

Optics for physicists

©M. Saffman
University of Wisconsin-Madison

March 7, 2019

Contents

Preface	2
1 Optical waves	3
1.1 Maxwell equations	4
1.2 Wave equation in linear media	4
1.2.1 Solutions of the scalar wave equation	5
1.3 Energy and momentum in optical fields	6
1.4 Refractive Index	8
1.5 Polarization	10
1.6 Polarizers and waveplates	14
1.7 Jones calculus	14
1.8 Stokes vectors and Mueller matrices	15
1.9 Angular momentum of light	16
1.10 Fresnel coefficients	17
1.11 The beamsplitter	20
1.12 Brewster's angle	22
1.13 Total internal reflection	23
1.14 Reflection from metals	26
1.15 Prisms	27
1.16 Calcite beam displacer	27
2 Coherence and Noise	29
2.1 Broadband fields	29
2.2 Temporal coherence	30
2.3 Frequency and Phase noise	31
2.4 Degree of temporal coherence	32
2.5 Spatial coherence	33
3 Interference and Interferometers	37
3.1 Interference of optical waves	38
3.1.1 Plane wave interference	38
3.1.2 Interference of plane and spherical waves	39
3.1.3 Interference of two spherical waves	40
3.1.4 Multibeam interference	41
3.2 Fabry-Pérot Interferometer	42
3.2.1 Transmitted intensity	44

3.2.2	Reflected intensity	45
3.2.3	Finesse	47
3.2.4	Quality factor	50
3.2.5	Phase shifts	50
3.2.6	Finite beam size effects	51
3.2.7	Sensitivity to length changes	51
3.2.8	Sensitivity to pressure	52
3.2.9	Determination of mode matching coefficient	52
3.3	Etalon	53
3.3.1	Tilted etalon	55
3.4	Other interferometers	56
3.4.1	Sagnac interferometer	57
3.5	Fourier transform spectroscopy	57
3.6	Multilayer coatings	59
3.6.1	Quarterwave anti-reflection coating	59
3.6.2	Quarterwave high-reflection coating	60
3.6.3	Multilayer coatings	61
3.6.4	Multilayer high reflectivity coating	63
4	Geometrical optics	67
4.1	Reflection and refraction	67
4.1.1	Paraxial Ray Matrices	67
4.1.2	Sign conventions	69
4.1.3	Ray matrix for a concave mirror	70
4.1.4	Imaging with a thin lens	71
4.1.5	Negative lens	72
4.1.6	Compound lens	73
4.1.7	Principal Planes	74
4.1.8	Lens Maker's formula	76
4.1.9	Optical plates	77
4.1.10	Brightness	79
4.2	Aberrations	79
4.2.1	Primary monochromatic aberrations	80
4.2.2	Chromatic aberration	82
4.3	Image forming instruments	84
4.3.1	The eye	84
4.3.2	Microscope	85
4.3.3	Telescope	85
4.3.4	Binoculars	87
4.4	Solid Angle	87
4.5	Eikonal equation	89
4.5.1	Ray equation	91

5 Fourier optics	93
5.1 Paraxial propagation and diffraction	94
5.1.1 Fresnel Diffraction	96
5.1.2 Fraunhofer Diffraction	98
5.1.3 Diffraction from an aperture	99
5.1.4 Beyond Fresnel diffraction	102
5.1.5 Vectorial diffraction theory	103
5.1.6 Fourier transforming properties of lenses	103
5.1.7 Fields with radial symmetry	104
5.1.8 More general transformations	105
5.1.9 Summary of diffraction formulae	106
5.2 Applications	107
5.2.1 Airy disk	107
5.2.2 Spot of Arago	108
5.2.3 Fresnel zones	109
5.2.4 Talbot effect	111
5.2.5 Bottle beam lenses	112
5.3 Fourier optical image processing	116
5.3.1 Repeated objects	116
5.3.2 Periodic apertures	117
5.3.3 Fourier filtering	118
5.3.4 Beam shaping	120
5.4 Diffraction gratings	124
6 Gaussian beams	126
6.1 Lowest order Gaussian mode	126
6.1.1 Higher order Hermite-Gaussian beams	129
6.1.2 Laguerre-Gaussian beams	131
6.1.3 Ince-Gaussian beams	132
6.2 Nondiffracting beams	132
6.3 Gaussian beam transformation by a lens	133
6.3.1 Gaussian beam focusing	135
6.3.2 Fiber collimation	137
6.3.3 Refocusing with a mirror	139
6.3.4 Zoom lenses for Gaussian beams	139
6.3.5 Two-lens zoom	139
6.3.6 Three-lens zoom	141
6.4 Apertured Gaussian Beams	143
6.4.1 Apertured beam in intermediate plane	145
7 Optical resonators	147
7.1 Resonator eigenmodes	147
7.2 Linear resonator	148
7.3 Linear cavity with internal optical element	150
7.4 Resonant frequencies	151

7.5	Mode matching	152
7.6	Ring cavity	154
7.6.1	Ring cavity with angled crystal	155
8	Light and Matter	158
8.1	Einstein's 1917 theory	158
8.1.1	Optical gain	160
8.2	Absorption area law	161
8.3	Resonant absorption due to narrowband radiation	162
8.3.1	Saturation effects	164
8.3.2	Scattering cross section and absorption	165
8.4	Refractive index	166
8.5	Summary	168
8.6	Microscopic model of the refractive index	169
8.6.1	Fields from accelerated charges	169
8.6.2	The radiated field from a collection of bound charges	173
9	Light Scattering	178
9.1	Scattering geometry	178
9.2	Rayleigh scattering	180
10	Dispersive and nonlinear pulse propagation	181
10.1	The wave equation	182
10.1.1	Linear propagation in inhomogeneous media	183
10.2	Propagation in linear dispersive media	184
10.3	Wave equation in "Kerr" media	187
10.3.1	Kerr effect in atomic media	188
10.4	Optical solitons	188
10.5	Amplitude equation for dispersive nonlinear pulse propagation	189
10.6	Filamentation instability	190
11	Quadratic nonlinearities	193
11.1	Parabolic equations describing propagation of parametrically coupled waves .	194
11.1.1	Second harmonic generation	194
11.1.2	Cascading	195
11.2	Conversion efficiency	196
11.3	Boyd-Kleinman theory	197
11.3.1	Boyd-Kleinman theory with walk-off	199
11.4	Quasi-Phase Matching	200
11.5	Intracavity conversion efficiency	201
11.6	Normalized equations	202
11.6.1	Constant pump solution	203
11.6.2	Effect of a phase shift	204
11.6.3	Optical parametric amplification	206
11.6.4	Combined harmonic generation and parametric amplification	206
11.6.5	Constant pump solution with parametric amplification	208

12 Optical waveguides and fiber optics	209
12.1 One-dimensional waveguide modes	209
12.2 Guided modes in a two-dimensional waveguide	209
13 Electro-optics	211
13.1 Photodetection and signal to noise ratio	211
13.1.1 Thermal noise	214
13.1.2 Relative intensity noise	215
13.1.3 Shot noise	215
13.1.4 Noise equivalent power	216
13.2 Detector performance	217
13.2.1 Photovoltaic detector	217
13.2.2 Photoconducting detector	219
13.2.3 Heterodyne detection	221
13.2.4 Photomultiplier	223
13.2.5 Noise in electron multiplying detectors	224
13.3 Optical Modulators	227
13.3.1 Piezoelectric device	227
13.3.2 Acousto-optic modulator	228
13.3.3 Time bandwidth product	230
13.3.4 Electro-optic modulator	230
13.4 Laser stabilization	233
13.4.1 Pound-Drever-Hall locking	233
13.4.2 Frequency response of PDH error signal	236
13.4.3 Tilt-Locking	241
13.4.4 Littrow laser scanning	242
13.4.5 Frequency comb stability	244
14 Electronics	245
14.1 Linear components	245
14.1.1 Low and high pass filters	246
14.1.2 Phase advance	247
14.2 Operational amplifiers	247
14.2.1 Inverting configuration	247
14.2.2 Electro-optical servo loops	249
14.3 Analysis of a noise eater	251
14.3.1 Measuring RIN and Power spectrum units	252
A Fourier analysis	253
A.1 Fourier transforms	253
A.2 Self-reciprocal functions	256
A.3 Fourier identities	258
A.3.1 Transforms on the full line	258
A.3.2 Transforms on the half line	259
A.3.3 Delta function	259
A.3.4 Definite integrals:	261

B Window deformation	263
C Books on optics	266

List of Figures

1.1	Wave propagation in right and left handed media.	10
1.2	Polarization ellipse with semimajor and semiminor axes a, b and rotation angle α	11
1.3	Reflection of light from a mirror. The coordinate system for the incident wave is $(\epsilon_x, \epsilon_y, \epsilon_z)$ and for the reflected wave $(\epsilon_{x'}, \epsilon_{y'}, \epsilon_{z'})$. To make both coordinate systems right handed we set $\epsilon_{x'} = \epsilon_x, \epsilon_{y'} = -\epsilon_y, \epsilon_{z'} = -\epsilon_z$	13
1.4	Fresnel coefficients r_s, r_p and t_s, t_p determine the amplitude of the reflected and transmitted waves.	18
1.5	Fresnel coefficients for three different values of $n = n_2/n_1$	19
1.6	The beamsplitter.	20
1.7	Perfect transmission of p-polarized light occurs at Brewster's angle.	23
1.8	(left) TIR at 45 deg. incidence angle with $n_1 > n_2$. (right) Frustrated TIR. The transmitted wave decreases exponentially with increasing gap width d	23
1.9	Real and imaginary parts of $\theta_2 = \sin^{-1}[n \sin(\theta_1)]$ for $n = 1.5$	24
1.10	Phase shift in TIR for $n = 1.50, 1.51, 1.52$	26
1.11	Some different prisms. Image from Wikipedia.	27
1.12	Calcite beam displacer. Image from Thorlabs.	28
2.1	Spectral filtering of an extended thermal source prepares an optical field with temporal coherence that is spatially incoherent.	34
2.2	Two-slit interferometer for characterizing spatial coherence.	35
3.1	Interference of two plane waves with $\lambda = 1, \theta = 11.48$ deg. and $\Lambda = 5$	38
3.2	Interference of a plane wave and a spherical wave with $\lambda = 1, L = 30$, and $\theta = 11.48$ deg. The right hand plot shows the intensity as a function of x for $y = 0$ and $\theta = 0$	40
3.3	Interference of two spherical waves displaced along z with $\lambda = 1, L = 60$, and $d = 10$. The right hand plot shows the intensity as a function of x for $y = 0$	41
3.4	Interference of multiple plane waves with angles uniformly distributed.	42
3.5	Fabry-Pérot cavity.	43
3.6	Transmission (solid line) and reflectivity (dashed line) of a Fabry-Pérot cavity with $R_{\text{in}} = .9$, $R_{\text{out}} = 0.8$, and $R_{\text{cav}} = 0.99$. The inset zooms in near the resonance.	46
3.7	Fabry-Pérot cavity finesse(left) and build up factor (right) as a function of round trip internal loss for input coupler transmission of $T_{\text{in}} = 0.05, 0.02, 0.01$ (solid, dashed, dotted lines) and $R_{\text{out}} = 0.999$	48

3.8	Transmission (solid line) and reflection (dashed line) phase shifts for $R_{\text{in}} = R_{\text{out}} = 0.95$, $R_{\text{cav}} = 0.99$. The finesse is $\mathcal{F} = 55.8$.	50
3.9	Etalon of thickness d and refractive index n .	53
3.10	Etalon at normal incidence (solid line) and for tilted rays (dashed lines).	55
3.11	Main types of interferometers.	56
3.12	Fourier transform spectroscopy with a Michelson interferometer. Moving one of the mirrors changes the differential path length d leading interference at the output port.	57
3.13	Fourier transform spectroscopy signal for an input spectrum of two Gaussians centered at ω_1 and ω_2 . The left two plots show the output with only one of the Gaussians present and the right hand plot shows the output when both input frequencies are present which leads to interference.	58
3.14	Reflection from a thin film of thickness d . The external medium has index n_0 , the film has index n_1 and the substrate has index n_2 .	59
3.15	Reflectance of quarter wave film with index n_1 in between air ($n_0 = 1$) and glass ($n_2 = 1.5$).	61
3.16	Wave propagation through a multilayer film.	61
3.17	High reflectivity coating made of alternating quarterwave layers of high and low index.	63
3.18	High reflectivity coating made of alternating quarterwave layers of high and low index with air on the left and a substrate with index n_s on the right.	64
3.19	High reflectivity coating without (blue dots) and with (yellow dots) inclusion of edge layers for $n_L = 1.38$, $n_H = 2.30$ and $n_{\text{air}} = 1$, and $n_s = 1.5$.	65
4.1	Reflection and transmission at an interface.	67
4.2	Fractional error in approximating $\sin(\theta)$ by θ .	68
4.3	Ray propagation through a thin lens.	69
4.4	Ray matrices in the tangential and sagittal planes. The relative refractive index is $n_r = n_2/n_1$. In the fourth and fifth lines $R > 0$ for the interface concave to the right. The matrices for transmission through the tilted interface can be found in [21].	70
4.5	Geometry for imaging with a concave mirror.	71
4.6	Imaging with a thin lens. The object and image distances are d_1 and d_2 , respectively. The distances measured from the principal foci are z_1 and z_2 .	72
4.7	A negative lens forms a virtual image to the left of the lens.	73
4.8	Compound lens composed of two thin lenses. The principal planes are labeled PP.	74
4.9	Principal plane analysis. A composite optical system (top) is equivalent to an effective thin lens with the propagation distances measured to and from the principal planes (bottom).	75
4.10	Thick lens.	77
4.11	A tilted slab of index n_2 and thickness t .	77
4.12	A tilted mirror with radius of curvature R , thickness t , and substrate index n_2 .	78
4.13	Seidel aberrations.	80
4.14	Spherical aberration of a plano-convex lens.	82

4.15 Refractive index of BK-7 (left) and fused silica (right) as a function of wavelength. Plots taken from Thorlabs catalog.	82
4.16 Optics of the human eye.	84
4.17 Microscope with objective lens and eyepiece.	85
4.18 Refracting telescope with objective lens of focal length f_o and eyepiece of focal length f_e	86
4.19 Binocular optics.	87
4.20 Calculation of solid angle of window or lens.	87
4.21 Fractional solid angle as a function of cone angle θ , $f_{\#}$, and NA . The dashed line is the cone angle as a function of the $f_{\#}$ or NA	88
4.22 Ray propagation in a medium with continuously varying index $n(\mathbf{r})$	90
4.23 Periodic ray trajectories in a gradient index fiber. The numerical solutions are for $n(x) = 1.2 \operatorname{sech}(x)$, and $\theta(0) = 0.1, 0.79$	91
5.1 Phase delay for off-axis ray propagation.	93
5.2 Huygens' construction for describing propagation of wavefronts.	94
5.3 Coordinate system for optical propagation.	95
5.4 Phase delay for off-axis ray propagation. The effective phase between points separated by z along the \hat{z} axis is larger for axial propagation with $\hat{k} = \hat{z}$ (left) than for off-axis propagation with $\hat{k} = \hat{\theta}$ (right).	95
5.5 Paraxial evolution can be solved by convolution with the impulse response in the spatial domain or multiplication by the transfer function in the Fourier domain.	97
5.6 Fresnel integrals $C(s), S(s)$ and Cornu spiral.	100
5.7 Diffracted intensity using Fresnel diffraction (blue solid line) and Fraunhofer diffraction (green dashed line) from a slit of width $L = 20. \mu\text{m}$, with $\lambda = 1. \mu\text{m}$ at different distances $z/z_{\text{Fresnel}} = 0.1, 0.5, 1, 5, 10$. The intensity has been normalized to the incident intensity at the aperture plane. The aperture parameters give $z_{\text{Fresnel}} = 54. \mu\text{m}$ and $z_{\text{Fraunhofer}} = 400. \mu\text{m}$	101
5.8 Diffracted intensity from Fresnel diffraction theory (blue solid line) and Rayleigh-Sommerfeld theory Eq. (5.17) (green dashed line) from a square aperture of width $L = 20. \mu\text{m}$, with $\lambda = 1. \mu\text{m}$ at different distances $z/z_{\text{Fresnel}} = 0.1, 0.5, 1$. The intensity is shown as a function of x_2 for $y_2 = 0$ and has been normalized to the incident intensity at the aperture plane. The aperture parameters give $z_{\text{Fresnel}} = 54. \mu\text{m}$	103
5.9 Field transformation with a lens.	104
5.10 Airy diffraction rings formed by uniform illumination of a lens. The circular lens parameters are diameter $d = 100 \mu\text{m}$, focal length $f = 1 \text{ cm}$, $\lambda = 1 \mu\text{m}$. The annulus has inner diameter $0.6d$ and outer diameter d . The intensity is normalized to the incident intensity at the lens.	108
5.11 Spot of Arago calculated for $\lambda = 1. \mu\text{m}$, $a = 10. \mu\text{m}$, giving $z_{\text{Fresnel}} = 54 \mu\text{m}$. The diffracted intensity normalized to the incident intensity is shown for $z/z_{\text{Fresnel}} = 1, 2, 10, 20$	109
5.12 Fresnel zone plate for lensless focusing.	110

5.13 Gaussian beam illuminating a lens with a π phase in the inner region of radius b	112
5.14 BoB profiles for $w_0 = 82.76 \mu\text{m}$, $w_3 = 3 \mu\text{m}$, $b = 69 \mu\text{m}$ and $s = 1$ (left). These parameters correspond to $\lambda = 0.78 \mu\text{m}$ and $f = 1 \text{ mm}$. The plotted intensity is normalized to the peak of the input Gaussian with waist w_0 . The inset shows that the on-axis intensity does not vanish for any b . On the right the transverse profiles are shown at axial displacements up to $50 \mu\text{m}$ from the focal plane.	115
5.15 Three dimensional intensity distribution about the focus. Same parameters as in Fig. 5.14.	115
5.16 Fourier filtering setup. The filter function $t(\rho_2)$ is positioned in the Fourier plane between the lenses.	118
5.17 Fourier filtering setup for beam shaping starting with a uniform intensity input.	120
5.18 Near Gaussian intensity profile I_2 obtained by Fourier filtering of the Airy pattern resulting from uniform illumination of a circular aperture.	121
5.19 Fourier filtering setup for creating an array of Airy-Gauss beams. Each aperture in the input plane has radius a as in Fig. 5.17.	122
5.20 On axis intensity as a function of axial coordinate z computed by Fresnel diffraction. The profiles are an Airy-Gauss beam (I_{AG}), a dark Airy-Gauss beam ($ 1 - E_{AG} ^2$), a Gaussian beam (I_G), and a dark Gaussian beam ($ 1 - E_G ^2$). Parameters for numerical calculations were $a = b = 1.0 \mu\text{m}$, $\lambda = 0.825 \mu\text{m}$, $f = 2 \mu\text{m}$ for the Airy-Gauss beam and $w_G = 0.974a$ for the Gaussian beam.	124
5.21 Interference of N plane waves emitted from apertures separated by h	124
 6.1 Gaussian beam of wavelength $\lambda = 1$, and waist $w_0 = 2\lambda$. The inset shows intensity contours.	128
6.2 Intensity of Hermite-Gaussian beams in the focal plane $z = 0$ with $\lambda = 0.5$, $w_0 = 1$ in regions of size 6×6	130
6.3 Intensity of Laguerre-Gaussian beams in the focal plane $z = 0$ with $\lambda = 0.5$, $w_0 = 1$ in regions of size 6×6	131
6.4 Intensity of $LG_{0,3} + 0.75LG_{0,0}$ for (left to right) $z = 0, L_R/2, L_R$ with $\lambda = 0.5$, $w_0 = 1$ in regions of size 6×6	132
6.5 Bessel beam intensity profiles for $n = 0, 1, 2, 3$ and $k_\rho = 2\pi$	133
6.6 Gaussian beam transformation by a lens of focal length f . The incident waist w_{10} is positioned a distance z_1 in front of the lens. The output waist w_{20} occurs a distance z_2 after the lens.	134
6.7 Output waist size (left) and output waist position (right) for transformation by a single lens.	135
6.8 Numerical aperture for a given focusing ratio at aperture parameter $p = 4$. The dashed line shows a large r approximation.	136
6.9 Output waist (left) and waist position (right) for collimation of the light emitted by a single mode fiber.	137

6.10 Width on a screen at distance z_s vs. lens position z_1 . The curves are labeled with $z_s = 1, 2, 5, 10$ m.	138
6.11 Beam width as z_1 is varied (left) and as z_2 is varied (right). The flat line in the left hand graph is an exact result for all z_1 . The flat line in the right hand graph is an approximate result for z_2 similar to f	138
6.12 Refocusing of light in a crystal with a curved mirror	139
6.13 Layout of a two lens zoom.	140
6.14 Layout of a three lens zoom. The three lens zoom is afocal so we include a fixed f_4 to transform the output. The nominal distances are $z_{10}, z_{20}, z_{30}, z_{40}$. The lenses move by amounts $\delta_1, \delta_2, \delta_3$ during zoom operation giving actual distances $z_1 = z_{10} + \delta_1$, $z_2 = z_{20} + \delta_2 - \delta_1$, $z_3 = z_{30} + \delta_3 - \delta_2$, and $z_4 = z_{40} - \delta_3$	141
6.15 Three lens zoom as in Fig. 6.14 with $f_1 = 5.$, $f_2 = -2.5$, $f_3 = 5.$, $f_4 = 9..$ The actual magnification is shown on the left and the lens displacements on the right.	142
6.16 Gaussian beam filtered with a circular aperture.	143
6.17 Intensity profiles in back focal plane for $w_1 = w_2 = 0.126$ mm, $\lambda = 0.5$ μm , $f = 100.$ mm and $D/w_1 = 1, 2, 3, 4, 5$ (left). The profile for $D/w_1 = 5$ is indistinguishable from that with $D = \infty$ shown in purple. The right hand plot shows the weak oscillating tails for $D/w_1 = 4, 4.25, 4.5.$	144
6.18 Airy ring structure of the intensity profiles in the back focal plane for $D/w_1 = 0.5, 1, 2$, and all other parameters the same as in Fig. 6.17. The red curves are for a Gaussian beam with no aperture, the blue curves are the apertured Gaussian and the green curves are from the Airy formula (5.27).	145
6.19 Intensity profiles for $D/w_1 = 4$, and all other parameters the same as in Fig. 6.17 for $z = (0, 0.1, 0.5, 1, 2)L_R$ with $L_R = \pi w_2^2/\lambda = 200.$ mm.	146
 7.1 Linear resonator with spherical mirrors.	148
7.2 Stability diagram for a linear resonator. The shaded regions are stable, and everywhere else is unstable. The points marked with dots are symmetric plane-plane ($g_1 = g_2 = 1, R_1 = R_2 = \infty$), confocal ($g_1 = g_2 = 0, L = R_1 = R_2$), and concentric ($g_1 = g_2 = -1, L = 2R_1 = 2R_2$). The figure also shows asymmetric hemispherical and concave-convex resonators.	149
7.3 Linear resonator containing an optical crystal of length L_c and refractive index n_c	150
7.4 Power overlap between a Gaussian beam and a Gaussian cavity mode. The line plots show the variation with s and ζ	152
7.5 Mode matching to a linear resonator. Without the input coupler (top) the focusing lens produces a waist w_0 at position z_0 . With the input coupler in place (bottom) the waist is transformed to w_2 at position z_2	153
7.6 Ring bowtie cavity. The fields are drawn for a harmonic generation configuration.	155
7.7 Fundamental mode waist for a bowtie cavity with $R = 4$ cm, $L_c = 3$ cm, $n_c = 2.3$, and $\lambda = 0.96$ μm	155

