reduced from 27 for  $\chi_{ijk}^{(2)}$  to 18 by contracting the last two indices ( $j, k$ ) to a single index  $l$ , that is,  $d_{ijk} \rightarrow d_{il}$ . So because of the intrinsic permutation symmetry, we have

$$\begin{array}{ccccccccc} jk : & 11 & 22 & 33 & 23, 32 & 31, 13 & 12, 21 \\ l : & 1 & 2 & 3 & 4 & 5 & 6. \end{array}$$

For example, the matrix equation describing frequency doubling reads with the  $d_{ij}$  tensor

$$\begin{pmatrix} P_x(2\omega) \\ P_y(2\omega) \\ P_z(2\omega) \end{pmatrix} = 2\epsilon_0 \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{pmatrix} \begin{pmatrix} E_x(\omega)^2 \\ E_y(\omega)^2 \\ E_z(\omega)^2 \\ 2E_y(\omega)E_z(\omega) \\ 2E_x(\omega)E_z(\omega) \\ 2E_x(\omega)E_y(\omega) \end{pmatrix}. \quad (15.6)$$

#### 15.2.3.2 Kleinman Symmetry

Often the resonance frequencies of a nonlinear material are much higher than those of the driving fields. Then the susceptibilities – which typically have forms similar to our classical model of Eq. (15.1) – depend only weakly on the frequency and are subject to the approximate *Kleinman symmetry*. If furthermore the susceptibility does not even depend on the frequency, the Cartesian indices can be permuted without permuting the corresponding frequencies at the same time.

The Kleinman symmetry reduces the maximum number of independent matrix elements from 18 to 10.

#### 15.2.4 Crystal Symmetry

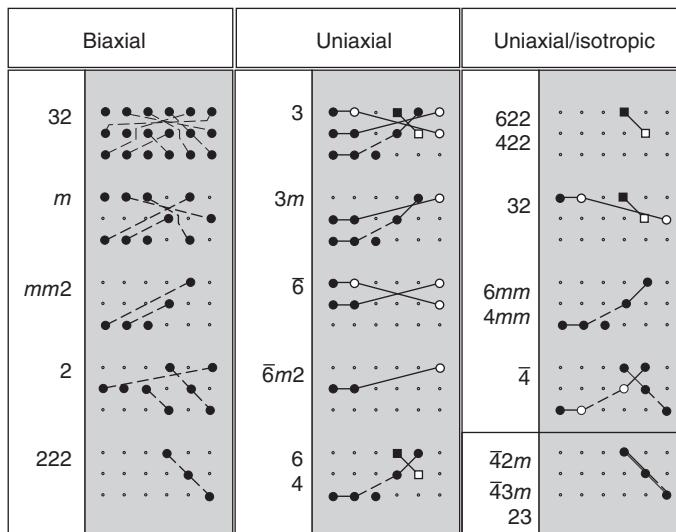
A crystal with inversion symmetry cannot show any susceptibility of second order at all. With the inversion of all coordinates, the sign of the field amplitude changes as well as that of the polarization:

$$P_i(\mathbf{r}) = d_{ijk} E_j(\mathbf{r}) E_k(\mathbf{r}) \xrightarrow{\mathbf{r} \rightarrow -\mathbf{r}} -P_i(\mathbf{r}) = d_{ijk} E_j(-\mathbf{r}) E_k(-\mathbf{r}).$$

Thus the inversion symmetry leads to  $d_{ijk} = \chi_{ijk}^{(2)}/2 = 0$ , and from 32 crystal classes, those 11 exhibiting inversion symmetry are eliminated. The symmetry properties of the remaining crystal classes significantly reduce the number of nonvanishing nonlinear  $d$  coefficients that are independent of each other. In Figure 15.4 the nonzero coefficients for the different crystal classes are given in the standard notation.

#### 15.2.5 Effective Value of the Nonlinear $d$ Coefficient

In general nonlinear crystals (NLCs) are anisotropic and birefringent; indeed we are going into detail that the asymmetry of birefringence really makes their efficient application possible. Frequently one finds quotations of effective values  $d_{\text{eff}}$  depending on the so-called phase matching angles  $\theta$  and  $\phi$  given for the  $d_{il}$  coefficients, which are also tabulated [218].



**Figure 15.4** Nonvanishing second-order coefficients  $d_{\text{eff}}$  for crystal classes without inversion symmetry, represented by their standard point symmetry group notation [215, 216]; for the crystal classification scheme, see [217]. Identical coefficients are connected by lines (dashed: only for Kleinman symmetry). Full and open symbols indicate opposite signs; square symbols vanish at Kleinman symmetry.

### 15.3 Wave Propagation in Nonlinear Media

In order to understand the propagation of waves in a nonlinear medium [219], we first consider again the general form of the wave equation in matter:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}, t) + \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \mathbf{E}(\mathbf{r}, t) = -\frac{1}{\epsilon_0 c^2} \frac{\partial^2}{\partial t^2} \mathbf{P}(\mathbf{r}, t).$$

The first term of the vector identity  $\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$  cannot be removed in nonlinear optics as easily as for linear isotropic media because  $\nabla \cdot \mathbf{E} = 0$  can no longer be inferred from  $\nabla \cdot \mathbf{D} = 0$ . Fortunately the first term can be neglected in many cases of interest, especially for the limiting case of planar waves:

$$\left( \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{E}(\mathbf{r}, t) = \frac{1}{\epsilon_0 c^2} \frac{\partial^2}{\partial t^2} \mathbf{P}(\mathbf{r}, t).$$

The polarization contains linear and nonlinear parts,  $\mathbf{P} = \mathbf{P}^{(1)} + \mathbf{P}^{\text{NL}}$ . The linear contribution has an effect only on one or more fundamental waves  $\mathbf{E}^F$  driving the process and is taken into account through the refraction coefficient  $n^2 = 1 + \chi^{(1)}$ , that is,  $\mathbf{P}^{(1)} = \epsilon_0(n^2 - 1)\mathbf{E}^F$ . Then a new wave equation is obtained driven by the nonlinear polarization,  $\mathbf{P}^{\text{NL}}$ :

$$\left( \nabla^2 - \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{E}(\mathbf{r}, t) = \frac{1}{\epsilon_0 c^2} \frac{\partial^2}{\partial t^2} \mathbf{P}^{\text{NL}}(\mathbf{r}, t).$$

If  $\mathbf{P}^{\text{NL}}(\mathbf{r}, t)$  vanishes, the already known equation for the propagation of a wave in a dielectric medium is found. In a dispersive medium, the refraction coefficient depends on the frequency,  $n = n(\omega)$ . We now consider again each frequency component  $\omega_i$  separately and also split the oscillating part from the positive and negative polarization components:

$$\mathbf{P}^{\text{NL}}(\mathbf{r}, t) = \sum_i [\tilde{\mathbf{P}}_i(\mathbf{r}) e^{-i\omega_i t} + \tilde{\mathbf{P}}_i^*(\mathbf{r}) e^{i\omega_i t}] / 2.$$

With this notation the wave equation separates into single-frequency Helmholtz equations and can be written as

$$\left( \nabla^2 + \frac{n(\omega)^2 \omega_i^2}{c^2} \right) \mathcal{E}_i(\mathbf{r}) e^{i\mathbf{k}\mathbf{r}} = -\frac{\omega_i^2}{\epsilon_0 c^2} \tilde{\mathbf{P}}_i(\mathbf{r}). \quad (15.7)$$

#### 15.3.1 Coupled Amplitude Equations

To simplify Eq. (15.8) we first consider only planar waves propagating in the  $z$  direction. Additionally, it is generally realistic to assume again that the amplitudes of the waves change only slowly compared with the wavelength or that the curvature of the amplitude is much smaller than the curvature of the wave:

$$\left| \frac{\partial^2 \mathcal{E}(z)}{\partial z^2} \right| \ll k \left| \frac{\partial \mathcal{E}(z)}{\partial z} \right|.$$

Then with

$$\frac{\partial^2}{\partial z^2} [\mathcal{E}(z) e^{ikz}] \simeq e^{ikz} \left[ 2ik \frac{\partial}{\partial z} - k^2 \right] \mathcal{E}(z),$$

the wave equation is reduced to an approximate form

$$\left[ 2ik \frac{\partial}{\partial z} - k^2 + \frac{n^2(\omega)\omega^2}{c^2} \right] \mathcal{E}(z) = -\frac{\omega^2}{\epsilon_0 c^2} \tilde{\mathcal{P}}(\omega) e^{-ikz}.$$

With  $k^2 = n^2(\omega)\omega^2/c^2$  we can furthermore identify the wave vector of propagation in a dielectric medium and arrive at

$$\frac{d}{dz} \mathcal{E}(z) = \frac{\omega^2}{\epsilon_0 c^2} \frac{i}{2k} \tilde{\mathcal{P}}(\omega) e^{-ikz}. \quad (15.8)$$

Incidentally, a more exact consideration shows that not only a forward-running but also a backward-running wave is generated, but only the forward-running wave couples significantly to the fundamental wave ([172], Chapter 33).

For each of the complicated wave equations from (15.7), we can therefore draw up a simpler equation according to (15.8) replacing the polarization by its explicit form, for example, according to (15.5). The most important problems of nonlinear optics can be solved with this standard method.

### 15.3.2 Coupled Amplitudes for Three-Wave Mixing

The nonlinear polarization is the origin of the coupling between the fundamental waves  $\mathbf{E}_1(\omega_1)$  and  $\mathbf{E}_2(\omega_2)$  and their mixing product  $\mathbf{E}_3(\omega)$ . For symmetry reasons the same  $\chi^{(2)}$  coefficient is valid for the nonlinear susceptibility in all three cases; see Eq. (15.5). We introduce the abbreviation

$$\Delta k = k - k_1 - k_2,$$

and write the polarization in the form

$$\tilde{\mathcal{P}}_3(z) e^{-ikz} = 4\epsilon_0 d_{\text{eff}} \mathcal{E}_1 e^{ik_1 z} \mathcal{E}_2 e^{ik_2 z} e^{-ikz} = 4\epsilon_0 d_{\text{eff}} \mathcal{E}_1 \mathcal{E}_2 e^{-i\Delta k z}.$$

It is applied to  $\mathcal{P}_{1,2}$  correspondingly. The factor 4 occurs here because we have to sum over all contributions to the polarization according to Eq. (15.2).

In order to study the evolution of amplitudes  $\mathcal{E}_{1-3}$ , these polarizations are inserted into Eq. (15.8), where for the sake of improved transparency we use the complex conjugate of the equations for  $\mathcal{E}_{1,2}$ :

$$\begin{aligned} \frac{d}{dz} \mathcal{E}_3(\omega) &= \frac{2i\omega d_{\text{eff}}}{cn(\omega)} \mathcal{E}_1 \mathcal{E}_2 e^{-i\Delta k z}, \\ \frac{d}{dz} \mathcal{E}_1^*(\omega_1) &= \frac{-2i\omega_1 d_{\text{eff}}}{cn(\omega_1)} \mathcal{E}_3^* \mathcal{E}_2 e^{-i\Delta k z}, \\ \frac{d}{dz} \mathcal{E}_2^*(\omega_2) &= \frac{-2i\omega_2 d_{\text{eff}}}{cn(\omega_2)} \mathcal{E}_1 \mathcal{E}_3^* e^{-i\Delta k z}. \end{aligned} \quad (15.9)$$

In principle with these equations, all relevant  $\chi^{(2)}$  processes can be addressed. For passive processes, including frequency doubling, sum and difference frequency mixing, and optical rectification, initial conditions  $\mathcal{E}_1, \mathcal{E}_2 \neq 0$  and  $\mathcal{E}_3 = 0$  have to be considered; see Sections 15.4 and 15.5. It is also possible to understand the parametric oscillator (Section 15.6) using this set of equations. It resembles an active medium in a way similar to the laser, where the initial conditions now have the form  $\mathcal{E}_1, \mathcal{E}_2 = 0$  and  $\mathcal{E}_3 \neq 0$ .

### 15.3.3 Energy Conservation

The intensity  $I$  of a linearly polarized wave in a dielectric medium with refraction coefficient  $n(\omega)$  is

$$I = \frac{1}{2} n(\omega) c \epsilon_0 |E|^2.$$

By multiplying Eq. (15.9) from the left side with the respective conjugate amplitudes  $n(\omega_i) c \epsilon_0 \mathcal{E}_i^*/2$ , the *Manley–Rowe relation* is obtained:

$$\frac{1}{\omega} \frac{d}{dz} I_3(\omega) = -\frac{1}{\omega_1} \frac{d}{dz} I_1(\omega_1) = -\frac{1}{\omega_2} \frac{d}{dz} I_2(\omega_2).$$

This describes the conservation of energy, since the expression is equivalent to

$$I_3(\omega) + I_1(\omega_1) + I_2(\omega_2) = 0,$$

which is straightforwardly shown by using  $\omega = \omega_1 + \omega_2$ . This relation is also called “photon conservation,” relating to an interpretation where in this process two photons with frequencies  $\omega_1$  and  $\omega_2$  are combined to make one photon with frequency  $\omega$ . One should keep in mind, however, that this is just another expression for energy conservation. Nonlinear optics does not at all have to invoke quantum physics for theoretical explanations here.

Nonetheless the conservation of photon number is convenient and helps to transform Eq. (15.9) to normalized amplitudes:

$$\mathcal{A}_i = \sqrt{\frac{n(\omega_i)}{\omega_i}} \mathcal{E}_i.$$

The amplitude of the electromagnetic wave is now

$$I = c \epsilon_0 \omega |\mathbf{A}(\mathbf{r}, t)|^2,$$

and we find

$$\begin{aligned} \frac{d}{dz} \mathcal{A}_3(\omega) &= i\kappa \mathcal{A}_1 \mathcal{A}_2 e^{-i\Delta kz}, \\ \frac{d}{dz} \mathcal{A}_1^*(\omega_1) &= -i\kappa \mathcal{A}_3^* \mathcal{A}_2 e^{-i\Delta kz}, \\ \frac{d}{dz} \mathcal{A}_2^*(\omega_2) &= -i\kappa \mathcal{A}_1 \mathcal{A}_3^* e^{-i\Delta kz}, \end{aligned} \quad (15.10)$$

with the material-dependent coupling coefficient

$$\kappa = \frac{2d_{\text{eff}}}{c} \sqrt{\frac{\omega \omega_1 \omega_2}{n(\omega) n(\omega_1) n(\omega_2)}}. \quad (15.11)$$

## 15.4 Frequency Doubling

The first important special case of the coupled amplitude equations (15.9) is frequency doubling. It has a particularly great significance because with this method coherent harmonics of a fundamental wave can be generated. By this means, for example, coherent ultraviolet radiation becomes available at wavelengths where no tunable laser system exists. Eqs. (15.9) are in this case

reduced to two equations due to the degeneracy of  $\omega_1$  and  $\omega_1$ . We once again recapitulate the form for the field intensity of the fundamental wave  $\mathcal{E}_{\text{FUN}}$  and for the second harmonic  $\mathcal{E}_{\text{SHG}}$ :

$$\begin{aligned}\frac{d}{dz}\mathcal{E}_{\text{SHG}}(2\omega) &= \frac{i2\omega}{cn(2\omega)}d_{\text{eff}}\mathcal{E}_{\text{FUN}}^2(\omega)e^{-i\Delta kz}, \\ \frac{d}{dz}\mathcal{E}_{\text{FUN}}(\omega) &= \frac{i\omega}{cn(\omega)}2d_{\text{eff}}\mathcal{E}_{\text{SHG}}(2\omega)\mathcal{E}_{\text{FUN}}^*(\omega)e^{i\Delta kz}.\end{aligned}$$

Because of the degeneracy, the term for frequency doubling appears only once in Eq. (15.2), and therefore the first equation is smaller by a factor of 2 than in Eq. (15.9). The so-called phase mismatch,

$$\Delta k = k_{2\omega} - 2k_\omega = \frac{2\omega}{c}(n_{2\omega} - n_\omega), \quad (15.12)$$

apparently depends on the difference of the refraction coefficients of the fundamental and harmonic waves. Owing to the fact that dispersion of common materials is always present, we have  $n_{2\omega} \neq n_\omega$ . For simplification we again use normalized equations (15.10) with  $\mathcal{A}_{\text{FUN}}(\omega) = (n_\omega/\omega)^{1/2}\mathcal{E}_{\text{FUN}}(\omega)$  and  $\mathcal{A}_{\text{SHG}}(\omega) = (n_{2\omega}/\omega)^{1/2}\mathcal{E}_{\text{SHG}}$ :

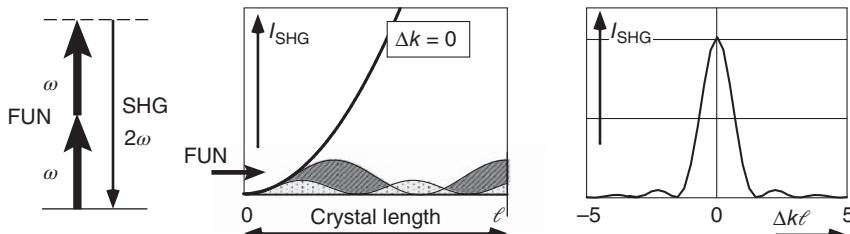
$$\begin{aligned}\frac{d}{dz}\mathcal{A}_{\text{SHG}} &= i\kappa\mathcal{A}_{\text{FUN}}^2 e^{-i\Delta kz}, \\ \frac{d}{dz}\mathcal{A}_{\text{FUN}} &= i\kappa\mathcal{A}_{\text{SHG}}\mathcal{A}_{\text{FUN}}^* e^{i\Delta kz}.\end{aligned} \quad (15.13)$$

The coupling coefficient,  $\kappa = (2d_{\text{eff}}/c)/(\omega^3/n_{2\omega}n_\omega^2)^{1/2}$ , here is also slightly modified compared to (15.11).

### 15.4.1 Weak Conversion

Usually only the fundamental wave enters a crystal of length  $\ell$ , that is, we have  $\mathcal{A}_{\text{SHG}}(z = 0) = 0$ . In the weak conversion approximation, we assume that the fundamental wave is only slightly weakened, that is,  $\mathcal{A}_{\text{FUN}} \simeq \text{const}$ . Then we need to solve only the first equation of the system (15.13) and finally achieve the harmonic amplitude at  $z = \ell$  (Figure 15.5):

$$\mathcal{A}_{\text{SHG}} = \kappa\ell\mathcal{A}_{\text{FUN}}^2(\omega)e^{i\Delta k\ell/2} \frac{\sin(\Delta k\ell/2)}{\Delta k\ell/2}.$$



**Figure 15.5** Evolution of the intensity  $I_{\text{SHG}}$  of the second harmonic for the limiting case of weak conversion. Only for perfect phase matching can continuous growth of the nonlinear product be achieved. Otherwise the radiation power oscillates between the fundamental and the harmonic waves like the wave in the middle picture.

The quantities that depend on the material are combined into the conversion coefficient  $\Gamma$ :

$$\Gamma^2 = \frac{\kappa^2}{c\epsilon_0\omega} = \frac{4d_{\text{eff}}^2\omega^2}{c^3\epsilon_0 n_\omega^2 n_{2\omega}}. \quad (15.14)$$

Further, the intensity of the harmonic wave depends only on the crystal length  $\ell$ , the incident intensity, and the phase mismatch  $\Delta k$ :

$$I_{\text{SHG}} = \Gamma^2 \ell^2 I_{\text{FUN}}^2 \frac{\sin^2(\Delta k \ell / 2)}{(\Delta k \ell / 2)^2}.$$

Depending on the magnitude of the phase mismatch  $\Delta k$ , it obviously oscillates between the fundamental radiation field and the harmonic wave when propagating through the crystal.

The phase mismatch according to Eq. (15.12) is  $|n_\omega - n_{2\omega}| \simeq 10^{-2}$  in typical crystals with normal dispersion. That is why the intensity of the harmonic oscillates with a period of a few  $10 \mu\text{m}$ , which is called the “coherence length”:

$$\ell_{\text{coh}} = \frac{\pi}{\Delta k} = \frac{\lambda}{4(n_{2\omega} - n_\omega)}. \quad (15.15)$$

Only in the case of perfect “phase matching” at  $(n_\omega - n_{2\omega}) = 0$  does the intensity grow continuously with crystal length:

$$I_{\text{SHG}} = \Gamma^2 I_{\text{FUN}}^2 \ell^2. \quad (15.16)$$

This relation suggests that for frequency doubling it is worthwhile to increase the intensity by focusing and to make the crystal longer. Though by focusing, as we know from the description of Gaussian beams (Section 2.3), a constant intensity, a quasi-planar wave, is generated only in a narrow range around the focus, so that a compromise between the demand for strong focusing and long crystals has to be found.

### 15.4.2 Strong Conversion

In the extreme case of strong conversion, the decrease of the intensity of the pumping wave cannot be neglected anymore. We consider the case of perfect phase matching  $\Delta k = 0$ . In order to get real equations, let us introduce the quantities  $\tilde{\mathcal{A}}_{\text{SHG}} = \mathcal{A}_{\text{SHG}} e^{i\phi_{2\omega}}$  and  $\tilde{\mathcal{A}}_{\text{FUN}} = \mathcal{A}_{\text{FUN}} e^{i\phi_\omega}$  with real amplitudes  $\tilde{\mathcal{A}}$ . Then we have

$$\frac{d}{dz} \tilde{\mathcal{A}}_{\text{SHG}} = i\kappa \tilde{\mathcal{A}}_{\text{FUN}}^2 e^{-i(2\phi_\omega - \phi_{2\omega})},$$

and we are now free to choose the relative phase of the amplitudes, for instance,  $e^{-i(2\phi_\omega - \phi_{2\omega})} = -i$ . As a result of energy conservation, we have

$$\frac{d}{dz} (|\mathcal{A}_{\text{SHG}}|^2 + |\mathcal{A}_{\text{FUN}}|^2) = 0.$$

Then for the case of a harmonic vanishing at the input facet at  $\mathcal{A}_{\text{SHG}}(z = 0) = 0$  and  $\mathcal{A}_{\text{FUN}}(z = 0) = \mathcal{A}_0$  (we immediately remove the  $\sim$  marks), the real equations are

$$\frac{d}{dz} \mathcal{A}_{\text{SHG}} = \kappa \mathcal{A}_{\text{FUN}}^2 = \kappa |\mathcal{A}_{\text{FUN}}|^2 = \kappa (\mathcal{A}_0^2 - \mathcal{A}_{\text{SHG}}^2),$$

$$\frac{d}{dz} \mathcal{A}_{\text{FUN}} = -\kappa \mathcal{A}_{\text{SHG}} \mathcal{A}_{\text{FUN}}.$$

The first equation can be solved by standard techniques and results in

$$\mathcal{A}_{\text{SHG}}(z) = \mathcal{A}_{10} \tanh(\kappa \mathcal{A}_0).$$

Thus in principle 100% conversion efficiency can be achieved for frequency doubling since at the end of a long crystal, the harmonic intensity

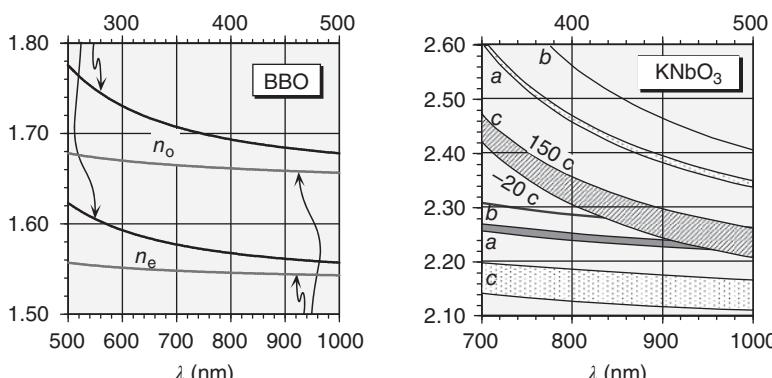
$$I_{\text{SHG}}(z) = I_0 \tanh^2(\Gamma I_0^{1/2} z)$$

should be found. This result is important particularly for frequency doubling with powerful pulsed lasers since it promises very efficient second harmonic generation.