7.8	Ring cavity for an optically pumped laser with angled crystal. The geometrical round trip distance L_T is $L_T = L_2 + L_{d1} + L_{d2} + L_u$. The path length inside the crystal is L_c . The detail in the upper right shows the geometry of the Brewster cut crystal. The beam path in the lower right with positions 1 – 6 is used for generating a detailed optical layout. Position 0 is at the center of the crystal.	156
7.9	Calculated waists at the center of the Ti:Sa crystal (left) and between the flat mirrors (right). Parameters: $\lambda = 0.685 \mu\text{m}$, $L_c = 1 \text{ cm}$, $n_c = 1.77$, $R = 2.5 \text{ cm}$, $\theta = 19 \text{ deg.}$, $L_T = 15 \text{ cm}$, $h = 2.3 \text{ cm}$	157
7.10	Layout of the Ti:Sa cavity for the parameters of Fig. 7.9.	157
8.1	Two level atom and transitions due to Einstein A, B coefficients.	158
8.2	Broad band light spectrum interacting with a narrowband resonance.	161
8.3	Scattering of an electric field $E(\omega)$ from a two-level atom with excited state spontaneous decay rate γ	162
8.4	a) Change in refractive index from Eq. (8.20) for parameters corresponding to a low density gas of Cs atoms on the D2 transition: $N = 10^{19} \text{ m}^{-3}$, $\lambda = 0.852 \mu\text{m}$, $k = 2\pi/\lambda$, $\sigma_0 = \frac{3}{2\pi}\lambda^2$. The detuning is $\Delta = \omega - \omega_{21}$ and with the assumption that $ \Delta \ll \omega_{21}$ we put $\omega/\omega_{21} = 1$. b) Absorption coefficient from Eq. (8.16) for the same parameters as in a).	168
8.5	Geometry for calculation of field from moving source charge.	170
8.6	Radiation from an accelerated charge.	173
8.7	Plate with bound charges of density N and surface density $\eta = LN$. The charges are driven by the incident field \mathcal{E}_{in}	174
9.1	Scattering geometry.	178
9.2	Scattered intensity averaged over a circular aperture with half cone angle of 0.1 deg. for a water drop with $n = 1.33$ in air, $d = 5 \mu\text{m}$, and $\lambda = 0.63 \mu\text{m}$	180
10.1	Moving reference frame for describing pulse propagation.	184
10.2	Modulation instability gain for Kerr media from Eq. (10.34a) for $2sA_0^2 = 1$.(left) Numerical calculation (center). Observation in a photorefractive crystal from [89] (right).	191
11.1	Polarization P_i is induced in a medium with quadratic nonlinearity due to fields E_j, E_k . a) the polarization is positive. b) after inverting the fields and the crystal the polarization is negative. c) inverting a centrosymmetric crystal does not change the polarization.	193
11.2	Birefringent phase matching. Light at λ_1 is polarized along \hat{x} and light at λ_2 is polarized along \hat{y}	197
11.3	Integral appearing in Boyd-Kleinman theory describing harmonic generation as a function of the dimensionless parameters $\Delta k L_c$ and L_{R1}/L_c for $L_c = 1$	198
11.4	A quasi-phase matched crystal is periodically poled with period Λ which results in changing the sign of the $\chi^{(2)}$ coefficient every half period.	200

11.5 Intracavity SHG for different linear loss L and input coupler transmission T . Parameters are for a 2 cm LBO crystal at .821 μm and optimum focusing which gives $E_{\text{NL}} = 0.00015$.	202
12.1 Attenuation of fused silica fibers from E. F. Schubert, Light Emitting Diodes, Cambridge University Press.	210
13.1 Circuit for analyzing shot and thermal noise.	212
13.2 Shot noise RIN level.	216
13.3 Transimpedance photodetector circuit.	217
13.4 Noise model of transimpedance photodetector circuit. The components are: J_{sig} = detector signal current, J_d = detector dark current, C_d = detector capacitance, R_{sh} = detector shunt resistance, R_s = detector series resistance, J_{tn} = detector Johnson noise current, J_{fn} = feedback resistor Johnson noise current, J_{an} = op amp input current noise, V_{an} = op amp input voltage noise, R_a = amplifier input resistance, G = gain of ideal amplifier, R_f = feedback resistor.	218
13.5 Output noise voltage for a Thorlabs FDS100 detector with LF356(green), AD8675 (orange), OPA209(red) opamps.	219
13.6 Biasing circuit for a photoconductive detector followed by two gain stages.	219
13.7 A weak signal and strong local oscillator are combined for heterodyne detec- tion.	221
13.8 Signal amplification in a photomultiplier.	224
13.9 Examples of piezoelectric modulators. Three point control of a mirror holder (left), angular tilt of a mirror (center), a long bimorph actuator for mm scale motion (right).	227
13.10 Bragg diffraction in an acousto-optic deflector with acoustic wavelength Λ . A picture of a real device is shown on the right.	228
13.11 Double pass geometry for frequency shifting without angular deflection.	229
13.12 Relative intensity in EOM order m as a function of the maximum phase shift η .	232
13.13 Arrangement for Pound-Drever-Hall stabilization of a laser to a reference res- onator.	234
13.14 PDH error signal for $R_{\text{in}} = R_{\text{out}} = 0.95$, $R_{\text{cav}} = 1$. giving a cavity finesse of $F = 61$, $(\Omega/2\pi)/\nu_{\text{FSR}} = 0.1$. The curves are labeled with the local oscillator phase ϕ .	236
13.15 Response of PDH detector to frequency fluctuations.	239
13.16 Response of PDH detector normalized to $\bar{V} = 4\pi\eta_d \mathcal{E}_{\text{in}} ^2 J_0(\epsilon)J_1(\epsilon)$ for $\mathcal{F} =$ 500, 1000, 5000 as a function of modulation frequency. The vertical dashed lines show the corresponding values of $\omega_c/2$.	240
13.17 Error signal as a function of cavity tuning.	242
13.18 Geometrical layout of a Littrow laser cavity.	242
14.1 Low and high pass filters.	246
14.2 Amplitude and phase response of low and high pass filters.	246

14.3 Phase advance circuit. The response is shown for $R_1 = 3000$. Ohms, $R_2 = 300$. Ohms, $C = 5$ nF giving $\omega_0/2\pi = 9650$ Hz.	247
14.4 Inverting opamp. The calculated response is for $R_f = 10$ k Ω , $C_f = 1$ nF, $R_1 = 1$ k Ω	248
14.5 Inverting opamp with phase advance input. The response is shown for $R_f = 10$. k Ω , $C_f = 1$ nF, $R_{1a} = 1$. k Ω , $R_{2a} = 100$. Ω , $C_a = 1$ nF, giving $\omega_f/2\pi = 15.9$ kHz, $\omega_a/2\pi = 144.7$ kHz.	248
14.6 Generic closed loop feedback circuit.	249
14.7 a) Arrangement for locking a laser beam to an optical cavity using a dither lock. b) shows the H and G circuit components.	249
14.8 Noise eater response with closed loop transfer function (left) and GH (right). Parameters: $G_0 = -10.$, $R_1 = 1$. k Ω , $R_f = 100$. k Ω , $C_f = 0.2$ nF, giving $\omega_0/2\pi = 8$. kHz and $\omega_G/2\pi = 100$ kHz.	250
14.9 Gain and phase margins.	250
14.10 Noise eater circuit.	251
14.11 Closed loop response of fast noise eater circuit. The actuator is modeled as a low pass filter with $\omega_G/2\pi = 50$ kHz and dc gain $G_0 = -1$. The logarithmic frequency scale is referenced to $\omega_0 = 2\pi \times 1$ kHz.	251
A.1 Fourier series representation of a square wave for 1, 2, 3, 4 spectral components.	254
A.2 Gaussian function $f(t)$ and the Fourier transform $g(\omega)$ for $\sigma = 1/2, 1, 2$	256
B.1 Maximal deformation of a pyrex vacuum window.	264
B.2 Minimum window thickness such that the peak intensity of a focused Gaussian beam is at least half of the unaberrated value.	265

Preface

Optics is an old subject with origins in the earliest of human scientific explorations. Many books have been written on the topic (a sampling is provided in Appendix C) and the reader may justifiably wonder if there is a need for yet another one. Despite the range of available texts, as a teacher of optics to undergraduate and graduate physics students at the University of Wisconsin-Madison for several decades, I have lacked a single text that covered the topics I felt were important for a physicist interested in optics. This book is an attempt to fill that lack with the goal of being adequately comprehensive yet concise, and detailed yet accessible.

Roughly the first half of the book deals with linear optics and covers the basics of optical waves, interferometry, geometrical optics, Fourier optics, Gaussian beams, and optical resonator design. The material includes descriptions of all the basic optical components used in research. The second half of the book concerns topics related to light matter interactions including scattering and absorption, dispersive and nonlinear pulse propagation, nonlinear materials, frequency conversion, and instabilities. The book concludes with a survey of electro-optical devices for detection, modulation, and stabilization of optical fields.

Chapter 1

Optical waves

Electromagnetic waves cover a wide range of frequencies and wavelengths:

description	frequency	wavelength
am radio waves	1 MHz	300 m
fm radio waves	100 MHz	3 m
microwaves, cell phones	1-100 GHz	30 - 0.3 cm
mm waves	300 GHz	1 mm
infrared light	300 THz	1 μ m
visible light	375-750 THz	0.8-0.4 μ m
x-rays	3×10^{17} Hz	1 nm

All of these different regimes have a unified description as electromagnetic wave solutions of the Maxwell equations. The wavelength range that constitutes optical waves is not well defined. The word light is most often used to refer to the range from about 0.1 - 10 μ m with visible light a subset of this range. Although the physics of wave propagation in vacuum is independent of the wavelength, phenomena associated with the interaction between electromagnetic waves and matter vary greatly in different wavelength regions. As a consequence the devices and technology used to generate, modulate, and detect electromagnetic waves are very different for different wavelengths. The technology is highly developed for some spectral regions and less so for others. For example sub mm waves with frequencies ranging from a few hundred GHz to a few THz are relatively difficult to work with.

One of the primary applications of optical waves is imaging. We do this everyday with visible light but imaging is also routinely performed with x-rays (medical diagnostics), infrared light (thermal imaging), and cm scale waves. The wavelength λ of the radiation used for imaging governs the resolution that can be achieved, although there are many techniques for surpassing the nominal limit of λ .

1.1 Maxwell equations

In SI units the Maxwell equations are

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (1.1a)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (1.1b)$$

$$\nabla \cdot \mathbf{D} = \rho, \quad (1.1c)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (1.1d)$$

\mathbf{E} is the electric field, \mathbf{B} is the magnetic field, \mathbf{D} is the electric displacement, and \mathbf{H} is the magnetic field strength.

These are supplemented by the constitutive relations

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad (1.2a)$$

$$\mathbf{H} = \frac{1}{\mu_0} \mathbf{B} - \mathbf{M}. \quad (1.2b)$$

In general the electric and magnetic polarization of the medium \mathbf{P} and \mathbf{M} are nonlinear and anisotropic functions of the fields. In the simplest case of linear and isotropic media the constitutive relations can be written

$$\mathbf{D} = \epsilon \mathbf{E}, \quad \mathbf{H} = \mathbf{B}/\mu.$$

In vacuum there is no medium to be polarized so $\mathbf{P} = 0$, $\mathbf{M} = 0$ and $\epsilon = \epsilon_0$, $\mu = \mu_0$.

1.2 Wave equation in linear media

Assuming linear and isotropic media we have

$$\nabla \times \nabla \times \mathbf{E} = -\frac{\partial}{\partial t} \nabla \times \mathbf{B} = -\frac{\partial}{\partial t} \nabla \times \mu \mathbf{H} = -\frac{\partial}{\partial t} (\nabla \mu \times \mathbf{H} + \mu \nabla \times \mathbf{H}).$$

The left hand side is

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = \nabla(\nabla \cdot \mathbf{D}/\epsilon) - \nabla^2 \mathbf{E} = \nabla\left(\frac{1}{\epsilon}\nabla \cdot \mathbf{D} + \nabla\left(\frac{1}{\epsilon}\right) \cdot \mathbf{D}\right) - \nabla^2 \mathbf{E}$$

If ϵ and μ are uniform in space then $\nabla(1/\epsilon) = 0$ and $\nabla \mu = 0$. In addition if there are no free charges and no currents then $\nabla \cdot \mathbf{D} = 0$ and $\mathbf{J} = 0$ and we get

$$-\nabla^2 \mathbf{E} = -\mu \frac{\partial}{\partial t} \nabla \times \mathbf{H} = -\epsilon \mu \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

or

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{1}{\epsilon \mu} \nabla^2 \mathbf{E} = 0.$$

The wave speed is $v = 1/\sqrt{\epsilon\mu}$ and the wave equation is

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - v^2 \nabla^2 \mathbf{E} = 0. \quad (1.3)$$

In vacuum $\epsilon\mu = \epsilon_0\mu_0$ and the wave speed is $1/\sqrt{\epsilon_0\mu_0} \equiv c$ the speed of light. It is convenient to introduce the refractive index $n = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}}$ and write $v = c/n = 1/\sqrt{\epsilon\mu}$. The refractive index has the value 1 in vacuum and can be greater than or less than 1 inside media. Remarkably it is also possible to have negative n which leads to some unusual and very interesting effects.

1.2.1 Solutions of the scalar wave equation

Equation (1.3) is a vector equation for the three spatial components of \mathbf{E} . Each component satisfies a scalar equation

$$\frac{\partial^2 E}{\partial t^2} - v^2 \nabla^2 E = 0. \quad (1.4)$$

In the approximation that E only depends on t and one spatial coordinate, say z , we have

$$\frac{\partial^2 E}{\partial t^2} - v^2 \frac{\partial^2 E}{\partial z^2} = 0.$$

The general solution of this equation is

$$E = c_1 f_1(kz - \omega t) + c_2 f_2(kz + \omega t) \quad (1.5)$$

where c_1, c_2 are constants, $k = 2\pi/\lambda$ is the wavenumber, λ is the wavelength, $\omega = 2\pi\nu$ is the angular frequency, ν is the frequency, and f_1, f_2 are arbitrary functions. This ansatz satisfies the wave equation provided $\omega = \pm vk$. Taking ω to be positive the first term proportional to f_1 represents a wave propagating towards positive z and the second term proportional to f_2 represents a wave propagating towards negative z . The solution (1.5) is a plane wave solution since the amplitude only depends on z and is constant in planes perpendicular to the direction of propagation. As we show in Sec. 1.4 plane wave solutions of the Maxwell equations are polarized perpendicular to the propagation direction.

Other types of scalar solutions are also possible. Consider a point source of waves that radiates uniformly in all directions. The field amplitude will then only depend on the distance to the origin $r = |\mathbf{r}|$. In spherical polar coordinates the radial part of the Laplacian is $\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \frac{\partial E}{\partial r})$ and the wave equation takes the form

$$\frac{\partial^2 E}{\partial t^2} - v^2 \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial E}{\partial r} \right) = 0. \quad (1.6)$$

The wave energy scales as E^2 and the surface area of a sphere grows as r^2 we thus anticipate that a solution which conserves energy is of the form $E = f(kr - \omega t)/r$ with f an unknown function. Plugging into (1.6) we find

$$\omega^2 \frac{f''}{r} - k^2 v^2 \frac{f''}{r} = 0.$$

We again find a solution for arbitrary f which now takes the form

$$E = \frac{f(kr - \omega t)}{r}.$$

This is a spherical wave with amplitude decreasing as $1/r$. In general there may be both outgoing and incoming waves so the general solution is

$$E = c_1 \frac{f_1(kr - \omega t)}{r} + c_2 \frac{f_2(kr + \omega t)}{r}.$$

This solution is valid for $r > 0$.

1.3 Energy and momentum in optical fields

From the Maxwell equations (1.1a-1.1d) we have that

$$\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{H} \cdot (\nabla \times \mathbf{E}) = \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} + \mathbf{J} \cdot \mathbf{E},$$

which can be rewritten as

$$\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} + \nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\mathbf{J} \cdot \mathbf{E}. \quad (1.7)$$

Assuming that the medium is linear in its electric and magnetic properties we can rewrite Eq. (1.7) in the form

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{S} = -\mathbf{J} \cdot \mathbf{E}, \quad (1.8)$$

where the energy density is

$$u = \frac{1}{2} (\mathbf{E} \cdot \mathbf{D} + \mathbf{H} \cdot \mathbf{B}) \quad \left[\frac{\text{J}}{\text{m}^3} \right], \quad (1.9)$$

and the energy flow is given by the Poynting vector

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad \left[\frac{\text{J}}{\text{m}^2 \text{s}} \right]. \quad (1.10)$$

The rate at which the electromagnetic fields do work on a current distribution \mathbf{J} in a volume V is given by

$$\int_V d^3x \mathbf{J} \cdot \mathbf{E}, \quad (1.11)$$

from which we can identify $-\mathbf{J} \cdot \mathbf{E}$ as the local rate at which charges do work on the fields. Equation (1.8) can therefore be interpreted as an energy continuity equation.

We now wish to find the intensity of a traveling electromagnetic wave. The Maxwell equations are written in terms of real fields, whereas it is convenient to use a complex

notation for electromagnetic waves. Consider a component of the real electric field

$$\begin{aligned}
 E &= |\mathcal{E}| \cos(\omega t - \theta) \\
 &= \frac{|\mathcal{E}|}{2} (e^{-i(\omega t - \theta)} + e^{i(\omega t - \theta)}) \\
 &= \frac{|\mathcal{E}| e^{i\theta}}{2} e^{-i\omega t} + c.c. \\
 &= \frac{\mathcal{E}}{2} e^{-i\omega t} + c.c. .
 \end{aligned} \tag{1.12}$$

Specializing to scalar, monochromatic traveling plane waves we can write

$$\mathbf{E} = \hat{x} \frac{\mathcal{E}}{2} e^{i(kz - \omega t)} + c.c. , \tag{1.13a}$$

$$\mathbf{H} = \hat{y} \frac{\mathcal{H}}{2} e^{i(kz - \omega t)} + c.c. . \tag{1.13b}$$

The time averaged Poynting vector is then

$$\langle \mathbf{S} \rangle = \hat{z} \frac{1}{4} (\mathcal{E} \mathcal{H}^* + \mathcal{E}^* \mathcal{H}) . \tag{1.14}$$

Using the Maxwell equations, the expression for the speed of light $c = (\epsilon_0 \mu_0)^{-1/2}$, and assuming the medium is non-magnetic so $\mu = \mu_0$, we find that $\mathcal{H} = \epsilon_0 n c \mathcal{E}$, where n is the index of refraction ($c/n = \omega/k$). Hence

$$\langle \mathbf{S} \rangle = \hat{z} \frac{\epsilon_0 n c}{2} |\mathcal{E}|^2 . \tag{1.15}$$

The intensity is thus

$$I = \frac{\epsilon_0 n c}{2} |\mathcal{E}|^2 \text{ [W/m}^2\text{]}, \tag{1.16}$$

where the field \mathcal{E} has units of V/m. In cgs units $[\mathcal{E}] = \text{statvolts/cm}$ and the corresponding intensity is

$$I = \frac{n c}{8\pi} |\mathcal{E}|^2 \text{ [dyne/cm}^2\text{]}. \tag{1.17}$$

Although the cgs expression is free of factors of ϵ_0 the field units of statvolts/cm are not as convenient for practical calculations as are SI units where we have V/m. Although the choice of units is arbitrary, there has been a strong trend towards adoption of SI in the last decades and we will follow the trend.

As an example of an optical field strength consider a beam with power $P = 1 \text{ mW}$ and a diameter of $d = 1 \text{ mm}$. This corresponds to a low power laser pointer. The intensity is $I = P/A = 4P/\pi d^2 = 1270 \text{ W/m}^2$. Using $n = 1$, $c = 299792458 \text{ m/s}$, and $\epsilon_0 = 8.85 \times 10^{-12}$ we find a field strength $\mathcal{E} = 980 \text{ V/m} = 0.98 \text{ V/mm}$. The (oscillating) potential difference across the beam is just under 1 V.

Light beams also carry momentum. A single photon has energy $\hbar\omega$ and momentum $\mathbf{p} = \hbar\mathbf{k}$. The number of photons per second per unit area in a beam is

$$N = \frac{I}{\hbar\omega} \left[\frac{1}{\text{m}^2 \text{s}} \right].$$

The momentum transported per second per unit area is

$$N\hbar k = \frac{I}{\hbar\omega} \hbar k = I \frac{k}{\omega} = \frac{nI}{c} = \frac{\epsilon_0 n^2}{2} |\mathcal{E}|^2 \quad \left[\frac{\text{kg}}{\text{m s}^2} \right].$$

Although the photon momentum is small on macroscopic scales it is sufficient to slow down and cool atoms, molecules, and small mechanical objects.

1.4 Refractive Index

Let's take a closer look at the refractive index. In linear isotropic media with no free charges or currents the Maxwell equations are

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (1.18a)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \quad (1.18b)$$

$$\nabla \cdot \mathbf{D} = 0, \quad (1.18c)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (1.18d)$$

and $\mathbf{D} = \epsilon \mathbf{E}$, $\mathbf{H} = \frac{1}{\mu} \mathbf{B}$. Let's write the fields in complex form as plane waves propagating in the direction \hat{k}

$$\mathbf{E} = \frac{\mathcal{E}}{2} e^{i\theta} \hat{E} + c.c. \quad (1.19)$$

with $\theta = \mathbf{k} \cdot \mathbf{r} - \omega t$, $\mathbf{k} = |k| \hat{k}$, and similarly for $\mathbf{D}, \mathbf{B}, \mathbf{H}$. Here \hat{E} is a unit vector in the direction of \mathbf{E} and we will assume $\omega > 0$. Note that k is the wavenumber inside the medium and it is related to the vacuum wavenumber k_0 by $k = nk_0$. Therefore $\omega/k = c/n$ and $\omega/k_0 = c$.

It is instructive to examine more closely the propagation phase $\theta = \mathbf{k} \cdot \mathbf{r} - \omega t = nk_0 \hat{k} \cdot \mathbf{r} - \omega t$, with \hat{k} a unit vector. In media with gain or loss the refractive index is complex and can be written as $n = n_r + i\kappa$ where n_r refers to the real part and κ is the imaginary part. In such media the complex field amplitude propagates as

$$\mathbf{E} \sim e^{i(nk_0 \hat{k} \cdot \mathbf{r} - \omega t)} \hat{E} \sim e^{i(n_r k_0 \hat{k} \cdot \mathbf{r} - \omega t)} e^{-k_0 \kappa \hat{k} \cdot \mathbf{r}} \hat{E}.$$

We introduced the refractive index $n = \sqrt{\frac{\epsilon \mu}{\epsilon_0 \mu_0}} = c\sqrt{\epsilon \mu}$ at the end of Sec. 1.2. The permittivity and permeability for oscillating fields are in general complex quantities¹. If we limit ourselves to the simplest situation of both ϵ and μ real there are four possible cases

$$\epsilon > 0, \mu > 0 \quad n_r = \pm c\sqrt{\epsilon \mu}, \quad \kappa = 0, \quad (1.20a)$$

$$\epsilon > 0, \mu < 0 \quad n_r = 0, \quad \kappa = \pm c\sqrt{\epsilon \mu}, \quad (1.20b)$$

$$\epsilon < 0, \mu > 0 \quad n_r = 0, \quad \kappa = \pm c\sqrt{\epsilon \mu}, \quad (1.20c)$$

$$\epsilon > 0, \mu > 0 \quad n_r = \pm c\sqrt{\epsilon \mu}, \quad \kappa = 0. \quad (1.20d)$$

¹This is consistent with the fact that the Maxwell equations are strictly real. A complex valued material response is a convenient way of representing frequency dependent phases in the material response. The theory can also be formulated with explicitly real permittivity and permeability, but then takes on a less compact mathematical form.

When ϵ and μ have opposite signs the wave propagates $\sim e^{\mp k_0 c \sqrt{\epsilon\mu} \hat{k} \cdot \mathbf{r}}$. In passive media the wave cannot grow exponentially and the only physically sensible solution is $\kappa = +c\sqrt{\epsilon\mu}$. We see that we must take the positive square root for κ in (1.20). The wave is exponentially attenuated in this type of medium, which is called a single negative medium.

When ϵ and μ have the same sign $\kappa = 0$ so the wave propagates without attenuation and $n_r = \pm c\sqrt{\epsilon\mu}$. There is an ambiguity as to which sign should be used for the square root. To resolve this question we use the Maxwell equations. The differential operators acting on the representation (1.19) take on a simple form:

$$\nabla \times \dots \rightarrow i\mathbf{k} \times \dots, \quad \frac{\partial}{\partial t} \dots \rightarrow -i\omega \dots$$

With these rules the first two Maxwell equations can be written as

$$\mathbf{k} \times \mathbf{E} = \mu\omega\mathbf{H}, \quad \mathbf{k} \times \mathbf{H} = -\epsilon\omega\mathbf{E}. \quad (1.21)$$

With ω positive we see that when $\epsilon > 0$ and $\mu > 0$ then $\mathbf{k} \times \mathbf{E} \sim \mathbf{H}$, $\mathbf{k} \times \mathbf{H} \sim -\mathbf{E}$ and $\mathbf{E} \times \mathbf{H} \sim \mathbf{k}$. Using (1.21), the formula for the vector triple product, and $\mathbf{k} = |k|\hat{k}$ we can show that the Poynting vector is

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} = \frac{|k|}{\mu\omega} E^2 \hat{k}.$$

When $k > 0$ then $|k| = k$ and substituting with $k/\omega = n/c$ we get

$$\mathbf{E} \times \mathbf{H} = \frac{n}{\mu c} E^2 \hat{k}.$$

The refractive index is

$$n = +\sqrt{\epsilon\mu/(\epsilon_0\mu_0)} \quad (\epsilon > 0, \quad \mu > 0).$$

Since $\mathbf{E} \times \mathbf{H}$ is parallel to $+\hat{k}$ when $\epsilon > 0$ and $\mu > 0$ we must have n positive which means taking the positive square root. In this case $\mathbf{E}, \mathbf{H}, \mathbf{k}$ form a right handed coordinate system and the Poynting vector is parallel to \mathbf{k} . This is also called a right handed medium as shown in Fig. 1.1.

Question: We have shown that plane waves are polarized perpendicular to \mathbf{k} . On the other hand Eq. (1.3) has nonzero solutions for a component of \mathbf{E} polarized parallel to the direction of propagation. How do you explain the apparent contradiction?

It is interesting to consider the possibility of $n < 0$ [1]. When $\epsilon < 0$ and $\mu < 0$ then $\mathbf{k} \times \mathbf{E} \sim -\mathbf{H}$, $\mathbf{k} \times \mathbf{H} \sim \mathbf{E}$ and $\mathbf{E} \times \mathbf{H} \sim -\mathbf{k}$. In this case $\mathbf{E}, \mathbf{H}, \mathbf{k}$ form a left handed coordinate system and the Poynting vector is parallel to $-\mathbf{k}$ as seen in Fig. 1.1. We still have

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} = \frac{|k|}{\mu\omega} E^2 \hat{k}$$

except now $\mu < 0$ and to satisfy $\mathbf{E} \times \mathbf{H} \sim -\mathbf{k}$ we must have $n < 0$ so that

$$\mathbf{E} \times \mathbf{H} = -\frac{n}{\mu c} E^2 \hat{k} \sim -\hat{k}$$

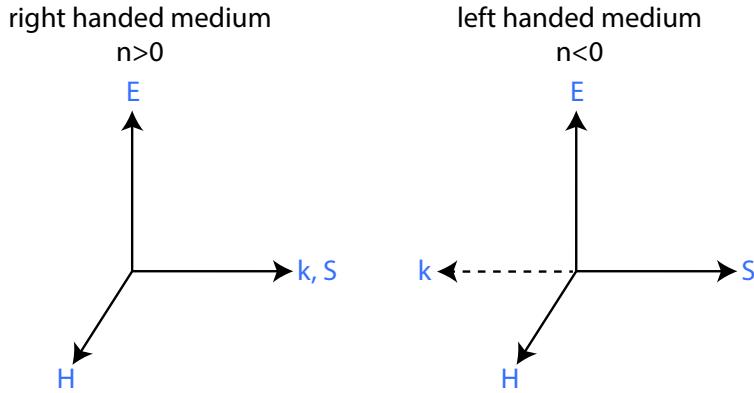


Figure 1.1: Wave propagation in right and left handed media.

and

$$n = -\sqrt{\epsilon\mu/(\epsilon_0\mu_0)} \quad (\epsilon < 0, \mu < 0).$$

We see that for consistency we have to take the negative square root. This is also called a left handed medium. It is possible to design and fabricate materials that have a negative refractive index. This leads to many unusual and remarkable properties including perfect lenses[2], negative Doppler shift, invisibility cloaks, etc.

1.5 Polarization

The most general polarization of a plane electromagnetic wave propagating along ϵ_z is²

$$\mathbf{E} = E_{0x} \cos(kz - \omega t + \phi_x) \epsilon_x + E_{0y} \cos(kz - \omega t + \phi_y) \epsilon_y$$

The field amplitudes are E_{0x}, E_{0y} and the phases are ϕ_x, ϕ_y . This is the most general form of the polarization of a plane wave. The Cartesian components are

$$\begin{aligned} \frac{E_x}{E_{0x}} &= \cos(kz - \omega t) \cos(\phi_x) - \sin(kz - \omega t) \sin(\phi_x), \\ \frac{E_y}{E_{0y}} &= \cos(kz - \omega t) \cos(\phi_y) - \sin(kz - \omega t) \sin(\phi_y). \end{aligned}$$

It can be easily verified that

$$\left(\frac{E_x}{E_{0x}} \right)^2 + \left(\frac{E_y}{E_{0y}} \right)^2 - 2 \frac{E_x}{E_{0x}} \frac{E_y}{E_{0y}} \cos(\phi) = \sin^2(\phi) \quad (1.22)$$

where $\phi = \phi_y - \phi_x$ is the phase difference between the two field components. This is the equation of an ellipse in the coordinates E_x, E_y that is followed by the electric field vector as the light propagates. In general E_{0x} and E_{0y} are not equal to each other so it is convenient

²We will sometimes use $\epsilon_x, \epsilon_y, \epsilon_z$ to denote the Cartesian unit vectors $\hat{x}, \hat{y}, \hat{z}$.

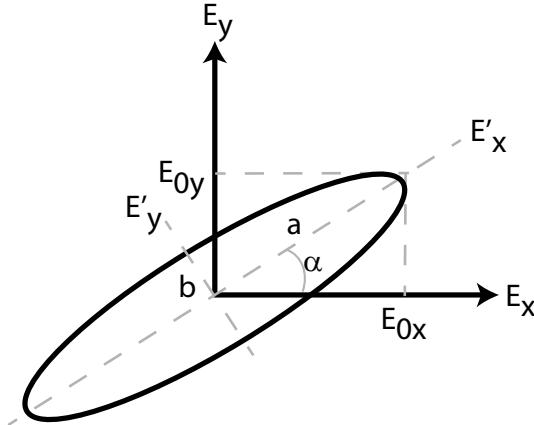


Figure 1.2: Polarization ellipse with semimajor and semiminor axes a, b and rotation angle α .

to introduce $E_{\text{mag}} = \sqrt{E_{0x}^2 + E_{0y}^2}$ and write $E_{0x} = E_{\text{mag}} \cos(\theta)$, $E_{0y} = E_{\text{mag}} \sin(\theta)$. The ratio of the amplitudes is then $E_{0y}/E_{0x} = \tan(\theta)$ and

$$\left(\frac{E_x}{\cos(\theta)} \right)^2 + \left(\frac{E_y}{\sin(\theta)} \right)^2 - 2 \frac{E_x E_y}{\cos(\theta) \sin(\theta)} \cos(\phi) = E_{\text{mag}}^2 \sin^2(\phi). \quad (1.23)$$

The polarization ellipse is shown in Fig. 1.2.

We can put this equation into normal form by using the rotated coordinates

$$E'_x = \cos(\alpha)E_x + \sin(\alpha)E_y, \quad E'_y = -\sin(\alpha)E_x + \cos(\alpha)E_y.$$

Inserting the expressions for E'_x, E'_y into (1.23) we find that the cross term vanishes if we choose α to satisfy

$$\tan(2\alpha) = \tan(2\theta) \cos(\phi).$$

It can be shown after some algebra that the semimajor and semiminor axes of the ellipse are

$$a = E_{\text{mag}} \left[\frac{1 + \sqrt{1 - \sin^2(2\theta) \sin^2(\phi)}}{2} \right]^{1/2}, \quad b = E_{\text{mag}} \left[\frac{1 - \sqrt{1 - \sin^2(2\theta) \sin^2(\phi)}}{2} \right]^{1/2}.$$

This general polarization state simplifies for particular values of the parameters. When $\theta = \pi/4$ and $\phi = \pm\pi/2$ we have $E_{0x} = E_{0y} = E_{\text{mag}}/\sqrt{2} = a = b$ and the light is circularly polarized. When $\phi = n\pi$ with n an integer we have $a = E_{\text{mag}}$, $b = 0$ and the light is linearly polarized.

Complex unit vectors

An alternative way of analyzing the polarization is to define a complex, unit polarization vector as

$$\epsilon = \frac{\mathcal{E}_x \epsilon_x + \mathcal{E}_y \epsilon_y}{|\mathcal{E}|}$$

the vector field can be written as

$$\mathbf{E} = |\mathcal{E}| \left(\frac{e^{i(kz-\omega t)}}{2} \boldsymbol{\epsilon} + \frac{e^{-i(kz-\omega t)}}{2} \boldsymbol{\epsilon}^* \right).$$

If we pick a fixed reference plane, say $z = 0$, the time dependence of the field is

$$\mathbf{E}(z = 0) = |\mathcal{E}| \left(\frac{e^{-i\omega t}}{2} \boldsymbol{\epsilon} + \frac{e^{i\omega t}}{2} \boldsymbol{\epsilon}^* \right).$$

The most general complex unit vector in the $x - y$ plane can be written as

$$\boldsymbol{\epsilon} = \cos \theta e^{i\chi_x} \boldsymbol{\epsilon}_x + \sin \theta e^{i\chi_y} \boldsymbol{\epsilon}_y.$$

The field then takes the form

$$\mathbf{E} = |\mathcal{E}| [\cos \theta (\cos \chi_x \cos \omega t - \sin \chi_x \sin \omega t) \boldsymbol{\epsilon}_x + \sin \theta (\cos \chi_y \cos \omega t - \sin \chi_y \sin \omega t) \boldsymbol{\epsilon}_y].$$

In general the magnitude of the field is time dependent and oscillates between maximum and minimum values. The magnitude is

$$|\mathbf{E}| = [\cos^2(\theta) \cos^2(\chi_x + \omega t) + \sin^2(\theta) \cos^2(\chi_y + \omega t)]^{1/2}$$

There are several basic states of polarization. When $\chi_y = \chi_x \pm \pi$ then

$$\begin{aligned} \mathbf{E} &= |\mathcal{E}| [\cos \theta (\cos \chi_x \cos \omega t - \sin \chi_x \sin \omega t) \boldsymbol{\epsilon}_x - \sin \theta (\cos \chi_x \cos \omega t - \sin \chi_x \sin \omega t) \boldsymbol{\epsilon}_y] \\ &= |\mathcal{E}| (\cos \theta \boldsymbol{\epsilon}_x - \sin \theta \boldsymbol{\epsilon}_y) \cos(\omega t + \chi_x). \end{aligned}$$

and $E_y/E_x = -\tan \theta = \text{constant}$ at all times so the polarization is linear.

When $\chi_y = \chi_x \pm \pi/2$ we have

$$\begin{aligned} \mathbf{E} &= |\mathcal{E}| [\cos \theta (\cos \chi_x \cos \omega t - \sin \chi_x \sin \omega t) \boldsymbol{\epsilon}_x \pm \sin \theta (\sin \chi_x \cos \omega t + \cos \chi_x \sin \omega t) \boldsymbol{\epsilon}_y] \\ &= |\mathcal{E}| [\cos \theta \cos(\omega t + \chi_x) \boldsymbol{\epsilon}_x \pm \sin \theta \sin(\omega t + \chi_x) \boldsymbol{\epsilon}_y] \end{aligned}$$

so

$$E_y/E_x = \pm \tan \theta \tan(\omega t + \chi_x).$$

The light is elliptically polarized since the field vector traces out an ellipse in the $x - y$ plane with period $2\pi/\omega$.

When $\theta = \pi/4$ so $\tan \theta = 1$ the light is circularly polarized and choosing the time origin such that $\chi_x = 0$ the electric field is

$$\mathbf{E} = |\mathcal{E}| \frac{\cos(\omega t) \boldsymbol{\epsilon}_x \pm \sin(\omega t) \boldsymbol{\epsilon}_y}{\sqrt{2}}. \quad (1.24)$$

The plus(minus) signs correspond to an observer looking parallel with the direction of light propagation (towards positive z) seeing the electric field rotating about the $\boldsymbol{\epsilon}_z$ axis in a clockwise(counterclockwise) direction. These two cases are referred to as left hand circular

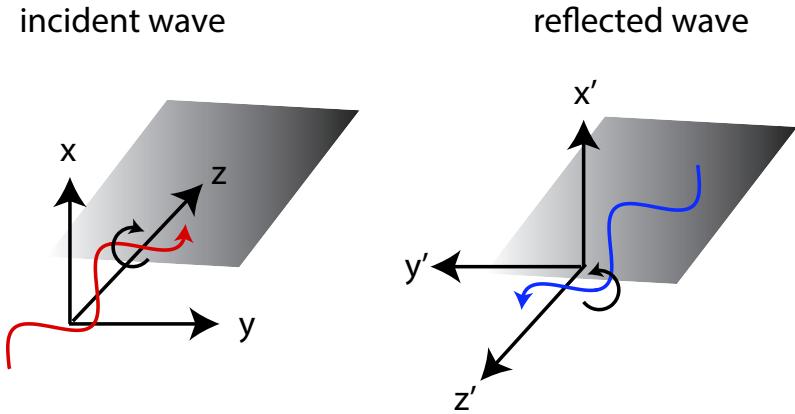


Figure 1.3: Reflection of light from a mirror. The coordinate system for the incident wave is $(\epsilon_x, \epsilon_y, \epsilon_z)$ and for the reflected wave $(\epsilon_{x'}, \epsilon_{y'}, \epsilon_{z'})$. To make both coordinate systems right handed we set $\epsilon_{x'} = \epsilon_x, \epsilon_{y'} = -\epsilon_y, \epsilon_{z'} = -\epsilon_z$.

(lhc) and right hand circular (rhc) states of polarization³. The electric field for circular polarization is thus

$$\mathbf{E}_{\text{rhc}} = |\mathcal{E}| \frac{\cos(\omega t)\epsilon_x - \sin(\omega t)\epsilon_y}{\sqrt{2}}, \quad \mathbf{E}_{\text{lhc}} = |\mathcal{E}| \frac{\cos(\omega t)\epsilon_x + \sin(\omega t)\epsilon_y}{\sqrt{2}}.$$

and the complex basis vectors are

$$\epsilon_{\text{rhc}} = \frac{\epsilon_x - i\epsilon_y}{\sqrt{2}}, \quad \epsilon_{\text{lhc}} = \frac{\epsilon_x + i\epsilon_y}{\sqrt{2}}.$$

When using complex basis vectors orthogonality is determined using a complex inner product. We define the overlap of two polarization states 1, 2 as

$$f = \epsilon_2^* \cdot \epsilon_1.$$

With this definition we see that

$$\epsilon_{\text{rhc}}^* \cdot \epsilon_{\text{rhc}} = \epsilon_{\text{lhc}}^* \cdot \epsilon_{\text{lhc}} = 1, \quad \epsilon_{\text{rhc}}^* \cdot \epsilon_{\text{lhc}} = \epsilon_{\text{rhc}} \cdot \epsilon_{\text{lhc}}^* = 0.$$

One may ask what happens to the handedness of circular polarization upon reflection from a mirror. The geometry is shown in Fig. 1.3. At normal reflection from a conducting surface the parallel component of the field must vanish so $E_{xr} = -E_x, E_{yr} = -E_y$, where r refers to the reflected beam. Due to the inversion of the y axis to keep the coordinate system right handed we find

$$E_{x'r} = E_{xr} = -E_x$$

³This convention is traditional, but is opposite to the usual right hand rule for vector cross products. Note that if we define the sense of field rotation as that seen by an observer looking towards negative z (opposite to the direction of light propagation) then a clockwise(counterclockwise) sense of rotation is right(left) hand circular polarization.

but

$$E_{y'r} = -E_{yr} = E_y.$$

Thus $E_{y'r}/E_{x'r} = -E_y/E_x$ which implies that the handedness of the light is inverted upon reflection.

1.6 Polarizers and waveplates

Polarizers are optical elements that transmit light polarized along the axis of the polarizer and block light polarized in the perpendicular direction. If the incident light is polarized at an angle θ with respect to the axis of the polarizer the amplitude of the transmitted light is $\cos(\theta)$.

Question: A beam of circularly polarized light with unit intensity is incident on a polarizer. What is the intensity of the transmitted light?

Waveplates are used to change the polarization state of light. They can be made from birefringent media which have different indices of refraction for different polarizations. The two most common waveplates are the half wave plate and quarter wave plate.

The half wave plate imparts a half wave or π phase difference to light polarized parallel and perpendicular to the axis of the waveplate. If linearly polarized light is incident with angle θ between the polarization vector and the axis of the half wave plate the plane of polarization of the output light is rotated by 2θ .

The quarter wave plate imparts a quarter wave or $\pi/2$ phase difference to light polarized parallel and perpendicular to the axis of the waveplate. This converts linearly polarized light to circular, and vice versa.

Starting with linear polarization any desired polarization state can be synthesized using a half and quarter wave plate. Combinations of polarizers and waveplates can be used to synthesize other components. For example a linear polarizer together with waveplates can be used to make a polarizer that transmits say right handed circularly polarized light and blocks left handed polarized light.

1.7 Jones calculus

The polarization of a plane wave can be described with two complex coefficients written as a column vector. Transformation of the polarization due to optical elements can be described with a 2×2 complex matrix. The state of polarization and transformation operators are sometimes referred to as Jones vectors and matrices[3] and they provide a convenient formalism for following the polarization through a sequence of optical elements.

The Jones vectors for basic polarization states are

linearly polarized at angle θ to x	$\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$
right circularly polarized	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$
left circularly polarized	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}.$

The Jones matrices for some common optical elements are

linear polarizer at angle θ to x	$\begin{pmatrix} \cos^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin^2(\theta) \end{pmatrix}$
half wave plate at angle θ to x	$\begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}$
quarter wave plate at angle θ to x	$\begin{pmatrix} \cos^2(\theta) + i\sin^2(\theta) & (1-i)\cos(\theta)\sin(\theta) \\ (1-i)\cos(\theta)\sin(\theta) & \sin^2(\theta) + i\cos^2(\theta) \end{pmatrix}.$

1.8 Stokes vectors and Mueller matrices

An arbitrary polarization state depends on three real parameters E_x, E_y the field amplitudes along x and y and ϕ their phase difference. Recall the equation for the polarization ellipse (1.22)

$$\left(\frac{E_x(t)}{E_{0x}}\right)^2 + \left(\frac{E_y(t)}{E_{0y}}\right)^2 - 2\frac{E_x(t)}{E_{0x}}\frac{E_y(t)}{E_{0y}}\cos(\phi) = \sin^2(\phi).$$

Take the time average of both sides, use $\langle E_x(t)^2 \rangle = E_{0x}^2/2$, $\langle E_y(t)^2 \rangle = E_{0y}^2/2$, $\langle E_x(t)E_y(t) \rangle = E_{0x}E_{0y}\cos(\phi)/2$, and multiply through by $4E_{0x}^2E_{0y}^2$ to get

$$4E_{0x}^2E_{0y}^2 - 4E_{0x}^2E_{0y}^2\cos^2(\phi) = 4E_{0x}^2E_{0y}^2\sin^2(\phi).$$

This can be rewritten as

$$(E_{0x}^2 + E_{0y}^2)^2 = (E_{0x}^2 - E_{0y}^2)^2 + (2E_{0x}E_{0y}\cos(\phi))^2 + (2E_{0x}E_{0y}\sin(\phi))^2.$$

The four factors from left to right that are each squared are the Stokes parameters[4] S_0, S_1, S_2, S_3 defined as

$$\begin{aligned} S_0 &= E_x^2 + E_y^2 \\ S_1 &= E_x^2 - E_y^2 \\ S_2 &= 2E_xE_y\cos(\phi) \\ S_3 &= 2E_xE_y\sin(\phi). \end{aligned}$$

Since $S_0^2 = S_1^2 + S_2^2 + S_3^2$ there are only three independent parameters. The parameter S_0 is proportional to the wave intensity while S_1, S_2, S_3 characterize the polarization state. This can be represented as a point on the surface of a sphere known as the Poincaré sphere with

coordinates S_1, S_2, S_3 . The Stokes parameters are particularly useful for describing partially polarized light, but we will not discuss this further here.

Just as Jones matrices are used to propagate a Jones vector through an optical system Mueller matrices are used to propagate Stokes vectors. The Mueller matrices are 4×4 with

16 components. If we define a Stokes vector $\mathbf{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix}$ a transformed vector is

$$\mathbf{S}_{\text{out}} = \mathbf{MS}$$

where \mathbf{M} is a Mueller matrix. The Mueller matrix for a polarizer at angle θ is

$$\mathbf{M} = \begin{pmatrix} 1 & \cos(2\theta) & \sin(2\theta) & 0 \\ \cos(2\theta) & \cos^2(2\theta) & \cos(2\theta)\sin(2\theta) & 0 \\ \sin(2\theta) & \cos(2\theta)\sin(2\theta) & \sin^2(2\theta) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

1.9 Angular momentum of light

Light carries angular momentum due to the polarization (spin) state and the spatial structure of the mode. We will consider here only the polarization contribution for the simplest case of an electromagnetic field propagating in vacuum.

The classical expression for angular momentum is $\mathbf{J} = \mathbf{r} \times \mathbf{P}$. The energy density can be written as $\frac{1}{c}\langle \mathbf{S} \rangle = \frac{1}{c}\langle \mathbf{E} \times \mathbf{H} \rangle$ so the momentum density can be written as

$$\mathbf{p} = \frac{\hbar k}{\hbar\omega} \frac{\langle \mathbf{S} \rangle}{c} = \frac{1}{c^2} \langle \mathbf{S} \rangle = \epsilon_0 \mu_0 \langle \mathbf{S} \rangle = \epsilon_0 \langle \mathbf{E} \times \mathbf{B} \rangle.$$

The angular momentum density is thus

$$\mathbf{j} = \epsilon_0 \mathbf{r} \times \langle \mathbf{E} \times \mathbf{B} \rangle. \quad (1.25)$$

This expression accounts for the angular momentum due to the spatial structure of the field, which is called orbital angular momentum, as well as the angular momentum due to polarization or spin of the photons. When the field is a plane wave with planes of constant \mathbf{E} and \mathbf{B} there is no orbital angular momentum. In this case Eq. (1.25) describes the spin angular momentum.

The magnetic field can always be written as $\mathbf{B} = \nabla \times \mathbf{A}$ with \mathbf{A} the vector potential. It can furthermore be shown that when the angular momentum is due to photon spin Eq. (1.25) is equivalent to $\mathbf{j} = \epsilon_0 \mathbf{E} \times \mathbf{A}$ so the spin angular momentum of a beam is

$$\mathbf{J} = \epsilon_0 \int_{\text{beam}} d\mathbf{r} \mathbf{E} \times \mathbf{A}$$

where \mathbf{A} is the vector potential. The electric field is related to the vector potential by $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}$. The vector potential for the field (1.24) is

$$\mathbf{A} = \frac{|\mathcal{E}|}{\omega} \frac{\sin(\omega t) \mathbf{\epsilon}_x \mp \cos(\omega t) \mathbf{\epsilon}_y}{\sqrt{2}} \quad (1.26)$$

giving

$$\mathbf{J} = \mp \frac{\epsilon_0}{2\omega} \epsilon_z \int_{\text{beam}} d\mathbf{r} |\mathcal{E}|^2 = \mp \frac{1}{c\omega} \epsilon_z \int_{\text{beam}} d\mathbf{r} I,$$

with I the optical intensity. We then write $\int_{\text{beam}} d^3\mathbf{r} I = L \int_{\text{area}} d^2\mathbf{r} I = LP$ where P is the optical power and $L = c\Delta t$ is the effective length of a photon mode. Combining the factors gives

$$\mathbf{J} = \mp \frac{P\Delta t}{\omega} \epsilon_z.$$

The energy transported by the beam in a time Δt is $U = P\Delta t = N\hbar\omega$ where N is the number of photons. Thus the angular momentum per photon is

$$\frac{\mathbf{J}}{N} = \mp \hbar \epsilon_z.$$

We see that a photon with right(left) hand circular polarization carries angular momentum $-(+)\hbar$ directed along the direction of propagation. This is the spin angular momentum of the photon which is a spin 1 particle. Note that when light is reflected normally from a mirror the handedness of circular polarization changes, but the direction of propagation also changes, so the angular momentum per photon does not change.