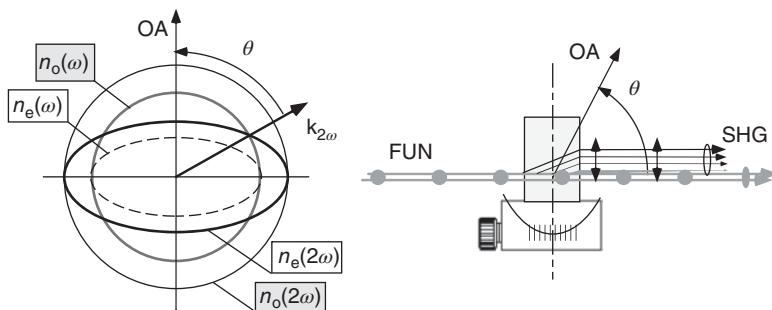
### 15.4.3 Phase Matching in Nonlinear and Birefringent Crystals

We have already seen in Eq. (15.15) that frequency conversion takes place only over a certain length depending on the dispersion  $n(\omega)$ . Birefringent crystals such as BBO or KNbO<sub>3</sub> (Figure 15.6) with properties already introduced in Section 3.4.1 make it possible to realize  $\ell_{\text{coh}} \rightarrow \infty$  by choosing a direction of propagation in which the refraction coefficients of fundamental and harmonic wave are identical,  $n(\omega) = n(2\omega)$ . Also, in Section 15.4.6, we shall discuss the method of “quasi-phase matching,” which has arisen more recently as a successful method of outwitting dispersion.

The simplest situation occurs in uniaxial crystals. For the ordinary beam, polarization and propagation direction are perpendicular to the optical axis, and the



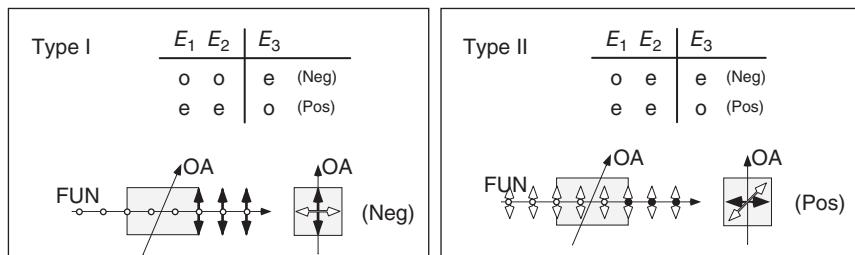
**Figure 15.6** Refractive index for BBO and KNbO<sub>3</sub> as a function of the wavelength. In the uniaxial BBO crystal, the ordinary refraction coefficient ( $n_o$ ) of a fundamental wave is found in between the ordinary and the extraordinary coefficient ( $n_e$ ) of half the wavelength, which corresponds to the frequency-doubled wave and makes the angle phase matching possible. In the triaxial KNbO<sub>3</sub> (refractive indices  $n_a$ ,  $n_b$ ,  $n_c$ ), phase matching can be achieved by temperature tuning.



**Figure 15.7** Phase matching by angle tuning (“critical phase matching”). On the left, the “indicatrix” for a uniaxial crystal is presented. In order to realize angle phase matching, there has to be an intercept between the ellipsoids of the refractive indices for the ordinary ( $n_o$ ) and extraordinary ( $n_e$ ) beams. The fundamental must propagate at the phase matching angle with respect to the optical axis. On the right a typical setup is shown, in which the crystal angle can be adjusted. Fundamental and harmonic waves deviate from each other since they correspond to the ordinary and extraordinary beams, respectively. The angle between fundamental and harmonic waves is called the walk-off angle.

phase velocity is characterized by the linear ordinary refraction coefficient  $n_o(\omega)$ . Since frequency conversion usually takes place in crystals with normal dispersion for the harmonic wave, the smaller refractive index must always be chosen, that is, in a *negatively* uniaxial crystal ( $n_e < n_o$ ), the harmonic has to be chosen as the extraordinary beam, while in a *positively* uniaxial crystal ( $n_o < n_e$ ) as the ordinary one (Figure 15.7). Then phase matching can be achieved by choosing the polarization of the fundamental wave orthogonal to the harmonic (“type I phase matching”). Alternatively, according to Eq. (15.6), the polarization of the fundamental wave can as well be spread over ordinary and extraordinary beams (i.e., incidence under 45° to the crystal axes) using the “type II phase matching,” so that the four alternatives of Figure 15.8 are available.

$$\frac{1}{n_e(\theta)} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2}.$$



**Figure 15.8** Polarization directions of fundamental and harmonic waves for phase matching with birefringent crystals. In a crystal with negative (positive) birefringence, the shortest wavelength has to propagate on the extraordinary (ordinary) beam. For type I matching all polarization directions are orthogonal. For type II matching one polarization direction is used to achieve equally strong projections onto the optical principal axes.

### 15.4.3.1 Angle or Critical Phase Matching

As we have already investigated in Section 3.4.1, the refraction coefficient  $n_e(\theta)$  of the extraordinary beam depends on the angle between the optical axis and the beam direction according to the “indicatrix” since the polarization has components parallel as well as perpendicular to the optical axis (Eq. (3.29)).

Phase matching can now be achieved by choosing appropriately the angle between the fundamental wave and the optical axis. For a negatively (positively) uniaxial crystal, the angles of phase matching for type I/II are determined from the conditions

$$\begin{array}{lll} \text{type I} & \text{neg} & n_e(\theta, 2\omega) = n_o(\omega), \\ & \text{pos} & n_e(\theta, \omega) = n_o(2\omega), \\ \text{type II} & \text{neg} & n_e(\theta, 2\omega) = \frac{1}{2}[n_o(\omega) + n_e(\theta, \omega)], \\ & \text{pos} & n_o(2\omega) = \frac{1}{2}[n_o(\omega) + n_e(\theta, \omega)]. \end{array}$$

From Eq. (3.29) we have for the case of negative type I phase matching

$$\frac{\cos^2\theta_m}{n_0(2\omega)} + \frac{\sin^2\theta_m}{n_e(2\omega)} = \frac{1}{n_0(\omega)},$$

from which we can deduce the phase matching angle

$$\sin^2\theta_m = \frac{n_o^{-2}(\omega) - n_o^{-2}(2\omega)}{n_e^{-2}(2\omega) - n_o^{-2}(2\omega)}.$$

Similar relations are deduced for the other cases.

For applications NLCs are cut with respect to the input facet (see Figure 15.7) in such a way that for normal incidence, propagation occurs near to the ideal phase matching angle from the beginning. In order to minimize losses, these facets are frequently antireflection coated as well, sometimes even for both the fundamental and harmonic waves. An example of a crystal in frequent use for angle-tuned phase matching is BBO; see Figure 15.6.

Once angle phase matching is achieved, the *walk-off* problem occurs, since the ordinary and extraordinary beams propagate with the same phase velocity but not in the same direction. The *walk-off angle*  $\rho$  has already been discussed with Eq. (3.30) for uniaxial crystals:

$$\tan \rho = \frac{n^2(\theta)}{2} \left( \frac{1}{n_o^2} - \frac{1}{n_e^2} \right) \sin 2\theta.$$

The harmonic wave therefore leaves the NLC with an elliptical beam profile. Furthermore, the intensity no longer grows quadratically with the crystal length like in Eq. (15.16) but just linearly once the harmonic beam has walked off by a full diameter of the fundamental beam.

### 15.4.3.2 Noncritical or 90° Phase Matching

The disadvantages of angle phase matching can be avoided if one succeeds in matching the ordinary and extraordinary refraction coefficients under the condition  $\theta = 90^\circ$ . This situation is realized in special crystals where one of the two refractive indices can be tuned over a quite large range by controlling the temperature. Because of the long interaction length, this method provides

particularly large conversion efficiency. For this reason KNbO<sub>3</sub> (Figure 15.6) is a very important nonlinear material: it has a large nonlinear coefficient, and it allows 90° phase matching in the important near-infrared range. In Figure 15.6 the refractive indices for the three axes (*a*, *b*, *c*) were presented. It shows that for the *a* cut phase matching can be achieved for frequency doubling from 840 to 960 nm and for the *b* cut from 950 to 1060 nm. Of course, the methods of angle phase matching can also be employed with these crystals.

Besides “90° phase matching,” the terms *temperature* and *noncritical* phase matching are used for this type of phase matching.

#### 15.4.4 Frequency Doubling with Gaussian Beams

Having looked at the principle of phase matching from the point of view of plane waves, we now have to study the influence of realistic laser beams. The conversion efficiency increases with the intensity of the fundamental wave, and so focusing is an obvious choice. On the other hand, too strong focusing leads to large divergence and reduces the effect again (Figure 15.9). So intuitively an optimum effect is anticipated if the Rayleigh length roughly corresponds to the crystal length.

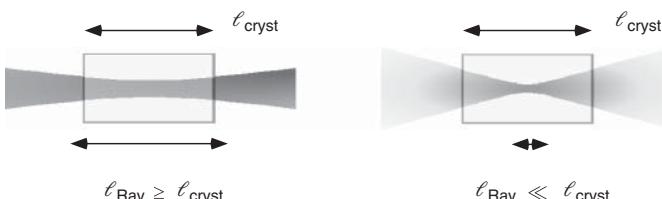
A Gaussian beam (details about wave optics can be found in Section 2.3) in the TEM<sub>00</sub> mode has radial intensity distribution and total power given by

$$\begin{aligned}\mathcal{E}(r) &= \mathcal{E}_0 e^{-(r/w_0)^2}, \\ P &= \frac{\pi c \epsilon_0}{2} 2\pi \int_0^\infty dr r |\mathcal{E}(r)|^2 = I_0 \frac{\pi w_0^2}{2}\end{aligned}$$

near the beam waist, where the characteristic parameters are

$$\begin{aligned}w_0 &= \left( \frac{b\lambda}{4\pi n_\omega} \right)^{1/2} && \text{radius of the beam waist,} \\ b &= 2z_0 && \text{confocal parameter,} \\ \theta_{\text{div}} &= \frac{\lambda}{\pi w_0 n_\omega} && \text{divergence angle of the Gaussian mode.}\end{aligned}$$

Boyd and Kleinman [220] gave a detailed discussion of this problem in the 1960s and worked out suitable mathematical tools for its treatment. In the limiting case of weak conversion and weak focusing (i.e.,  $b \ll \ell$ ), it can be dealt with by simple radial integration. At the end of a crystal of length  $\ell$  and with perfect



**Figure 15.9** Focusing of a fundamental wave into a nonlinear crystal. If the Rayleigh zone of the Gaussian beam is larger than the crystal length, an almost planar wave propagates in the crystal volume. If the focusing is too tight, phase matching is again violated in the strongly divergent sections of the beam.

phase matching  $\Delta k = 0$ , the following field strength can be found ( $\kappa$  according to Eq. (15.11)):

$$\mathcal{E}_{\text{SHG}}(r) = i\kappa \mathcal{E}_{\text{FUN}}^2 \ell.$$

With  $w_{\text{SHG}}^2 = w_{\text{FUN}}^2/2$ , the beam waists of fundamental and harmonic waves, the total output power depends on the fundamental input power and parameters such as the material constant  $\Gamma$  from Eq. (15.14) and the crystal length  $\ell$ :

$$P_{\text{SHG}} = \Gamma^2 \ell^2 I_0^2 \frac{\pi w_{\text{SHG}}^2}{2} = \Gamma^2 \ell^2 P_{\text{FUN}}^2 \frac{1}{\pi w_{\text{FUN}}^2}. \quad (15.17)$$

This corresponds to the already known result of Eq. (15.16). Besides, it can be calculated easily that the fundamental and harmonic waves have the same confocal parameter  $b_{\text{SHG}} = b_{\text{FUN}}$  under these circumstances.

Boyd *et al.* have extended this analysis, initially derived for the case of  $90^\circ$  phase matching, to the angle phase matching situation. For this, normalized coordinates for the propagation direction ( $z \rightarrow t$ ) and the walk-off direction (walk-off angle  $\rho$  and  $x \rightarrow u$ ) are introduced,

$$t = \frac{\sqrt{2\pi} z}{\ell_a} \quad \text{with} \quad \ell_a = \sqrt{\pi} w_{\text{FUN}} / \rho,$$

$$u = \frac{\sqrt{2}(x - \rho\ell)}{w_{\text{FUN}}},$$

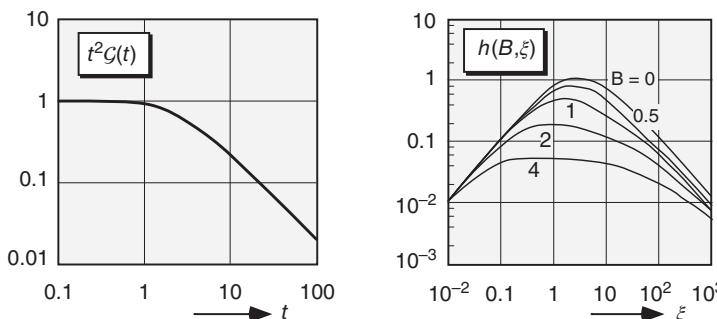
and two new functions are defined by (Figure 15.10)

$$\mathcal{F}(u, t) = \frac{1}{t} \int_0^t e^{-(u+\tau)^2} d\tau,$$

$$\mathcal{G}(t) = \int_{-\infty}^{\infty} \mathcal{F}^2(u, t) du.$$

The length  $\ell_a$  is called the “aperture length” and indicates when the harmonic beam has left the volume of the fundamental wave by walk-off. As a result, Eq. (15.17) is modified by the function  $\mathcal{G}(t) \leq 1$ :

$$P_{\text{SHG}} = \frac{\Gamma^2 \ell^2 P_{\text{FUN}}^2}{\pi w_{\text{FUN}}^2} \mathcal{G}(t).$$



**Figure 15.10** Graphical representation of the functions  $\mathcal{G}(t)$  and  $h(B, \xi)$  (according to Boyd and Kleinman [220]).

This describes the reduction of the output power caused by the walk-off compared to beams that propagate with perfect overlap. In order also to describe the influence of focusing, it is common to introduce the parameters

$$\begin{aligned} h(B, \xi) & \quad \text{Boyd-Kleinman reduction factor,} \\ B = \frac{1}{2}\rho(kl)^{1/2} & \quad \text{birefringence parameter,} \\ \xi = \ell/b & \quad \text{normalized crystal length.} \end{aligned} \quad (15.18)$$

The result is

$$P_{\text{SHG}} = \frac{\Gamma^2 \ell^2 P_{\text{FUN}}^2}{\pi u_{\text{SHG}}^2} \frac{1}{\xi} h(B, \xi).$$

For 90° phase matching, we have  $B = 0$ , and for  $\xi = \ell/b < 0.4$ , we can approximate  $h(0, \xi) = h_0(\xi) \simeq \xi$ , so that the previous result of Eq. (15.17) can be reproduced. One generally finds

$$h(0, \xi) = h_0(\xi) \simeq 1 \quad \text{for } 1 \leq \xi \leq 6,$$

and the maximum value

$$h_0(\xi) = 1.068 \quad \text{at } \xi = 2.84$$

is realized for a crystal length corresponding to nearly three times the Rayleigh length.

Another useful approximation for  $h(B, \xi)$  can be obtained by drawing on the birefringence parameter  $B$  (Eq. (15.18)). For  $1 \leq \xi \leq 6$  where  $h(B, \xi) \simeq h_M(B)$  and  $h_M(0) \simeq 1$  holds, the approximation

$$h_M(B) \simeq \frac{h_M(0)}{1 + (4B^2/\pi)h_M(0)} \simeq \frac{1}{1 + \ell/\ell_{\text{eff}}} \quad \text{at } \ell/\ell_{\text{eff}} \gg 1$$

is found. Here the effective crystal length  $\ell_{\text{eff}}$  has been introduced:

$$\ell_{\text{eff}} = \frac{\pi}{k\rho^2 h_M(0)} \simeq \frac{\pi}{k\rho^2}.$$

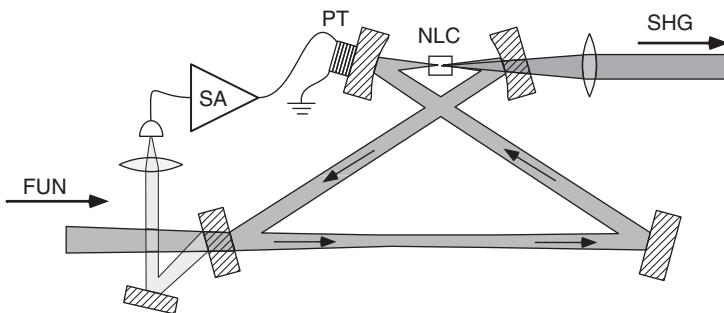
### 15.4.5 Resonant Frequency Doubling

The low conversion efficiency of NLCs can be used in a better way if the light is recycled after passing through the crystal. This can be achieved in passive resonators, some of the essential features of which we shall now describe. Alternatively, some nonlinear components can be inserted into active resonators. An important example for *intra-cavity frequency doubling* is the powerful frequency-doubled neodymium laser from Section 8.4.2.

#### 15.4.5.1 Passive Resonators

To the losses of the resonator by transmission ( $T$ ) and absorption ( $A$ ), we now also have to add the conversion of the radiation power of the fundamental into the harmonic wave. Ashkin *et al.* [221] have elaborated that the maximum power of the harmonic wave can be determined from the implicit equation

$$P_{2\omega} = \frac{16T^2 \eta_{\text{SP}} P_\omega}{[2 - \sqrt{1-T}(2-A-\sqrt{\eta_{\text{SP}} P_{2\omega}})]^4}. \quad (15.19)$$



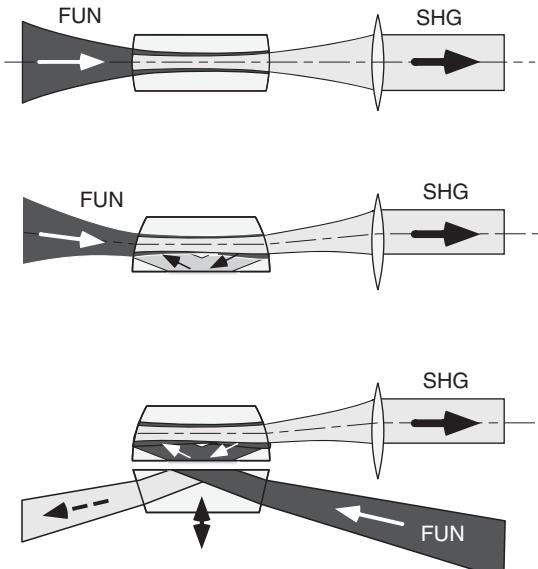
**Figure 15.11** Frequency doubling in a “bowtie” ring resonator. FUN: fundamental wave; SHG: second harmonic wave; SA: servo amplifier; NLC: nonlinear crystal; PT: piezo translator.

Here the single-pass conversion efficiency (i.e., by single passage of the fundamental wave) for the crystal is indicated by  $\eta_{SP} = P_{2\omega}^0 / P_\omega^2$ .

In Figure 15.11 a ring resonator to enhance the fundamental wave intensity is presented. The NLC is positioned at the focus of the resonator. For optimum results, the fundamental wave has to be precisely matched to the Gaussian mode of the resonator by external optical elements. Furthermore, one of the resonator mirrors can be adjusted by a piezo translator (PT). It is controlled by a servo amplifier (SA) and ensures that the resonator length is resonantly matched to the fundamental wave (FUN). The error signal may, for instance, be obtained from the properties of the light reflected off the input mirror.

In the best case  $P_{2\omega}$  in Eq. (15.19) can be maximized by adjusting the transmission  $T$ . This is not possible when mirrors with a fixed reflectivity are used, but the frustrated total reflection can be used to achieve variable coupling of a resonator with a driving field (see Figure 15.12).

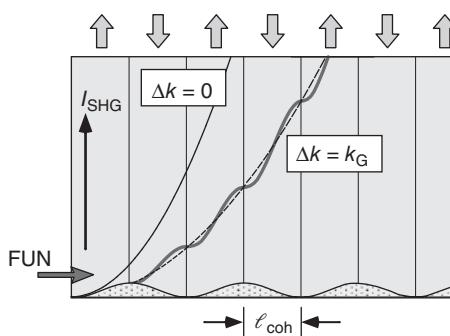
**Figure 15.12** Frequency doubling in monolithic resonators. In the lower ring resonator, the fundamental wave (FUN) is coupled by frustrated total internal reflection (FTIR). The coupling strength is controlled by varying the separation of the monolithic resonator and the coupling prism.



A compact layout for frequency conversion is offered by external resonators, which are directly made from the NLC, that is, “monolithically” manufactured (Figure 15.12). They are well suited for temperature-controlled phase matching (Section 15.4.3.2). The mirrors are integrated through thin layers deposited onto the end facets of the NLC or by means of total internal reflection. The coupling into the ring can advantageously be achieved by frustrated total internal reflection (FTIR) since the transmission is set by varying the separation, and therefore optimum conversion conditions according to Eq. (15.19) can be obtained.

#### 15.4.6 Quasi-phase Matching

For frequency conversion, the proper materials always have to be used under special conditions, for example, angle phase matching, since the generally small electro-optical coefficients do not allow any large tolerances. The low conversion efficiency of a laser beam at a single pass through a nonlinear material has driven the search for better materials (i.e., especially with higher electro-optical coefficients) or improved methods, like the resonator-enhanced frequency doubling from the previous chapter. The search for new materials, though, is laborious, and the effort for servo controls is quite high with resonant methods.



**Figure 15.13** Quasi-phase matching in nonlinear crystals. The orientation of the ferroelectric domains is inverted after each coherence length  $\ell_c$  (“periodic poling,” PP). The wave in the lower part shows the effect of the crystal without PP. FUN: Input of fundamental wave. (According to [224].)

the electrodes and generates “periodic poling” of the ferroelectric domains of certain NLCs.<sup>2</sup> As a consequence a periodic sign change occurs in the coupling of fundamental and harmonic wave. Successfully used crystals include  $\text{LiNbO}_3$  and  $\text{KTiOPO}_4$  (KTP). For the periodically manipulated form, also new abbreviations, for example, PPLN, for periodically poled  $\text{LiNbO}_3$  are used.

We have already investigated the role of the coherence length  $\ell_{\text{coh}}$  (Eq. (15.15)) in Section 15.4.1 for a material where the phase evolution of fundamental and

An alternative route has been opened through successful generation of so-called “periodically poled” materials where existing and reliable nonlinear materials are tailored in such a way that they allow efficient frequency conversion. The principle of quasi-phase matching is presented in Figure 15.13. It was already suggested shortly after the invention of the laser [222] but has led to reproducible and robust results only with the manufacturing methods of microelectronics [223].