1.10 Fresnel coefficients

Wide optical beams (plane waves) propagating along the direction \mathbf{k} are polarized perpendicularly to \mathbf{k} . When the beam is reflected from an interface we define a plane of incidence containing the incident beam with wavevector \mathbf{k}_i and the reflected beam with wavevector \mathbf{k}_r . When the incident beam is polarized parallel to the plane of incidence we call the beam p polarized, and when the beam is polarized perpendicular to the plane of incidence we call the beam s polarized⁴. The geometry is shown in Fig. 1.4

The coefficients for transmission and reflection of optical waves at interfaces between different media are called the Fresnel coefficients. The reflected and transmitted fields E_r, E_t are related to the incident field E_i by

$$\begin{aligned} E_{rs} &= r_s E_{is}, & E_{rp} &= r_p E_{ip}, \\ E_{ts} &= t_s E_{is}, & E_{tp} &= t_p E_{ip}. \end{aligned}$$

Here s, p refer to the s and p polarized components of the fields. These coefficients can be found using the boundary conditions for electromagnetic fields at surfaces separating media with different electric and magnetic properties. The boundary conditions are continuity of the tangential components of \mathbf{E} and the normal components of \mathbf{B} . The presence of surface charges or currents results in discontinuity of the normal component of \mathbf{D} or the tangential component of \mathbf{H} .

⁴s stands for senkrecht, the German word for perpendicular.

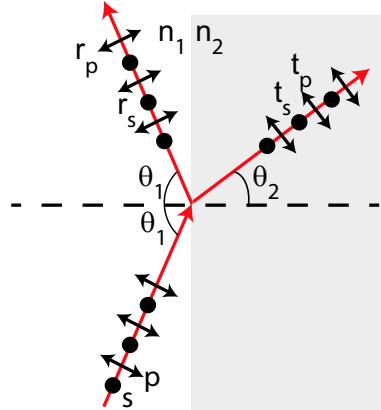


Figure 1.4: Fresnel coefficients r_s, r_p and t_s, t_p determine the amplitude of the reflected and transmitted waves.

For normal incidence the relations connecting the electric fields of the incident, transmitted and reflected waves, E_i, E_t , and E_r are

$$E_t = \frac{2n_1}{n_1 + n_2} E_i \quad (1.27)$$

$$E_r = -\frac{n_2 - n_1}{n_2 + n_1} E_i. \quad (1.28)$$

Here n_1 is the index of the medium containing the incident wave, and n_2 is the index of the medium containing the transmitted wave.

The Fresnel coefficients for light incident at angle θ_1 in a medium with index n_1 and refracted at angle θ_2 into a medium with index n_2 are (angles are measured from the surface normal)

$$r_s = -\frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \quad (1.29a)$$

$$t_s = \frac{2 \cos(\theta_1) \sin(\theta_2)}{\sin(\theta_1 + \theta_2)} \quad (1.29b)$$

$$r_p = -\frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \quad (1.29c)$$

$$t_p = \frac{2 \cos(\theta_1) \sin(\theta_2)}{\sin(\theta_1 + \theta_2) \cos(\theta_1 - \theta_2)}. \quad (1.29d)$$

The subscripts s and p refer to polarization perpendicular and parallel to the plane of incidence respectively. A derivation of these expressions can be found many places, for example in [5].

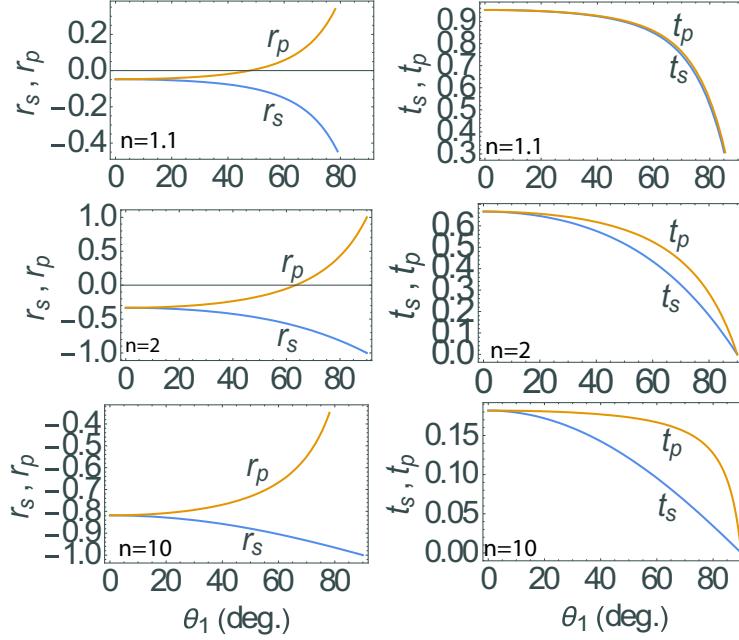


Figure 1.5: Fresnel coefficients for three different values of $n = n_2/n_1$.

If we eliminate θ_2 using Snell's law $n_1 \sin \theta_1 = n_2 \sin \theta_2$ we get

$$r_s = \frac{\cos(\theta_1) - \sqrt{n^2 - \sin^2(\theta_1)}}{\cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}} \quad (1.30a)$$

$$t_s = \frac{2 \cos(\theta_1)}{\cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}} \quad (1.30b)$$

$$r_p = \frac{n^2 \cos(\theta_1) - \sqrt{n^2 - \sin^2(\theta_1)}}{n^2 \cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}} \quad (1.30c)$$

$$t_p = \frac{2n \cos(\theta_1)}{n^2 \cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}}. \quad (1.30d)$$

where we have introduced the relative index $n = n_2/n_1$. Some plots of the coefficients are shown in Fig. 1.5. For $n > 1$ the coefficients are always real valued. For $n < 1$ the coefficients are complex for certain angles. We will look at this in more detail in a later section.

The fraction of the incident power that is reflected from an interface is called the reflectance and is given by $R_s = |r_s|^2$ or $R_p = |r_p|^2$ for s or p polarizations. Energy conservation requires that the transmittance is given by $T_s = 1 - R_s$ or $T_p = 1 - R_p$. It is important to note that $T_s \neq |t_s|^2$, $T_p \neq |t_p|^2$. Instead we have the relations

$$|r_s|^2 + \frac{n_2 \cos(\theta_2)}{n_1 \cos(\theta_1)} |t_s|^2 = 1, \quad (1.31a)$$

$$|r_p|^2 + \frac{n_2 \cos(\theta_2)}{n_1 \cos(\theta_1)} |t_p|^2 = 1. \quad (1.31b)$$

These expressions account for the dependence of intensity on refractive index and on the change of beam cross section at an interface proportional to $1/\cos \theta$. It can be verified that Eqs. (1.29) satisfy these relations. Only when $n_2 = n_1$ and $\theta_2 = \theta_1$ do we get the simpler relation $|r|^2 + |t|^2 = 1$.

It is useful to define the Fresnel coefficients for light propagating backwards from medium 2 to medium 1. These are

$$r'_s = \frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \quad (1.32a)$$

$$t'_s = \frac{2 \cos(\theta_2) \sin(\theta_1)}{\sin(\theta_1 + \theta_2)} \quad (1.32b)$$

$$r'_p = \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \quad (1.32c)$$

$$t'_p = \frac{2 \cos(\theta_2) \sin(\theta_1)}{\sin(\theta_1 + \theta_2) \cos(\theta_1 - \theta_2)}. \quad (1.32d)$$

We see that $r'_s = -r_s$, $r'_p = -r_p$ while t_s, t'_s and t_p, t'_p are related by angle dependent factors. It can be verified that the following mixed conservation relations are satisfied for arbitrary angles

$$|r_s|^2 + t_s t'_s = 1, \quad (1.33a)$$

$$|r_p|^2 + t_p t'_p = 1. \quad (1.33b)$$

These energy conservation relations will be useful in analysis of Fabry-Pérot resonators.

1.11 The beamsplitter

A generic beam splitter is shown in Fig. 1.6. There are two input ports (0 and 1) and two output ports (2 and 3). Without making any assumptions about the internal structure of the beamsplitter, apart from the requirement that it is lossless, the transmission and reflection

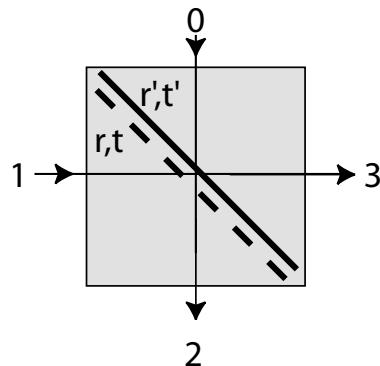


Figure 1.6: The beamsplitter.

coefficients must satisfy the following relations (Stokes, 1849)

$$|r'| = |r|, \quad (1.34a)$$

$$|t'| = |t|, \quad (1.34b)$$

$$|r|^2 + |t|^2 = |r'|^2 + |t'|^2 = 1, \quad (1.34c)$$

$$rt'^* + r'^*t = 0. \quad (1.34d)$$

In this context the Fresnel coefficients r, t, r', t' refer to the effective behavior of the beam splitter for a given state of polarization. Relations (1.34) are thus separately valid for both s and p polarizations. It is important to keep in mind that Eqs. (1.34) describe a situation where the index of refraction is the same at all external ports of the beamsplitter. If this were not the case, then we would not find e.g. $|t| = |t'|$.

These reciprocity relations can be derived as follows. The incident intensity is

$$I_{\text{in}} = (\mathcal{E}_0, \mathcal{E}_1) \cdot (\mathcal{E}_0, \mathcal{E}_1)^\dagger = |\mathcal{E}_0|^2 + |\mathcal{E}_1|^2 = I_0 + I_1.$$

The output fields are

$$\begin{pmatrix} \mathcal{E}_2 \\ \mathcal{E}_3 \end{pmatrix} = \begin{pmatrix} t' & r \\ r' & t \end{pmatrix} \begin{pmatrix} \mathcal{E}_0 \\ \mathcal{E}_1 \end{pmatrix},$$

and the output intensity is

$$\begin{aligned} I_{\text{out}} &= (\mathcal{E}_2, \mathcal{E}_3) \cdot (\mathcal{E}_2, \mathcal{E}_3)^\dagger \\ &= (|r'|^2 + |t'|^2)I_0 + (|r|^2 + |t|^2)I_1 + (rt'^* + r'^*t)\mathcal{E}_0^*\mathcal{E}_1 + (r^*t' + r't^*)\mathcal{E}_0\mathcal{E}_1^*. \end{aligned}$$

Conservation of intensity for arbitrary input fields gives the conditions

$$|r|^2 + |t|^2 = 1, \quad (1.35a)$$

$$|r'|^2 + |t'|^2 = 1, \quad (1.35b)$$

$$rt'^* + r'^*t = 0. \quad (1.35c)$$

The additional conditions $|r| = |r'|$ and $|t| = |t'|$ follow from these relations.

We could also derive Eqs. (1.34) from the formal requirement that a lossless beamsplitter results in a unitary transformation of the fields. The scattering matrix of the beamsplitter is

$$S = \begin{pmatrix} t' & r \\ r' & t \end{pmatrix}$$

and unitarity means $SS^\dagger = S^\dagger S = I$. These conditions result in

$$|r|^2 + |t|^2 = 1, \quad |r'|^2 + |t'|^2 = 1, \quad |r|^2 + |t'|^2 = 1, \quad |r'|^2 + |t|^2 = 1, \quad (1.36a)$$

$$rt'^* + r'^*t = 0, \quad rt^* + r'^*t' = 0. \quad (1.36b)$$

The relations (1.36a) are equivalent to (1.34a - 1.34c). The second equality in (1.36b) looks different, but is equivalent to (1.34d). To see this write $r = |r|e^{i\phi_r}$, $r' = |r'|e^{i\phi'_r}$, $t = |t|e^{i\phi_t}$, $t' = |t'|e^{i\phi'_t}$. Using the amplitude relations $|r| = |r'|$, $|t| = |t'|$ Eqs. (1.36b) both lead to

$$\phi_r + \phi'_r - \phi_t - \phi'_t = (2m + 1)\pi$$

with m an integer. Different phase choices are possible. For example we can set $\phi_t = \phi'_t = 0$ so $t = t'$ and both quantities are positive. With this choice it follows that $\phi_r + \phi'_r = (2m+1)\pi$, which gives $r = |r|e^{i\phi_r}$ and $r' = -|r|e^{-i\phi_r}$. A common choice for calculations is to take $\phi_r = 0$ so $r > 0$, $r' = -r$, and $t = t' > 0$.

The beamsplitter has some remarkable properties when used with weak optical fields containing single photons. Suppose a single photon is incident at port 0. For a 50/50 beamsplitter with $|r|^2 = |t|^2 = 1/2$ the photon will emerge at port 2 half the time and at port 3 half the time. Similarly if the photon is incident at port 1 it will emerge at port 2 half the time and at port 3 half the time. Based on these probabilities we might expect that if two photons were incident, one at port 0 and one at port 1 then half the time we would observe one photon at each output port, and a quarter of the time we would observe two photons at port 2 and a quarter of the time two photons at port 3.

However, what is observed is that two photons emerge at port 2 half the time and two photons emerge at port 3 half the time, but we never see one photon at each output port. This is known as the Hong-Ou-Mandel effect after the people who first demonstrated it experimentally[6]. The reason is that for two identical photons we must add the amplitudes for the different reflection and transmission paths and these amplitudes cancel out for a lossless beamsplitter for the case when one photon exits each port.

1.12 Brewster's angle

Light polarized in the plane of incidence (p-polarization) experiences zero reflectance and perfect transmission at Brewster's angle. This can be seen from the geometrical construction in Fig. 1.7. The incident light interacts with the second medium to drive an oscillating dipole which radiates into the reflected and transmitted beams. The oscillating dipole is parallel to the electric field in the second medium, and there must be a non-zero component perpendicular to the \mathbf{k} vector of the reflected light in order to observe a non-zero reflection. When $\theta_i + \theta_t = \pi/2$ the oscillating dipole is parallel to \mathbf{k}_r and the reflection vanishes.

The zero reflection condition is thus $\theta_t = \pi/2 - \theta_i$ or, using Snell's law,

$$n_i \sin(\theta_i) = n_t \sin(\theta_t) = n_t \sin(\pi/2 - \theta_i) = n_t \cos(\theta_i)$$

and

$$\theta_i = \tan^{-1}(n_t/n_i) = \theta_B.$$

For an air-glass interface $n_i = 1$, $n_t = 1.5$ we have $\theta_B = 56.3$ deg.

Light that is s-polarized has an electric field component that is perpendicular to \mathbf{k}_r for all incident angles, so there is no Brewster's angle effect for this polarization. Repeated passage of a light beam through a stack of plates at Brewster's angle will attenuate the s-polarized light while transmitting all of the p-polarized light. This can be used to create a highly polarized beam. Brewster's angle is also widely used in gain cells in laser cavities as a way of obtaining a very low loss interface for p-polarized light.

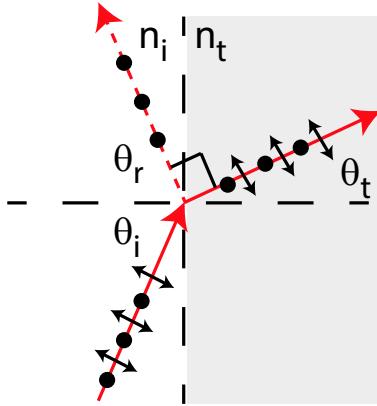


Figure 1.7: Perfect transmission of p-polarized light occurs at Brewster's angle.

1.13 Total internal reflection

Light is internally reflected on the hypotenuse of the prism shown in Fig. 1.8. The incidence angle is $\theta_1 = 45$ deg. The refractive index of the prism is n_1 and outside the prism $n_2 = 1$. Let us find the condition on n_1 such that 100% of the light energy is reflected. This is referred to as total internal reflection (TIR). From Snell's law $n_1 \sin \theta_1 = n_1/\sqrt{2} = n_2 \sin \theta_2$. Thus $\theta_2 = \pi/2$ results in $n_1 = \sqrt{2}n_2 = \sqrt{2}$ which marks the onset of TIR at this incidence angle.

The Fresnel coefficients at the onset of TIR are $r_s = 1, t_s = 2, r_p = -1, t_p = 2\sqrt{2}$. The t_p and t_s coefficients have the values found from Eqs. (1.30) but they do not imply that energy is transmitted into the second region since $\theta_2 = \pi/2$. There is only an evanescent, exponentially decaying wave beyond the interface. It may appear strange that $t_s, t_p > 1$ but we can verify the energy conservation relations (1.31)

$$\begin{aligned}|r_s|^2 + \frac{n_2 \cos(\theta_2)}{n_1 \cos(\theta_1)} |t_s|^2 &= 1, \\ |r_p|^2 + \frac{n_2 \cos(\theta_2)}{n_1 \cos(\theta_1)} |t_p|^2 &= 1,\end{aligned}$$

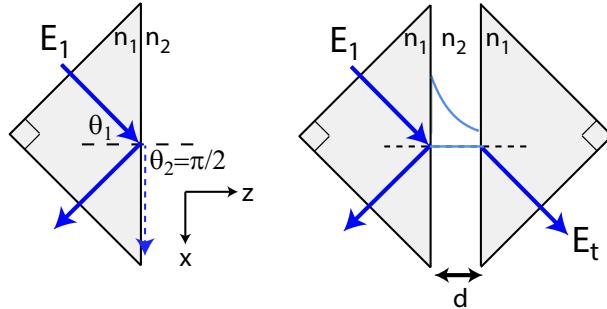


Figure 1.8: (left) TIR at 45 deg. incidence angle with $n_1 > n_2$. (right) Frustrated TIR. The transmitted wave decreases exponentially with increasing gap width d .

which evaluate to

$$1 + 0 = 1.$$

Let's calculate the transmitted field amplitude. For s polarization we have $E_{ts} = t_s E_{is}$ and

$$t_s = \frac{2 \cos(\theta_1) \sin(\theta_2)}{\sin(\theta_1 + \theta_2)}.$$

Putting $n_1 = n$ and $n_2 = 1$ the critical angle at the onset of TIR is $\sin(\theta_c) = 1/n$. When $\theta_1 > \theta_c$ use $\sin(\theta_2) = n \sin(\theta_1) > 1$ which implies that θ_2 is a complex quantity and

$$\cos(\theta_2) = \sqrt{1 - \sin^2(\theta_2)} = i\sqrt{n^2 \sin^2(\theta_1) - 1}.$$

The real and imaginary parts of θ_2 are shown in Fig. 1.9. We see that past the critical angle θ_c the real part is pegged at 90 deg. and the imaginary part increases. This corresponds to stronger and stronger exponential decay of the transmitted field as we will show below.

The Fresnel coefficient for transmission is

$$t_s = \frac{2n \cos(\theta_1) \sin(\theta_1)}{n \cos(\theta_1) \sin(\theta_1) + i \sin(\theta_1) \sqrt{n^2 \sin^2(\theta_1) - 1}} = \frac{2n \cos(\theta_1)}{n \cos(\theta_1) + i \sqrt{n^2 \sin^2(\theta_1) - 1}}.$$

The Poynting vector of the transmitted wave is $\mathbf{S} = \mathbf{E} \times \mathbf{H}$. From Sec. 1.4 we have the expressions

$$\mathbf{E} = -\frac{1}{\epsilon_0 \omega} \mathbf{k}_t \times \mathbf{H}, \quad \mathbf{H} = \frac{1}{\mu_0 \omega} \mathbf{k}_t \times \mathbf{E}$$

where \mathbf{k}_t is the wavevector of the light outside the prism and ω is the frequency. Thus

$$\mathbf{S} = -\frac{1}{\epsilon_0 \omega} (\mathbf{k}_t \times \mathbf{H}) \times \mathbf{H} = \frac{1}{\epsilon_0 \omega} H^2 \mathbf{k}_t.$$

We see that the energy flows in the direction \mathbf{k}_t which is given by

$$\mathbf{k}_t = k_0 \cos(\theta_2) \boldsymbol{\epsilon}_z + k_0 \sin(\theta_2) \boldsymbol{\epsilon}_x = k_0 n \sin(\theta_1) \boldsymbol{\epsilon}_x + i k_0 \sqrt{n^2 \sin^2(\theta_1) - 1} \boldsymbol{\epsilon}_z.$$

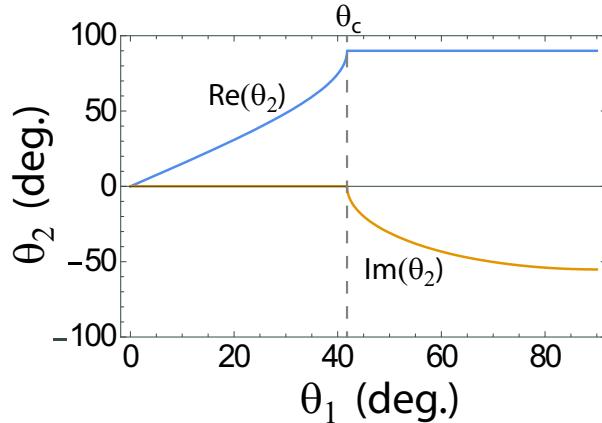


Figure 1.9: Real and imaginary parts of $\theta_2 = \sin^{-1}[n \sin(\theta_1)]$ for $n = 1.5$.

We see that the Poynting vector has a real component along ϵ_x corresponding to the energy propagating along the interface and an imaginary component perpendicular to the interface. The imaginary component corresponds to an exponentially decaying wave. This becomes apparent by looking at the transmitted field

$$\begin{aligned} E_t &\sim e^{-\omega t} e^{ik_0 n \sin(\theta_1)x} e^{ik_0 n \sqrt{n^2 \sin^2(\theta_1) - 1} z} \\ &= e^{-\omega t} e^{ik_0 n \sin(\theta_1)x} e^{-k_0 n \sqrt{n^2 \sin^2(\theta_1) - 1} z} \end{aligned}$$

which decays exponentially along z perpendicular to the interface. Nevertheless, if a second prism is arranged as shown in Fig. 1.8 there will be a transmitted beam with

$$\frac{E_t^2}{E_1^2} \simeq |t_s|^2 = \frac{4n^2 \cos^2(\theta_1)}{n^2 - 1}.$$

This is valid for very small separation d . For larger d there is an additional exponential decay factor.

There are also other curious effects associated with total internal reflection. We saw that the transmitted beam propagates along the edge of the prism. This results in a lateral displacement of the reflected beam. This cannot be seen with plane waves since a lateral displacement of an ideal plane wave does not change the beam. On the other hand for finite beams a lateral displacement is seen. This is known as the Goos-Hänchen effect[7]. A calculation of the size of the shift is best done with Fourier optics so we will defer a discussion until later on. ♠??

An interesting and useful feature of TIR is that the reflected s and p polarization components experience a differential phase shift which can be used to convert the polarization state of the light. We can calculate this phase shift as follows. The critical angle beyond which TIR occurs is $\theta_c = \sin^{-1}(n_2/n_1)$. The Fresnel reflection coefficient for s polarization is from (1.30a)

$$r_s = \frac{n' \cos(\theta_1) - \sqrt{1 - n'^2 \sin^2(\theta_1)}}{n' \cos(\theta_1) + \sqrt{1 - n'^2 \sin^2(\theta_1)}} = \frac{n' \cos(\theta_1) - i\sqrt{n'^2 \sin^2(\theta_1) - 1}}{n' \cos(\theta_1) + i\sqrt{n'^2 \sin^2(\theta_1) - 1}}$$

with $n' = n_1/n_2 > 1$. This is an expression of the form $r_s = (a - ib)/(a + ib) = e^{-i2\tan^{-1}(b/a)}$ which has unit modulus. For p polarization we have

$$r_p = -\frac{\cos(\theta_1) - n' \sqrt{1 - n'^2 \sin^2(\theta_1)}}{\cos(\theta_1) + n' \sqrt{1 - n'^2 \sin^2(\theta_1)}} = -\frac{\cos(\theta_1) - in' \sqrt{n'^2 \sin^2(\theta_1) - 1}}{\cos(\theta_1) + in' \sqrt{n'^2 \sin^2(\theta_1) - 1}}.$$

The relative phase shift of the p and s components is thus

$$\delta\phi = \phi_p - \phi_s = -2 \left[\tan^{-1} \left(\frac{n' \sqrt{n'^2 \sin^2(\theta_1) - 1}}{\cos(\theta_1)} \right) - \tan^{-1} \left(\frac{\sqrt{n'^2 \sin^2(\theta_1) - 1}}{n' \cos(\theta_1)} \right) \right].$$

Let's look at this for a glass prism in air so $n = 1/1.5$ and $n' = 1/n = 1.5$. Figure 1.10 shows the differential phase as a function of incidence angle. We see that for $\theta_1 \sim 54$ deg. the phase shift is very close to $\pi/4$ and changes by only a few degrees for changes in the refractive index which correspond to hundreds of nm worth of dispersion in glass. Thus two internal bounces will give a $\pi/2$ shift which will convert linearly polarized light to circular. This forms the basis for an achromatic polarization conversion device, known as the Fresnel rhomb[8].