For the main fabrication method, a periodic pattern of alternating electrodes is deposited onto the crystal. A high-voltage pulse then reverses the orientation underneath

2 Of the 18 crystal groups allowing phase matching by using birefringence, only 10 are suitable for this method due to symmetry reasons.

harmonic waves is not matched with the help of birefringence. The coherence length  $\ell_c$  sets the scale for the period of choice for the artificially induced domain change. At the domain walls, phase reversal of the coupling between fundamental and harmonic waves takes place because of the change of the sign of the  $d$  coefficient. Thus the retroactive conversion found in a homogeneous material is suppressed by the periodic poling structure, and the harmonic wave instead continues to grow.

The theoretical description of frequency doubling in homogeneous materials can straightforwardly be extended to the situation of periodic poling, where the modulation of the sign of the  $d$  coefficient is taken into account by a Fourier series:

$$d(z) = d_{\text{eff}} \sum_{m=-\infty}^{\infty} G_m e^{-ik_m z} \quad \text{and} \quad G_m = \frac{2}{m\pi} \sin(m\pi\ell/\Lambda). \quad (15.20)$$

Especially  $k_m = 2\pi m/\Lambda$  indicates the reciprocal vector of the domain lattice, where  $\Lambda$  indicates the geometrical length of the period. In the end, only one of the Fourier components plays a significant role, all the others contributing only weak conversion similar to the mismatched situation without periodic poling lattice. The important coefficient fulfills the “quasi-phase matching condition”  $\Delta k = k_m$ , and it is used in the orders  $m = 1, 3, \dots$ . Furthermore one finds [224] that the effective  $d$  coefficient is reduced by the Fourier coefficient ( $|G_m| < 1$ ):

$$d_Q = d_{\text{eff}} G_m.$$

We can now adapt the coupled amplitude equations (15.13) to the new situation by the replacements  $\Delta k \rightarrow \Delta k_Q = \Delta k - k_m$  and  $d_{\text{eff}} \rightarrow d_Q$  and  $\kappa \rightarrow \kappa_Q$ , respectively, in Eq. (15.11). The largest coefficient occurs in first order  $m = 1$ ; thus  $d_Q/d_{\text{eff}} = 2/\pi$  from Eq. (15.20); higher orders though allow longer periods and therefore reduce the tolerance requirements in manufacturing.

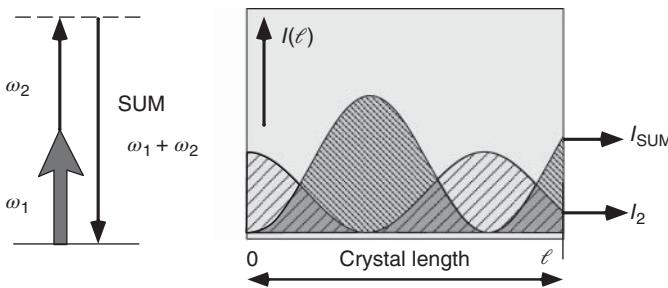
The quasi-phase matching concept causes a reduction of the nonlinear coefficients, but, more importantly, it offers efficient conversion largely independent of the special crystal properties of birefringence. Successful operation of the continuous parametric oscillators, the topic of the following chapter, has been stimulated very much by structured materials with periodic poling [225].

## 15.5 Sum and Difference Frequency

### 15.5.1 Sum Frequency

For this case we have to consider the full set of equations (15.10). In the case of sum frequency generation, there are already two fields with intensities  $I_{1,2}(z = 0) = I_{10,20}$  (and thus  $\mathcal{A}_1 \gg \mathcal{A}_2$ ) present at the entrance of a crystal. For the special case of a very strong pumping field  $I_{10} \gg I_{20}$  and perfect phase matching ( $\Delta k = 0$ ), Eqs. (15.10) are greatly simplified:

$$\begin{aligned} \text{(i)} \quad & \frac{d}{dz} \mathcal{A}_{\text{SUM}} = i\kappa \mathcal{A}_1 \mathcal{A}_2, \\ \text{(ii)} \quad & \frac{d}{dz} \mathcal{A}_1 \simeq 0, \\ \text{(iii)} \quad & \frac{d}{dz} \mathcal{A}_2 = i\kappa \mathcal{A}_1^* \mathcal{A}_{\text{SUM}}. \end{aligned} \quad (15.21)$$



**Figure 15.14** Weak conversion limit of sum frequency generation as a function of the crystal length. Radiative power oscillates between  $I_{\text{SUM}}$  and the weaker of the two input components, here  $I_2$ .

The solutions can be easily found by inserting (15.21 (iii)) into (15.21 (i)),

$$\frac{d^2}{dz^2} \mathcal{A}_{\text{SUM}} = -\kappa^2 |\mathcal{A}_1|^2 \mathcal{A}_{\text{SUM}},$$

and by applying initial conditions  $\mathcal{A}_{1,2}(z=0) = \mathcal{A}_{10,20}$ . With the inverse scaling length,

$$K = \sqrt{\frac{\kappa^2 I_{10}}{c\epsilon_0 \omega_1}},$$

the normalized amplitudes and intensity evolve along  $z$  according to (Figure 15.14)

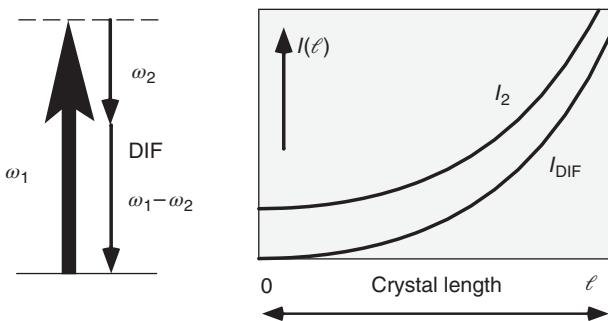
$$\begin{aligned} \mathcal{A}_2(z) &= \mathcal{A}_{20} \cos(Kz), & I_2(z) &= I_{20} \sin^2(Kz), \\ \mathcal{A}_{\text{SUM}}(z) &= \mathcal{A}_{20} \sin(Kz), & I_{\text{SUM}}(z) &= (\omega_{\text{SUM}}/\omega_2) I_{20} \sin^2(Kz). \end{aligned}$$

The intensity of the sum frequency wave has naturally to be larger by the factor  $\omega_{\text{SUM}}/\omega_2$  than  $I_2$  because energy is drawn from both pump waves. When the weaker input component is entirely converted, difference frequency generation (at the initial frequency  $\omega_2$ ) occurs, until all radiation power is used up again. Thus the intensity oscillates between the sum frequency and the weaker of the two components  $I_{1,2}$ .

### 15.5.2 Difference Frequency and Parametric Gain

Let us again consider the case in which a third wave is generated from a strong pumping wave (normalized amplitude  $d\mathcal{A}_1/dz \approx 0$ ) by difference frequency mixing with a second weaker wave. Then in analogy to Eq. (15.21) and in the case of perfect phase matching ( $\Delta k = 0$ ), the coupled amplitude equations are approximately

$$\begin{aligned} \text{(i)} \quad & \frac{d}{dz} \mathcal{A}_{\text{DIF}} = i\kappa \mathcal{A}_1 \mathcal{A}_2^*, \\ \text{(ii)} \quad & \frac{d}{dz} \mathcal{A}_1 \simeq 0, \\ \text{(iii)} \quad & \frac{d}{dz} \mathcal{A}_2 = i\kappa \mathcal{A}_{\text{DIF}}^* \mathcal{A}_1. \end{aligned} \tag{15.22}$$



**Figure 15.15** Parametric gain for difference frequency generation.

The corresponding solutions are (Figure 15.15)

$$\begin{aligned} \mathcal{A}_2(z) &= \mathcal{A}_{20} \cosh(Kz), & I_2(z) &= I_{20} \cosh^2(Kz), \\ \mathcal{A}_{\text{DIF}}(z) &= -i\mathcal{A}_{20} \sinh(Kz), & I_{\text{DIF}}(z) &= (\omega_1/\omega_2)I_{10} \sinh^2(Kz). \end{aligned}$$

For  $Kz \gg 1$  the intensity dependence shows an interesting behavior,

$$I_1(z) \simeq I_{20} e^{2Kz} \quad \text{and} \quad I_2(z) \simeq (\omega_2/\omega_1)I_{10} e^{2Kz},$$

where both waves are amplified in this “parametric process” at the expense of the pumping wave! A more general solution for Eq. (15.22) with coefficients  $\alpha, \beta$  to match initial conditions is given by

$$\mathcal{A}_1(z) = \alpha \sinh(Kz) + \beta \cosh(Kz).$$

## 15.6 Optical Parametric Oscillators

Nonlinear generation of coherent radiation is not only interesting for short-wavelength production. In principle it promises generation of coherent and fully *tunable* radiation over very wide wavelength ranges. For this purpose the *optical parametric oscillator* (OPO) has been suggested and investigated for a long time. This nonlinear process, more than all others, is affected by loss processes that we have completely neglected up to now but will investigate more deeply here.

First we phenomenologically add the losses  $\gamma$  suffered by the waves passing through the crystal to the coupled amplitude equations from Eq. (15.9) and introduce the specific terms *signal wave* and *idler wave* of the parametric oscillator:

$$\begin{aligned} \left( \frac{d}{dz} + \gamma \right) \mathcal{A}_P(\omega) &= i\kappa \mathcal{A}_S \mathcal{A}_I e^{-i\Delta kz} && \text{pumping wave,} \\ \left( \frac{d}{dz} + \gamma_S \right) \mathcal{A}_S(\omega_S) &= i\kappa \mathcal{A}_P \mathcal{A}_I^* e^{i\Delta kz} && \text{signal wave,} \\ \left( \frac{d}{dz} + \gamma_I \right) \mathcal{A}_I(\omega_I) &= i\kappa \mathcal{A}_S^* \mathcal{A}_P e^{i\Delta kz} && \text{idler wave.} \end{aligned}$$

Additionally, we again assume the intensity of the pumping wave to be constant ( $d\mathcal{A}_P/dz \simeq 0$ ). From the ansatz  $\mathcal{A}_S(z) = \tilde{\mathcal{A}}_S e^{(\Gamma+i\Delta k/2)z}$  and  $\mathcal{A}_I(z) = \tilde{\mathcal{A}}_I e^{(\Gamma+i\Delta k/2)z}$ ,

with constant amplitudes  $\tilde{\mathcal{A}}_{S,I}$ , the condition

$$\left[ \left( \Gamma + \gamma_S + i \frac{\Delta k}{2} \right) \left( \Gamma + \gamma_I - i \frac{\Delta k}{2} \right) - \kappa^2 |\mathcal{A}_P|^2 \right] \mathcal{A}_S = 0 \quad (15.23)$$

can be obtained. For constant  $\mathcal{A}_S \neq 0$  this is exactly fulfilled if the term in square brackets vanishes. This is the case for

$$\Gamma_{\pm} = -\frac{\gamma_I + \gamma_S}{2} \pm \frac{1}{2} \sqrt{(\gamma_I - \gamma_S - i\Delta k)^2 + 4\kappa^2 |\mathcal{A}_P|^2}.$$

To simplify the interpretation, we consider the special case  $\gamma = \gamma_S = \gamma_I$  where the relation for  $\Gamma_{\pm}$  becomes particularly simple,

$$\Gamma_{\pm} = -\gamma \pm g, \quad g = \frac{1}{2} \sqrt{-\Delta k^2 + 4\kappa^2 |\mathcal{A}_P|^2}. \quad (15.24)$$

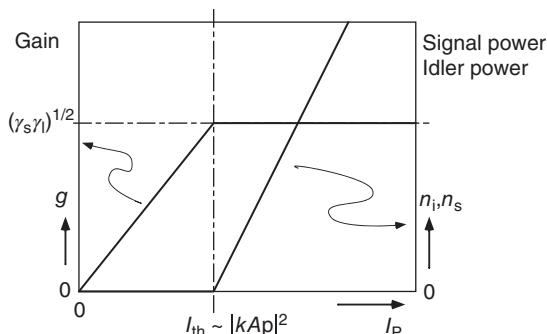
The general solution for the coupled waves is

$$\begin{aligned} \mathcal{A}_S &= (\mathcal{A}_{S+} e^{gz} + \mathcal{A}_{S-} e^{-gz}) e^{-\gamma z} e^{-i\Delta kz/2}, \\ \mathcal{A}_I^* &= (\mathcal{A}_{I+}^* e^{gz} + \mathcal{A}_{I-}^* e^{-gz}) e^{-\gamma z} e^{-i\Delta kz/2}, \end{aligned}$$

and obviously for  $g > \gamma$ , gain is expected. If at the entrance of the crystal there are the amplitudes  $\mathcal{A}_{S,I}(z=0) = \mathcal{A}_{S0,I0}$ , then for the limiting case of weak conversion, that is,  $d\mathcal{A}_P/dz \approx 0$ , we find the following field strengths at the end at  $z = \ell$  (Figure 15.16):

$$\begin{aligned} \mathcal{A}_S(\ell) &= [\mathcal{A}_{S0} \cosh(g\ell) - (i/g)(\Delta k \mathcal{A}_{S0} + i\kappa \mathcal{A}_P \mathcal{A}_{I0}^*) \sinh(g\ell)] e^{-gl} e^{i\Delta k\ell/2}, \\ \mathcal{A}_I(\ell) &= [\mathcal{A}_{I0} \cosh(g\ell) - (i/g)(\Delta k \mathcal{A}_{I0} + i\kappa \mathcal{A}_P \mathcal{A}_{S0}^*) \sinh(g\ell)] e^{-gl} e^{i\Delta k\ell/2}. \end{aligned} \quad (15.25)$$

For perfect phase matching ( $\Delta k = 0$ ) and for  $\mathcal{A}_{S0} = 0$ , we reproduce the old result from the difference frequency generation. How the incident fields are really amplified obviously depends on their phase position at the entrance. If there is only one incident field, then the second wave "searches" the right phase position for optimum gain. The solutions of (15.25) depend on the condition that there is at least one field already present at the crystal entrance. In close analogy to the laser, the fulfillment of condition (15.23) can also be understood as a threshold condition. If the parametric gain is generated in a resonator (Figure 15.17), then the parametric amplifier becomes a parametric oscillator.



**Figure 15.16** Gain and power of the parametrically generated fields in a parametric oscillator (see Figure 9.1).

**Table 15.1** Comparison of laser and optical parametric oscillator (OPO).

	Laser	OPO
Process	$\chi^{(1)}$ , resonant	$\chi^{(2)}, \chi^{(3)}, \dots$ , nonresonant
Mechanism	Occupation number inversion	Nonlinear polarization
Pump process	Incoherent, energy storables	Coherent, not storables

Like the laser the OPO starts spontaneously if the gain  $g$  overcomes the losses  $\sqrt{\gamma_I \gamma_S}$ . Parametric oscillators can be operated singly, doubly, or even triply resonant, to keep the threshold as low as possible, though again at the expense of large efforts for servo-controlling the optical resonator. It is of course not surprising that according to Eq. (15.24), the gain is proportional to the pumping intensity.

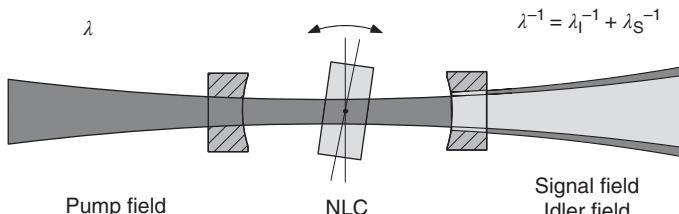
In the operation of tunable lasers (e.g., Ti-sapphire laser, dye laser), inversion is commonly provided by powerful pump lasers. In contrast to the OPO, a *coherent* pumping field is, however, not essential for the laser process. In fact incoherent processes, for example, decay from the pump level, typically take part in the occupation of the upper laser level.

Since the gain depends on the phase mismatch  $\Delta k$  according to Eq. (15.24), the wavelengths of signal and idler wave,  $\lambda_S$  and  $\lambda_I$ , which have to fulfill the equation

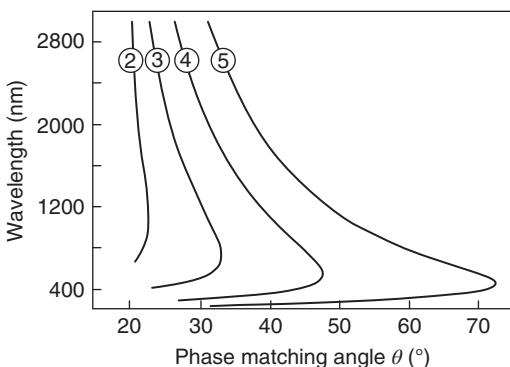
$$\lambda_p^{-1} = \lambda_I^{-1} + \lambda_S^{-1}$$

due to energy conservation, can be tuned by varying the angle or the temperature of the birefringent and NLC. If the pump wavelength in the *degenerate optical parametric oscillator* (DOPO) is decomposed exactly into two photons at  $\omega_S = \omega_I = \omega_p/2$ , the reverse process of frequency doubling, then the corresponding phase matching condition has to be valid again,  $n_{2\omega}(\omega_p) = n_\omega(\omega_p/2)$ . If ordinary dispersion,

$$n_\omega(\omega_{S,I}) \approx n_\omega(\omega_p/2) + n^{(1)}(\omega_{S,I} - \omega_p/2) + \dots ,$$



**Figure 15.17** OPO with linear resonator. The matching of signal and idler wave is achieved by turning the crystal axis if the phase matching is achieved by angular matching. A multiply resonant setup is in principle difficult to get.



**Figure 15.18** Tunability of a parametric oscillator driven with a BBO crystal: wavelength of signal and idler waves. The OPO is pumped by the second (532 nm), third (355 nm), fourth (266 nm), or even fifth (213 nm) harmonic of a Nd laser at 1064 nm.

is assumed, then a quadratic form for the phase matching condition of the signal and idler frequency is expected near to the degeneracy point:

$$c\Delta k = 0 = n_{2\omega}(\omega_p)\omega_p - [n_\omega(\omega_p/2)\omega_p + n^{(1)}(\omega_s - \omega_i)^2 + \dots].$$

On the other hand, the difference of the refraction coefficient depends approximately linearly on the angle or the temperature, so that the quadratic behavior can also be found in the experimental dependence (Figure 15.18).

## Problems

- 15.1 Frequency doubling with KDP** (a) Determine the angle for type I phase matching at  $\lambda = 1 \mu\text{m}$ . The indices of refraction are  $n_o^\omega = 1.496044$  and  $n_o^{2\omega} = 1.514928$  for the ordinary beam and  $n_e^\omega = 1.460993$  and  $n_e^{2\omega} = 1.472486$  for the extraordinary beam. (b) Sketch the index ellipsoid and the direction of propagation within the crystal with respect to the optical axis. How would you cut the crystal? (c) Carry out the same estimates for type II phase matching. Which index of refraction is relevant for the harmonic wave?
- 15.2 Temperature phase matching with KNbO<sub>3</sub>** Use the data for the indices of refraction from Figure 15.6 in order to estimate the wavelength regions for which KNbO<sub>3</sub> can be used for frequency doubling by temperature tuning.
- 15.3 Frequency doubling with short pulse** We consider a pulsed plane wave with Gaussian envelope and center frequency  $\omega$ :

$$\mathbf{E}_1(z, t) = \frac{1}{2} \{ \mathbf{e}_1 A_1(z, t) e^{-i(\omega t - kz)} + \text{c.c.} \}, A_1(0, t) = A_0 e^{-(t^2/2\delta)}.$$

The index of refraction of the fundamental wave is  $n$ . Furthermore we assume that the phase matching condition is fulfilled for a crystal of length  $\ell$ . We neglect losses and wave front deformations, and the group velocities of fundamental and harmonic wave are called  $v_{g1}$  and  $v_{g2}$ .

(a) Express  $A_1(z, t)$  as a function of  $A_1(0, t)$  and  $v_{g1}$ . (b) Within the *slowly varying envelope approximation*, the wave equation within the crystal reads

$$\frac{\partial A_2}{\partial z} + \frac{1}{v_{g2}} \frac{\partial A_2}{\partial z} = i \frac{2\omega}{2nc} \chi_{\text{eff}} A_1^2(z, t).$$

Here  $\chi_{\text{eff}}$  is the effective susceptibility. Substitute  $u = t - z/v_{g1}$  and  $v = t - z/v_{g2}$ , and introduce  $\beta = 1/v_{g1} + 1/v_{g2}$ . (c) Solve the wave equation for  $A_2(u, v)$ . (Hint:  $\text{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp(-u^2) du$ .) Give  $A_2(z, t)$  for the initial condition  $A_2(0, t) = 0$  for all  $t$ . (d) Transform to the rest frame that moves at the group velocity of the harmonic wave  $v_{g2}$ . Sketch how  $A_2$  develops in this rest frame as a function of time. Give the amplitude at the end of the crystal. (e) For which value  $t_0$  is  $A_2(\ell, t)$  maximal? What is the shape of  $A_2(\ell, t_0)$ ? What happens at  $\beta\ell/\delta \gg 1$ ?

## 16

### Nonlinear Optics II: Four-Wave Mixing

In analogy to the three-wave mixing processes of Section 15.2, it is not difficult to compile a typology for four-wave phenomena. Three of the four waves generate a polarization

$$P_i(\omega) = \epsilon_0 \chi_{ijk\ell}^{(3)}(\omega; \omega_1, \omega_2, \omega_3) E_j(\omega_1) E_k(\omega_2) E_\ell(\omega_3), \quad (16.1)$$

which is now characterized by the third-order susceptibility. This fourth-rank tensor describing *four-wave mixing* (FWM) has up to 81 independent components and therefore is not even to be subjected to general symmetry considerations, which could still be described with limited effort for the second-order susceptibility. Instead it is important from the beginning to consider special cases. For formal consideration, there are basically no new aspects compared to three-wave mixing –only the number of coupled amplitude equations is increased by one in FWM.

#### 16.1 Frequency Tripling in Gases

It is obvious in analogy to frequency doubling to ask for frequency tripling by means of the  $\chi^{(3)}$  nonlinearity. In Figure 16.1 it can be seen that *third harmonic generation* (THG) is one of numerous special cases of FWM.

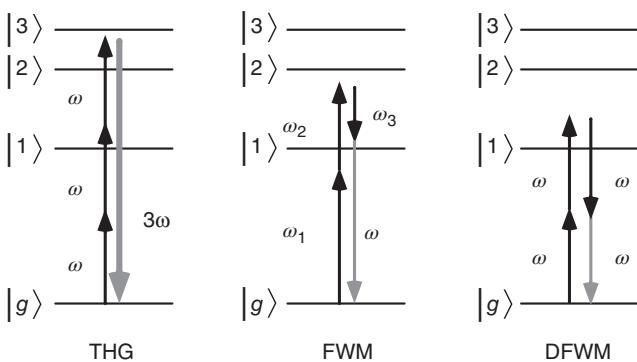
For practical reasons, this  $\chi^{(3)}$  process is really only used when frequencies lying very deep in the ultraviolet (UV) spectral range are to be reached. While nonlinear crystals are transparent (i.e., at wavelengths  $\lambda > 200$  nm), it is an advantage to use frequency doubling and consecutive summation in a two-step  $\chi^{(2)}$  process (Figure 16.2). For example, the 1064 nm line of the Nd laser is transformed, preferably with KTP and LBO materials (systematic names are potassium titanyl phosphate ( $\text{KTiOPO}_4$ ) and Lithium triborate ( $\text{LiB}_3\text{O}_5$ )), to the wavelengths 532 and 355 nm. For this a conversion efficiency of 30% using pulsed light is a matter of routine. The UV radiation generated at 355 nm in this way is very suitable to pump dye lasers in the blue spectral range.