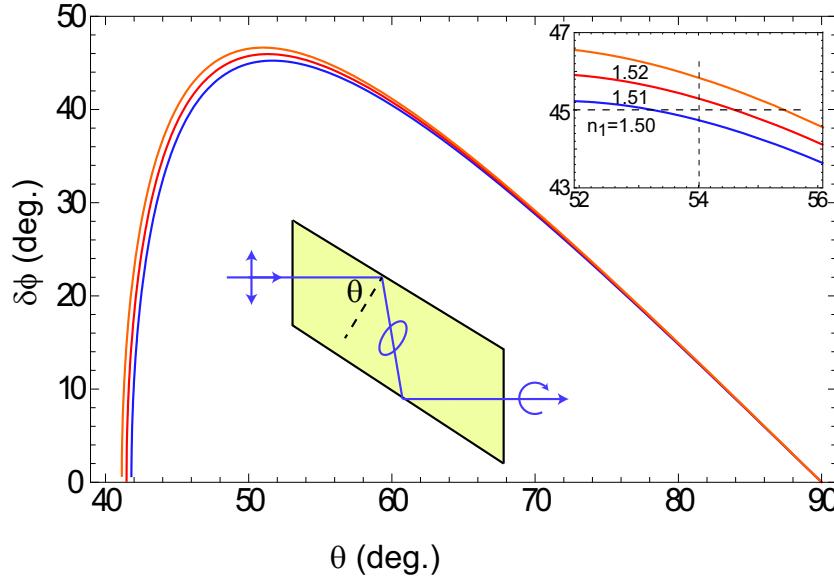


Figure 1.10: Phase shift in TIR for $n = 1.50, 1.51, 1.52$.

1.14 Reflection from metals

An ideal metallic conductor is a perfect reflector since the electric field vanishes inside the conductor. For real metals with finite conductivity the reflectivity is less than unity and is accompanied by a phase shift.

Let us write the refractive index as $n = n_r + i\kappa$. A plane wave propagates as

$$\mathcal{E}(z) \sim e^{ikz} = e^{ik_0 nz} = e^{ik_0 n_r z} e^{-k_0 \kappa z}.$$

We see that the attenuation depends on κ leading to an exponential intensity loss $I(z) = I(0)e^{-\alpha z}$ with absorption coefficient $\alpha = 2k_0\kappa$.

Snell's law and the expressions for Fresnel reflection coefficients remain valid with complex \tilde{n} . Equations (1.30) take the form

$$r_s = \frac{\cos(\theta_1) - \sqrt{n^2 - \sin^2(\theta_1)}}{\cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}} \quad (1.37a)$$

$$t_s = \frac{2 \cos(\theta_1)}{\cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}} \quad (1.37b)$$

$$r_p = \frac{n^2 \cos(\theta_1) - \sqrt{n^2 - \sin^2(\theta_1)}}{n^2 \cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}} \quad (1.37c)$$

$$t_p = \frac{2n \cos(\theta_1)}{n^2 \cos(\theta_1) + \sqrt{n^2 - \sin^2(\theta_1)}}. \quad (1.37d)$$

There is still a Brewster's angle with a minimum in the reflectivity for p-polarization, although the reflectivity does not go to zero.

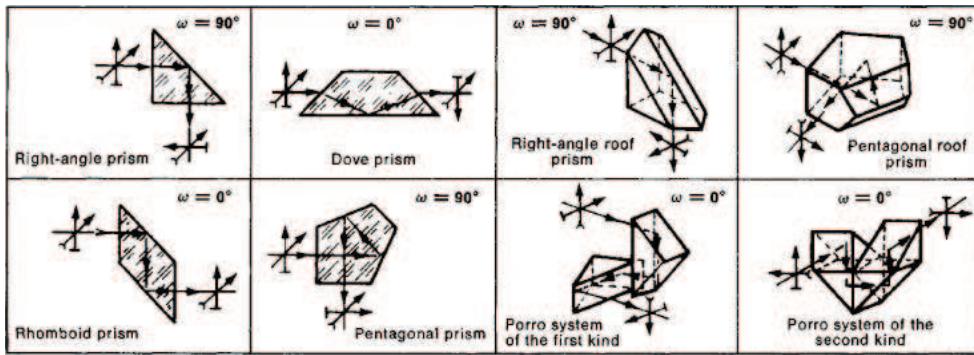


Figure 1.11: Some different prisms. Image from Wikipedia.

Let's evaluate the reflectivity for the simplest case of normal incidence, $\theta_1 = 0$. We have

$$r_s = \frac{1 - \sqrt{n^2}}{1 + \sqrt{n^2}} = \frac{1 - n}{1 + n} \quad (1.38a)$$

$$r_p = -\frac{n^2 - \sqrt{n^2}}{n^2 + \sqrt{n^2}} = -\frac{n - 1}{n + 1} = \frac{1 - n}{1 + n}. \quad (1.38b)$$

We see $r_s = r_p$ as expected at normal incidence. We can separate the real and imaginary parts as

$$r_s = \frac{(1 - n)(1 + n^*)}{(1 + n)(1 + n^*)} = \frac{1 - |n|^2 - 2i\text{Im}(n)}{1 + |n|^2 + 2\text{Re}(n)} = \frac{1 - n_r^2 - \kappa^2 - 2i\kappa}{1 + n_r^2 + \kappa^2 + 2n_r}.$$

Silver at visible wavelengths has $n_r = 0.13$ and $\kappa = 4.05$ which gives $|r_s|^2 = 0.971$ and $\arg(r_s) = -2.66 = -152$. deg. The attenuation coefficient at $\lambda = 0.5 \mu$ is

$$\alpha = 2k_0\kappa = 1.0 \times 10^8 \text{ m}^{-1} = 100 \text{ } \mu\text{m}^{-1}.$$

The intensity is attenuated by a factor of $e^{-100} = 6. \times 10^{-45}$ at a distance of $1 \mu\text{m}$ below the surface.

1.15 Prisms

Another category of optical component is the prism. There are many different types of prisms with different functionalities. They can be used as mirrors as retroreflectors, as wavelength dispersing elements, and as image rotators.

1.16 Calcite beam displacer

The calcite beam displacer is a birefringent component that splits an incident beam into two parallel beams, one of which is p polarized and one of which is s polarized. The s polarized component is undisplaced while the p component receives a transverse displacement d . For a displacer of length L , and internal angle α the transverse displacement is

$$d = L \tan \alpha = L \frac{(n_e^2 - n_0^2) \tan \theta}{n_e^2 + n_0^2 \tan^2 \theta}$$

where θ is the angle of the optical axis from the crystal end plane.

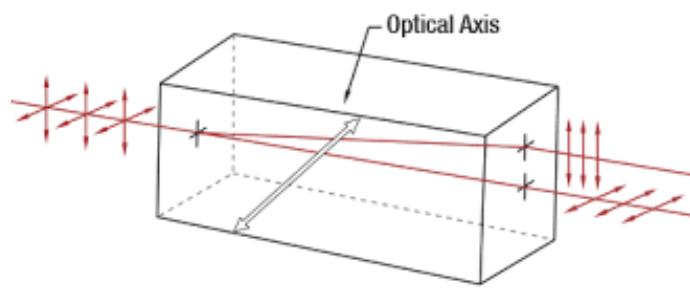


Figure 1.12: Calcite beam displacer. Image from Thorlabs.

Chapter 2

Coherence and Noise

2.1 Broadband fields

When the optical field consists of a continuum of frequencies with a spectral width that is not small compared to the mean frequency a description in terms of a single carrier frequency ω as in Eqs. (1.13) is no longer convenient. Instead we allow for a continuum of frequencies with each frequency having a spectral amplitude $\tilde{\mathcal{E}}(\omega)$. The time domain field is then given by the Fourier representation

$$\mathcal{E}(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \tilde{\mathcal{E}}(\omega) e^{i[k(\omega)z - \omega t]} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \tilde{\mathcal{E}}(\omega) e^{i\frac{\omega}{c}(z - ct)}.$$

Fourier analysis is reviewed in Appendix A. In the second equality we have used $k(\omega) = \omega/c$ which is valid for a nondispersive medium. The spectral amplitude of the field is

$$\tilde{\mathcal{E}}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \mathcal{E}(z, t) e^{-i\frac{\omega}{c}(z - ct)}.$$

The spectral amplitude has units of V × s/m. The spectral intensity is

$$\tilde{I}(\omega) = \frac{\epsilon_0 c}{2} |\tilde{\mathcal{E}}(\omega)|^2.$$

The spectral intensity has units of (W/m²) × s².

The real electric field is

$$\begin{aligned} \mathcal{E}(z, t) &= \frac{\mathcal{E}(z, t)}{2} + \frac{\mathcal{E}^*(z, t)}{2} \\ &= \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \tilde{\mathcal{E}}(\omega) e^{i\frac{\omega}{c}(z - ct)} + \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \tilde{\mathcal{E}}^*(\omega) e^{-i\frac{\omega}{c}(z - ct)}. \end{aligned}$$

It is easily verified that if $\mathcal{E}(z, t)$ is real then $E(z, t) = \mathcal{E}(z, t)$ and

$$\tilde{\mathcal{E}}(\omega) = \tilde{\mathcal{E}}^*(-\omega).$$

Using this relation we can write the real field as a single sided integral

$$E(z, t) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} d\omega \left[\tilde{\mathcal{E}}(\omega) e^{i\frac{\omega}{c}(z - ct)} + \tilde{\mathcal{E}}^*(\omega) e^{-i\frac{\omega}{c}(z - ct)} \right].$$

This shows that it is sufficient to represent $E(z, t)$ using only positive frequency spectral amplitudes.

The time averaged intensity is

$$\begin{aligned}\langle I \rangle_T &= \frac{1}{T} \frac{\epsilon_0 c}{2} \int_{-T/2}^{T/2} d\tau \mathcal{E}(z, t + \tau) \mathcal{E}^*(z, t + \tau) \\ &= \frac{1}{T} \frac{\epsilon_0 c}{2} \int_{-T/2+t}^{T/2+t} d\tau \mathcal{E}(z, \tau) \mathcal{E}^*(z, \tau) \\ &= \frac{1}{T} \frac{\epsilon_0 c}{2} \frac{1}{2\pi} \int_{-T/2+t}^{T/2+t} d\tau \int_0^\infty d\omega \tilde{\mathcal{E}}(\omega) e^{i\frac{\omega}{c}(z-c\tau)} \int_0^\infty d\omega' \tilde{\mathcal{E}}^*(\omega') e^{-i\frac{\omega'}{c}(z-c\tau)} \\ &= \frac{1}{T} \frac{\epsilon_0 c}{2} \int_0^\infty d\omega \int_0^\infty d\omega' \tilde{\mathcal{E}}(\omega) \tilde{\mathcal{E}}^*(\omega') e^{i(\omega-\omega')z/c} \frac{1}{2\pi} \int_{-T/2+t}^{T/2+t} d\tau e^{-i(\omega-\omega')\tau}.\end{aligned}$$

To proceed we assume the intensity is statistically stationary so that $\langle I \rangle_T$ is independent of t , and put $t = 0$. The integral over τ is written as

$$\frac{1}{2\pi} \int_{-T/2}^{T/2} d\tau e^{-i(\omega-\omega')\tau} = \delta_T(\omega - \omega').$$

When the spectral support is band limited, i.e. $\mathcal{E}(\omega) \simeq 0$ for $|\omega| > \omega_{\max}$ and $T \gg 1/\omega_{\max}$ we can make the approximation $\delta_T(\omega - \omega') \simeq \delta(\omega - \omega')$ to arrive at

$$\langle I \rangle_T = \frac{1}{T} \frac{\epsilon_0 c}{2} \int_0^\infty d\omega |\tilde{\mathcal{E}}(\omega)|^2.$$

For sufficiently long averaging times we define an average integrated intensity

$$\bar{I} = \lim_{T \rightarrow \infty} (T \langle I \rangle_T) = \frac{\epsilon_0 c}{2} \int_0^\infty d\omega |\tilde{\mathcal{E}}(\omega)|^2.$$

The units of \bar{I} are $(\text{W}/\text{m}^2) \times \text{s}$. We then introduce a normalized spectral intensity $S(\omega)$ by

$$S(\omega) = \frac{\tilde{I}(\omega)}{\bar{I}} \quad [\text{s}]. \quad (2.1)$$

The normalized spectral intensity integrated over all frequencies is

$$\int_0^\infty d\omega S(\omega) = 1.$$

With these definitions the frequency dependent field amplitude has units of $[\tilde{\mathcal{E}}(\omega)] = \text{V}/\text{Hz}^{1/2}\text{m}$ and for the spectral intensity $[S(\omega)] = 1/\text{Hz}$.

2.2 Temporal coherence

Consider a monochromatic scalar wave $E = E_0 \cos(\omega t)$ with real amplitude E_0 and frequency ω . At time $t = \tau$ the average or expected value of the field is

$$\langle E(\tau) \rangle = \langle E_0 \cos(\omega\tau) \rangle.$$

The angular brackets indicate a time or ensemble average over the parameters E_0 and ω . If these parameters are fixed we find immediately that

$$\langle E(\tau) \rangle = E_0 \cos(\omega\tau)$$

so the field has an amplitude that ranges between $-E_0$ and E_0 no matter how long the time τ . Such a field has perfect temporal coherence.

Real optical fields do not have perfect coherence. They have amplitude, frequency, and phase noise. Let's look more closely at the case of phase noise assuming the amplitude and frequency are fixed parameters. We take the field to be

$$E = E_0 \cos[\omega t + \phi(t)], \quad (2.2)$$

with ω a fixed carrier frequency and $\phi(t)$ a time dependent phase. The expectation value of the field at time τ is

$$\begin{aligned} \langle E(\tau) \rangle &= E_0 \langle \cos(\omega\tau) \cos(\phi(\tau)) - \sin(\omega\tau) \sin(\phi(\tau)) \rangle \\ &= E_0 \cos(\omega\tau) \langle \cos(\phi(\tau)) \rangle - E_0 \sin(\omega\tau) \langle \sin(\phi(\tau)) \rangle \end{aligned} \quad (2.3)$$

If $\phi(\tau)$ is a uniformly distributed random variable then $\langle \cos(\phi(\tau)) \rangle = \langle \sin(\phi(\tau)) \rangle = 0$ and $\langle E(\tau) \rangle = 0$. This is fully incoherent light that has no coherence. The typical situation is that the light has some coherence, neither zero nor perfect, which is called partial coherence.

2.3 Frequency and Phase noise

Light is often used to perform measurements of physical quantities. The precision and accuracy of measurements depend, among other things, on the noise characteristics of the light source. In this section we will study frequency and phase noise of quasimonochromatic light, such as that emitted by a laser.

As the starting point for our discussion we assume a phase modulated field as in Eq. (2.2) but written in the form

$$E(t) = E_0 \cos[2\pi\nu_0 t + \phi(t)], \quad (2.4)$$

with ν_0 the fixed carrier frequency in Hz, and E_0 the real amplitude. The complex representation of this field is

$$E(t) = \frac{\mathcal{E}_0}{2} e^{i[2\pi\nu_0 t + \phi(t)]} + \frac{\mathcal{E}_0^*}{2} e^{-i[2\pi\nu_0 t + \phi(t)]}. \quad (2.5)$$

The so-called analytic signal corresponding to the field E is defined as

$$V(t) = \mathcal{E}_0 e^{i[2\pi\nu_0 t + \phi(t)]}.$$

The autocorrelation function of the analytic signal is

$$\begin{aligned} R(\tau) &= \frac{1}{2} \langle V(t) V^*(t + \tau) \rangle \\ &= \frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} dt V(t) V^*(t + \tau) \\ &= \frac{|\mathcal{E}_0|^2}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} dt e^{i[\phi(t) - \phi(t+\tau)]}. \end{aligned}$$

The actual form of the autocorrelation depends on $\phi(t)$.

Gaussian noise is an important case that can be dealt with analytically. If $\phi(t)$ is determined by a Gaussian probability distribution then the odd order moments of ϕ vanish and the even order moments are all related to the 2nd order moment by

$$\langle [\phi(t) - \phi(t + \tau)]^{2n} \rangle = \frac{(2n)!}{2^n n!} \langle [\phi(t) - \phi(t + \tau)]^2 \rangle^n$$

so

$$\begin{aligned} R(\tau) &= \frac{|\mathcal{E}_0|^2}{2} \left\langle \sum_{n=0}^{\infty} \frac{i^n [\phi(t) - \phi(t + \tau)]^n}{n!} \right\rangle \\ &= \frac{|\mathcal{E}_0|^2}{2} \sum_{n=0}^{\infty} \frac{i^{2n} \langle [\phi(t) - \phi(t + \tau)]^{2n} \rangle}{(2n)!} \\ &= \frac{|\mathcal{E}_0|^2}{2} \sum_{n=0}^{\infty} \frac{(-1)^n \frac{(2n)!}{2^n n!} \langle [\phi(t) - \phi(t + \tau)]^2 \rangle^n}{(2n)!} \\ &= \frac{|\mathcal{E}_0|^2}{2} e^{-\frac{\langle [\phi(t) - \phi(t + \tau)]^2 \rangle}{2}} \\ &= \frac{|\mathcal{E}_0|^2}{2} e^{-\frac{\langle \Delta \phi^2(\tau) \rangle}{2}}. \end{aligned}$$

The second moment of the phase fluctuations can be expressed in terms of the power spectrum of frequency noise as

$$\dots \quad (2.6)$$

2.4 Degree of temporal coherence

The degree of coherence can be measured with a Michelson interferometer. We will show that if the path length difference of the arms of the interferometer is L , such that the differential time delay is $\tau = 2L/c$, then the time integrated output intensity is

$$S_{\text{out}}(\tau) = \int_{-\infty}^{\infty} dt I_{\text{out}}(t) = 2S [1 + \text{Re}[\gamma(\tau)]] \quad (2.7)$$

where $S = \int_{-\infty}^{\infty} dt I(t)$ is the time integrated input intensity and the complex degree of coherence is

$$\gamma(\tau) = \frac{\int_{-\infty}^{\infty} d\omega \tilde{I}(\omega) e^{-i\omega\tau}}{\int_{-\infty}^{\infty} d\omega \tilde{I}(\omega)}$$

with $\tilde{I}(\omega) = \frac{\epsilon_0 c}{2} |\tilde{\mathcal{E}}(\omega)|^2$ the spectral intensity. This result is derived in the P & W book p.203.

The coherence time is the characteristic time after which the degree of coherence substantially decays. A precise definition is

$$\tau_c = \int_{-\infty}^{\infty} d\tau |\gamma(\tau)|^2 = 2 \int_0^{\infty} d\tau |\gamma(\tau)|^2.$$

The coherence length is the distance traveled by light in a coherence time and is given by $l_c = c\tau_c$.

The temporal coherence can be measured using a scanning Michelson interferometer as described in Sec. 3.5. As the path length difference of the interferometer arms is scanned the signal $S_{\text{out}}(\tau)$ will vary. For small changes of τ , near any particular value, maxima and minima of $S_{\text{out}}(\tau)$ are observed. The visibility of the signal is defined as

$$V(\tau) = \frac{S_{\text{out},\text{max}}(\tau) - S_{\text{out},\text{min}}(\tau)}{S_{\text{out},\text{max}}(\tau) + S_{\text{out},\text{min}}(\tau)}.$$

It is not hard to show that $V(\tau) = |\gamma(\tau)|$.

As an example consider a Gaussian pulse with spectrum

$$\tilde{I}(\omega) = \frac{\epsilon_0 c}{2} |\mathcal{E}_0|^2 T^2 e^{-T^2(\omega-\omega_0)^2}.$$

Here ω_0 is the central, or carrier, frequency of the light, T is the temporal width, and \mathcal{E}_0 is the peak amplitude. A short calculation shows that

$$\gamma(\tau) = e^{-\tau^2/(4T^2)} e^{-i\omega_0\tau}$$

so

$$V(\tau) = e^{-\tau^2/(4T^2)}.$$

The coherence time is

$$\tau_c = \int_{-\infty}^{\infty} d\tau |\gamma(\tau)|^2 = \sqrt{2\pi} T.$$

Another example is white light with $\tilde{I}(\omega) = I_0$, independent of frequency. Although it may not be feasible to produce such a spectrum extending to all frequencies a blackbody source approximates this over a finite frequency range. The degree of coherence is

$$\gamma(\tau) = \frac{\int_{-\infty}^{\infty} d\omega \tilde{I}(\omega) e^{-i\omega\tau}}{\int_{-\infty}^{\infty} d\omega \tilde{I}(\omega)} = \frac{2\pi\delta(\tau)}{2\pi\delta(0)} = \frac{\delta(\tau)}{\delta(0)}.$$

We see that $\gamma(0) = 1$ and $\gamma(\tau \neq 0) = 0$. White light has no temporal coherence. This is called incoherent light.

At the opposite extreme is a monochromatic source $\tilde{I}(\omega) = I_0\delta(\omega - \omega_0)$. We find $\gamma(\tau) = e^{-i\omega_0\tau}$. The magnitude of the coherence is $|\gamma(\tau)| = 1$ at all delay times and there is perfect temporal coherence.

2.5 Spatial coherence

Light may have temporal coherence and be spatially incoherent or coherent. Consider a thermal source, such as an incandescent light bulb, that is spectrally filtered by a Fabry-Pérot resonator, as shown in Fig. 2.1. If the spectral width of the source is smaller than the free spectral range of the resonator a single frequency will be transmitted. The light is temporally coherent, but spatially incoherent since the fields due to different points on the source filament do not have a stable relative phase.

Spatial coherence can be characterized using a two-slit interferometer as shown in Fig. 2.2. A source point x' gives fields at observation point x

$$\begin{aligned}\mathcal{E}_1(x, x') &= \mathcal{E}_0(x') e^{ik(r_1+d_1)} e^{-i\omega t} \\ \mathcal{E}_2(x, x') &= \mathcal{E}_0(x') e^{ik(r_2+d_2)} e^{-i\omega t}.\end{aligned}$$

Here subscripts 1, 2 refer to light that passes through the upper or lower opening in the screen and \mathcal{E}_0 is the field amplitude at x' . We can characterize an extended source by dividing it up into small regions x'_j so the total field at point x is

$$\mathcal{E}(x) = \sum_j [\mathcal{E}_1(x, x'_j) + \mathcal{E}_2(x, x'_j)].$$

The intensity is $I(x) = \frac{\epsilon_0 c}{2} |\mathcal{E}(x)|^2$ and

$$|\mathcal{E}(x)|^2 = \sum_{m,n} [\mathcal{E}_{1m} \mathcal{E}_{1n}^* + \mathcal{E}_{2m} \mathcal{E}_{2n}^* + 2\text{Re}(\mathcal{E}_{1m} \mathcal{E}_{2n}^*)].$$

We have used a short hand notation $\mathcal{E}_{jm} = \mathcal{E}_j(x, x'_m)$. The first two terms give

$$\begin{aligned}\sum_{m,n} \mathcal{E}_{1m} \mathcal{E}_{1n}^* &= \sum_{m,n} \mathcal{E}_0(x'_m) \mathcal{E}_0^*(x'_n) e^{ik[r_1(x'_m) - r_1(x'_n)]}, \\ \sum_{m,n} \mathcal{E}_{2m} \mathcal{E}_{2n}^* &= \sum_{m,n} \mathcal{E}_0(x'_m) \mathcal{E}_0^*(x'_n) e^{ik[r_2(x'_m) - r_2(x'_n)]}.\end{aligned}$$

To proceed we assume the source plane exhibits partial phase coherence such that $\mathcal{E}_0(x'_m) = \mathcal{E}_0 e^{i\phi_m}$. If ϕ_m is a random, uniformly distributed phase then $\sum_{m,n \neq m} \rightarrow 0$ for the proceeding two equations. However,

$$\sum_{m,n=m} \mathcal{E}_{1m} \mathcal{E}_{1m}^* = \sum_m |\mathcal{E}_0(x'_m)|^2 = \frac{1}{\frac{\epsilon_0 c}{2}} \sum_m I_m,$$



Figure 2.1: Spectral filtering of an extended thermal source prepares an optical field with temporal coherence that is spatially incoherent.

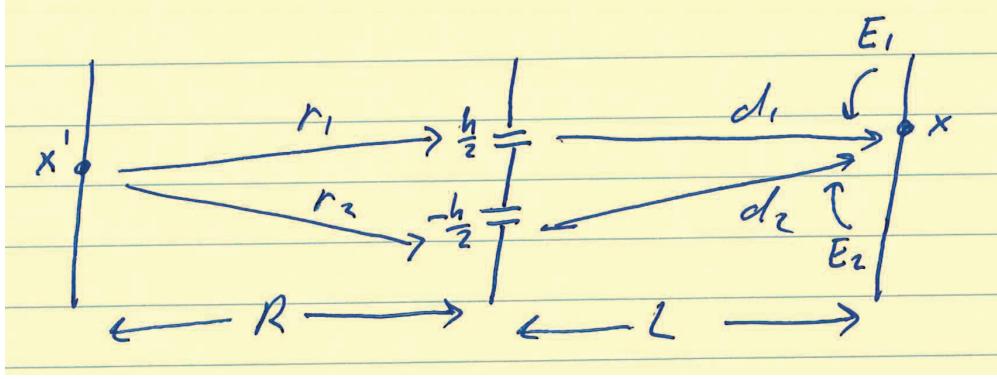


Figure 2.2: Two-slit interferometer for characterizing spatial coherence.

where $I_m = I(x'_m)$ is the source intensity at x'_m . Therefore the intensity at observation point x can be written as

$$\begin{aligned}\langle I \rangle &= \left\langle \frac{\epsilon_0 c}{2} |\mathcal{E}(x)|^2 \right\rangle \\ &= 2 \sum_m I_m + 2 \operatorname{Re} \left[\sum_m I_m e^{ik[r_1(x'_m) - r_2(x'_m)]} e^{ik(d_1 - d_2)} \right].\end{aligned}$$

When $R \gg x'_m, h$ and $L \gg x, h$ we are in the paraxial limit of small angles and we can write $r_1(x'_m) - r_2(x'_m) \simeq -hx'_m/R$, $d_1 - d_2 \simeq -hx/L$. This results in

$$\begin{aligned}\langle I \rangle &= 2 \sum_m I_m + 2 \operatorname{Re} \left(\sum_m I_m e^{-ikhx'_m/R} e^{-ikhx/L} \right) \\ &= \left(2 \sum_m I_m \right) \left[1 + \operatorname{Re} \left(\frac{\sum_m I_m e^{-ikhx'_m/R} e^{-ikhx/L}}{\sum_m I_m} \right) \right].\end{aligned}$$

We can rewrite this in a form analogous to Eq. (2.7) for the temporal coherence as

$$\langle I \rangle = 2 \sum_m I_m [1 + \operatorname{Re}(\gamma(h))]$$

with

$$\gamma(h) = \frac{\sum_m I_m e^{-ikhx'_m/R}}{\sum_m I_m} e^{-ikhx/L}.$$

We have written the result in terms of a discrete sum over source elements. For a continuous source this becomes

$$\langle I \rangle = \left[2 \int dx' I(x') \right] [1 + \operatorname{Re}(\gamma(h))]$$

with

$$\gamma(h) = \frac{\int dx' I(x') e^{-ikhx'/R}}{\int dx' I(x')} e^{-ikhx/L}.$$

Measurement of the spatial degree of coherence was used by Michelson and Pease in 1921 to measure the diameter of a star[9]. To do so they used the result that for a uniform intensity circular disk which subtends an angle θ at the observation plane $\gamma(h) = 0$ when

$$h = 1.22 \frac{\lambda}{\theta}$$

where λ is the wavelength of the light. Then using $\theta = D/R$, with R the known distance to the star, the diameter D can be determined from measurements of $\gamma(h)$.

Chapter 3

Interference and Interferometers

The coherence properties of optical fields can be studied using interferometers which rely on interference between light waves. When observing interference we usually use an intensity detector such as photographic film, or a CCD camera, or our eyes. The intensity pattern shows the vector interference so if $\mathbf{E}_1 = E_{1x}\boldsymbol{\epsilon}_x + E_{1y}\boldsymbol{\epsilon}_y$, $\mathbf{E}_2 = E_{2x}\boldsymbol{\epsilon}_x + E_{2y}\boldsymbol{\epsilon}_y$, the total field is $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ and the intensity is

$$I \sim \mathbf{E} \cdot \mathbf{E} = (E_{1x} + E_{2x})^2 + (E_{1y} + E_{2y})^2.$$

In terms of the complex amplitudes introduced earlier $\mathbf{E}_1 = (\frac{\mathcal{E}_1}{2}\boldsymbol{\epsilon}_1 + c.c.)$, $\mathbf{E}_2 = (\frac{\mathcal{E}_2}{2}\boldsymbol{\epsilon}_2 + c.c.)$, $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2$ are unit polarization vectors, possibly complex, and the intensity is

$$\begin{aligned} I &= \frac{\epsilon_0 c}{2} |\mathcal{E}_1 \boldsymbol{\epsilon}_1 + \mathcal{E}_2 \boldsymbol{\epsilon}_2|^2 \\ &= \frac{\epsilon_0 c}{2} (|\mathcal{E}_1|^2 + |\mathcal{E}_2|^2) + \frac{\epsilon_0 c}{2} (\mathcal{E}_1 \mathcal{E}_2^* \boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2^* + \mathcal{E}_1^* \mathcal{E}_2 \boldsymbol{\epsilon}_1^* \cdot \boldsymbol{\epsilon}_2). \end{aligned}$$

Here we are suppressing the explicit time and space dependence of the amplitudes $\mathcal{E}_1, \mathcal{E}_2$.

If the two fields are orthogonally polarized, $\boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2^* = 0$, then the interference term, which depends on the relative phases of the fields, vanishes. When the fields have some degree of parallel polarization there will be nonzero interference. Consider the case of perfectly parallel polarization, $\boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2^* = 1$, and put $\mathcal{E}_2 = \alpha e^{i\phi} \mathcal{E}_1$ with $0 \leq \alpha \leq 1$ and ϕ a relative phase. The intensity is then

$$\begin{aligned} I &= \frac{\epsilon_0 c}{2} (|\mathcal{E}_1|^2 + |\mathcal{E}_2|^2) + \frac{\epsilon_0 c}{2} (\mathcal{E}_1 \mathcal{E}_2^* + \mathcal{E}_1^* \mathcal{E}_2) \\ &= \frac{\epsilon_0 c}{2} |\mathcal{E}_1|^2 (1 + \alpha^2) + \epsilon_0 c |\mathcal{E}_1|^2 \alpha \cos(\phi) \\ &= \frac{\epsilon_0 c}{2} |\mathcal{E}_1|^2 [1 + \alpha^2 + 2\alpha \cos(\phi)]. \end{aligned}$$

The intensity ranges from a maximum to a minimum value depending on the phase ϕ . It is useful to characterize the interference by the visibility or fringe contrast, defined as

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\alpha}{1 + \alpha^2}.$$

This simple expression has some interesting consequences. The relative intensity of the two fields is α^2 . Even if the second field is very weak, say $\alpha^2 = 0.01$, the interference visibility is

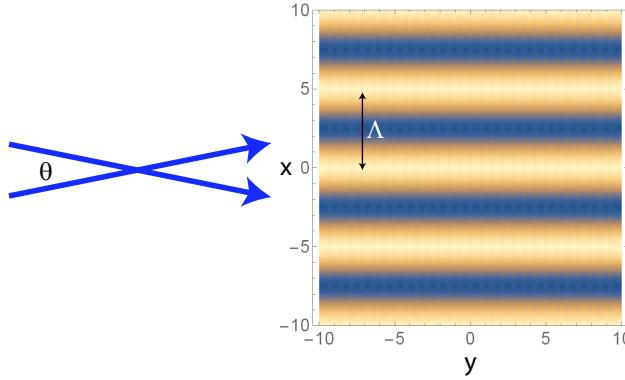


Figure 3.1: Interference of two plane waves with $\lambda = 1$, $\theta = 11.48$ deg. and $\Lambda = 5$.

0.198. In other words a 1% intensity effect leads to $\sim 20\%$ visibility. This is a useful result if the goal is to detect a weak beam using an interference effect. Conversely it is annoying if we are trying to get rid of interference fringes in an imaging system.

3.1 Interference of optical waves

Let's continue by considering several basic geometries where interference occurs.

3.1.1 Plane wave interference

Consider two fields propagating in the $x-z$ plane at angles $\pm\theta/2$ from the ϵ_z axis. Explicitly $\mathbf{k}_{1,2} = k[\pm \sin(\theta/2)\epsilon_x + \cos(\theta/2)\epsilon_z]$ with $k = 2\pi/\lambda$. We will take the fields to be polarized along ϵ_y which makes this effectively a scalar problem. The total electric field is

$$\mathcal{E} \sim \mathcal{E}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} + \mathcal{E}_2 e^{i\mathbf{k}_2 \cdot \mathbf{r}} \sim \mathcal{E}_1 e^{ik[\sin(\theta/2)x + \cos(\theta/2)z]} + \mathcal{E}_2 e^{ik[-\sin(\theta/2)x + \cos(\theta/2)z]}.$$

At a fixed longitudinal position, say $z = 0$, the intensity is

$$\begin{aligned} I &\sim |\mathcal{E}_1 e^{ik \sin(\theta/2)x} + \mathcal{E}_2 e^{-ik \sin(\theta/2)x}|^2 \\ &= |\mathcal{E}_1|^2 + |\mathcal{E}_2|^2 + \mathcal{E}_1 \mathcal{E}_2^* e^{i2k \sin(\theta/2)x} + \mathcal{E}_1^* \mathcal{E}_2 e^{-i2k \sin(\theta/2)x}. \end{aligned}$$

Introduce a relative phase of the amplitudes by $\mathcal{E}_1 \mathcal{E}_2^* = |\mathcal{E}_1| |\mathcal{E}_2| e^{i\phi}$ so

$$I \sim |\mathcal{E}_1|^2 + |\mathcal{E}_2|^2 + 2|\mathcal{E}_1||\mathcal{E}_2| \cos[2k \sin(\theta/2)x + \phi].$$

Maxima in the intensity are separated by the period Λ with

$$\Lambda = \frac{\lambda}{2 \sin(\theta/2)} \sim \frac{\lambda}{\theta},$$

where the last approximation holds for small θ . The interference pattern is shown in Fig. 3.1.

3.1.2 Interference of plane and spherical waves

An outgoing spherical wave has electric field

$$\mathbf{E}(\mathbf{r}, t) = E \frac{\cos(\mathbf{k} \cdot \mathbf{r} - \omega t)}{f(r)} \boldsymbol{\epsilon}(\mathbf{k}).$$

The polarization $\boldsymbol{\epsilon}(\mathbf{k})$ depends on the propagation direction \mathbf{k} . Energy conservation for a wave expanding into 4π steradians requires that $f(r) = |\mathbf{r}| = r$. Using complex amplitudes the wave is

$$\mathbf{E}(\mathbf{r}, t) = \frac{\mathcal{E}}{2} \frac{e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}}{r} \boldsymbol{\epsilon}(\mathbf{k}) + c.c. .$$

Note that the symbol \mathcal{E} now has units of field times length. For an outgoing spherical wave we define \mathbf{k} and \mathbf{r} with respect to the same origin so $\mathbf{k} \cdot \mathbf{r} = kr$ and

$$\mathbf{E}(\mathbf{r}, t) = \frac{\mathcal{E}}{2} \frac{e^{i(kr - \omega t)}}{r} \boldsymbol{\epsilon}(\mathbf{k}) + c.c. .$$

The interference of a plane wave $\mathbf{E}_1 = \frac{\mathcal{E}_1}{2} e^{i(\mathbf{k}_1 \cdot \mathbf{r})} \boldsymbol{\epsilon}_1 + c.c.$ and a spherical wave \mathbf{E}_2 gives an intensity

$$I \sim |\mathcal{E}_1|^2 + \frac{|\mathcal{E}_2|^2}{r^2} + \left[\frac{\mathcal{E}_1 \mathcal{E}_2^*}{r} e^{i(\mathbf{k}_1 \cdot \mathbf{r} - kr)} \boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2^* + c.c. \right].$$

Let the plane wave propagate in the $x - z$ plane so $\mathbf{k}_1 = k[\sin(\theta/2)\boldsymbol{\epsilon}_x + \cos(\theta/2)\boldsymbol{\epsilon}_z]$, $r = \sqrt{x^2 + y^2 + z^2}$ and

$$I \sim |\mathcal{E}_1|^2 + \frac{|\mathcal{E}_2|^2}{r^2} + \left[\frac{\mathcal{E}_1 \mathcal{E}_2^*}{r} e^{ik[\sin(\theta/2)x + \cos(\theta/2)z - \sqrt{x^2 + y^2 + z^2}]} \boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2^* + c.c. \right].$$

At a fixed longitudinal position, say $z = L$, the intensity is

$$I \sim |\mathcal{E}_1|^2 + \frac{|\mathcal{E}_2|^2}{x^2 + y^2 + L^2} + \left[\frac{\mathcal{E}_1 \mathcal{E}_2^*}{\sqrt{x^2 + y^2 + L^2}} e^{ik[\sin(\theta/2)x + \cos(\theta/2)L - \sqrt{x^2 + y^2 + L^2}]} \boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2^* + c.c. \right].$$

We can simplify this by assuming a small angular range so that $\boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2^*(\mathbf{k}_2) \sim \text{constant}$, and set the constant equal to unity. A small angular range implies that $x, y \ll L$. Expanding to second order in x and y we get

$$I \sim |\mathcal{E}_1|^2 + \frac{|\mathcal{E}_2|^2}{L^2} + \left[\frac{\mathcal{E}_1 \mathcal{E}_2^*}{L} e^{ik[\cos(\theta/2) - 1]L} e^{ik[\sin(\theta/2)x - \frac{x^2 + y^2}{2L}]} + c.c. \right].$$

We have neglected the dependence of the amplitude, but not the phase, on x^2 and y^2 . This is our first encounter of what we will refer to as a paraxial approximation. The intensity is shown in Fig. 3.2 and takes the form of a bull's eye pattern centered in the direction θ and with the fringe period decreasing away from the center. Notice that the center of the bull's eye is not a maximum of intensity.

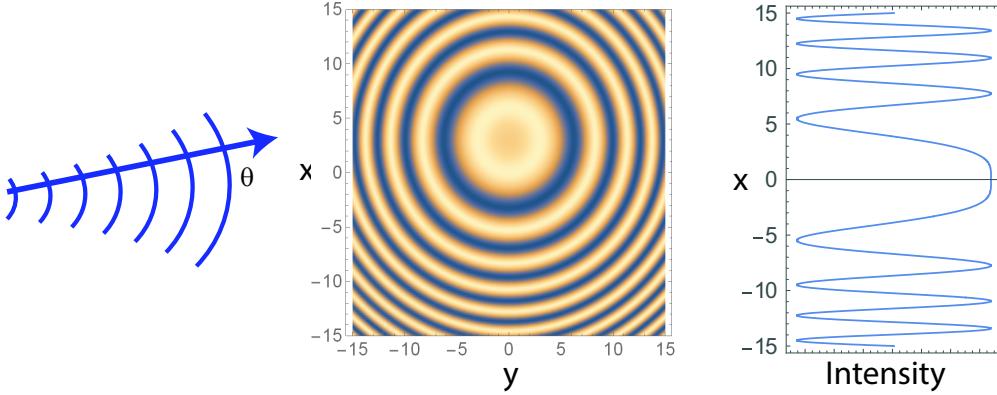


Figure 3.2: Interference of a plane wave and a spherical wave with $\lambda = 1$, $L = 30$, and $\theta = 11.48$ deg. The right hand plot shows the intensity as a function of x for $y = 0$ and $\theta = 0$.

It is instructive to measure x along the direction perpendicular to \mathbf{k}_1 which corresponds to putting $\theta = 0$. The intensity simplifies to

$$I \sim |\mathcal{E}_1|^2 + \frac{|\mathcal{E}_2|^2}{L^2} + \left[\frac{\mathcal{E}_1 \mathcal{E}_2^*}{L} e^{-i\frac{kx^2}{2L}} + c.c. \right]$$

which is shown in the right hand plot in Fig. 3.2. The interference term is now proportional to $\cos(kx^2/2L)$. Neighboring intensity maxima are found for

$$\frac{kx_2^2}{2L} - \frac{kx_1^2}{2L} = 2\pi$$

which we solve for the period

$$\Lambda_{21} = \frac{\lambda L}{\bar{x}}.$$

Here $\Lambda_{21} = x_2 - x_1$ is the distance between neighboring maxima, and $\bar{x} = (x_2 + x_1)/2$ is the average x position. Away from the origin the fringe period decreases $\sim 1/\bar{x}$, i.e. the interference is chirped. The first off-axis maximum is at $\Lambda_{21} = x_{\max}$, $\bar{x} = x_{\max}/2$ so $x_{\max} = \sqrt{2\lambda L}$ and the first intensity minimum is at $x_{\min} = \sqrt{\lambda L}$. When we get to a detailed discussion of diffraction phenomena we will encounter this type of scaling often, and will refer to a length scale $\sim \sqrt{\lambda L}$ as diffractive scaling.

3.1.3 Interference of two spherical waves

Consider two spherical waves with one originating at $\mathbf{r} = 0$ and the other one originating at $\mathbf{R} = d\epsilon_z$. The total field is

$$\mathbf{E} = \frac{\mathcal{E}_1}{2} \frac{e^{ikr_1}}{r_1} \epsilon_1 + \frac{\mathcal{E}_2}{2} \frac{e^{ikr_2}}{r_2} \epsilon_2 + c.c. .$$

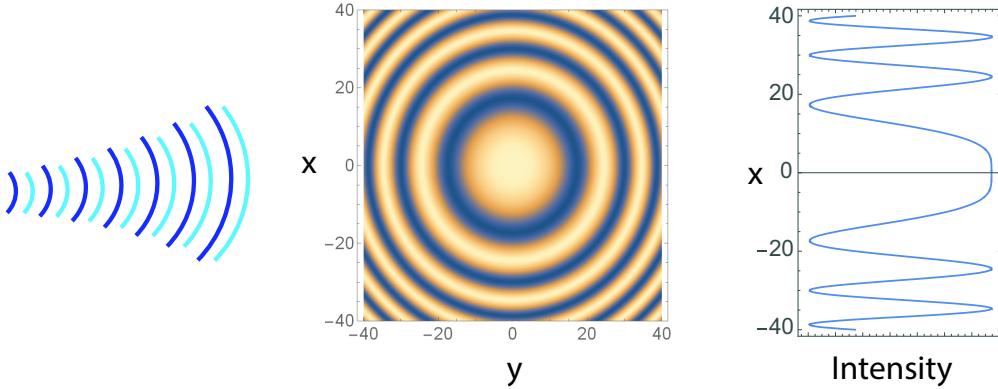


Figure 3.3: Interference of two spherical waves displaced along z with $\lambda = 1$, $L = 60$, and $d = 10$. The right hand plot shows the intensity as a function of x for $y = 0$.

We will again make a small angle approximation and put $\epsilon_1 \cdot \epsilon_2 = 1$, $r_1 \simeq L + \frac{x^2+y^2}{2L}$, $r_2 \simeq (L-d) + \frac{x^2+y^2}{2(L-d)}$ to get

$$I \sim \frac{|\mathcal{E}_1|^2}{L^2} + \frac{|\mathcal{E}_2|^2}{(L-d)^2} + \left[\frac{\mathcal{E}_1 \mathcal{E}_2^*}{L(L-d)} e^{ikd} e^{ik\left[\frac{x^2+y^2}{2L} - \frac{x^2+y^2}{2(L-d)}\right]} + c.c. \right].$$

The interference term at $z = L \gg x, y$ is proportional to $\cos\left[kd + k(x^2 + y^2)(\frac{1}{2L} - \frac{1}{2(L-d)})\right] = \cos\left[kd - \frac{kd(x^2 + y^2)}{2L(L-d)}\right]$. We again have a chirped interference pattern with the fringe period decreasing like $1/x$ as shown in Fig. 3.3. Taking $kd = q2\pi$, with q an integer, and $y = 0$ the first minimum is at $x_{\min} = \sqrt{\frac{\lambda L(L-d)}{d}}$ and the first maximum is at $x_{\max} = \sqrt{2}x_{\min}$. The interference is azimuthally symmetric about the z axis so we get a “bull’s eye” pattern.

3.1.4 Multibeam interference

Sharper interference fringes can be obtained when more than two beams simultaneously interfere. Consider N plane waves propagating in the $x - z$ plane at angles $\theta_j/2$ from the ϵ_z . If all beams have the same amplitude \mathcal{E} the intensity is

$$I \sim |\mathcal{E}|^2 \left| \sum_{j=1}^N e^{ik[\sin(\theta_j/2)x + \cos(\theta_j/2)z]} \right|^2.$$

Taking a fixed axial plane $z = 0$ we find

$$\begin{aligned} I &\sim N|\mathcal{E}|^2 + |\mathcal{E}|^2 \sum_{j=1}^N \sum_{k>j}^N e^{i[\sin(\theta_j/2) - \sin(\theta_k/2)]kx} + e^{-i[\sin(\theta_j/2) - \sin(\theta_k/2)]kx} \\ &= N|\mathcal{E}|^2 + 2|\mathcal{E}|^2 \sum_{j=1}^N \sum_{k>j}^N \cos[(\sin(\theta_j/2) - \sin(\theta_k/2))kx]. \end{aligned}$$

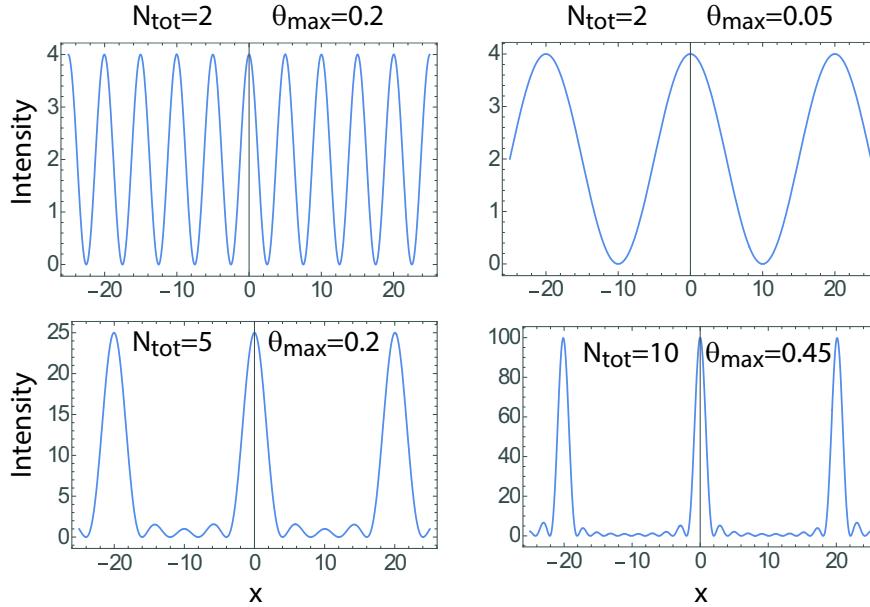


Figure 3.4: Interference of multiple plane waves with angles uniformly distributed.

Suppose the angles θ_j are evenly distributed between $-\theta_{\max}$ and $+\theta_{\max}$ so $\theta_j = -\theta_{\max} + 2\frac{j-1}{N-1}\theta_{\max}$. We find for the intensity the behavior shown in Fig. 3.4. The left hand column shows that if we increase the number of beams while keeping the total angular spread the same the spatial period increases and the width of each high intensity lobe stays constant. In the right hand column the number of beams and the angular spread are adjusted to maintain the same spatial period. Doing so we see that using more beams the peak intensity can be increased and the width of the intensity peak can be made narrower while keeping the same spatial period.

3.2 Fabry-Pérot Interferometer

An optical cavity can be used to enhance the intensity of an incident beam and has useful frequency dependent reflection and transmission properties. The presence of a gain medium inside a resonant cavity leads to laser operation. An interferometer is an instrument that uses interference from surfaces for analyzing optical properties of light. There are several different types of interferometer including Fabry-Pérot, Michelson, Mach-Zehnder, Fizeau, and Sagnac. We will start with a discussion of the Fabry-Pérot geometry [10].

Consider a general linear cavity with an input mirror, an output mirror, and internal losses. The amplitude of the internal field after transmission through the input coupler is $E_0 = t_{\text{in}}E_{\text{in}}$ where the intensity transmission of the input coupler is $T_{\text{in}} = |t_{\text{in}}|^2$. After each cavity round trip the internal field must be multiplied by a complex factor $g = |g|e^{i\phi(\nu)}$. The magnitude of g accounts for coupling and internal losses and the phase shift depends on the optical frequency ν and the cavity parameters.