If we neglect geometry effects, the polarization of third order is

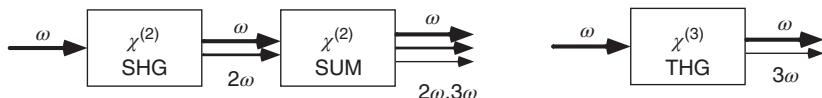
$$\mathcal{P}^{3\omega} = \epsilon_0 \chi^{(3)}(3\omega; \omega, \omega, \omega) \mathcal{E}^3.$$

The condition for phase matching, which is

$$\Delta k = k_{3\omega} - 3k_\omega,$$



**Figure 16.1** Selected  $\chi^{(3)}$  processes for which the state of the nonlinear material is preserved: third harmonic generation (THG), an example of four-wave mixing (FWM) and degenerate four-wave mixing (DFWM).



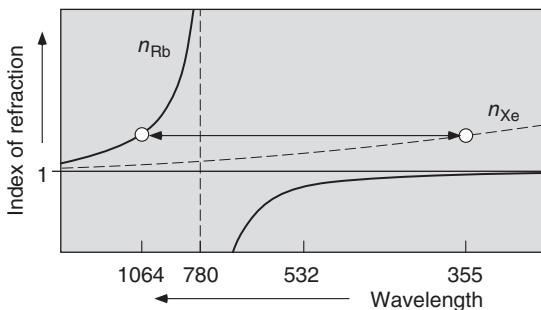
**Figure 16.2** Frequency tripling with two-step  $\chi^{(2)}$  and one-step  $\chi^{(3)}$  processes.

in this case has to be obtained by adjusting the refraction coefficients of the fundamental wave and harmonic wave as for the generation of the second harmonic (Section 15.4). As already mentioned, crystals are of limited use for frequency tripling due to their very small  $\chi^{(3)}$  coefficients, poor transparency, and the danger of optically induced damage caused by extreme input power and strong absorption of the UV harmonic wave. Gases, however, have a high threshold of destruction and good transparency below the threshold of photoionization, which is at  $\lambda \approx 50$  nm for several noble gases.

The disadvantage of low density in a gas can be compensated by enhancing the nonlinear process using a suitable molecular or atomic resonance in the vicinity of the fundamental wave. Therefore, for the generation of UV light at very short wavelengths, often alkali vapors are used, which allow near-resonant amplification due to their transition frequencies at wavelengths in the visible and near UV range. They also exhibit a relatively rapidly varying refractive index with normal or anomalous dispersion depending on the position of the fundamental frequency. The resonance lines of the noble gases are in the deep UV ( $< 100$  nm) and mostly in the range of normal dispersion. By adding the 100–10 000-fold amount of noble gas atoms to an alkali vapor, the phase velocity of the harmonic wave can be adjusted. Figure 16.3 shows a qualitative example of phase matching for frequency tripling of the 1064 nm line of a Nd laser: xenon gas is added to rubidium vapor.

Even if the generation of extreme ultraviolet (XUV) radiation in a gas container were successful, the transport to the planned application still raises special problems since the atmosphere and even the best-known window material, cooled LiF, lose their transparency slightly below 100 nm. That is why very short-wave coherent radiation has in general to be generated very close to the experiment.

**Figure 16.3** Matching refractive indices for frequency tripling of the 1064 nm radiation in rubidium vapor (D2 resonance line at 780 nm) by adding xenon gas.



## 16.2 Nonlinear Refraction Coefficient (Optical Kerr Effect)

In the third order a nonlinear contribution to the polarization at the fundamental wave itself also arises. This is a special case of the *degenerate four-wave mixing* (DFWM), which obviously occurs with well-matched phase propagation because  $\Delta\mathbf{k} = \mathbf{k} + \mathbf{k} - \mathbf{k} = \mathbf{k}$  from the beginning. In analogy to the traditional electro-optical Kerr effect, where the refraction coefficient depends on the external electric field (see Section 3.5.1), nonlinear materials showing this effect are often called *Kerr media*.

The contribution to the polarization of the fundamental wave at the driving frequency  $\omega$  is

$$\mathcal{P}^{\text{KE}}(\omega) = \epsilon_0 \chi_{\text{eff}}^{(3)}(\omega; \omega, \omega, -\omega) |\mathcal{E}(\omega)|^2 \mathcal{E}(\omega),$$

so that the total polarization is<sup>1</sup>

$$\mathcal{P}(\omega) = \epsilon_0 [\chi^{(1)} + \chi_{\text{eff}}^{(3)} |\mathcal{E}(\omega)|^2] \mathcal{E}(\omega) = \epsilon_0 \chi_{\text{eff}} \mathcal{E}(\omega).$$

The total polarization clearly depends on the intensity, and it is convenient to describe this phenomenon transparently anyway by an intensity-dependent refraction coefficient,

$$n = n_0 + n_2 I,$$

with  $n_0$  the common linear refractive index and  $n_2$  a new material constant describing this nonlinearity. By comparing to  $n^2 = 1 + \chi_{\text{eff}}$  and with  $I = n_0 \epsilon_0 c |\mathcal{E}|^2 / 2$ ,

$$n_2 \simeq \frac{1}{n_0^2 c \epsilon_0} \chi_{\text{eff}}^{(3)}.$$

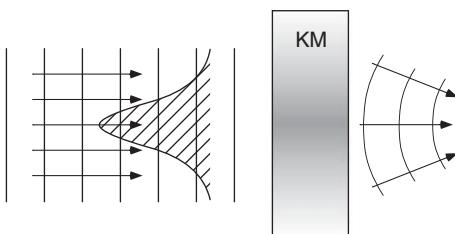
The nonlinear coefficient  $n_2$  naturally depends on the material. Its value varies over a large range and is, for example, just  $10^{-16}\text{--}10^{-14} \text{ cm}^2 \text{ W}^{-1}$  for common glasses. However, it can be larger by several orders of magnitude in special materials, for example, in doped glasses. The propagation of the light field will then

<sup>1</sup> There are several definitions of the susceptibility used, which differ from each other mainly by geometry and factors accounting for degeneracy. Here we use an effective susceptibility neglecting such details.

strongly depend on the intensity distribution in both space and time. Owing to this nonlinearity, transverse intensity variations of a light beam cause distortions of optical wave fronts, leading, for instance, to self-focusing. In Section 3.3.2 we have already seen that the self-modulation of the phase caused by longitudinal variations of the intensity, for example, in a laser pulse, can lead to the generation of solitons under certain conditions.

### 16.2.1 Self-Focusing

The transverse Gaussian profile of the TEM<sub>00</sub> mode is certainly the best-known and most important intensity distribution of all light beams. If the intensity is sufficiently large, for example, in a short intense laser pulse, then in a Kerr medium, it causes an approximately quadratic variation of the refraction coefficient and thus a lens effect, which acts like a converging lens for  $n_2 > 0$  and like a diverging lens for  $n_2 < 0$  (Figure 16.4).



**Figure 16.4** Self-focusing of a planar wave in a Kerr medium (KM). The intensity profile of a Gaussian beam causes a parabolic transverse variation of the refraction coefficient and therefore acts like a lens.

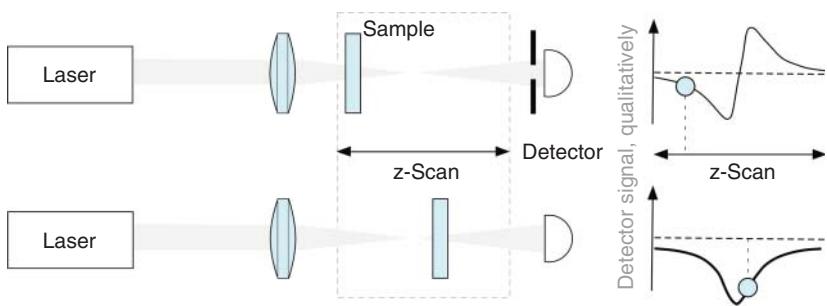
The focal length depends on the maximum intensity. This effect is actually related to the *thermal lens* where the refractive index variation is caused by an inhomogeneous temperature profile. The change of the refraction coefficient is caused by local temperature variation there and by a nonlinearity here. Temperature modifications can also be generated by a laser beam (e.g., through absorption), but thermal changes are usually very slow (milliseconds)

compared to the very fast optical Kerr effect (femto-to nanoseconds) and thus generally not desirable from the practical point of view.

#### Example: Z-scan technique for measuring $\chi^{(3)}$ coefficients

Applications of nonlinear optical phenomena relies on the availability of suitable materials with good – that is, large –  $\chi^{(3)}$  coefficients. The so-called *z-scan* method employs self-focusing to measure these material constants. The basic and simple setup is shown in Figure 16.5: a sample with nonlinear optical properties is swept through the focus of a laser beam (“z-scan”), typically for five times the Rayleigh length  $z_0$  (Section 2.3.1.1). The nonlinear effect increases with local flux density of the light beam. In the closed aperture configuration, the detector effectively monitors the central maximum intensity of the beam. The Kerr nonlinearity modifies the focusing properties of the beam, and hence a dispersive signal caused by the intensity-dependent index of refraction change  $\Delta n = n_2 \cdot I$  occurs.

Monitoring the full beam on the detector amounts to a measurement of the change in absorption,  $\Delta\alpha$ . Extraction of material constants requires a model for Gaussian beam propagation in the sample. Since all nonlinear processes occur simultaneously, interpretation is not always simple [226].

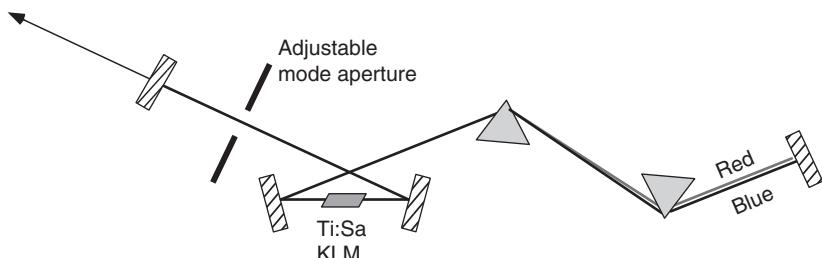


**Figure 16.5** Experimental setup for the z-scan technique. Upper (lower) row: closed (open) aperture arrangement. The position of the sample on the transmission signals is indicated by  $\odot$ .

### 16.2.1.1 Kerr Lens Mode Locking

One of the most important applications of self-focusing at present is the so-called Kerr lens mode locking (KLM), which has made the construction of laser sources for extremely short pulses nearly straightforward (see Section 9.5). The self-mode-locking concept was discovered in 1991 [143] with a Ti–sapphire laser, which could be switched from cw to stable pulsed operation by small mechanical disturbances. The laser (Figure 16.6) consists of just the laser crystal, the mirrors, and a pair of prisms for the compensation of the crystal dispersion in the laser crystal and the laser components (see also Section 9.5.3 and Figure 9.16).

The trick of self-mode-locking is to align the laser resonator in such a way that during pulsed operation – at which only the induced Kerr lens is active – the resonator field suffers from less losses than during cw operation. One can control these losses, for instance, by means of an additional aperture in the resonator. In order to switch the laser from cw to pulsed mode, intensity fluctuations must be induced, for example, by relaxation oscillations. It is sometimes sufficient to cause mechanical disturbance – the laser operating mode is so to say induced with a punch.

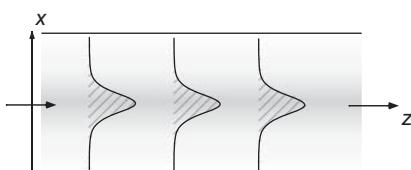


**Figure 16.6** Ti-sapphire laser with intrinsic Kerr lens mode locking (KLM) in the laser crystal. The pair of prisms compensates the dispersion. This simple layout generates typical pulse lengths of 50–100 fs. With the mode aperture, cw operation is put at an energetic disadvantage against the pulsed operation.

For stable pulsed and mode-locked operation, it is necessary that the shape of the pulse reproduces itself after one period of revolution. The short pulse has a correspondingly wide frequency spectrum and is lengthened due to the dispersion of the laser crystal itself as well as other resonator components. Thus prism pairs are used for compensation (Figure 16.6). In materials with normal dispersion, red wavelengths travel faster than blue components. Compensation of this delay is achieved by the double prism compensator, which refracts the blue waves stronger than the red components, resulting in longer geometric paths for red than for blue waves. Further modifications of the pulse are caused by self-phase modulation; see Section 16.3.

### 16.2.1.2 Spatial Solitons

Another consequence of self-focusing ought to be mentioned as well. As we have investigated in the section about quadratic index media, optical waves are guided in axial media such as a gradient fiber. In a nonlinear medium it is possible that an intense light beam causes “self-wave guiding” through the nonlinearity of the optical Kerr effect. A propagating beam tends to diverge, as described by Gaussian beam optics, as a result of diffraction. In a Kerr medium, however, the intensity distribution may at the same time cause a quadratic transverse index variation prompting a lensing effect. If it exactly compensates diffraction, it allows stable and self-guided beam propagation [227, 228].



**Figure 16.7** Propagation of a spatial soliton in a Kerr medium. Note that confinement in the second transverse direction must be achieved by other means.

We introduce the intensity-dependent transverse variation of the refraction coefficient in the  $x$  direction,

$$n(x) = n_0 + n_2 I(x) = n_0 + \frac{2n_2 |\mathcal{A}(x)|^2}{cn_0 \epsilon_0}, \quad (16.2)$$

similar to the paraxial Helmholtz equation (2.35). For the sake of clarity, we introduce  $\kappa = 2k^2 n_2 / cn_0^2 \epsilon_0$ , obtaining the *nonlinear Schrödinger equation* (see Section 3.3.2),

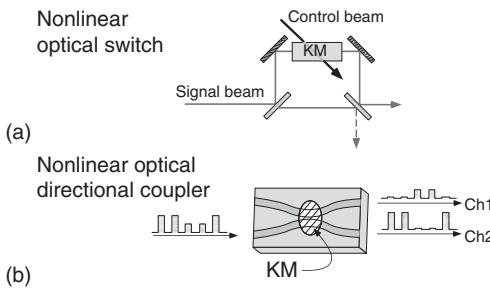
$$\left( \frac{\partial^2}{\partial x^2} + 2ik \frac{\partial}{\partial z} + \kappa^2 |\mathcal{A}|^2 \right) \mathcal{A} = 0,$$

which of course has only its mathematical structure in common with quantum mechanics. It is known that this equation has self-consistent solutions of the form

$$\mathcal{A}(x, z) = \mathcal{A}_0 \operatorname{sech} \left( \frac{x}{w_0} \right) \exp \left( \frac{iz}{4z_0} \right).$$

The properties of this wave are similar to the Gaussian modes with a “beam waist”  $w_0^2 = (\kappa \mathcal{A}_0)^2 / 2$  and a “Rayleigh length”  $z_0 = kw_0^2 / 2$ . The wave propagates along the  $z$  direction and is called a *spatial soliton*.<sup>2</sup> In contrast to the Gaussian beam (see Section 2.3) in a homogeneous medium, the beam parameters ( $w_0, z_0$ )

<sup>2</sup> The “optical” solitons varying in time discussed in Section 3.3.2 are more widely known though.



**Figure 16.8** Applications of the nonlinear optical Kerr effect. (a) A Kerr medium (KM) can be used in order to direct a signal beam into one of two exits of a Mach–Zehnder interferometer. In this “all-optical” switch, the refraction coefficient of the Kerr cell causes a phase delay depending on the status of the control beam. (b) In a directional coupler (e.g., realized through surface wave guides in  $\text{LiNbO}_3$ ), an incoming signal is distributed into two output channels (Ch1 and Ch2). The coupling efficiency can depend on the input intensity and thus separate pulses of different intensity.

now depend on the amplitude  $A_0$ ! The self-stabilizing mode does not propagate divergently either, but keeps its form undamped over large distances.

Note that from the beginning with Eq. (16.2), we have considered a one-dimensional variation (in  $x$ ) of the index of refraction only. It turns out that the two-dimensional analogue with variations in both  $x$  and  $y$  does not yield stable solutions. Two-dimensionally stable modes of propagation can, of course, be obtained if an additional wave guiding effect in the second direction is applied, introduced by, for example, saturation phenomena or other additional nonlinearities.

### 16.2.1.3 Nonlinear Optical Devices

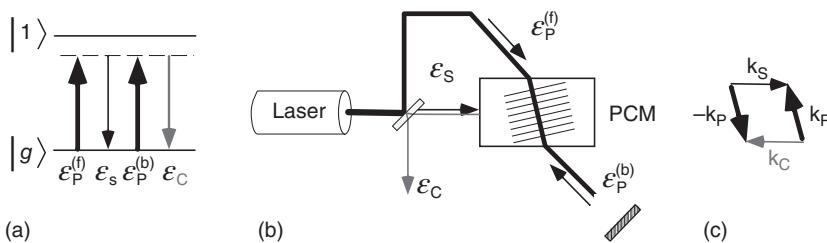
The nonlinear optical Kerr effect is quite interesting for certain applications, for example, in optical communications. Two examples are presented in Figure 16.8. A nonlinear switch is realized by changing the optical length in one branch of a Mach–Zehnder interferometer through a control beam using the Kerr effect. In this way the signal beam can be switched between the two exits. In a nonlinear directional coupler, the coupling efficiency depends on the intensity of the input signal so that pulse sequences with two different intensities can be multiplexed into two channels.

### 16.2.2 Phase Conjugation

Phase conjugation (or “wave front reversal”) occurs as a special case of DFWM(Figures 16.1 and 16.9). The phase adjustment is fulfilled intrinsically and ideally since only one optical frequency is involved. The polarization is again calculated according to Eq. (16.1):

$$\mathcal{P}^{\text{PC}}(\omega_S) = \epsilon_0 \chi_{\text{eff}}^{(3)}(\omega_S; \omega_p, \omega_p, -\omega_S) \mathcal{E}_p^{(\text{f})} \mathcal{E}_p^{(\text{b})} \mathcal{E}_S^*.$$

Because  $\sum_i \mathbf{k}_i = 0$  the phase matching condition is always fulfilled in a trivial way if two waves (in Figure 16.9, the forward-running ( $\mathcal{E}_p^{(\text{f})}$ ) and backward-running



**Figure 16.9** (a) Phase conjugation as a special case of degenerate four-wave mixing.  
(b) Simple setup for phase conjugation (PCM = phase conjugating medium, e.g., BaTiO<sub>3</sub>, CS<sub>2</sub>).  
(c) The phase matching condition is always fulfilled in a trivial way.

( $\mathcal{E}_P^{(b)}$ ) pump waves) counter-propagate each other. The phase conjugating process can be strongly enhanced by choosing a wavelength in the vicinity of a one-photon resonance.

We now study a simplified theoretical description of phase conjugation, the result of which differs only slightly from the more exact method. In this the nonlinear change of the refraction coefficient for the pumping waves is also taken into account. We especially assume that the intensity of the pump waves does not change,  $d\mathcal{E}_P/dz \approx 0$ . Then only two waves instead of four need to be considered:

$$\mathcal{P}_C = \epsilon_0 \chi_{\text{eff}}^{(3)} \mathcal{E}_P^2 \mathcal{E}_S^*,$$

$$\mathcal{P}_S = \epsilon_0 \chi_{\text{eff}}^{(3)} \mathcal{E}_P^2 \mathcal{E}_C^*.$$

We set  $\kappa = \omega \chi_{\text{eff}}^{(3)} / 2nc\mathcal{E}_P^2$  and consider the signal and conjugate waves propagating in the positive and negative  $z$  directions:

$$\mathcal{A}_C = \mathcal{A}_{C0} e^{ikz} \quad \text{and} \quad \mathcal{A}_S = \mathcal{A}_{S0} e^{-ikz}.$$

They have to fulfill the differential equations

$$\frac{d}{dz} \mathcal{A}_{S0} = i\kappa \mathcal{A}_{C0}^*, \quad \mathcal{A}_{S0}(z=0) = \mathcal{A}(0),$$

$$\frac{d}{dz} \mathcal{A}_{C0} = -i\kappa \mathcal{A}_{S0}^*, \quad \mathcal{A}_{C0}(z=\ell) = 0.$$

The boundary conditions at the end of the crystal assume that a signal wave exists at the front end (at  $z = 0$ ) of the crystal but no conjugated wave yet at the rear ( $z = \ell$ ). Here the origin of the phase conjugation is clearly identified as the newly generated conjugated wave  $\mathcal{A}_{C0}$ , which is driven by the conjugate amplitude  $\mathcal{A}_{S0}^*$ .

The solutions are found straightforwardly. For the signal wave as well as for the conjugated one, there is amplification:

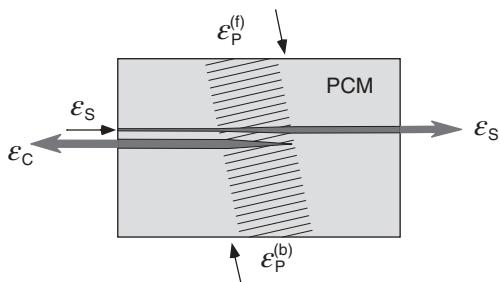
$$\mathcal{A}_{S0} = \frac{\mathcal{A}(0)}{\cos(|\kappa|\ell)} \quad \text{and} \quad \mathcal{A}_{C0} = \frac{i\kappa}{|\kappa|} \tan(|\kappa|\ell) \mathcal{A}^*(0).$$

The phase conjugation has a fascinating application for *wave front reconstruction* or *wave front reversal*. Before we study this phenomenon in more detail, we introduce an alternative interpretation derived from conventional holography, which we have discussed already in Section 6.8. In holography, a conjugated wave is known to occur as well!

The interference of a pump wave with the signal wave causes a periodic modulation of the intensity and thus of the refraction coefficient in the phase conjugating medium (PCM in Figure 16.11) with reciprocal lattice vector  $\mathbf{K}$ :

$$\mathbf{K} = \mathbf{k}_p - \mathbf{k}_s \quad \text{and}$$

$$\Lambda = \frac{1}{2} \lambda \sin(\theta/2).$$



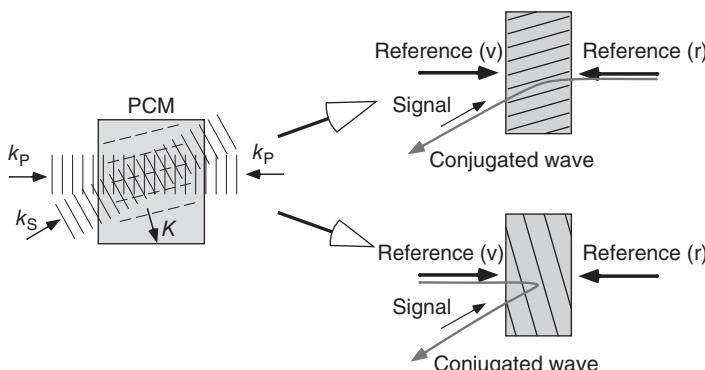
**Figure 16.10** Signal and conjugated waves within a phase conjugating medium (PCM). Both waves are amplified.

The counter-propagating pump wave exactly fulfills the Bragg condition,

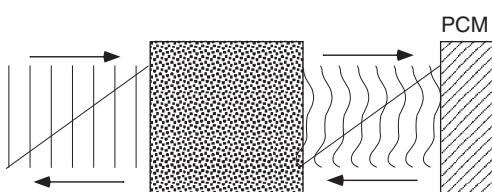
$$\sin(\theta/2) = \frac{\lambda}{2\Lambda},$$

and is diffracted by this phase grating into the direction precisely opposing the signal wave.

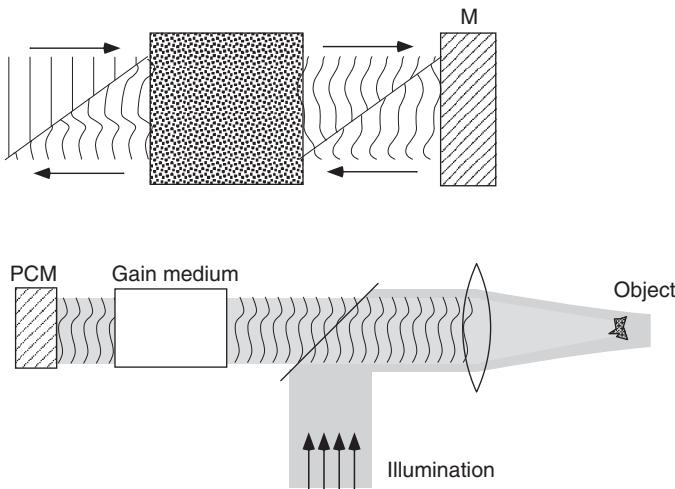
Wave front reversal is shown in Figure 16.12 by comparison to a conventional mirror. While a conventional mirror turns a wave front around on reflection, a phase conjugating mirror (PCM) reverses the wave vector of propagation while maintaining the shape of the wave front. Any distortion caused by inhomogeneous but linear media will be reversed, and hence a laser beam with initially smooth wave fronts will emerge from a PCM with identical shape. A possible application is efficient focusing of intense laser radiation onto an object with a surface inappropriately matched to conventional, that is, Gaussian-shaped, laser beams.



**Figure 16.11** Real-time holography and phase conjugation. Left: geometry of relevant waves. Upper right: the forward-propagating pump wave forms a grating by superposition with the signal wave. The backward-running pump wave fulfills the Bragg condition for this grating and is scattered in the direction of the signal wave. Lower right: a similar argument can be used for interference of the backward-propagating pump wave and the signal wave.



**Figure 16.12** Wave front reversal or reconstruction using a phase conjugating mirror (PCM) and a conventional one (M).



**Figure 16.13** Application of a phase conjugating mirror for focusing of intense laser radiation onto an optically inappropriate adjusted object.

### 16.3 Self-Phase Modulation

The nonlinear modification of the refraction coefficient takes effect not only on the spatial wave fronts of laser light but also on the time-variant structure. These nonlinear phenomena are not only important for short pulse lasers because of their extreme peak intensities but also used for relevant applications. Consider a light pulse with Gaussian amplitude distribution and characteristic pulse length  $\tau$ ,

$$E(t) = E_0 e^{-(t/\tau)^2/2} e^{-i\omega t} \quad \text{and} \quad I(t) = I_0 e^{-(t/\tau)^2},$$

during passage through a nonlinear medium. The phase of the light pulse at the end of a sample of length  $\ell$  develops according to

$$\begin{aligned} \Phi(t) &= nkz|_{\ell} = n(t)k\ell \\ &= (n_0 + n_2 I_0 e^{-(t/\tau)^2}) ckt. \end{aligned}$$

The instantaneous frequency is then

$$\omega(t) = \frac{d}{dt}\Phi(t) = [n_0 - n_2 I_0 2(t/\tau) e^{-(t/\tau)^2}] ck.$$

During the pulse this represents a shift from blue to red frequencies or vice versa, depending on the sign of  $n_2$ . This phenomenon is generally called a *frequency chirp*. In the center at  $\exp[-(t/\tau)^2] \simeq 1$ , a linear variation can be found:

$$\omega(t) \simeq \omega_0 - 2\beta t \quad \text{with} \quad \beta = \omega_0 \frac{n_2 I_0}{n_0} \frac{\ell}{\tau}.$$

Frequency chirp is not unusual; in fact, the laser pulses emitted by the simplest versions of Kerr lens mode-locked lasers (Section 16.2.1.1) always tend to exhibit such frequency variations. In previous chapters we have furthermore encountered other situations where self-phase modulation is important. It is the origin of soliton propagation in optical fibers described in Section 3.3.2. The stretchers and compressors introduced in Section 9.5.5 can be used to control, remove, or enhance the chirp.

## Problems

- 16.1 Generation of the third harmonic wave** Consider a gas in a cell bounded with planes at  $z = 0$  and  $z = \ell$ . A monochromatic, plane wave propagates in the  $z$  direction,  $\mathbf{E}(z, t) = \frac{1}{2}[E\mathbf{e}_x \exp[-i(\omega t - kz)]]$ . (a) Describe qualitatively the spectrum of the nonlinear polarization of the gas. What is the role of the anisotropy of the system? (b) Give a scalar expression for the nonlinear polarization at  $3\omega$ . (c) Calculate the field radiated at  $3\omega$ . Take the initial condition  $A_{3\omega}(z = 0) = 0$  for the amplitude of the harmonic wave. How does the intensity grow with  $z$ ? (d) Vary the density of gas inside the cell. How does the intensity of the harmonic wave at the exit of the cell depend on the density?
- 16.2 Phase conjugating mirror I** Imagine you were looking at yourself in a phase conjugating mirror. What would you see?
- 16.3 Phase conjugating mirror II** In a phase conjugating medium of effective length  $\ell$ , we consider the amplitude of the conjugated wave with  $z$  dependence  $A_C(z) = A_C(0) \cos(|\kappa|(z - \ell)) / \cos(|\kappa|\ell)$  and the signal wave with  $A_S(z) = i\kappa^* A_S^*(0) \sin(|\kappa|(z - \ell)) / \cos(|\kappa|\ell)$ . The coupling constant of the interaction with the pump waves reads  $\kappa = (\omega/2nc)\chi^{(3)}A_{P1}A_{P2}$ . (a) Sketch the spatial dependence of the amplitudes and intensities inside the crystal for the conjugated and for the signal wave. How does the system evolve for  $\pi/4 < |\kappa|\ell = \pi/2$ ? (b) Calculate and interpret the reflectivity of the conjugated beam, which is defined by  $R = |A_C(0)/A_S(0)|^2$ . Study the special case  $|\kappa|\ell = \pi/2$ .
- 16.4 Dispersion compensation** Describe qualitatively the effect of the prisms on the dispersion of the resonator in Figure 16.6. How can one tune the compensation?

# A

## Mathematics for Optics

### A.1 Spectral Analysis of Fluctuating Measurable Quantities

The Fourier transformation is the “natural” method to describe the evolution of an optical wave since in the end all optical phenomena can be considered the summation of the action of elementary waves according to Huygens’ principle. Exactly this action is calculated with the help of the Fourier transformation.

By *fluctuations* of a physical quantity, we understand its irregular variations in time. Deterministic physical predictions can be made not about the actual behavior of a time-variant quantity but about the probability distribution of its possible values, for example, the amplitude distribution of a signal voltage. From the theory of probability, it is known that the distribution of a stochastic quantity  $V(t)$  is completely determined when all of its moments are known. By this, the averages  $\langle V \rangle$ ,  $\langle V^2 \rangle$ ,  $\langle V^3 \rangle, \dots$  are understood. Often a certain distribution is known – or assumed – for example, a Gaussian normal distribution for *random* events. Then it is sufficient to give the leading moments of the distribution, for example, the average value  $\langle V \rangle$  and the *variance*  $\langle (V - \langle V \rangle)^2 \rangle$ . The square root of the variance is called the *root mean square deviation* or in short r.m.s. value  $V_{\text{rms}}$ :

$$V_{\text{rms}}^2 = \frac{1}{T} \int_0^T [V(t) - \langle V(t) \rangle]^2 dt = \langle V^2(t) \rangle - \langle V(t) \rangle^2. \quad (\text{A.1})$$

In experiments, nearly all measurable quantities are ultimately converted into electrical signals reflecting their properties. For processing dynamic electrical signals, *filters* play a very special role since their use allows the desired and undesired parts of a signal to be separated from each other. The action of a filter or a combination of filters can be understood most simply by the effect on a sinusoidal or *harmonic* quantity with a varying frequency  $f = \omega/2\pi$ . Thus it is important for both theoretical and practical reasons to characterize the fluctuations of a measurable quantity not only in the time domain but also in the frequency domain, that is, by a spectral analysis.

In physics and engineering sciences, the description of a time-dependent quantity by its frequency or *Fourier components* has proven to be invaluable for a long time. The complex voltage  $V(t)$ , for example, can be decomposed into partial waves and described in frequency space:

$$V(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{V}(\omega) e^{-i\omega t} d\omega = \int_{-\infty}^{\infty} \mathcal{V}(f) e^{-2\pi ift} df. \quad (\text{A.2})$$

We can interpret  $\mathcal{V}(f) df$  as the amplitude of a partial wave at the frequency  $f$  and with frequency bandwidth  $df$ . The *amplitude spectrum* has the unit V Hz<sup>-1</sup>, and as a complex quantity it also contains information about the phase angle of the Fourier components. The functions  $V(t)$  and  $\mathcal{V}(\omega)$  constitute a *Fourier transform pair* with the inverse transformation

$$\mathcal{V}(\omega) = \int_{-\infty}^{\infty} V(t) e^{i\omega t} dt. \quad (\text{A.3})$$

The effect of a simple system of filters, for example, low or high pass, on a harmonic excitation can often be given by a *transfer function*  $T(\omega)$ . The advantages of the frequency or Fourier decomposition according to Eq. (A.2) show up in the simple linear relation between the input and the output of such a network:

$$V'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} T(\omega) \mathcal{V}(\omega) e^{-i\omega t} d\omega.$$

This method delivers satisfactory results for numerous technical applications. This is especially valid in the case when the signal is periodic and the relation between time and frequency domains is exactly known. A noisy signal varies sometimes rapidly, sometimes slowly, and consequently it has contributions from both low and high frequencies. Thus the mathematical relation according to Eq. (A.2) cannot be given since an infinitely expanded measurement interval would be necessary. From a rigorous mathematical point of view, even a very large time interval cannot be considered a sufficiently good approximation since there is not even some information about the boundedness of the function and thus about the convergence properties of the integral transformation.

On the other hand, the Fourier component of an arbitrary signal<sup>1</sup> can indeed be measured with an appropriate narrowband filter by measuring its average transmitted power. In every *spectrum analyzer* the signal strength  $V^2$  transmitted through a filter with tunable center frequency  $f$  and bandwidth  $\Delta f$  is measured. The square is generated by electronic hardware, for example, by rectification and analogue quadrature of the input. Let us take  $P_V(t) = V^2(t)$  as the *generalized power* of an arbitrary signal  $V(t)$ .

For the formal treatment, we introduce the Fourier integral transform of the function  $V(t)$  on a finite measurement interval of length  $T$ :

$$\mathcal{V}_T(f) = \int_{-T/2}^{T/2} V(t) e^{i2\pi ft} dt. \quad (\text{A.4})$$

---

<sup>1</sup> In most cases the signal will be an electric current or voltage. In optics the currents or voltages are generated with photodetectors.

The average total power in this interval is

$$\langle V^2 \rangle_T = \frac{1}{T} \int_{-T/2}^{T/2} V^2(t) dt.$$

We can introduce the Fourier integral transform according to Eq. (A.4) and exchange the order of integration (we leave out the index  $\langle \rangle_T$  in the following since there cannot be any confusion):

$$\begin{aligned} \langle V^2 \rangle &= \frac{1}{T} \int_{-T/2}^{T/2} [V(t) \int_{-\infty}^{\infty} \mathcal{V}_T(f) e^{-2\pi i ft} df] dt \\ &= \frac{1}{T} \int_{-\infty}^{\infty} [\mathcal{V}_T(f) \int_{-T/2}^{T/2} V(t) e^{-2\pi i ft} dt] df. \end{aligned}$$

The variable  $\langle V^2 \rangle$  is very useful because with its help we can calculate the variance  $\Delta V^2 = \langle V^2 \rangle - \langle V \rangle^2$  and thus the second moment of the distribution of the quantity  $V(t)$ , at least within the restricted interval  $[-T/2, T/2]$ . Since  $V(t)$  is a real quantity, we have  $\mathcal{V}_T(-f) = \mathcal{V}_T^*(f)$  according to (A.2), and we can write

$$\langle V^2 \rangle = \frac{1}{T} \int_{-\infty}^{\infty} [\mathcal{V}_T(f) \mathcal{V}_T(-f)] df = \frac{1}{T} \int_{-\infty}^{\infty} |\mathcal{V}_T(f)|^2 df.$$

Owing to the symmetry of  $\mathcal{V}_T(f)$ , it is sufficient to carry out the single-sided integration  $0 \rightarrow \infty$ . We define the power spectral density  $S_V(f)$ ,

$$S_V(f) = \frac{2|\mathcal{V}_T(f)|^2}{T}, \quad (\text{A.5})$$

obtaining a relation that may be interpreted as

$$\langle V^2 \rangle = \int_0^{\infty} S_V(f) df. \quad (\text{A.6})$$

According to this,  $S_V(f) df$  is exactly the contribution of the average power of a signal  $V(t)$  transmitted by a linear filter with center frequency  $f$  and bandwidth  $\Delta f$ . Toward higher frequencies the *power spectrum*  $S_V(f)$  usually drops off with  $1/f^2$  or faster so that the total noise power remains finite.

Often the formal and unphysical notation  $\sqrt{S_V(f)}$  with units  $\text{V Hz}^{-1/2}$  is used, which again gives a noise amplitude. This always refers to a noise power, however. For optical detectors the noise amplitudes of voltage and current in units of  $(V^2 \text{ Hz}^{-1})^{1/2}$  and  $(A^2 \text{ Hz}^{-1})^{1/2}$ , respectively, are most important and are thus given separately once again:

$$i_n(f) = \sqrt{S_I(f)}, \quad e_n(f) = \sqrt{S_U(f)}. \quad (\text{A.7})$$

Then, the r.m.s. values of noise current and voltage in a detector bandwidth  $B$  are  $I_{\text{rms}} = i_n \sqrt{B}$  and  $U_{\text{rms}} = e_n \sqrt{B}$ , respectively. In a rather sloppy way, they are often simply called “current noise” and “voltage noise,” but one has to be aware of the fact that in calculations always only the squared values  $i_n^2 B$  and  $e_n^2 B$  are used. Also, for applications of this simple relation, one assumes that the noise properties are more or less constant within the frequency interval of width  $B$ .

### A.1.1 Correlations

The fluctuations of measurable quantities can alternatively be described by means of correlation functions. With correlation functions, one investigates how the value of a quantity  $V(t)$  evolves away from an initial value:

$$C_V(t, \tau) = \langle V(t)V(t + \tau) \rangle_T = \frac{1}{T} \int_{-T/2}^{T/2} V(t)V(t + \tau) dt.$$

In this case we have already assumed a realistic finite time interval  $T$  for the measurement. In general, we will investigate stationary fluctuations, which do not themselves depend on time, so that the correlation function does not explicitly depend on time either. Often useful physical information is given by the normalized correlation function,

$$g_V(\tau) = \frac{\langle V(t)V(t + \tau) \rangle}{\langle V \rangle^2},$$

which for  $\tau \rightarrow 0$  results in

$$g_V(\tau \rightarrow 0) = \frac{\langle V(t)^2 \rangle}{\langle V \rangle^2} = 1 + \frac{\Delta V^2}{\langle V \rangle^2}.$$

Here the term  $\Delta V(t)^2 = \langle (\langle V \rangle - V(t))^2 \rangle$  results exactly in the variance. This directly allows one to assess the fluctuations.

We can build a valuable relation with the spectral power density by using the bounded Fourier transforms according to Eq. (A.4) and exchanging again the order of time and frequency integrations:

$$C_V(\tau) = \frac{1}{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-T/2}^{T/2} \mathcal{V}_T(f') \mathcal{V}_T(f) e^{-i2\pi f' t} e^{-i2\pi f(t+\tau)} df df' dt.$$

For very long times  $T \rightarrow \infty$ , we may replace the time integration by the Fourier transform of the delta function,  $\delta(f) = \int_{-\infty}^{\infty} e^{i2\pi ft} dt$ , yielding

$$\begin{aligned} C_V(\tau) &= \frac{1}{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{V}_T(f') \mathcal{V}_T(f) \delta(f + f') e^{-i2\pi f \tau} df df' \\ &= \int_0^{\infty} \frac{2|\mathcal{V}_T(f)|^2}{T} e^{-i2\pi f \tau} df. \end{aligned}$$

With the help of Eq. (A.5), we can immediately justify the Wiener–Khinchin theorem, which establishes a relation between the correlation function and the power spectral density of a fluctuating quantity:

$$C_V(\tau) = \int_0^{\infty} S_V(f) e^{-i2\pi f \tau} df \tag{A.8}$$

and

$$S_V(f) = \int_0^{\infty} C_V(\tau) e^{i2\pi f \tau} d\tau. \tag{A.9}$$

### A.1.2 Schottky Formula

One of the most important and fundamental forms of noise is the so-called shot noise. It arises if a measurable quantity consists of a flow of particles being registered by the detector at random times, for example, the photon flow of a laser beam or the photoelectrons in a photomultiplier or a photodiode.

Let us consider a flow of particles that are registered by a detector as needlelike sharp electrical impulses at random times. We are interested in the power spectrum of this current of random events. If  $N_T$  particles are registered during a measurement interval of length  $T$ , the current amplitude can be given as a sequence of discrete pulses registered at individual instants  $t_k$ :

$$I(t) = \sum_{k=1}^{N_T} g(t - t_k). \quad (\text{A.10})$$

The function  $g(t)$  accounts for the finite rise time  $\tau$  of a real detector, which would give a finite length even to an infinitely sharp input pulse. At first we determine the Fourier transform

$$\mathcal{I}(f) = \sum_{k=1}^{N_T} \mathcal{G}_k(f),$$

with the Fourier transform of the  $k$ th individual event  $\mathcal{G}_k(f) = e^{i2\pi f t_k} \mathcal{G}(f)$ :

$$\mathcal{G}(f) = \int_{-\infty}^{\infty} g(t) e^{i2\pi f t} dt. \quad (\text{A.11})$$

Any single event has to be normalized according to  $\int_{-\infty}^{\infty} g(t) dt = 1$ . If the events are shaped like pulses of typical length  $\tau = f_G/2\pi$ , then the spectrum has to be continuous at frequencies far below the cutoff frequency  $f_G$ ,  $\mathcal{G}(f \ll f_G) \simeq 1$ .

By definition of the power spectrum (A.5), we have  $S_I(f) = 2\langle |\mathcal{I}_T(f)|^2 \rangle / T$ . Thus one calculates

$$\begin{aligned} |\mathcal{I}_T(f)|^2 &= |\mathcal{G}(f)|^2 \sum_{k=1}^{N_T} \sum_{k'=1}^{N_T} e^{i2\pi f(t_k - t_{k'})} \\ &= |\mathcal{G}(f)|^2 \left( N_T + \sum_{k=1}^{N_T} \sum_{k'=1, \neq k}^{N_T} e^{i2\pi f(t_k - t_{k'})} \right). \end{aligned}$$

Averaging over an ensemble makes the second term in the lower row vanish, and  $N_T$  is replaced by the average value  $\bar{N}$ . Thus the power density of the noise is

$$S_I(f) = \frac{2\bar{N}|\mathcal{G}(f)|^2}{T}, \quad (\text{A.12})$$

which depends only on the spectrum  $|\mathcal{G}(f)|^2$  of an individual pulse.

For “needlelike” pulses with a realistic length  $\tau$ , we anticipate an essentially flat spectrum, that is, a *white spectrum* in the frequency range  $f \leq \tau/2\pi$ . For random uncorrelated pulses, we expect not the amplitudes but the intensities to add.

If it is also taken into account that  $S_I(f)$  is obtained by single-sided integration (Eq. (A.5)), we can interpret all factors in Eq. (A.12).

In the special case of an electric current, the relation with the noise power spectral density is called the *Schottky formula*, which is valid for Fourier frequencies below the cutoff frequency of the detector  $f_G$ ,

$$S_I(f) = 2e\bar{I}, \quad (\text{A.13})$$

where we have used  $\bar{I} = e\bar{N}/T$ .

If the amplitude of the individual event fluctuates as well, for example, if we have  $\int_{-\infty}^{\infty} g(t - t_k) dt = \eta_k$ , then Eq. (A.12) is replaced by

$$S_I(f) = \frac{2\bar{N}\langle\eta^2\rangle|\mathcal{G}(f)|^2}{T}. \quad (\text{A.14})$$

Now the average current is  $\bar{I} = \bar{N}\bar{e}\bar{\eta}/T$ , with an average charge  $\bar{e}\bar{\eta}$ . In the Schottky formula (A.13), an additional *excess noise* factor  $F_e = \langle\eta^2\rangle/\langle\eta\rangle^2$  is introduced:

$$S_I(f) = 2\langle e\eta \rangle \langle I \rangle \frac{\langle \eta^2 \rangle}{\langle \eta \rangle^2}. \quad (\text{A.15})$$

This variant is important for photomultipliers and avalanche photodiodes subject to intrinsically fluctuating amplification.

Let us finally consider the special case of an amplitude distribution that has only the random values  $\eta = 0$  and  $\eta = 1$ . In this case we have  $F_e = 1$ , so that events not being registered do not contribute to the noise.