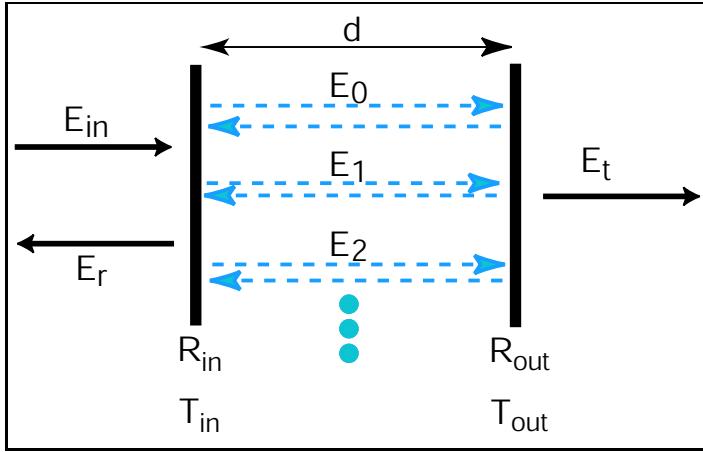


Figure 3.5: Fabry-Pérot cavity.

Parameters	
R_{in}	intensity reflectivity of input mirror, $T_{in} = 1 - R_{in}$
R_{out}	intensity reflectivity of output mirror, $T_{out} = 1 - R_{out}$
R_{cav}	effective internal cavity reflectivity neglecting losses at mirrors ($R_{cav} = 1 - \mathcal{L}_{cav}$ accounts for round trip absorption \mathcal{L}_{cav})
$\mathcal{R} = R_{in}R_{out}R_{cav}$	total reflectivity
E_{in}	amplitude of incident electric field
E_t	amplitude of transmitted electric field
E_r	amplitude of reflected electric field
E_{cav}	amplitude of electric field internal to the cavity

Table 3.1: Parameters of an optical cavity.

We will express the amplitude and phase as

$$\begin{aligned} |g| &= \sqrt{R_{in}R_{out}R_{cav}} = \sqrt{R_{in}R_{out}(1 - \mathcal{L}_{cav})}, \\ e^{i\phi(\nu)} &= e^{i\phi_0} e^{i\phi_1(\nu)} \frac{r'_{in}}{|r'_{in}|} \frac{r'_{out}}{|r'_{out}|}. \end{aligned}$$

Here R_{cav} accounts for the round trip internal losses neglecting the input and output mirrors. For example if the medium inside the cavity has (intensity) absorption coefficient α then $R_{cav} = e^{-\alpha 2d}$, where d is the length of the cavity. The primed coefficients refer to the case where the field is incident on a mirror from inside the cavity. The static phase accounts for any interfaces inside the cavity, and the frequency dependent term is

$$\phi_1(\nu) = kn(\nu)2d = \frac{4\pi d}{c}\nu n(\nu)$$

where we have allowed for a possible dispersion of the index of the intracavity medium.

The total internal field due to the sum of the circulating fields is then

$$E_{cav} = E_0 + E_1 + E_2 + \dots,$$

where $E_{j+1} = gE_j$. Thus

$$E_{\text{cav}} = \sum_{j=0}^{\infty} E_0 g^j = E_0 \frac{1}{1-g} = E_{\text{in}} \frac{t_{\text{in}}}{1-g}. \quad (3.1)$$

Referring to Fig. 3.5 E_{cav} is the field just to the right of the input mirror that is propagating to the right.

It is instructive to derive Eq. (3.1) in a different way that does not involve summing an infinite series of circulating waves. Looking at Fig. 3.5 we see that the internal field to the right of the input mirror is given by the sum of the transmitted input field plus the reflected cavity field. We can write $E_{\text{cav}} = t_{\text{in}}E_{\text{in}} + E_{\text{cav}}\sqrt{R_{\text{in}}R_{\text{cav}}R_{\text{out}}}e^{i\phi}$ and solving for E_0 we find

$$E_{\text{cav}} = E_{\text{in}} \frac{t_{\text{in}}}{1 - \sqrt{R_{\text{in}}R_{\text{cav}}R_{\text{out}}}e^{i\phi}} = E_{\text{in}} \frac{t_{\text{in}}}{1-g}$$

which is the same as (3.1).

3.2.1 Transmitted intensity

With the cavity field known we can write the transmitted field as

$$\begin{aligned} E_t &= E_{\text{cav}}\sqrt{(1+R_{\text{cav}})/2} t'_{\text{out}} \\ &= E_{\text{in}} t_{\text{in}} t'_{\text{out}} \sqrt{(1+R_{\text{cav}})/2} \frac{1}{1-g}, \end{aligned} \quad (3.2)$$

from which we find for the transmitted intensity

$$\begin{aligned} \frac{I_t}{I_{\text{in}}} &= T_{\text{in}} T_{\text{out}} ((1+R_{\text{cav}})/2) \left| \frac{1}{1-g} \right|^2, \\ &= \frac{T_{\text{in}} T_{\text{out}} ((1+R_{\text{cav}})/2)}{1+\mathcal{R}} \left[1 - \frac{2\sqrt{\mathcal{R}}}{1+\mathcal{R}} \cos[\phi_0 - \theta_{r_{\text{in}}} - \theta_{r_{\text{out}}} + 4\pi d\nu n(\nu)/c] \right]^{-1}, \end{aligned} \quad (3.3)$$

where $\mathcal{R} = |g|^2 = R_{\text{in}}R_{\text{out}}R_{\text{cav}}$. Assuming that the internal cavity phase and the phase shifts due to mirror reflections cancel we can write

$$\frac{I_t}{I_{\text{in}}} = \frac{T_{\text{in}} T_{\text{out}} ((1+R_{\text{cav}})/2)}{\left(1-\sqrt{\mathcal{R}}\right)^2} \left[1 + \frac{4\sqrt{\mathcal{R}}}{\left(1-\sqrt{\mathcal{R}}\right)^2} \sin^2(2\pi d\nu n(\nu)/c) \right]^{-1}, \quad (3.4)$$

which is the well known Airy formula showing a sequence of periodic maxima in the transmitted intensity as can be seen in Fig. 3.6. Note that when there are no internal losses, $R_{\text{cav}} = 1$, and the mirrors are matched, $R_{\text{in}} = R_{\text{out}}$, the peak cavity transmission is unity.

When there are finite internal losses the peak transmission is less than unity. To study the cavity transmission when the mirrors have high reflectivity we can write the maximum transmission from (3.4) as

$$\left. \frac{I_t}{I_{\text{in}}} \right|_{\text{max}} = \frac{1+R_{\text{cav}}}{2} \frac{T_{\text{in}} T_{\text{out}}}{\left(1 - \sqrt{(1-T_{\text{in}})(1-T_{\text{out}})R_{\text{cav}}}\right)^2}.$$

Up to this point we have assumed that $R + T = 1$ at each mirror and accounted for real losses due to absorption and scattering by introducing an effective round trip cavity reflectivity $R_{\text{cav}} < 1$. To proceed we will put $R_{\text{cav}} = 1 - \mathcal{L}_{\text{cav}}$ where \mathcal{L}_{cav} is the round trip loss excluding the mirror transmission. We find

$$\begin{aligned} \frac{I_t}{I_{\text{in}}} \Big|_{\text{max}} &= \frac{2 - \mathcal{L}_{\text{cav}}}{2} \frac{T_{\text{in}}T_{\text{out}}}{\left(1 - \sqrt{(1 - T_{\text{in}})(1 - T_{\text{out}})(1 - \mathcal{L}_{\text{cav}})}\right)^2} \\ &\simeq \frac{2 - \mathcal{L}_{\text{cav}}}{2} \frac{4T_{\text{in}}T_{\text{out}}}{(T_{\text{in}} + T_{\text{out}} + \mathcal{L}_{\text{cav}})^2}. \end{aligned}$$

For a symmetric cavity $T_{\text{in}} = T_{\text{out}} = T$ this simplifies to

$$\frac{I_t}{I_{\text{in}}} \Big|_{\text{max}} = (1 - \mathcal{L}_{\text{cav}}/2) \frac{T^2}{(T + \mathcal{L}_{\text{cav}}/2)^2}. \quad (3.5)$$

Suppose the mirror losses are p times larger than the mirror transmission so $\mathcal{L}_{\text{cav}} = 2pT$. We then find

$$\frac{I_t}{I_{\text{in}}} \Big|_{\text{max}} = (1 - pT) \frac{1}{(1 + p)^2}. \quad (3.6)$$

To reach 90% transmission in a cavity with good mirrors so $1 - pT \approx 1$ we need $p < 0.054$.

A small value of p is increasingly difficult to achieve as the mirror reflectivity is increased. To put this in the context of mirror coating technology a good dielectric mirror produced by electron beam evaporation¹ has absorption and scattering losses of $20 + 150 = 170$ ppm. Lower loss coatings are possible with ion beam sputtering (IBS). Mirror coatings with losses of 3 ppm have been produced and characterized[11] and state of the art mirrors used at the Laser Interferometer Gravitational-Wave Observatory (LIGO) have losses < 1 ppm.

As an example take $T = 1 - 0.9997$ which gives a cavity finesse of $\mathcal{F} = 10,500$ (see Sec. 3.2.3) in the absence of additional losses. This reflectivity corresponds to a transmission of $T = 0.0003$ or 300 ppm. A standard electron beam evaporated mirror will have loss of 170 ppm giving $p = 0.57$ and peak transmission of 41%.

3.2.2 Reflected intensity

The reflected field is the sum of the field directly reflected from the input mirror and the cavity field that is transmitted back through the input mirror which gives

$$\begin{aligned} E_r &= E_{\text{in}}r_{\text{in}} + E_{\text{cav}}gt'_{\text{in}}/r'_{\text{in}} \\ &= E_{\text{in}}r_{\text{in}} \left(1 + \frac{t_{\text{in}}t'_{\text{in}}}{r_{\text{in}}r'_{\text{in}}} \frac{g}{1-g}\right) \\ &= E_{\text{in}} \frac{r_{\text{in}}}{R_{\text{in}}} \frac{R_{\text{in}} - g}{1-g}. \end{aligned} \quad (3.7)$$

¹Values quoted on Laseroptik.de .

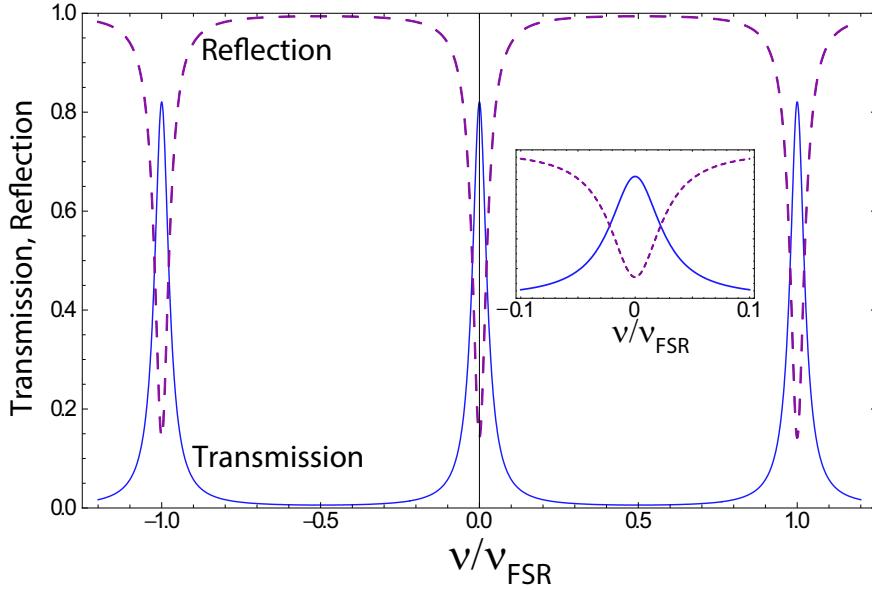


Figure 3.6: Transmission (solid line) and reflectivity (dashed line) of a Fabry-Pérot cavity with $R_{\text{in}} = .9$, $R_{\text{out}} = 0.8$, and $R_{\text{cav}} = 0.99$. The inset zooms in near the resonance.

In passing from the second to the third lines we have used $r_{\text{in}}r'_{\text{in}} = -R_{\text{in}}$ and, from Eqs. (1.33), $t_{\text{in}}t'_{\text{in}} = 1 - R_{\text{in}}^2$. Note that the interference between the directly reflected and cavity leakage beams, and hence the magnitude of the field, is independent of the mirror phase θ_r . A short calculation gives for the reflected intensity

$$\frac{I_r}{I_{\text{in}}} = \frac{R_{\text{in}} + \mathcal{R}/R_{\text{in}} - 2\mathcal{R}^{1/2} \cos[4\pi d\nu n(\nu)/c]}{1 + \mathcal{R} - 2\mathcal{R}^{1/2} \cos[4\pi d\nu n(\nu)/c]}.$$

Using $\cos 2x = 1 - 2\sin^2 x$ we can write this as

$$\frac{I_r}{I_{\text{in}}} = R_{\text{in}} \frac{(1 - \mathcal{R}^{1/2}/R_{\text{in}})^2 + 4(\mathcal{R}^{1/2}/R_{\text{in}}) \sin^2[2\pi d\nu n(\nu)/c]}{(1 - \mathcal{R}^{1/2})^2 + 4\mathcal{R}^{1/2} \sin^2[2\pi d\nu n(\nu)/c]}. \quad (3.8)$$

We see that for $R_{\text{in}} = 1$ the reflected intensity is equal to the incident intensity, independent of frequency, as it must be. When $R_{\text{in}} = \mathcal{R}^{1/2}$ or, equivalently, $R_{\text{in}} = R_{\text{cav}}R_{\text{out}}$ the cavity is “impedance matched” and the reflectivity is exactly zero on resonance. Put another way, impedance matching occurs when the input mirror transmission $T_{\text{in}} = 1 - R_{\text{in}} = 1 - R_{\text{cav}}R_{\text{out}} = \mathcal{L}$, where \mathcal{L} is the cavity round trip loss excluding the input coupler. Note that provided $R_{\text{cav}} = 1$ impedance matching occurs for $R_{\text{out}} = R_{\text{in}}$. For larger or smaller R_{in} the cavity is over or under coupled and the reflectivity never vanishes.

An interesting case is the symmetric cavity with $R_{\text{in}} = R_{\text{out}}$ which implies $\mathcal{R} = R_{\text{in}}^2 R_{\text{cav}}$

²Note that $|t_{\text{in}}| \neq |t'_{\text{in}}|$. This does not contradict the beamsplitter relations found in Sec. 1.11 where the index of refraction was the same on both sides of the beamsplitter, which is not generally true for the Fabry-Pérot resonator.

and

$$\frac{I_r}{I_{\text{in}}} = R_{\text{in}} \frac{\left(1 - R_{\text{cav}}^{1/2}\right)^2 + 4R_{\text{cav}}^{1/2} \sin^2[2\pi d\nu n(\nu)/c]}{\left(1 - R_{\text{in}}R_{\text{cav}}^{1/2}\right)^2 + 4R_{\text{in}}R_{\text{cav}}^{1/2} \sin^2[2\pi d\nu n(\nu)/c]}. \quad (3.9)$$

For a lossless cavity, $R_{\text{cav}} = 1$, and this last expression reduces to

$$\frac{I_r}{I_{\text{in}}} = \frac{\sin^2[2\pi d\nu n(\nu)/c]}{\left(\frac{1-R_{\text{in}}}{2R_{\text{in}}^{1/2}}\right)^2 + \sin^2[2\pi d\nu n(\nu)/c]}. \quad (3.10)$$

The symmetric lossless cavity is impedance matched so the reflected intensity vanishes on resonance, in agreement with the discussion above.

3.2.3 Finesse

It is customary to write the transmitted intensity as

$$\frac{I_t}{I_{\text{in}}} = \frac{T_{\text{in}}T_{\text{out}}((1 + R_{\text{cav}})/2)}{(1 - \mathcal{R}^{1/2})^2} \frac{1}{1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(2\pi d\nu n(\nu)/c)} \quad (3.11)$$

where the cavity finesse is given by

$$\mathcal{F} = \frac{\pi \mathcal{R}^{1/4}}{1 - \mathcal{R}^{1/2}}. \quad (3.12)$$

The reason for this choice of definition will be apparent in a moment³.

There are periodic maxima in the transmission at frequencies

$$\nu_N n(\nu_N) = \frac{c}{2d} N$$

with N an integer. The free spectral range, neglecting dispersion in the refractive index, is $\nu_{FSR} = \nu_{N+1} - \nu_N = c/(2dn)$. Defining the full width at half maximum of the transmission peaks as ν_{FWHM} we find that

$$\nu_{FWHM} = \frac{c}{\pi dn} \sin^{-1}\left(\frac{\pi}{2\mathcal{F}}\right)$$

so that when the finesse is large we have

$$\nu_{FWHM} \approx \frac{\nu_{FSR}}{\mathcal{F}}. \quad (3.13)$$

We see that defining the finesse by Eq. (3.12) the ratio between the longitudinal mode spacing and the full width of each mode is just the finesse.

³It is also possible to invert (3.12) for $\mathcal{R}(\mathcal{F}) = (2\mathcal{F}^4 + 4\pi^2\mathcal{F}^2 + \pi^4 - \pi\sqrt{16\mathcal{F}^6 + 20\pi^2\mathcal{F}^4 + 8\pi^4\mathcal{F}^2 + \pi^6})/(2\mathcal{F}^4)$ and use this in (3.4) to express I_t as a function of the finesse. This gives an unwieldy expression which is not convenient to work with.

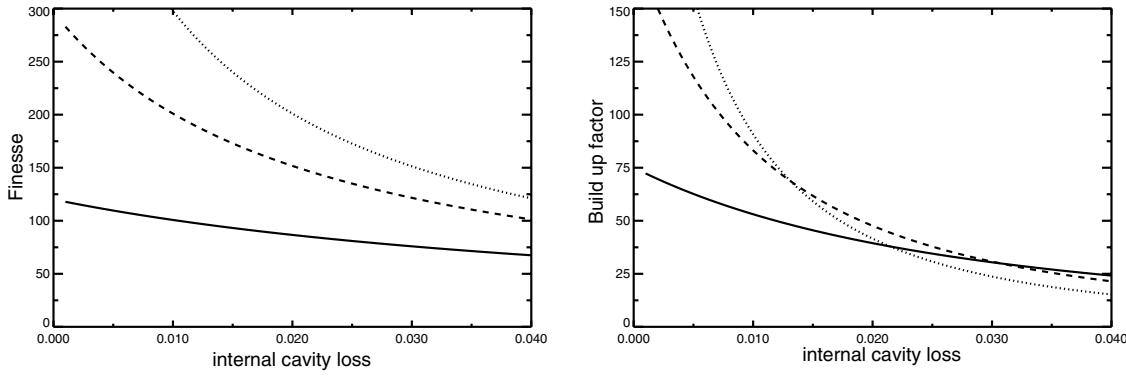


Figure 3.7: Fabry-Pérot cavity finesse(left) and build up factor (right) as a function of round trip internal loss for input coupler transmission of $T_{\text{in}} = 0.05, 0.02, 0.01$ (solid, dashed, dotted lines) and $R_{\text{out}} = 0.999$.

The finesse is large when the mirror transmission and internal losses are all small. In this limit $\mathcal{R} \simeq 1 - T_{\text{in}} - T_{\text{out}} - \mathcal{L}_{\text{cav}}$ where we have used $\mathcal{L}_{\text{cav}} = 1 - R_{\text{cav}}$. The finesse is then

$$\mathcal{F} \simeq \pi \frac{1 - (T_{\text{in}} + T_{\text{out}} + \mathcal{L}_{\text{cav}})/4}{(T_{\text{in}} + T_{\text{out}} + \mathcal{L}_{\text{cav}})/2} \simeq \frac{2\pi}{T_{\text{in}} + T_{\text{out}} + \mathcal{L}_{\text{cav}}}.$$

We identify the finesse as 2π divided by the total round trip loss including the mirror transmission out of the cavity.

Using the expressions for the finesse and the free spectral range resonance width we can write the transmitted and reflected intensity as

$$\frac{I_t}{I_{\text{in}}} = \frac{T_{\text{in}}T_{\text{out}}((1 + R_{\text{cav}})/2)}{(1 - \mathcal{R}^{1/2})^2} \frac{1}{1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(\pi\nu/\nu_{\text{FSR}})}, \quad (3.14a)$$

$$\frac{I_r}{I_{\text{in}}} = \frac{\frac{R_{\text{in}}(1 - \mathcal{R}^{1/2}/R_{\text{in}})^2}{(1 - \mathcal{R}^{1/2})^2} + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2[\pi\nu/\nu_{\text{FSR}}]}{1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2[\pi\nu/\nu_{\text{FSR}}]}. \quad (3.14b)$$

These equations are exact for any cavity reflectivities and losses with the definition $\mathcal{R} = R_{\text{in}}R_{\text{out}}R_{\text{cav}}$.

Photon lifetime

We can relate the FWHM of the cavity resonance to the photon lifetime inside the cavity as follows. Suppose in steady state the intensity inside the cavity is I_{cav} and the input field is suddenly turned off. The intensity will decay exponentially as $I_{\text{cav}}(t) = I_{\text{cav}}(0)e^{-\gamma_{\text{cav}}t}$, where γ_{cav} is the cavity decay rate. The change in intensity in one round trip is

$$\begin{aligned} \frac{\Delta I_{\text{cav}}}{\Delta t} &= \frac{I_{\text{cav}}\mathcal{R} - I_{\text{cav}}}{2dn/c} \\ &= -I_{\text{cav}}\nu_{\text{FSR}} [1 - (1 - T_{\text{in}})(1 - T_{\text{out}})(1 - \mathcal{L}_{\text{cav}})]. \end{aligned}$$

In the high finesse limit we have $\mathcal{F} \approx 2\pi/(T_{\text{in}} + T_{\text{out}} + \mathcal{L}_{\text{cav}})$ so that

$$\frac{dI_{\text{cav}}}{dt} \approx -I_{\text{cav}} \frac{2\pi\nu_{\text{FSR}}}{\mathcal{F}}$$

from which we see that

$$\gamma_{\text{cav}} \approx \frac{2\pi\nu_{\text{FSR}}}{\mathcal{F}} \approx 2\pi\nu_{\text{FWHM}} \equiv \omega_{\text{FWHM}}. \quad (3.15)$$

In other words the rate of energy decay from the cavity given by γ_{cav} is equal to ω_{FWHM} which is the width of the cavity resonance expressed in angular units.

The cavity decay rate affords an interpretation of the finesse in terms of the number of photon round trips in one cavity lifetime $\tau_{\text{cav}} = 1/\gamma_{\text{cav}}$. Let us say that in a time τ_{cav} a photon makes N cavity round trips. We have

$$N = \tau_{\text{cav}}/(2dn/c) = \tau_{\text{cav}}\nu_{\text{FSR}} = \tau_{\text{cav}}\mathcal{F}\nu_{\text{FWHM}} = \tau_{\text{cav}}\mathcal{F}\frac{\omega_{\text{FWHM}}}{2\pi} = \frac{\mathcal{F}}{2\pi}.$$

We can therefore express the finesse as

$$\mathcal{F} = 2\pi N.$$

Another useful quantity is the cavity build up factor which is the ratio of the on-resonance intensity inside the cavity to the input intensity. We have

$$\begin{aligned} I_{\text{cav,max}} &= I_{\text{in}} \frac{T_{\text{in}}}{\left(1 - \sqrt{\mathcal{R}}\right)^2} \\ &\approx I_{\text{in}} \frac{T_{\text{in}}\mathcal{F}^2}{\pi^2} \\ &\approx I_{\text{in}} \frac{4T_{\text{in}}}{(T_{\text{in}} + T_{\text{out}} + \mathcal{L}_{\text{cav}})^2} = I_{\text{in}} \frac{4}{T_{\text{in}}} \frac{1}{(1 + T_{\text{out}}/T_{\text{in}} + \mathcal{L}_{\text{cav}}/T_{\text{in}})^2}, \end{aligned}$$

where the last lines hold in the high finesse limit. Another useful relation is that when the cavity is impedance matched the buildup factor is $I_{\text{cav,max}}/I_{\text{in}} \simeq \mathcal{F}/\pi$. Figure 3.7 shows the finesse and build up factor as a function of the internal losses for fixed input coupler transmission.