## A.2 Time Averaging Formula

The planar wave is the most important and most simple limiting case that is treated for the propagation of optical waves. There the field vector at a defined position is described by a harmonic function of time:

$$\mathbf{F} = \mathbf{F}_0 e^{-i\omega t}.$$

Often averages of products of harmonically varying functions are required. For this, the Poynting theorem is very useful if physical quantities are described by the real part of a complex harmonic function. If  $\mathbf{F}$  and  $\mathbf{G}$  are two complex harmonic functions, then for arbitrary vector products  $\otimes$  we have for the average taken over a period

$$\langle \Re\{\mathbf{F}\} \otimes \Re\{\mathbf{G}\} \rangle = \frac{1}{2} \langle \Re\{\mathbf{F} \otimes \mathbf{G}^*\} \rangle. \quad (\text{A.16})$$

## B

### Supplements in Quantum Mechanics

#### B.1 Temporal Evolution of a Two-State System

##### B.1.1 Two-Level Atom

A hypothetical two-level atom has only one ground state  $|g\rangle$  and one excited state  $|e\rangle$  to which the raising and lowering operators

$$\sigma^\dagger = |e\rangle\langle g| \quad \text{and} \quad \sigma = |g\rangle\langle e|$$

belong. They are known as linear combinations of the Pauli operators:

$$\sigma^\dagger = \frac{1}{2}(\sigma_x + i\sigma_y), \quad \sigma = \frac{1}{2}(\sigma_x - i\sigma_y).$$

The Hamiltonian of the dipole interaction can be described by

$$H = \hbar\omega_0\sigma^\dagger\sigma + \hbar g e^{-i\omega t} + \hbar g^* e^{i\omega t}, \quad (\text{B.1})$$

with  $\omega_0 = (E_e - E_g)/\hbar$  and using the semiclassical approximation as well as the rotating-wave approximation (RWA). The dipole coupling rate  $g$  is derived from

$$V_{\text{dip}} = (\mathbf{d}^{(+)} + \mathbf{d}^{(-)}) \cdot (\mathbf{E}^{(+)} + \mathbf{E}^{(-)}),$$

where the operator of the dipole matrix element is  $q\mathbf{r} = \mathbf{d} = \mathbf{d}^{(+)} + \mathbf{d}^{(-)}$  and the electric field  $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}^{(+)} e^{-i\omega t} + \mathbf{E}^{(-)} e^{i\omega t}$ , and in general geometric factors accounting for the vectorial nature have to be taken into account [93]. For instance, the RWA is of no relevance for a  $\Delta m = \pm 1$  transition: because  $\mathbf{d}^{(+)} = \langle d \rangle (\mathbf{e}_x + i\mathbf{e}_y) e^{-i\omega_0 t}$  and  $\mathbf{E}^{(+)} = \mathcal{E}_0 (\mathbf{e}_x + i\mathbf{e}_y) e^{-i\omega_0 t}$ , we have exactly  $\mathbf{d}^{(+)} \cdot \mathbf{E}^{(+)} = \mathbf{d}^{(-)} \cdot \mathbf{E}^{(-)} = 0$  in this case.

##### B.1.2 Temporal Development of Pure States

In the interaction picture of quantum mechanics, the temporal development of a state is described according to the equation

$$|\Psi_I(t)\rangle = e^{-iH_I t/\hbar} |\Psi_I(0)\rangle, \quad (\text{B.2})$$

with the interaction Hamiltonian

$$\begin{aligned} H_I &= \hbar g \sigma^\dagger + \hbar g^* \sigma \\ &= \hbar |g| (\cos \phi \sigma_x + \sin \phi \sigma_y). \end{aligned} \quad (\text{B.3})$$

With the Rabi frequency  $\Omega_R = 2|g|$ , then

$$|\Psi_I(t)\rangle = e^{-i(\Omega_R t/2)(\cos \phi \sigma_x - \sin \phi \sigma_y)} |\Psi_I(0)\rangle. \quad (\text{B.4})$$

The state development can be taken from the matrix equation

$$\exp(-i\alpha \boldsymbol{\sigma} \cdot \mathbf{n}) = \mathbf{1} \cos \alpha - i \boldsymbol{\sigma} \cdot \mathbf{n} \sin \alpha. \quad (\text{B.5})$$

## B.2 Density Matrix Formalism

For the expert reader, for convenience, we here collect some results of quantum mechanics for the density operator, leading to the optical Bloch equations. The density matrix formalism allows one to treat an ensemble of two-level atoms.

In a basis of quantum states  $|i\rangle$ , the density operator has the spectral representation

$$\hat{\rho} = \sum_{ij} \rho_{ij} |i\rangle \langle j|.$$

The equations of motion of the discrete elements can then be obtained from the Heisenberg equation with the Hamiltonian  $\mathcal{H}$  under study:

$$i\hbar \frac{d}{dt} \hat{\rho} = [\mathcal{H}, \hat{\rho}].$$

For evaluation it is convenient to use the spectral representation of the Hamiltonian with elements  $H_{ij} = \langle i | \mathcal{H} | j \rangle$ :

$$\frac{d}{dt} \rho_{ij} = -\frac{i}{\hbar} \sum_k [H_{ik} \rho_{kj} - \rho_{ik} H_{kj}]. \quad (\text{B.6})$$

According to this the density matrix of a two-level atom consists of the expectation values

$$\begin{pmatrix} \langle \sigma^\dagger \sigma \rangle & \langle \sigma^\dagger \rangle \\ \langle \sigma \rangle & \langle \sigma \sigma^\dagger \rangle \end{pmatrix}.$$

The Hamiltonian for the states  $|g\rangle$  and  $|e\rangle$  contains the undisturbed operator of the free atom, and in semiclassical approximation the dipole term

$$\mathcal{V}_{\text{dip}} = -(d_{eg} \sigma^\dagger + d_{ge} \sigma) (E^{(+)} e^{-i\omega t} + E^{(-)} e^{i\omega t}),$$

so that

$$\mathcal{H} = \frac{1}{2} \hbar \omega_0 (\sigma^\dagger \sigma - \sigma \sigma^\dagger) + \frac{1}{2} (E_0^{(+)} e^{-i\omega_0 t} + E_0^{(-)} e^{i\omega_0 t}) (d_{eg} \sigma^\dagger + d_{ge} \sigma).$$

We will see that the expectation values  $\langle \sigma^\dagger \rangle$  and  $\langle \sigma \rangle$  oscillate with  $e^{i\omega_0 t}$  and  $e^{-i\omega_0 t}$ , respectively. In the vicinity of a resonance, we use the “RWA,” for which

the terms oscillating with  $\omega + \omega_0$  are neglected. We abbreviate  $g = -d_{\text{eg}}\mathcal{E}_0/2\hbar$  and find

$$\mathcal{H} = \hbar\omega_0\sigma^\dagger\sigma + \hbar g e^{-i\omega t}\sigma^\dagger + \hbar g^* e^{i\omega t}\sigma.$$

From this the equations of motions are obtained as

$$\begin{aligned}\dot{\rho}_{\text{ee}} &= ig^* e^{-i\omega t}\rho_{\text{eg}} - ig e^{i\omega t}\rho_{\text{ge}} = -\dot{\rho}_{\text{gg}}, \\ \dot{\rho}_{\text{eg}} &= i\omega_0\rho_{\text{eg}} + ig e^{-i\omega t}(\rho_{\text{ee}} - \rho_{\text{gg}}) = \dot{\rho}_{\text{ge}}^*.\end{aligned}$$

In the RWA it is moreover convenient to introduce “rotating” elements of the density matrix  $\rho_{\text{eg}} = \bar{\rho}_{\text{eg}} e^{-i\omega t}$  and  $\rho_{\text{ge}} = \bar{\rho}_{\text{ge}} e^{i\omega t}$ . Dropping the overbars for the sake of simplicity, we obtain (detuning  $\delta = \omega - \omega_0$ )

$$\begin{aligned}\dot{\rho}_{\text{ee}} &= -\dot{\rho}_{\text{gg}} = -ig\rho_{\text{ge}} + ig^*\rho_{\text{eg}}, \\ \dot{\rho}_{\text{eg}} &= -i(\omega - \omega_0)\rho_{\text{eg}} + ig(\rho_{\text{ee}} - \rho_{\text{gg}}) = -i\delta\rho_{\text{eg}} + ig(\rho_{\text{ee}} - \rho_{\text{gg}}).\end{aligned}$$

From this system of equations, the optical Bloch equations (7.36) can again be obtained by suitable replacements. After introducing phenomenological damping rates and  $\rho_{\text{eg}} = u + iv$ , for example, one obtains

$$\begin{aligned}\dot{u} &= \delta v - \frac{1}{2}\gamma u - 2\Im\{g\}w, \\ \dot{v} &= -\delta u - \frac{1}{2}\gamma v + 2\Re\{g\}w, \\ \dot{w} &= 2\Im\{g\}u - 2\Re\{g\}v - \gamma w.\end{aligned}\tag{B.7}$$

## B.3 Density of States

The calculation of the density of states (DOS)  $\rho(E) = \rho(\hbar\omega)$  as a function of energy is a standard problem of the physics of many-particle systems. It depends on the dispersion relation,

$$E = E(\mathbf{k}),$$

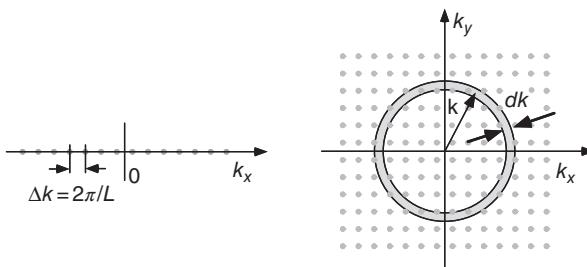
and on the dimension of the problem. In the general case it can be anisotropic as well, though we here limit ourselves to the isotropic case. Two important examples are the dispersion relations of the electron gas and photon gas:

$$\text{Electrons: } E(\mathbf{k}) = \frac{\hbar^2\mathbf{k}^2}{2m},$$

$$\text{Photons: } E(\mathbf{k}) = \hbar\omega = \hbar ck.$$

The DOS  $\rho(E) dE$  describes the number of states within an interval of width  $dE$  in energy space. It is calculated in  $n$  dimensions as

$$\begin{aligned}\rho(E) &= 2 \int_{V_k} d^n k \rho_{\mathbf{k}}(\mathbf{k}) \delta(E - E(\mathbf{k})) \\ &= 2 \frac{1}{(2\pi)^n} \int_{V_k} d^n k \delta(E - E(\mathbf{k})).\end{aligned}\tag{B.8}$$



**Figure B.1** Densities of states in 1D and 2D  $k$ -spaces.

In  $k$ -space we assume a constant density,  $\rho_{\mathbf{k}}(\mathbf{k}) = (1/2\pi)^n$ , within unit volume<sup>1</sup> (Figure B.1) and furthermore take into account the twofold degeneracy due to the polarization of electromagnetic waves and the electron spin, respectively. Then we obtain the densities of states from Table B.1.

**Table B.1** Densities of states in one, two, and three dimensions.

1D	2D	3D
Electromagnetic radiation field: $\omega = ck$ , $\rho(\omega)$		
$\frac{1}{\pi c} d\omega$	$\frac{\omega}{\pi c^2} d\omega$	$\frac{\omega^2}{\pi^2 c^3} d\omega$
Free electron gas: $E = \hbar^2 k^2 / 2m$ , $\rho(E)$		
$\frac{m}{\pi \hbar} (2mE)^{-1/2} dE$	$\frac{m}{\pi \hbar^2} dE$	$\frac{m}{\pi^2 \hbar^3} (2mE)^{1/2} dE$

<sup>1</sup> For the calculation of physically measurable quantities, it has to be summed over the volume of the many-particle system. Thus here we set  $L = 1$  for Eq. (B.8).

## Bibliography

- 1 Goos, F. and Hänchen, H. (1947) Ein neuer und fundamentaler versuch zur totalreflexion. *Ann. Phys. (Leipzig)*, **1**, 333–346.
- 2 Risset, C.A. and Vigoureux, J.M. (1992) An elementary presentation of the Goos-Hänchen effect. *Opt. Commun.*, **91**, 155–157.
- 3 Siegman, A. (1986) *Lasers*, University Science Books, Mill Valley, CA.
- 4 Orloff, J. (1997) *Handbook of Charged Particle Optics*, CRC Press, Boca Raton, FL.
- 5 Friedburg, H. and Paul, W. (1951) Optische abbildung mit neutralen atomen. *Naturwissenschaften*, **38**, 159–160.
- 6 Kaenders, W., Wynands, R., and Meschede, D. (1996) Ein diaprojektor für neutrale atome. *Phys. unserer Zeit*, **27** (1), 28–33.
- 7 Encyclopedia britannica cd98 (1998) CD-ROM Multimedia Edition. Table: The constant of the speed of light.
- 8 Einstein, A. (1905) Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Ann. Phys.*, **17**, 132–148.
- 9 Jackson, J.D. (1999) *Classical Electrodynamics*, 3rd edn, John Wiley & Sons, Inc., New York.
- 10 Abraham, M. (1903) Prinzipien der dynamik des elektrons. *Ann. Phys.*, **10**, 105–179.
- 11 Ueberholz, B. (1996) Ein kompakter titan-saphir-laser. Diplomarbeit, Mathematisch-naturwissenschaftliche Fakultät der Universität Bonn, Bonn.
- 12 Jones, R.C. (1941) A new calculus for the treatment of optical systems. *J. Opt. Soc. Am.*, **31**, 488–493.
- 13 Goldstein, D. (2010) *Polarized Light*, 3rd edn, Marcel Dekker, New York.
- 14 Hecht, E. (2016) *Optics*, 5th edn, Addison-Wesley, Boston, MA.
- 15 Humblet, J. (1943) Sur le moment d'impulsion d'une onde electromagnétique. *Physica*, **10** (7), 585–603.
- 16 Bergmann, K., Theuer, H., and Shore, B.W. (1998) Coherent population transfer among quantum states of atoms and molecules. *Rev. Mod. Phys.*, **70**, 1003–1025.
- 17 Bliokh, K.Y., Aiello, A., and Alonso, M.A. (2012) *The Angular Momentum of Light*, Cambridge University Press, Cambridge.
- 18 Beth, R.A. (1936) Mechanical detection and measurement of the angular momentum of light. *Phys. Rev.*, **50**, 115–125.

*Optics, Light and Lasers: The Practical Approach to Modern Aspects of Photonics and Laser Physics*, Third Edition. Dieter Meschede.

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2017 by Wiley-VCH Verlag GmbH & Co. KGaA.

- 19 Allen, L., Beijersbergen, M., Spreeuw, R., and Woerdman, J. (1992) Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A*, **45** (11), 8185–8189.
- 20 Mawardi, A., Hild, S., Widera, A., and Meschede, D. (2011) ABCD-treatment of a propagating doughnut beam generated by a spiral phase plate. *Opt. Exp.*, **19**, 21205.
- 21 Beijersbergen, M.W., Allen, L., van der Veen, H., and Woerdman, J.P. (1993) Astigmatic laser mode converters and transfer of orbital angular momentum. *Opt. Commun.*, **96**, 123–132.
- 22 Lauterborn, W., Kurz, T., and Wiesenfeldt, M. (1993) *Kohärente Optik*, Springer-Verlag, Berlin Heidelberg, New York.
- 23 Collins, S.A. Jr. (1970) Lens-system diffraction integral written in terms of matrix optics. *J. Opt. Soc. Am.*, **60**, 1168–1177.
- 24 Arfken, G.B. and Weber, H.J. (2001) *Mathematical Methods for Physicists*, Academic Press, London.
- 25 Johnson, P.B. and Christy, R.W. (1972) Optical constants of the noble metals. *Phys. Rev. B*, **6**, 4370–4379.
- 26 Palik, E.D. (1997) *Handbook of Optical Constants of Solids*, Academic Press.
- 27 Born, M. and Wolf, E. (1975) *Principles of Optics*, Pergamon Press, London.
- 28 Raether, H. (1988) *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer-Verlag, Berlin Heidelberg, New York.
- 29 Drazin, P.G. and Johnson, R.S. (1989) *Solitons: An Introduction*, Cambridge University Press, New York.
- 30 Saleh, B. and Teich, M. (1991) *Fundamentals of Photonics*, John Wiley & Sons, Inc., New York, Chichester, Brisbane.
- 31 Mollenauer, L.F., Stolen, R.H., and Islam, H. (1985) Experimental demonstration of soliton propagation in long fibers: loss compensated by Raman gain. *Opt. Lett.*, **10**, 229–231.
- 32 Brixner, T. and Gerber, G. (2003) Quantum control of gas-phase and liquid-phase femtochemistry. *Chem. Phys. Chem.*, **4**, 418–438.
- 33 Wynands, R., Diedrich, F., Meschede, D., and Telle, H.R. (1992) A compact tunable 60-dB faraday optical isolator for the near infrared. *Rev. Sci. Instrum.*, **63** (12), 5586–5590.
- 34 Kao, K.C. and Hockham, G.A. (1966) Fibre surface waveguides for optical frequencies. *Proc. IEE*, **113**, 1151.
- 35 Kim, J., Benson, O., and Yamamoto, Y. (1999) Single-photon turnstile device. *Nature*, **397**, 500.
- 36 Rybczynski, J., Hilgendorff, M., and Giersig, M. (2003) Nanosphere lithography – fabrication of various periodic magnetic particle arrays using versatile nanosphere masks, in *Low-Dimensional Systems: Theory, Preparation, and Some Applications*, NATO Science Series II, vol. **91** (eds L.M. Liz-Marzn and M. Giersig), Kluwer Academic Publishers, Dordrecht, pp. 163–172.
- 37 Matthias, S., Müller, F., Jamois, C., Wehrspohn, R.B., and Gösele, U. (2004) Large-area three-dimensional structuring by electrochemical etching and lithography. *Adv. Mater.*, **16**, 2166–2170.
- 38 Kittel, C. (2004) *Introduction to Solid State Physics*, 8th edn, John Wiley & Sons, Inc., Hoboken, NJ.

- 39 John, S. (1987) Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.*, **58**, 2486–2489.
- 40 Yablonovitch, E. (1987) Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.*, **58**, 2059.
- 41 Ioannopoulos, J.D., Meade, R.S., and Winn, J.N. (1995) *Photonic Crystals*, Princeton University Press, Princeton, NJ.
- 42 Sakoda, K., Ueta, T., and Ohtaka, K. (1997) Numerical analysis of eigenmodes localized at line defects in photonic lattices. *Phys. Rev. B*, **56**, 14905–14908.
- 43 Knight, J.C., Birks, T.A., Russell, P.St.J., and Atkin, D.M. (1996) All-silica single-mode optical fiber with photonic crystal cladding. *Opt. Lett.*, **21** (19), 1547–1549.
- 44 Bjarklev, A., Broeng, J., and Bjarklev, A.S. (2003) *Photonic Crystal Fibres*, Kluwer Academic Publishers, Boston, MA, Dordrecht, London.
- 45 Birks, T.A., Knight, J.C., and Russell, P.St.J. (1997) Endlessly single-mode photonic crystal fiber. *Opt. Lett.*, **22** (13), 961–963.
- 46 Pendry, J.B., Holden, A.J., Stewart, W.J., and Youngs, I. (1996) Extremely low frequency plasmons in metallic microstructures. *Phys. Rev. Lett.*, **76**, 4773.
- 47 Cai, W. and Shalaev, V. (2016) *Optical Metamaterials, Fundamentals and Applications*, 2nd edn, Springer-Verlag, Berlin Heidelberg, New York.
- 48 Veselago, V. (1968) The electrodynamics of substances with simultaneously negative values of  $\epsilon$  and  $\mu$ . *Sov. Phys. Usp.*, **10**, 509.
- 49 Pendry, J.B. (2004) Negative refraction. *Contemp. Phys.*, **45**, 191–202.
- 50 Fang, N., Lee, H., Sun, C., and Zhang, X. (2005) Sub-diffraction-limited optical imaging with a silver superlens. *Science*, **308**, 534.
- 51 Falk, D.S., Brill, D.R., and Stork, D.G. (1986) *Seeing the Light. Optics in Nature, Photography, Color, Vision, and Holography*, Harper & Row, Publishers, New York.
- 52 Kapitza, H.G. (1994) Mikroskopieren von anfang an. Firmenschrift Zeiss, Oberkochen.
- 53 Mansfield, S.M. and Kino, G.S. (1990) Solid immersion microscope. *Appl. Phys. Lett.*, **57**, 2615.
- 54 Vamivakas, A.N., Younger, R.D., Goldberg, B.B., Swan, A.K., Ünlü, M.S., Behringer, E.R., and Ippolito, S.B. (2008) A case study for optics: the solid immersion microscope. *Am. J. Phys.*, **76**, 758.
- 55 Wu, B. and Kumar, A. (2014) Extreme ultraviolet lithography and three dimensional integrated circuit – a review. *Appl. Phys. Rev.*, **1**, 011104.
- 56 International Technology Roadmap for Semiconductors 2.0. 2015 Edition. Executive Report. [http://www.semiconductors.org/clientuploads/Research\\_Technology/ITRS/2015/0\\_2015%20ITRS%202.0%20Executive%20Report%20\(1\).pdf](http://www.semiconductors.org/clientuploads/Research_Technology/ITRS/2015/0_2015%20ITRS%202.0%20Executive%20Report%20(1).pdf)
- 57 Goodman, J.W. (2005) *Introduction to Fourier Optics*, 3rd edn, McGraw-Hill, New York.
- 58 Neuhauser, W., Hoenstatt, M., Toschek, P., and Dehmelt, H. (1980) Localized visible  $ba^+$  mono-ion oscillator. *Phys. Rev. A*, **22**, 1137.
- 59 Moener, W. and Kador, L. (1989) Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.*, **62**, 2535.

- 60 Bobroff, N. (1986) Position measurement with a resolution and noise-limited instrument. *Rev. Sci. Instrum.*, **57**, 1152.
- 61 Thompson, R.E., Larson, D.R., and Webb, W.W. (2002) Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.*, **82**, 2775.
- 62 Betzig, E. (1995) Proposed method for molecular optical imaging. *Opt. Lett.*, **20**, 237.
- 63 Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H. (2006) Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, **313**, 1642.
- 64 Hell, S.W. and Wichmann, J. (1994) Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.*, **19**, 780.
- 65 Westphal, V. and Hell, S.W. (2005) Nanoscale resolution in the focal plane of an optical microscope. *Phys. Rev. Lett.*, **94**, 143903.
- 66 Peréz, J.-P. (1996) *Optique Fondements et Applications*, 3rd edn, Masson, Paris.
- 67 J. Rothenberg, Hubble Space Telescope: Engineering for Recovery, *Opt. Photon. News* **4** (11) 10.
- 68 Crocker, J.H. (1993) Engineering the costar. *Opt. Photon. News*, **4** (11), 22–26.
- 69 Fugate, R.Q. (1993) Laser beacon adaptive optics. *Opt. Photon. News*, **4** (6), 14–19.
- 70 Lipson, H.S., Lipson, S.G., and Tannhauser, D.S. (1981) *Optical Physics*, 2nd edn, Cambridge University Press, Cambridge.
- 71 Kogelnik, H. and Li, T. (1966) Laser beams and resonators. *Proc. IEEE*, **54** (10), 1312–1329.
- 72 Möllenstedt, G. and Düker, H. (1956) Beobachtungen und messungen an biprisma-interferenzen mit elektronenwellen. *Z. Phys.*, **145**, 377.
- 73 Carnal, O. and Mlynek, J. (1991) Young's double-slit experiment with atoms: a simple atom interferometer. *Phys. Rev. Lett.*, **66**, 2689.
- 74 Mandel, L. and Wolf, E. (1995) *Optical Coherence and Quantum Optics*, Cambridge University Press, Cambridge.
- 75 Adams, C.S. and Riis, E. (1997) Laser cooling and trapping of neutral atoms. *Prog. Quantum Electron.*, **21**, 1.
- 76 Kurtsiefer, C., Pfau, T., and Mlynek, J. (1997) Measurement of the Wigner function of an ensemble of helium atoms. *Nature*, **386**, 150.
- 77 Abbott, B.P. et al. (2016) LIGO scientific Collaboration, and VIRGO Scientific Collaboration. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, **116**, 061102.
- 78 ESA (2016) <http://sci.esa.int/lisa/> (accessed 20 September 2016).
- 79 Rauch, H. (2000) *Neutron Interferometry. Lessons in Experimental Quantum Mechanics*, Clarendon Press, Gloucestershire.
- 80 Noeckel, J.U. (1998) Mikrolaser als photonen-billard: wie chaos ans licht kommt. *Phys. J.*, **54** (10), 927–930.
- 81 Klein, M.V. and Furtak, T.E. (1986) *Optics*, 2nd edn, John Wiley & Sons, Inc., Boston, MA.

- 82 Hariharan, P. (1987) *Optical Holography*, Cambridge University Press, Cambridge.
- 83 Hinsch, K. (1992) Lasergranulation. *Phys. unserer Zeit*, **23** (2), 59–66.
- 84 Arecchi, F.T. (1965) Measurement of the statistical distribution of Gaussian and laser sources. *Phys. Rev. Lett.*, **15**, 912–916.
- 85 Loudon, R. (1983) *The Quantum Theory of Light*, Clarendon Press, Oxford.
- 86 Paul, H. (1995) *Photonen. Eine Einführung in die Quantenoptik*, B. G. Teubner, Stuttgart.
- 87 Panofsky, W.K.H. and Phillips, M. (1978) *Classical Electricity and Magnetism*, Addison Wesley, Reading, MA.
- 88 Weisskopf, V. and Wigner, E. (1930) Berechnung der natürlichen linienbreite auf grund der diracschen lichttheorie. *Z. Phys.*, **63**, 54.
- 89 Feynman, R.P., Leighton, R.B., and Sands, M. (1964) *The Feynman Lectures on Physics*, vols I–III, Addison-Wesley, Reading, MA.
- 90 Happer, W. (1972) Optical pumping. *Rev. Mod. Phys.*, **44** (2), 169–249.
- 91 Corney, A. (1988) *Atomic and Laser Spectroscopy*, Clarendon Press, Oxford.
- 92 Wineland, D., Itano, W.M., and Bergquist, J.C. (1987) Absorption spectroscopy at the limit: detection of a single atom. *Opt. Lett.*, **12** (6), 389–391.
- 93 Sobelman, I.I. (1977) *Atomic Spectra and Radiative Transitions*, Springer-Verlag, Berlin Heidelberg, New York.
- 94 Cohen-Tannoudji, C., Diu, B., and Laloe, F. (1977) *Quantum Mechanics*, vols I, II, John Wiley & Sons, Inc., New York.
- 95 Woodgate, K. (1980) *Elementary Atomic Structure*, Clarendon Press, Oxford.
- 96 Sargent, M. III, Scully, M.O., and Lamb, W.E. (1974) *Laser Physics*, Addison Wesley, Reading, MA.
- 97 Allen, L. and Eberly, J.H. (1975) *Optical Resonance and Two-Level Atoms*, Dover Publications, New York, reprinted 1987 edition.
- 98 NASA (2016) <https://science.nasa.gov/missions/cobe>.
- 99 Börner, G. (2005) Die dunkle energie: Kosmologie. *Phys. unserer Zeit*, **36**, 168.
- 100 Ladenburg, R. (1933) Dispersion in electrically excited gases. *Rev. Mod. Phys.*, **5** (4), 243–256.
- 101 Ladenburg, R. and Kopfermann, H. (1930) Untersuchungen über die anomale dispersion angeregter gase. v. teil. Negative dispersion in angeregtem neon. *Z. Phys.*, **65**, 167.
- 102 Kramers, H.A. and Heisenberg, W. (1925) Über die streuung von strahlung durch atome. *Z. Phys.*, **31**, 681.
- 103 Schawlow, A. and Townes, C.H. (1958) Infrared and optical masers. *Phys. Rev.*, **112**, 1940.
- 104 Bertolotti, M. (1983) *Masers and Lasers. An Historical Approach*, Adam Hilger, Bristol.
- 105 Townes, C.H. (1999) *How the Laser Happened*, Oxford University Press, New York.
- 106 Einstein, A. (1917) Zur quantentheorie der strahlung. *Phys. Z.*, **18**, 121.
- 107 Gordon, J.P., Zeiger, H.J., and Townes, C.H. (1955) The maser – new type of microwave amplifier, frequency standard, and spectrometer. *Phys. Rev.*, **99**, 1264.

- 108 Basov, N. and Prokhorov, A.M. (1954) 3-level gas oscillator. *Zh. Eksp. Teor. Fiz.*, **27**, 431.
- 109 Maiman, T.H. (1960) Stimulated optical radiation in ruby masers. *Nature*, **187**, 493.
- 110 Javan, A., Bennett, W.R., and Herriott, D.R. (1961) Population inversion and continuous optical maser oscillation in a gas discharge containing a He-Ne mixture. *Phys. Rev. Lett.*, **6**, 106.
- 111 White, A.D. and Rigden, J.D. (1962) Continuous gas maser operation in the visible. *Proc. IRE*, **50**, 1697.
- 112 Hecht, J. (2007) The breakthrough birth of the diode laser. *Opt. Photon. News*, **18** (7), 38.
- 113 Hanle, W. (1924) Über magnetische beeinflussung der polarisation der resonanzfluoreszenz. *Z. Phys.*, **30**, 93.
- 114 Moran, J. (1982) Maser action in nature, in *CRC Handbook of Laser Science and Technology*, vol. **1** (ed. M.J. Weber), CRC Press, Boca Raton, FL, pp. 483–506.
- 115 Schilke, P., Mehringer, D.M., and Menten, K. (2000) A Submillimetre HCN Laser in IRC 110216. *Astrophys. J.*, **528**, L37–L40.
- 116 Gubin, M.A. and Protsenko, E.D. (1997) Laser frequency standards based on saturated-dispersion lines of methane. *Quantum Electron.*, **27** (12), 1048–1062.
- 117 Strong, C.L. (1974) The amateur scientist: an unusual kind of gas laser that puts out pulses in the ultraviolet. *Sci. Am.*, **230** (6), 122–127.
- 118 Wöste, O. and Kühn, C. (1993) Wie baue ICH meinen laser selbst? *MNU-Z.*, **46/8**, 471–473.
- 119 Houldcroft, P. (1991) *Lasers in Materials Process*, Pergamon, Oxford.
- 120 Bass, M., van Stryland, E., Williams, D.R., and Wolfe, W.L. (eds) (1995) in *Handbook of Optics*, vols **I**, **II**, McGraw-Hill, Inc., New York.
- 121 Andreea, T. (1989) Entwicklung eines kontinuierlichen er:yalo lasers, Diplomarbeit, Sektion Physik der Ludwig-Maximilians-Universität München, München.
- 122 Henderson, B. and Imbusch, G.F. (1989) *Optical Spectroscopy of Inorganic Solids*, Monographs on the Physics and Chemistry of Materials, Oxford University Press, Oxford.
- 123 Kaminskii, A.A. (1990) *Laser Crystals*, Springer Series in Optical Sciences, vol. **14**, Springer-Verlag, Berlin Heidelberg, New York.
- 124 Spectra Physics (1995) *Spectra Physics. Millenia-laser. Technical Description*, Spectra Physics, Inc.
- 125 Baer, T. (1986) Large-amplitude fluctuations due to longitudinal mode coupling in diode-pumped intracavity-doubled nd:yag lasers. *J. Opt. Soc. Am. B*, **3**, 1175.
- 126 Kane, T.J. and Byer, R.L. (1985) Monolithic, unidirectional single-mode ND:YAG ring laser. *Opt. Lett.*, **10** (2), 65–67.
- 127 Freitag, I., Rottengatter, P., Tünnermann, A., and Schmidt, H. (1993) Frequenzabstimmbare, diodengepumpte miniatur-ringlaser. *Laser Optoelektron.*, **25** (5), 70–75.

- 128 Desurvire, E. (1994) *Erbium Doped Fiber Amplifiers*, John Wiley & Sons, Inc., New York.
- 129 Snitzer, E. (1961) Optical maser action of  $\text{nd}^{+3}$  in a barium crown glass. *Phys. Rev. Lett.*, **7**, 444.
- 130 Digonnet, M.J.F. (2001) *Rare Earth Doped Fiber Lasers and Amplifiers (Optical Engineering)*, Marcel Dekker, New York.
- 131 Zellmer, H., Tünnermann, A., and Welling, H. (1997) Faser-laser. *Laser Optoelektron.*, **29** (4), 53–59.
- 132 Yeh, C. (1990) *Applied Photonics*, Academic Press, San Diego, CA.
- 133 Laming, R.I. and Loh, W.H. (1996) *Fibre Bragg Gratings; Application to Lasers and Amplifiers*, OSA Tops on Optical Amplification and Their Applications, vol. 5, Optical Society of America, Washington, DC.
- 134 Moulton, P.F. (1986) Spectroscopic and laser characteristics of  $\text{ti:al}_2\text{o}_3$ . *J. Opt. Soc. Am.*, **B3**, 125–133.
- 135 Gardiner, C. (1991) *Quantum Noise*, Springer-Verlag, Berlin Heidelberg, New York.
- 136 Walls, D.F. and Milburn, G.J. (1994) *Quantum Optics*, Springer-Verlag, Berlin Heidelberg, New York.
- 137 Yariv, A. (1991) *Quantum Electronics*, 4th aufl. edn, Holt Rinehart and Winston, Fort Worth, TX.
- 138 Haken, H. (1983) *Laser Theory*, Springer-Verlag, Berlin Heidelberg, New York.
- 139 Lax, M. and Louisell, W.H. (1967) Quantum Fokker-Planck solution for laser noise. *IEEE J. Quantum Electron.*, **QE3**, 47.
- 140 Louisell, W.H. (1973) *Quantum Statistical Properties of Radiation*, John Wiley & Sons, Inc., New York.
- 141 Yamamoto, Y. (ed.) (1991) *Coherence, Amplification and Quantum Effects in Semiconductor Lasers*, John Wiley & Sons, Inc., New York.
- 142 Bloch, I. (1996) Stimulierte lichtkr fte mit pikosekunden-laserpulsen. Diplomarbeit, Mathematisch-naturwissenschaftliche Fakultät der Universität Bonn, Bonn.
- 143 Spence, D.E., Kean, P.N., and Sibbett, W. (1991) 60-fsec pulse generation from a self-mode-locked Ti:sapphire laser. *Opt. Lett.*, **16** (1), 42.
- 144 Szipöcs, R., Ferencz, K., Spielmann, Ch., and Krausz, F. (1994) Sub-10-femtosecond mirror-dispersion controlled Ti:sapphire laser. *Opt. Lett.*, **20**, 602.
- 145 Strickland, D. and Mourou, G. (1985) Compression of amplified chirped optical pulses. *Opt. Commun.*, **56**, 219–221.
- 146 Wadsworth, W.J., Ortigosa-Blanch, A., Knight, J.C., Birks, T.A., Man, T.-P.M., and Russell, P.St.J. (2002) Supercontinuum generation in photonic crystal fibers and optical fiber tapers: a novel light source. *J. Opt. Soc. Am.*, **B19**, 2148–2155.
- 147 Reeves, W.H., Skryabin, D.V., Biancalana, F., Knight, J.C., Russell, P.St.J., Omenetto, F.G., Efimov, A., and Taylor, A.J. (2003) Transformation and control of ultrashort pulses in dispersion-engineered photonic crystal fibres. *Nature*, **424**, 511–515.
- 148 Russell, P. (2003) Photonic crystal fibers. *Science*, **299**, 358–362.

- 149 Ranka, J.K., Windeler, R.S., and Stentz, A.J. (2000) Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm. *Opt. Lett.*, **25**, 25.
- 150 Eckstein, J.N., Ferguson, A.I., and Hänsch, T.W. (1978) High-resolution 2-photon spectroscopy with picosecond light-pulses. *Phys. Rev. Lett.*, **40**, 847.
- 151 Diddams, A.A. (2010) The evolving optical frequency comb. *J. Opt. Soc. Am. B*, **27**, B51.
- 152 von Neumann, J. (1987) Notes on the photon-disequilibrium-amplification scheme (JVN), September 16, 1953. *IEEE J. Quantum Electron.*, **QE-23**, 659.
- 153 Nakamura, S. and Fasol, G. (1998) *The Blue Laser Diode*, Springer-Verlag, Berlin Heidelberg, New York.
- 154 Willardson, R. and Sugawara, M. (eds) (1999) *Self-Assembled InGaAs/GaAs Quantum Dots (Semiconductors & Semimetals)*, Academic Press, San Diego, CA.
- 155 Grundmann, M., Heinrichsdorff, F., Ledentsov, N.N., and Bimberg, D. (1998) Neuartige halbleiterlaser auf der basis von quantenpunkten. *Laser Optoelektron.*, **30** (3), 70–77.
- 156 Kazarinov, R.F. et al. (1971) Possibility of amplification of electromagnetic waves in a semiconductor with a superlattice. *Fiz. Tekh. Poluprovod.*, **5** (4), 797.
- 157 Faist, J., Capasso, F., Sivco, D.L., Sirtori, C., Hutchinson, A.L., and Cho, A.Y. (1994) Quantum cascade laser. *Science*, **264**, 553.
- 158 Faist, J. (2013) *Quantum Cascade Lasers*, Oxford University Press, Oxford.
- 159 Henry, C. (1982) Theory of the linewidth of semiconductor lasers. *IEEE J. Quantum Electron.*, **QE-18**, 259.
- 160 Dahmani, B., Hollberg, L., and Drullinger, R. (1987) Frequency stabilization of semiconductor lasers by resonant optical feedback. *Opt. Lett.*, **12**, 876.
- 161 Svelto, O. (1998) *Principles of Lasers*, Plenum Press, New York and London.
- 162 Chang-Hasnain, J. (1998) VCSELs: advances and future prospects. *Opt. Photon. News*, **5**, 34–39.
- 163 Jung, C., Jäger, R., Grabherr, M., Schnitzer, P., Michalzik, R., Weigl, B., Müller, S., and Ebeling, K.J. (1997) 4.8mW single mode oxide confined top-surface emitting vertical-cavity laser diodes. *Electron. Lett.*, **33**, 1790–1791.
- 164 Boyle, W.S. and Smith, G.E. (1970) Charge coupled semiconductor devices. *Bell Syst. Tech. J.*, **49**, 587–593.
- 165 Demtröder, W. (2003) *Laser Spectroscopy: Basic Concepts and Instrumentation*, 3rd edn, Springer-Verlag, Berlin Heidelberg, New York.
- 166 Meschede, D. (1992) Radiating atoms in confined space. from spontaneous to micromasers. *Phys. Rep.*, **211** (5), 201–250.
- 167 Dehmelt, H. (1981) Stored-ion spectroscopy, in *Advances in Laser Spectroscopy* (eds F.T. Arecchi, F. Strumia, and H. Walther), Plenum Press, New York and London, pp. 153–188.
- 168 Metcalf, H. and van der Straten, P. (1999) *Laser Cooling*, Springer-Verlag, Berlin Heidelberg, New York.
- 169 Letokhov, V.S. and Chebotayev, V.P. (1977) *Nonlinear Laser Spectroscopy*, Springer Ser. Opt. Sci., vol. 4, Springer-Verlag, Berlin Heidelberg, New York.

- 170 Schmidt, O., Knaak, K.-M., Wynands, R., and Meschede, D. (1994) Cesium saturation spectroscopy revisited: how to reverse peaks and observe narrow resonances. *Appl. Phys. B*, **59**, 167–178.
- 171 Lewis, G. (1926) The conservation of photons. *Nature*, **118**, 874.
- 172 Shen, Y.R. (1984) *The Principles of Nonlinear Optics*, John Wiley & Sons, Inc., New York.
- 173 Hänsch, T.W. and Walther, H. (1999) Laser spectroscopy and quantum optics. *Rev. Mod. Phys.*, **71**, 242.
- 174 Niering, M. et al. (2000) Measurement of the hydrogen 1S-2S transition frequency by phase coherent comparison with a microwave caesium fountain clock. *Phys. Rev. Lett.*, **84**, 5496.
- 175 Frisch, R. (1933) Experimenteller nachweis des einsteinschen strahlungsrückstosses. *Z. Phys.*, **86**, 42–48.
- 176 Lison, F., Schuh, P., Haubrich, D., and Meschede, D. (2000) High brilliance Zeeman-slowed cesium atomic beam. *Phys. Rev. A*, **61** (1) 013405.
- 177 Grimm, R. and Weidemüller, M. (2000) Optical dipole traps for neutral atoms. *Adv. At. Mol. Opt. Phys.*, **42**, 95.
- 178 Jessen, P. and Deutsch, I.H. (1996) Optical lattices, in *Advances in Atomic, Molecular, and Optical Physics*, vol. 37, (eds B. Bederson and H. Walther), Academic Press, Cambridge, pp. 95–138.
- 179 Meschede, D. and Metcalf, H. (2003) Atomic nanofabrication: atomic deposition and lithography by laser and magnetic forces. *J. Phys. D: Appl. Phys.*, **36**, R17–R38.
- 180 Padgett, M. and Allen, L. (1997) Optical tweezers and spanners. *Phys. World*, **10** (9), 35–38.
- 181 Cohen-Tannoudji, C. (1977) Atoms in strong resonant fields, in *Frontiers in Laser Spectroscopy, Les Houches 1975*, vol. 1 (eds R. Balian, S. Haroche, and S. Liberman), North-Holland, pp. 3–103.
- 182 Baumert, T., Grosser, M., Thalweiser, R., and Gerber, G. (1991) Femtosecond time-resolved molecular multiphoton ionization: the na<sub>2</sub> system. *Phys. Rev. Lett.*, **67** (27), 3753–3756.
- 183 Gaeta, Z.D., Nauenberg, M.W., Noel, W., and Stroud, C.R. Jr. (1994) Excitation of the classical-limit state of an atom. *Phys. Rev. Lett.*, **73** (5), 636–639.
- 184 Sakurai, J.J. (2000) *Advanced Quantum Mechanics*, Addison & Wesley, Reading, MA.
- 185 Cohen-Tannoudji, C., Dupont-Roc, J., and Grynberg, G. (1997) *Photons and Atoms: Introduction to Quantum Electrodynamics*, John Wiley & Sons, Inc., New York.
- 186 Dirac, P.A.M. (1927) The quantum theory of the emission and absorption of radiation. *Proc. R. Soc. London, Ser. A*, **114**, 243.
- 187 Kleppner, D. (1981) Inhibited spontaneous emission. *Phys. Rev. Lett.*, **47**, 233.
- 188 Berman, P. (ed.) (1994) *Cavity Quantum Electrodynamics*, Academic Press, San Diego, CA.
- 189 Hulet, R., Hilfer, E., and Kleppner, D. (1985) Inhibited spontaneous emission of a Rydberg atom. *Phys. Rev. Lett.*, **55**, 2137.

- 190 Jhe, W., Anderson, A., Hinds, E., Meschede, D., Moi, L., and Haroche, S. (1987) Suppression of spontaneous decay at optical frequencies: test of vacuum-field anisotropy in confined space. *Phys. Rev. Lett.*, **58**, 666.
- 191 Hartig, W., Pasmussen, W., Schieder, H., and Walther, H. (1976) Study of the frequency distribution of the fluorescent light induced by monochromatic radiation. *Z. Phys. A*, **278**, 205–210.
- 192 Mollow, B.R. (1969) Power spectrum of light scattered by two-level systems. *Phys. Rev.*, **188**, 1969–1975.
- 193 Glauber, R. (1963) Coherent and incoherent states of the radiation field. *Phys. Rev. Lett.*, **10**, 84.
- 194 Bachor, H. and Ralph, T. (2004) *A Guide to Experiments in Quantum Optics*, 2nd edn, Wiley-VCH Verlag GmbH, Weinheim.
- 195 Hanbury-Brown, R. and Twiss, R.Q. (1956) Correlations between photons in two coherent light beams. *Nature*, **177**, 27–29.
- 196 Arecchi, F.T., Gatti, E., and Sona, A. (1966) Time distributions of photons from coherent and Gaussian sources. *Phys. Lett.*, **20**, 27.
- 197 Purcell, E.M. (1956) The question of correlation between photons in coherent light rays. *Nature*, **178**, 1449.
- 198 Gomer, V. and Meschede, D. (2001) A single trapped atom: light–matter interaction at the microscopic level. *Ann. Phys. (Leipzig)*, **10**, 9–18.
- 199 Kimble, H.J., Dagenais, M., and Mandel, L. (1977) Photon antibunching in resonance fluorescence. *Phys. Rev. Lett.*, **39**, 691.
- 200 Diedrich, F. and Walther, H. (1987) Non-classical radiation of a single stored ion. *Phys. Rev. Lett.*, **58**, 203.
- 201 Kurtsiefer, C., Mayer, S., Zarda, P., and Weinfurter, H. (2000) Stable solid-state source of single photons. *Phys. Rev. Lett.*, **85** (85), 290.
- 202 Santori, C., Pelton, M., Solomon, G., Dale, Y., and Yamamoto, Y. (2001) Triggered single photons from a quantum dot. *Phys. Rev. Lett.*, **86**, 1502.
- 203 Shih, Y. (1999) Two-photon entanglement and quantum reality. *Adv. At. Mol. Phys.*, **41**, 1–42.
- 204 Hong, C.K., Ou, Z.Y., and Mandel, L. (1987) Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.*, **59**, 1903.
- 205 Kwiat, P.G., Mattle, K., Weinfurter, H., Zeilinger, A., Sergienko, A.V., and Shih, Y. (1995) New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.*, **75**, 4337–4341.
- 206 Einstein, A., Podolski, B., and Rosen, N. (1935) Can quantum-mechanical description of reality be considered complete? *Phys. Rev.*, **47**, 777–780.
- 207 Bell, J.S. (1964) On the Einstein–Podolski–Rosen paradox. *Physics*, **1**, 195–200.
- 208 Aspect, A., Grangier, P., and Roger, G. (1981) Experimental test of realistic local theories via bell's theorem. *Phys. Rev. Lett.*, **47**, 460.
- 209 Bohm, D. and Aharonov, Y. (1957) Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolski. *Phys. Rev.*, **108**, 1070.
- 210 Bohm, D. (1952) A suggested interpretation of the quantum theory in terms of hidden variables. I. *Phys. Rev.*, **85**, 166.

- 211 Clauser, J.F., Horne, M.A., Shimony, A., and Holt, R.A. (1969) Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, **23**, 880.
- 212 Dehlinger, D. and Mitchell, M. (2002) Entangled photons, nonlocality, and bell inequalities in the undergraduate laboratory. *Am. J. Phys.*, **70** (9), 903–910.
- 213 Scully, M. and Zubairy, M. (1997) *Quantum Optics*, Cambridge University Press, Cambridge.
- 214 Trojek, P., Schmid, Ch., Bourennane, M., Weinfurter, H., and Kurtsiefer, C. (2004) Compact source of polarization-entangled photon pairs. *Opt. Exp.*, **12**, 276.
- 215 Boyd, R.W. (2002) *Nonlinear Optics*, 2nd edn, Harlekin, San Diego, CA.
- 216 Zernike, F. and Midwinter, J.E. (1973) *Applied Nonlinear Optics*, John Wiley & Sons, Inc., New York.
- 217 Nye, J.F. (1985) *The Physical Properties of Crystals*, Clarendon Press, Oxford.
- 218 Dmitriev, V.G., Gurzadyan, G.G., and Nikogosyan, D.N. (1991) *Handbook of Nonlinear Optical Crystals*, Springer-Verlag, Berlin Heidelberg, New York.
- 219 Byer, R.L. (1977) Parametric oscillators and nonlinear materials, in *Nonlinear Optics*, Proceedings of the 16th Scottish Universities Summer School in Physics 1975 (eds P.G. Harper and B.S. Wherrett), Academic Press, London, pp. 47–160.
- 220 Boyd, G.D. and Kleinman, D.A. (1968) Parametric interaction of focused Gaussian light beams. *J. Appl. Phys.*, **39** (8), 3597–3639.
- 221 Ashkin, A., Boyd, G.D., and Dziedzic, J.M. (1966) Resonant optical second harmonic generation and mixing. *IEEE J. Quantum Electron.*, **QE-2** (6), 109–124.
- 222 Armstrong, J.A., Boembergen, N., Ducuing, J., and Perhan, P.S. (1962) Interactions between light waves in a nonlinear dielectric. *Phys. Rev.*, **127**, 1918–1939.
- 223 Magel, G.A., Fejer, M.M., and Byer, R.L. (1990) Quasi-phase-matched second harmonic generation of blue light in periodically poled linbo<sub>3</sub>. *Appl. Phys. Lett.*, **56**, 108–110.
- 224 Fejer, M.M., Magel, G.A., Hundt, D.H., and Byer, R.L. (1992) Quasi-phase-matched second harmonic generation: tuning and tolerances. *IEEE J. Quantum Electron.*, **28**, 2631–2654.
- 225 Schiller, S. and Mlynek, J. (1998) Continuous-wave optical parametric oscillators: materials, devices, and applications. *Appl. Phys. B*, **25** (6) 66. Special Issue.
- 226 Van Stryland, E.W. and Sheik-Bahae, M. (1998) Z-scan measurements of optical nonlinearities, in *Characterization Techniques and Tabulations for Organic Nonlinear Materials*, Marcel Dekker, Inc., pp. 655–692.
- 227 Kivshar, Y.S. and Stegeman, G.I. (2002) Spatial optical solitons. guiding light for future technology. *Opt. Photon. News*, **13**, 59–63.
- 228 Trillo, S. and Torruellas, W.E. (2001) *Spatial Solitons*, Springer-Verlag, Berlin Heidelberg, New York.

## Index

### **a**

- Abbe number 9
- Abbe sine condition 175
- Abbe's refractometer 25
- ABCD matrices 18
  - conventions 18
  - Gaussian modes 44
  - wave guides 23
- aberrations 169
  - chromatic 176
  - spherical 173
- Abraham–Lorentz equation 220, 222
- absorption 380
  - cross section 237
  - in optical materials 95
  - saturated 381
- achromats 177
- acousto-optical modulator (AOM) 114, 306
- AC Stark shift 410
- AC Stark splitting 408
- adiabatic elimination 292
- Adler equation 343
- afocal system 22, 166
- air laser 264
- Airy disc 70
- Airy function 199
- alpha parameter 288
- amplifier noise 358
- amplitude noise 298
- angle matching 471
- anharmonic oscillator 457
- anisotropic optical materials 103
- anomalous dispersion 221

- AOM (acousto-optical modulator) 114, 306
- aperture aberration 173
- argon laser 261
- astigmatism 174
- atomic beams 388
- atom optics 25, 187
- Autler-Townes doublet 408
- Autler-Townes splitting 410
- autocorrelation function 430
- avalanche photodiodes 374

### **b**

- Babinet's principle 74
- bandgap, photonic 136
- beam splitter 184
- Bell experiment 453
- Bell's inequality 450
- Bennett holes 256
- birefringence 103
  - beam walk-off 106
  - microscopic model 104
  - polarizers 109
  - strain 104
  - uniaxialcrystals 104
- birefringent filter 108, 282
- Bloch–Siegert shift 233
- Bloch's theorem 135
- Bloch waves 134
- bolometer 366
- Bragg diffraction 115
  - pendulum solution 116
- Brewster angle 87
- Brewster condition 87

**c**

candela 377  
 cat's-eye 8  
 cavity dumping 306  
 cavity QED 426  
 cavity ring down spectroscopy 217  
 celeritas 34  
 channeltron 370  
 chaotic field 435  
 chaotic light field 441  
 charge-coupled device (CCD)  
     sensors 375  
 chemical vapor deposition  
     (CVD) 12  
 chirped pulse amplification 313  
 chromatic aberration 176  
 cladding pumping 276  
 Clauser–Horne–Shimony–Holt (CHSH)  
     inequality 452  
 Clausius–Mossotti equation 228  
 coherence 181, 182  
     first-order 435  
     function 430  
     longitudinal 191  
     second-order 435, 436  
     temporal 192  
     transverse 186  
 coherent control 407  
 coherent states 439  
 coherent superposition 414  
 CO<sub>2</sub> laser 265  
 colour centre laser 280  
 coma 175  
 concave mirror 16  
     astigmatic aberration 16  
     astigmatism 16  
 confocal parameter 42  
 conjugated plate 27  
 continuity equation 33  
 contracted notation 462  
 corner cube reflector 8  
 corner frequency 358  
 Cornu spiral 73  
 correlation function 183, 428  
 cosmic background radiation 240  
 Coulomb gauge 32  
 coupled amplitude equations 465

crystal field splitting 270  
 current noise 499  
 curved mirrors 15  
 CVD (chemical vapor deposition) 12

**d**

dark current noise 357  
 dark field imaging 160  
 dark state 455  
 defect mode 139  
 density-matrix formalism 504  
 density of states 505  
 dephasing 413  
 depolarizing field 228  
 depopulation pumping 391  
 depth of focus 161  
 detectors  
     photovoltaic 372  
     pyro-electric 366  
     quantum 353  
     thermal 353  
 dielectric function 225  
 dielectric interfaces 83  
 dielectric media 30  
     optically dense 227  
     optically thin 225, 226  
 dielectric susceptibility 30, 225  
 difference frequency 478  
 diffraction 63, 84  
     Bragg domain 115  
     grating 188  
     Raman–Nath domain 115  
 digital light processing (DLP) 114  
 digital mirror device (DMD) 113  
 diode laser  
     amplitude modulation 340  
     arrays 349  
     dynamics 339  
     extended cavity 346  
     external cavity 346  
     heterostructure 330  
     high-power 348  
     linewidth 341  
     optical feedback 344  
     phase modulation 341  
     quantum well 336  
     tapered amplifier 349

- tunable 345  
wavelength 332  
wide stripe 349
- dioptr $e$  20  
dipole character 39  
dipole interaction 230  
dipole operator 230, 422  
dispersion 9, 380  
length 100  
relation 127  
distortion (aberration) 175  
distributed Bragg reflector (DBR) laser 346  
distributed feedback (DFB) laser 346  
Doppler effect  
broadening 255  
second order 393  
Doppler width 385  
double slit  
atomic beams 187  
electron beam 187  
Drude model 89  
dye laser 281
- e**  
Einstein coefficients 241, 242  
Einstein–Podolsky–Rosen (EPR) paradox 448  
electric polarizability 222  
electromagnetic field  
energy density 35  
momentum current density 37  
quantization 418  
electro-optical modulators (EOMs)  
110  
half-wave voltage 111  
erbium-doped fiber amplifiers (EDFAs)  
275  
erbium laser 275  
etalon 197  
intra-cavity 256  
evanescent wave field 88  
excimer laser 267  
eye 149, 151  
adaption 151  
eyepiece 153
- f**  
Fabry–Perot interferometer 197  
resolution 202  
Fabry–Perot spectrum  
analyser 259  
Faraday effect 117  
Faraday isolator 118  
Faraday rotators 117  
femto-chemistry 113  
Fermat’s principle 5  
fiber absorption 130  
fiber Bragg gratings (FBGs)  
277  
fibre laser 276  
field operator 419  
filter, spatial 131  
finesse 201  
coefficient 199  
flicker noise 358  
fluorescence microscopy 161  
focal point 15  
Fock state 439  
focus 15  
Fourier components 32  
Fourier optics 70, 159  
four-level system 243  
four-wave mixing (FWM) 485  
Fraunhofer diffraction 67  
circular aperture 69  
Gaussian transmitter 69  
single slit 68  
free induction decay (FID) 411  
free spectral range 200  
frequency chirp 100, 495  
frequency combs 315  
frequency doubling 466  
Gaussian beams 472  
resonant 474  
strong conversion 468  
weak conversion 467  
frequency modulation 111  
frequency tripling 485  
Fresnel diffraction 67, 71  
circular aperture 73  
straight edge 72  
Fresnel formulae 86  
Fresnel lenses 75

- Fresnel zones 75  
frustrated total internal reflection  
(FTIR) 88  
full width at half-maximum (FWHM)  
382
- g**  
gain  
saturated 250, 256  
small signal 256  
GaN laser 323  
gas lasers 261  
Gaussian beams 40  
ABCD rules 44  
beam radius 43  
beam waist 42  
confocal parameter 42  
divergence 43  
Gouy phase 43  
higher modes 47  
Gaussian principal mode 41  
Gaussian rays 41  
Gauss–Voigt profile 384  
Glan polarizers 109  
Glauber state 439  
graded-index (GRIN) fiber 129  
grating  
blazed 189  
holographic 189  
laser 345  
resolution 190  
gravity wave interferometer 194  
GRIN lens 14, 23  
group index of refraction 98  
group velocity 98–99
- h**  
Hagen–Rubens relation 91  
half-width at half maximum (HWHM)  
382  
halo 26  
Hanbury-Brown and Twiss experiment  
437  
Hanle effect 224  
HE/EH modes 128  
Helmholtz equation 35  
paraxial 47
- Hermitian–Gaussian modes 47  
Hermitian polynomials 48  
Hertzian dipole 39  
heterodyne detection 364  
heterostructures 329  
hidden variables 450  
hole burning 282, 388  
holes, spectral 256  
hologram  
in-line 211  
reconstruction 212  
holography 210  
recording 211  
homodyne detection 364  
homostructures 329  
Hong–Ou–Mandel dip 448  
Hong–Ou–Mandel interferometer  
446  
host crystals 268  
hot spot 75  
Hubble Space Telescope 168  
Huygens eyepiece 153  
Huygens' principle 63  
hybrid modes 126  
hydrogen atom, spectroscopy 392
- i**  
image amplifiers 377  
image converter 377  
image sensors 374  
image, stigmatic 15  
indexellipsoid 105  
index of refraction 3  
extraordinary 104  
inhomogeneous 6  
intensity-dependent 101  
negative 144  
ordinary 104  
indicatrix 105  
injection locking 342  
interband, intersubband transitions  
338  
interferometry 181  
Hong–Ou–Mandel 446  
intrinsic permutation symmetry 461  
inversion 231, 235, 236, 243

**j**

Jaynes–Cummings model 297, 422  
 Jones formalism, matrices 54  
 Jones vectors 52

**k**

Kerr effect 110  
 Kerr lens mode locking (KLM) 310, 489  
 Kirchhoff's integral theorem 65  
 Kleinman symmetry 462  
 $\text{KNbO}_3$  472  
 Köhler illumination 156  
 Kramers–Kronig relations 247

**l**

Laguerre–Gaussian modes 78  
 lambda half/quarter plates 107  
 lambda meter 193  
 Lamb dip 256  
 Lamb shift 418  
 lanthanides 269  
 Larmor formula 221, 242  
 Larmor frequency 223  
 laser 249  
     amplitude noise 302  
     cooling 394, 398  
     disc 278  
     end-pumped 273  
     fluctuations 302  
     gain profile 255  
     granulation 214  
     gyro 197  
     helium–neon 251  
     high-power 312  
     line selection 254  
     linewidth 258  
     mode selection 254  
     neodymium 271  
     noise 298  
     outcoupling 258, 291  
     power 257  
     pulsed 305  
     rate equations 291  
     RIN 303  
     single-atom 296  
     single-frequency 256

single-mode 256  
 speckle 214  
 spectroscopy 379  
 spiking 292  
 theory 285  
 threshold-less 295  
 transition-metal ions 279  
 types 251  
 white light 313  
 laser diode, inversion 325  
 laser-induced fluorescence (LIF) 379  
 laser resonator  
     bowtie 282  
     Z-shaped 273  
 laser threshold 290  
 law of refraction 3  
 least distance of distinct vision 151  
 lenses 15  
     aberrations 169  
     achromatic 177  
     afocal systems 22  
     biconvex 171  
     designs 170  
     diffraction limit 46  
     GRIN 23  
     magnetic 24  
     matrix 20  
     meniscus 171  
     periodic systems 22  
     planar convex 170  
     systems 21  
     thick 20  
     thin 20  
     types 170  
 LIF (laser-induced fluorescence) 379  
 light field, nonclassical 437  
 light forces 394  
 light propagation in matter 83, 121  
 light pulses 94  
     distortion 100  
     spectrum 95  
 light rays  
     extraordinary 106  
     ordinary 106  
 light sensors 353  
 light shift 410  
 line shapes 382

- linewidth 382
- Doppler 383
- heterodyne method 259
- homogeneous 383
- inhomogeneous 385
- natural 383
- phasor model 300
- pressure broadening 385
- time-of-flight 386
- liquid crystal (LC) modulator 113
- lock-in amplifier 404
- longitudinal relaxation 236
- Lorentz field 228
- Lorentz oscillator, in magnetic field 223
- Lorentz profile 221
- LP modes 128
- lumen 377
- Lyot filter 108, 282
  
- m**
- Mach–Zehnder interferometer 195
- magnifying glass 152
- Manley–Rowe relation 466
- Markov processes 431
- maser 249
  - natural 249
- master oscillator power amplifier (MOPA) 349
- material dispersion parameter 99
- matrix, lens 20
- matrix optics 17
  - conventions 18
- matter waves 187
- Maxwell–Bloch equations 287
- Maxwell–Lorentz equations 32
- Maxwell's equations 33, 35
- metal–oxide–semiconductor (MOS) capacitors 375
- metal-vapour lasers 263
- metamaterials 143
- meter, definition 35
- $M^2$  factor 348
- Michelson interferometer 191
- microchannel plate 370
- micro-laser 295
  
- microscope 154
  - Abbe theory 156
  - cover glass 155, 178
  - resolving power 155
- microscopy
  - confocal 162
  - scanning near-field optical 162
- minimum uncertainty states 440
- mirage 7
- miser 274
- mode dispersion 130
- mode locking 306
  - KLM 310
- mode matching 203
- mode pulling 256, 289
- molecular gas lasers 264
- Mollow triplet 428, 433
- momentum, angular 59
- monochromator 190
- monolithic miniature laser 274
- MTF (modulation transfer function) 159
- Müller matrices 54
- multimode fibers 130
- multiple-beam interference 197
  
- n**
- neodymium
  - amplifier 271
  - quantum states 270
- neodymium laser 271
  - frequency-doubled 273
- Newton equation 27, 150
- nitrogen laser 264
- noise amplitude 499
- noise properties of measurable quantities 497
- nonlinear optical switch 491
- nonlinear optics, crystal symmetries 461
- nonlinear polarization 459
- nonlinear products 459
- nonlinear Schrödinger equation 103
- normal dispersion 221
- number state 439
- numerical aperture 155
- optical fiber 13

***o***

obliquity factor 65  
 Onsager–Lax theorem 431  
 open quantum systems 426  
 optical angular momentum (OAM) 60  
 optical axis 104  
 optical Bloch equations 232, 236  
 optical cavities 202  
   concentric 207  
   confocal 206  
   damping 202  
   micro- 207  
   modes 203  
   plane parallel 205  
   resonance frequencies 204  
   symmetric 205  
 optical contact 108  
 optical diode 118, 275  
 optical fibers 122  
 optical Fourier transformation 70, 71  
 optical gain 244  
 optical grating 188  
 optical images 149  
 optical isolators 118  
 optical Kerr effect 487  
 optical lithography 158  
 optical mixer 364  
 optical modulators 110  
 optical molasses 398  
 optical parametric oscillator (OPO)  
   479  
 optical prisms 9  
 optical pumping 224, 229  
 optical resonator, coupling 200  
 optical spectra 428  
 optical spectral analysis 259  
 optical tweezers 403  
 optical wave guides 122  
 oscillator, spectrum 430  
 oscillator strength 228  
 OTF (optical transfer function) 159

***p***

parabolic mirrors 16  
 parametric fluorescence 445  
 parametric gain 478  
 paraxial approximation 17

particle optics 23  
 periodically poled materials 476  
 phase conjugation 491  
 phase-contrast imaging 160  
 phase diffusion 300  
 phase matching 467, 469  
   90° 471  
   non-critical 471, 472  
   temperature 472  
   types I and II 470  
 phase modulation 111  
 phase noise 298  
 phase, random walk 300  
 phase-sensitive rectification 404  
 phase velocity 34, 98  
 phasor model 300  
 photoactivated localization microscopy  
   (PALM) 165  
 photo-capacitors 375  
 photodiodes 372  
   biased 374  
   operation modes 373  
   photovoltaic 373  
 photon 418  
 photon echo 413  
 photonic bandgap (PBG) 136  
 photonic crystal fiber (PCF) 141  
 photonic crystals 132  
 photonic materials 132  
 photon number distribution 441  
 photon recoil 394  
 photorefraction 278  
 pin diodes 373  
 $\pi$  pulse 234, 411  
 plasma frequency, metallic 89  
 pn junction 321  
 Pockels cell 306  
 Pockels effect 110  
 Poincaré sphere 53  
 point spread function (PSF) 155, 169  
 polarization 50  
   dielectric 30  
   entanglement 452  
   macroscopic 30  
   microscopic 236  
 polarization-contrast imaging 160  
 polarizers 109

- Porro prism 8  
 position sensors 374  
 power spectral density 499  
 Poynting's theorem 37  
 Poynting vector 37  
 precision measurements 393  
 principal plane 151  
     principal point 27  
 principle of superposition 181  
 prism, minimum deflection  
     angle 9  
 propagation constant 123  
 pseudo spin 231, 234  
 pseudo thermal light 441  
 PSF (point spread function) 159  
 pulse compressor 313  
 pulse distortion 98  
 pulse length–bandwidth product 96  
 pulse propagation 94  
 pulse shaping 95, 113  
 pulse stretcher 313
- q**
- QED (quantum electrodynamics) 219, 417  
 Q factor 286, 382  
 Q-switch 305  
 quadrant detectors 374  
 quadratic index media 129  
 quantities, opto-electronic 357  
 quantum beats 414  
 quantum cascade laser 338  
 quantum confinement 338  
 quantum dots 337  
 quantum efficiency 354  
 quantum electrodynamics (QED) 219, 417  
 quantum electronics 219  
 quantum films 334  
 quantum fluctuations 431  
 quantum optics 219  
 quantum regression theorem 431  
 quantum sensors 353  
 quantum state reduction 422  
 quantum well 336  
 quantum wires 337  
 quasi-phase matching 476
- r**
- Rabi frequency 233  
     single-photon 288  
 Rabi nutation 234  
 Rabi splitting 410  
 radiation formula 240  
 radiation pressure 58  
 radiation reaction 221  
 radiative interaction, semiclassical  
     theory 219  
 rainbow 25  
 Ramsey spectroscopy 404  
 rare-earth ions 269  
 Rayleigh criterion 155  
 Rayleigh scattering 431  
 Rayleigh zone 42  
 ray tracing 151  
 reciprocal lattice 134  
 reference star 169  
 reflection  
     coefficient 85  
     dielectric 84  
     metallic 91  
 reflectivity 86  
 refraction coefficient, nonlinear 487  
 refractive index 3  
     complex 88  
     in conducting materials 89  
     effective 127  
     macroscopic 225  
     table 10  
 refractive power 20  
 relative intensity noise (RIN) 303  
 relaxation  
     longitudinal 236  
     oscillations 292  
     transverse 236  
 repopulation pumping 391  
 reservoirs 426  
 resonance fluorescence 427  
 resonator field, damping 286  
 resonator, unstable 217  
 retarder plates 107  
     zero order 107  
 retro-reflector 8  
 right, left-handed materials 37  
 ring laser 281

- ruby laser 268  
 Rydberg constant 392
- s**
- Sagnac interferometer 196  
 saturable absorber 310  
 saturated gain 289  
 saturation intensity 237  
 saturation parameter 237  
 saturation spectroscopy 388, 390  
 scanning near-field optical microscopy (SNOM) 162  
 Schawlow–Townes linewidth 258, 304, 341  
 Schmidt mirror 174  
 Schottky formula 501  
 second harmonic generation (SHG) 466  
 Seidel aberrations 172  
 self-focusing 488  
 self-injection locking 344  
 self-phase-modulation 494  
 Sellmeier equation 9  
 semiconductors 319
  - absorption of light 323
  - doped 320
  - emission of light 323
  - optical properties 322
 sensors
  - optical 353
  - quantum efficiency 354
  - sensitivity 354
 shot noise 501  
 shot-noise-limited detection 363  
 sidebands 112  
 signal-to-noise ratio 355  
 single lens 24  
 single-mode fibers 131  
 single molecule detection 164  
 single-photon source 445  
 skin effect
  - anomalous 90
  - normal 90
 slab laser 273  
 SLMs (spatial light modulators) 113  
 slowly varying envelope approximation (SVEA) 101
- small signal gain 244  
 Snell's law 3  
 SNOM (scanning near-field optical microscopy) 162  
 solid immersion lens 157  
 solid-state lasers 268  
 solitons, optical 101  
 spatial filter 49, 50  
 spatial light modulators (SLMs) 113  
 spatial soliton 490  
 speckle pattern 214  
 spectroscopy, Doppler-free 388  
 spherical aberration 173  
 spin echo 404  
 spontaneous emission 239, 242, 301, 418  
 interpretation 426  
 rate 242  
 suppression 425  
 spot diagram 171  
 stability criterion 23  
 stability diagram 23  
 step-index fibers 123
  - HE/EH modes 128
  - LP Modes 128
  - TE/TM modes 127
 stimulated absorption 239, 241  
 stimulated emission 239, 241, 292  
 stimulated emission depletion (STED) microscopy 165  
 Stokes' factor 65  
 Stokes matrices 52  
 Stokes parameter 52  
 strained quantum well 337  
 strong coupling 236, 296, 407  
 subpixel resolution 376  
 sum frequency 477  
 superposition principle 33  
 superradiance 265  
 surface plasmons, surface plasmon polaritons (SPPs), polaritons 92  
 SVEA (slowly varying envelope approximation) 101
- t**
- Taylor experiment 188  
 tele lens 179

- telescope 166
  - Galilean 166
  - HST 168
  - magnification 167
  - reflector 168
  - resolving power 166
- temperature, negative 242
- TE/TM modes 127
- theory of relativity 34
- thermal detectors 353, 364
- thermal field 435
- thermal lens 488
- thermal light field 441
- thermistor 366
- thermopile 365
- third harmonic generation 485
- three-wave mixing 459
- Ti–sapphire laser 279
- total internal reflection 87
  - frustrated 88
- transient phenomena 410
- transmission 86
- transmission coefficient 85
- transverse electric and magnetic (TEM) mode 41
- transverse relaxation 236
- triple mirror 8
- two-level atoms 229, 230
- two-level laser model 243
- two-photon spectroscopy 391
- two-wave polarization 462
- Tyndall effect 40
  
- v**
- vacuum fluctuations 421
- vacuum Rabi frequency 423
- vacuum Rabi splitting 297
  
- w**
- walk-off 106
- wave equation 33
  - with conductivity 89
  - for step-index fibers 123
- wave front reversal 492
- wave guides 122
  - absorption 130
  - planar 122
  - polarization maintaining 131
  - single-mode 131
  - weakly guiding 123
- waves
  - evanescent 88
  - meter 193
  - packet 415
  - planar 37, 39
  - spherical 38
- Weisskopf–Wigner theory 423
- Wiener–Khinchin theorem 299, 430, 500
  
- y**
- Yb laser 278
- Young’s double-slit 181, 185
  
- z**
- zoom lens 179
- z-scan 488