A quantity that is easily measured experimentally is the ratio of the maximum to minimum transmitted intensity as the cavity is scanned through resonance. This is given by

$$\frac{I_{t,\text{max}}}{I_{t,\text{min}}} = 1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2. \quad (3.16)$$

Note that if the finesse is very high the ratio becomes large and difficult to determine. In such a case it is more convenient to measure the cavity bandwidth in order to determine the finesse.

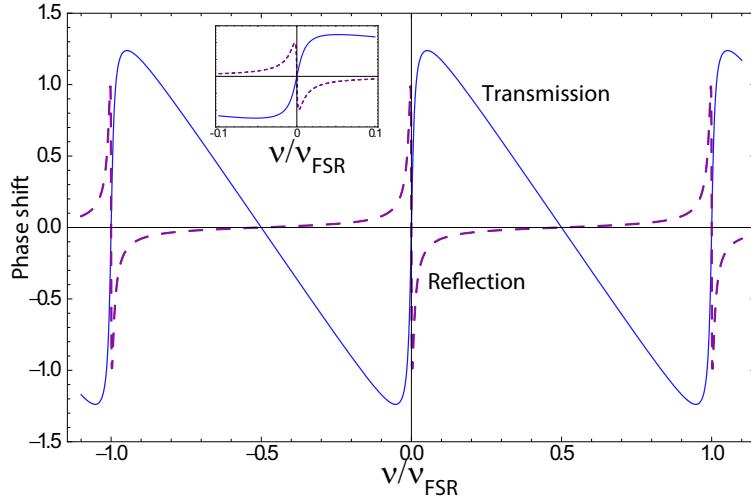


Figure 3.8: Transmission (solid line) and reflection (dashed line) phase shifts for $R_{\text{in}} = R_{\text{out}} = 0.95$, $R_{\text{cav}} = 0.99$. The finesse is $\mathcal{F} = 55.8$.

3.2.4 Quality factor

The quality factor is another quantity related to the finesse that is often used to characterize resonators. Particularly for low frequency waves, and in electrical circuits, one defines the quality factor as

$$Q_N = \frac{\omega_N}{\gamma_{\text{cav}}} \quad (3.17)$$

where $\omega_N = N\omega_{\text{FSR}}$ is the frequency of the N^{th} cavity resonance. Comparing with Eq. (3.13) we see that

$$Q_N = N\mathcal{F}.$$

The resonance number N can be written as

$$N = \frac{c/\lambda}{c/2d} = \frac{2d}{\lambda}$$

for a resonator of unit refractive index and length d . Thus for radiation of wavelength λ the quality factor is related to the finesse by

$$Q = \frac{2d}{\lambda} \mathcal{F}.$$

At optical frequencies we typically have $d \gg \lambda$ so $Q \gg \mathcal{F}$. At microwave frequencies d and λ are often comparable so $Q \approx \mathcal{F}$.

3.2.5 Phase shifts

The phase shift of the transmitted beam with respect to the input beam is

$$\theta_t = \arg \left(\frac{t_{\text{in}} t'_{\text{out}}}{1 - g} \right).$$

A little algebra gives the expression

$$\theta_t = \tan^{-1} \left[\frac{f \sin(2\pi\nu/\nu_{FSR})}{1 - f \cos(2\pi\nu/\nu_{FSR})} \right] \quad (3.18)$$

$$\approx \tan^{-1} \left[\frac{\sin(2\pi\nu/\nu_{FSR})}{1 - \cos(2\pi\nu/\nu_{FSR})} \right] \quad (3.19)$$

where $f = 1 + \frac{\pi^2}{2\mathcal{F}^2} - \frac{\pi}{\mathcal{F}}\sqrt{1 + \frac{\pi^2}{4\mathcal{F}^2}}$ and the second line holds in the high finesse limit.

The phase shift of the reflected beam is $\theta_r = \arg \left(r_{in} \frac{R_{in}-g}{1-g} \right)$ which gives the expression

$$\theta_r = \tan^{-1} \left[\frac{(R_{in} - 1)f \sin(2\pi\nu/\nu_{FSR})}{R_{in} + f^2 - (1 + R_{in})f \cos(2\pi\nu/\nu_{FSR})} \right] \quad (3.20)$$

$$\approx \tan^{-1} \left[\frac{(R_{in} - 1) \sin(2\pi\nu/\nu_{FSR})}{(1 + R_{in}) [1 - \cos(2\pi\nu/\nu_{FSR})]} \right] \quad (3.21)$$

where the second line is again applicable in the high finesse limit.

Figure 3.8 shows the phase shifts in transmission and reflection. The phase is an odd function of the detuning from cavity resonance and can be used to “lock” a laser beam to a cavity resonance. This is commonly done using the Pound-Drever-Hall method[12]. Although either transmission or reflection can be used the reflected phase has a larger slope, as can be seen from the inset in the figure, which provides for a more precise lock.

3.2.6 Finite beam size effects

If the beam has limited transverse size diffractive spreading will limit the attainable finesse irrespective of the mirror and cavity losses. This may be a limiting factor in the case of a planar mirror Fabry-Pérot.

We can estimate the necessary beam size to obtain a given finesse as follows. In the limit of low cavity loss and high finesse the effective number of cavity round trips is roughly the inverse of the single roundtrip loss, or $\mathcal{F}/(2\pi)$. The corresponding propagation distance is $L = 2d\mathcal{F}/(2\pi)$. If the input beam has waist radius w_0 then after propagation it will spread to $w = w_0\sqrt{1 + L^2/L_R^2}$, where $L_R = \pi w_0^2/\lambda$. Requiring that $w/w_0 \sim 1$ gives the condition $L^2 < L_R^2$ or

$$w_0 > \frac{\sqrt{\lambda d\mathcal{F}}}{\pi}. \quad (3.22)$$

For example a 5 cm long cavity with 1% losses so $\mathcal{F} \sim 600$ operated at $\lambda = 1 \mu\text{m}$ needs $w_0 > 1.7 \text{ mm}$.

3.2.7 Sensitivity to length changes

see hand written notes ♠??

3.2.8 Sensitivity to pressure

A Fabry-Pérot cavity with air inside is sensitive to pressure fluctuations due to the finite refractive index of air. In the visible part of the spectrum air at atmospheric pressure has refractive index $n \simeq 1 + 3 \times 10^{-4}$. In a cavity with length d the air induced phase shift in one round trip is

$$\delta\phi = 2dk(n - 1).$$

The refractive index is proportional to density which is proportional to pressure so the pressure dependent shift is

$$\delta\phi = 2dk \frac{P}{P_{\text{atm}}} (n - 1)$$

where P is the actual pressure and P_{atm} is atmospheric pressure.

Using $\lambda = .5 \mu\text{m}$, $d = 0.1 \text{ m}$, and a pressure of $10^{-6} \text{ mbar} \sim 10^{-9}$ of atmospheric, which corresponds to a good vacuum, but not UHV, gives $\delta\phi = 7.5 \times 10^{-7}$. A phase shift of 2π changes the cavity frequency by one FSR or $c/2d$. Thus the frequency shift at this pressure compared to the vacuum frequency is

$$\delta\nu = \frac{\delta\phi}{2\pi} \frac{c}{2d} = \frac{P}{P_{\text{atm}}} (n - 1)\nu = 180 \text{ Hz.}$$

Note that if the cavity were at atmospheric pressure we would be very sensitive to acoustic noise. In acoustics 0 dB, which is more or less the threshold of human perception, is defined as 2×10^{-4} microbar. A soft whisper may be +40 dB or 2×10^{-6} bar. Thus if the cavity were at atmospheric pressure and subjected to a 40 dB disturbance we would get a frequency shift of

$$\frac{P}{P_{\text{atm}}} (n - 1)\nu = (2 \times 10^{-6})(n - 1)\nu = 360 \text{ kHz.}$$

3.2.9 Determination of mode matching coefficient

The overlap of an incident beam with a resonant cavity mode can be measured as follows. Assume the intensity incident on the resonator is $I_{\text{in}} = I_0 + I_{\perp}$ with I_0 the mode matched intensity and I_{\perp} the intensity in orthogonal modes. The field reflected from the input mirror is approximately $E_r = r_{\text{in}}E_0 + E_t + r_{\text{in}}E_{\perp}$. Here E_t is the modematched intracavity field transmitted through the input mirror. We neglect the non modematched intracavity field since it is much weaker than the mode matched part in a high finesse cavity.

At a cavity resonance the sum $r_{\text{in}}E_0 + E_t$ leads to a reflected intensity for the mode matched light given by Eq. (3.8) as

$$I_0 \times \frac{1}{R_{\text{in}}} \frac{(R_{\text{in}} - \sqrt{\mathcal{R}})^2}{(1 - \sqrt{\mathcal{R}})^2} = mI_{\text{in}} \times \frac{1}{R_{\text{in}}} \frac{(R_{\text{in}} - \sqrt{\mathcal{R}})^2}{(1 - \sqrt{\mathcal{R}})^2}.$$

The directly reflected non modematched light gives a reflected intensity

$$I_{\perp} \times R_{\text{in}} = (1 - m)I_{\text{in}}R_{\text{in}}.$$

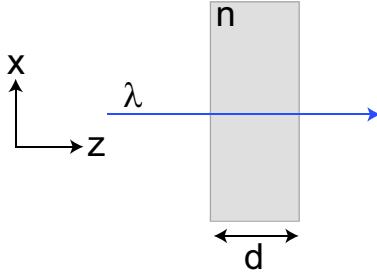


Figure 3.9: Etalon of thickness d and refractive index n .

The coefficient $m = I_0/I_{\text{in}}$ is the fraction of the intensity that is modematched to the cavity. Thus the observed reflected intensity at resonance will be given by

$$\frac{I_r}{I_{\text{in}}} = m \frac{1}{R_{\text{in}}} \frac{(R_{\text{in}} - \sqrt{\mathcal{R}})^2}{(1 - \sqrt{\mathcal{R}})^2} + (1 - m)R_{\text{in}}. \quad (3.23)$$

For an impedance matched cavity $R_{\text{in}} = \sqrt{\mathcal{R}}$, i.e. $R_{\text{in}} = R_{\text{cav}}R_{\text{out}}$ so the input transmission equals the cavity losses ($\mathcal{L} = 1 - R_{\text{cav}}R_{\text{out}}$) excluding the input coupler. In this case the reflected intensity vanishes on resonance when $m = 1$ but is finite for imperfect modematching.

To measure m we need three measurements. First determine R_{in} by measuring the input coupler reflectivity with the intracavity field blocked. Then measure the cavity finesse which determines \mathcal{R} , and finally measure the reflection dip I_r/I_{in} and use (3.23) to determine m .

3.3 Etalon

A short, fixed separation Fabry-Pérot is referred to as an etalon as shown in Fig. 3.9. The etalon is useful for spectral filtering of a light beam. A solid etalon is inherently stable consisting of a piece of transparent optical material with highly polished and parallel faces. If the faces are separated by a distance d , the material has an index n and the surfaces are coated with reflectivity R then the free spectral range is

$$\nu_{\text{FSR}} = \frac{c}{2nd}$$

and the transmission function from Eq. (3.14) is

$$\frac{I_t}{I_{\text{in}}} = \frac{T_{\text{in}}T_{\text{out}}((1 + R_{\text{cav}})/2)}{(1 - \mathcal{R}^{1/2})^2} \frac{1}{1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(\pi\nu/\nu_{\text{FSR}})} \quad (3.24)$$

with the definition $\mathcal{R} = R_{\text{in}}R_{\text{out}}R_{\text{cav}}$.

The contrast between the maximum and minimum transmitted intensity is

$$q = 1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2.$$

A moderate finesse of 100 gives a contrast of 4050 or 36 dB. Say we use fused silica of index $n = 1.45$ and thickness $d = 1$. mm then $\nu_{\text{FSR}} = 103$ (GHz) and the transmission bandwidth is $\nu_{\text{FWHM}} = \nu_{\text{FSR}}/\mathcal{F} = 1.03$ GHz. Neglecting loss in the fused silica ($R_{\text{cav}} = 1$) and $R_{\text{in}} = R_{\text{out}} = R$ we have $\mathcal{F} = \pi R^{1/2}/(1 - R)$ so $\mathcal{F} = 100$ implies $R = 0.97$. The peak transmission is then

$$I_{t,\max} = \frac{(1 - R)^2}{(1 - R)^2} = 1.$$

The detuning $\delta\nu$ at which the transmission has fallen to 0.99 of the peak is found from

$$\frac{1}{1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(\pi\delta\nu/\nu_{\text{FSR}})} = 0.99$$

which has the approximate solution $\delta\nu = 0.00049\nu_{\text{FSR}} = 50.4$ MHz.

An interesting question is whether or not a solid etalon is sufficiently stable to maintain a resonance to a desired accuracy. As an example suppose we wish to have a resonance that is stable to 50 MHz so that the transmission intensity varies by less than 1% for a fixed frequency optical wave. Let the temperature change by an amount δT giving a change $\delta\nu_{\text{FSR}}$ then the detuning relative to an etalon peak is

$$\delta\nu = -\frac{\nu}{\nu_{\text{FSR}}} \delta\nu_{\text{FSR}} = -\frac{c}{\lambda} \frac{\delta\nu_{\text{FSR}}}{\nu_{\text{FSR}}}.$$

The change in the free spectral range is

$$\delta\nu_{\text{FSR}} = -\nu_{\text{FSR}} \left(\frac{1}{n} \frac{dn}{dT} + \frac{1}{d} \frac{dd}{dT} \right) \delta T.$$

At room temperature and 587 nm wavelength fused silica has[13]

$$\frac{dn}{dT} = 0.87 \times 10^{-5}.$$

The coefficient of thermal expansion is

$$\frac{1}{d} \frac{dd}{dT} = 0.55 \times 10^{-6}$$

per degree C. Thus for a 1 C change in temperature which is probably typical for a quiet laboratory we find

$$\delta\nu_{\text{FSR}} = -\nu_{\text{FSR}} \times 6.6 \times 10^{-6}$$

so for a 0.5 μm wavelength

$$\delta\nu = \frac{c}{.5 \mu\text{m}} 6.6 \times 10^{-6} = 4.0 \text{ GHz.}$$

If the etalon is stabilized to 1 mK we get a drift of 4 MHz.

We can reduce the requirement on temperature stability by using a thinner etalon. Say we take $d = 0.25$ mm which is commercially available as a standard component. Then $\nu_{\text{FSR}} = 414$ GHz, $\nu_{\text{FWHM}} = 4.1$ GHz, and the frequency error giving 99% transmission is 202 MHz which corresponds to a temperature drift of 50 mK. Alternatively there are materials such as Zerodur, Invar and ULE glass that have smaller coefficients of thermal expansion.

3.3.1 Tilted etalon

Consider propagation through an etalon plate of index n and thickness d as in Fig. 3.10. At normal incidence the phase shift between the beam at the entrance point $(x, z) = (0, 0)$ and the exit point $(0, d)$ is $\phi = knd$, $k = 2\pi/\lambda$. There will be a transmission resonance when $2\phi = 2\pi m$ with m an integer. This gives a wavelength

$$\lambda_0 = \frac{2nd}{m}.$$

When the beam is incident at angle θ the phase shift between the same entrance and exit points is $\phi = n\mathbf{k}' \cdot \mathbf{r} = knd \cos \theta'$, with θ' the angle from the normal inside the slab. The wavelength at a transmission resonance is

$$\lambda = \frac{2nd \cos \theta'}{m}. \quad (3.25)$$

It may seem odd that the phase shift crossing the tilted etalon gets smaller, not larger, since the thickness apparently increases as we tilt the etalon. However, following a ray across the tilted etalon also gives a transverse displacement, whereas we should compare the phase at the same entrance and exit points as in the case of normal incidence. Doing so gives Eq. (3.25).

Writing (3.25) in terms of the input angle θ gives

$$\begin{aligned} \lambda &= \frac{2d}{m} [n^2 - \sin^2(\theta)]^{1/2} \\ &= \lambda_0 \left[1 - \frac{\sin^2(\theta)}{n^2} \right]^{1/2} \end{aligned} \quad (3.26)$$

We see that the resonant transmission wavelength λ decreases as θ is increased. Analogously, a fixed wavelength λ_0 will be resonant at angles θ_m satisfying

$$\sin^2(\theta_m) = n^2 - \left(\frac{\lambda_0}{2d} \right)^2 m^2.$$

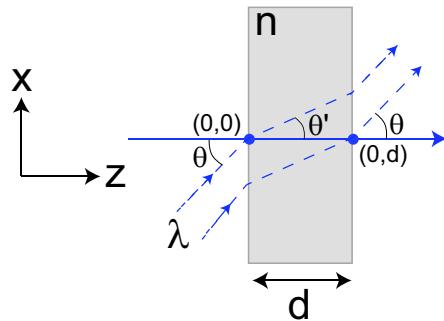


Figure 3.10: Etalon at normal incidence (solid line) and for tilted rays (dashed lines).

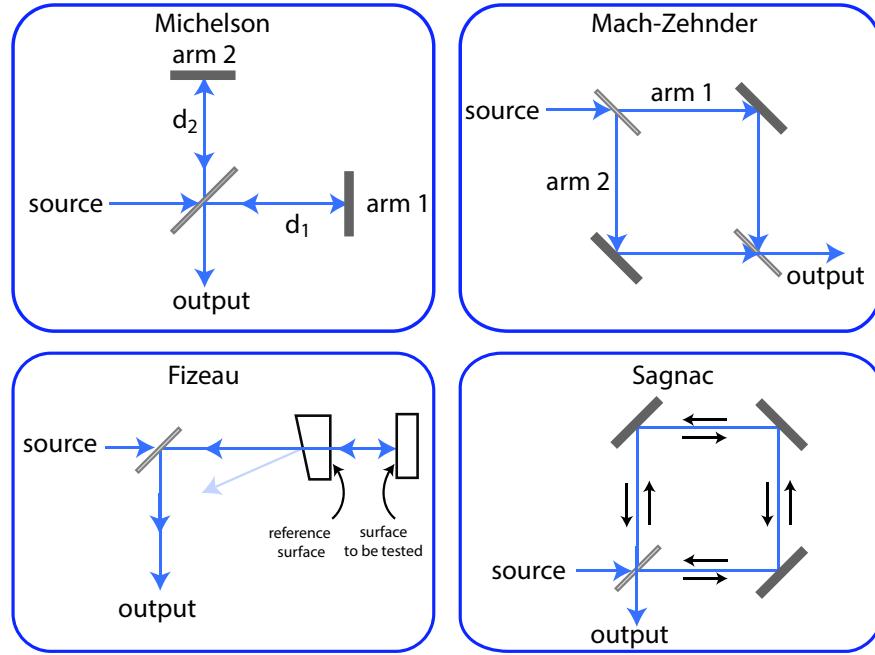


Figure 3.11: Main types of interferometers.

The axial resonance order is $m_0 = 2nd/\lambda_0 = \nu/\nu_{\text{FSR}}$ so we can write

$$\sin^2(\theta_m) = n^2 \left(1 - \frac{m^2}{m_0^2}\right)$$

and

$$\theta_m = \sin^{-1} \left[n \sqrt{1 - \frac{m^2}{m_0^2}} \right]$$

for $m < m_0$. As an example $\lambda_0 = 1$, $d = 100$, and $n = 1.5$ gives $m_0 = 300$ and resonant angles

$$\theta_{300} = 0., \quad \theta_{299} = 7.03 \text{ deg.}, \quad \theta_{298} = 9.96 \text{ deg.}, \quad \theta_{297} = 12.22 \text{ deg.}, \quad \text{etc.}$$

The implication is that illumination with a diverging spherical wave results in a transmitted intensity that takes the form of a bull's eye pattern as in Fig. 3.3.

3.4 Other interferometers

In addition to the Fabry-Pérot, the other main types of interferometer are the Michelson[14], Mach-Zehnder[15, 16], Fizeau[17], and Sagnac[18]. The geometries are shown in Fig. 3.11. The Michelson and Mach-Zehnder geometries are both used for measuring phase shifts or mirror displacements. The Fizeau interferometer is primarily used for testing optical surfaces and the Sagnac interferometer is used as a rotation sensor.

3.4.1 Sagnac interferometer

♣?? tba

3.5 Fourier transform spectroscopy

The Michelson interferometer can be used to measure the spectral content of an optical beam. When the input light has a single spectral component $I(\lambda)$ and the two arms have a path length difference d , as shown in Fig. 3.12 the output of the interferometer is

$$I(\lambda, d) = \frac{I(\lambda)}{2} [1 + \cos(kd)].$$

Here $d = 2(d_2 - d_1)$ is the path length difference between the arms. We could equivalently use the time delay variable $\tau = d/c$ and the angular frequency $\omega = 2\pi c/\lambda$ to write

$$\tilde{I}(\omega, \tau) = \frac{\tilde{I}(\omega)}{2} [1 + \cos(\omega\tau)].$$

Following the discussion in Sec. 2.1 we will work with a single sided spectral intensity distribution $\tilde{I}(\omega)$ with $\omega \geq 0$. The total intensity at fixed τ is then

$$\begin{aligned} I(\tau) &= \int_0^\infty d\omega \tilde{I}(\omega, \tau) \\ &= \frac{1}{2} \int_0^\infty d\omega \tilde{I}(\omega) + \frac{1}{2} \int_0^\infty d\omega \tilde{I}(\omega) \cos(\omega\tau). \end{aligned}$$

At zero delay

$$I(0) = \int_0^\infty d\omega \tilde{I}(\omega)$$

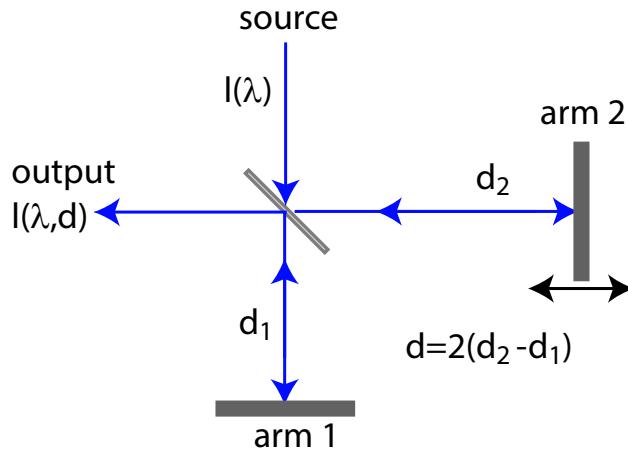


Figure 3.12: Fourier transform spectroscopy with a Michelson interferometer. Moving one of the mirrors changes the differential path length d leading interference at the output port.

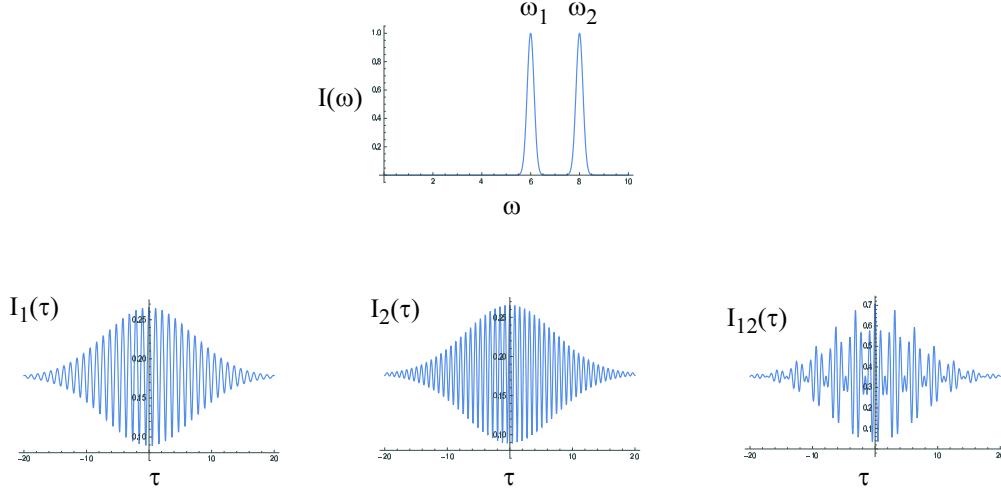


Figure 3.13: Fourier transform spectroscopy signal for an input spectrum of two Gaussians centered at ω_1 and ω_2 . The left two plots show the output with only one of the Gaussians present and the right hand plot shows the output when both input frequencies are present which leads to interference.

and

$$I(\tau) - \frac{I(0)}{2} = \sqrt{\frac{2}{\pi}} \int_0^\infty d\omega \frac{\pi^{1/2} \tilde{I}(\omega)}{2^{3/2}} \cos(\omega\tau).$$

Taking the inverse cosine transform we find the spectral intensity from

$$\tilde{I}(\omega) = \frac{4}{\pi} \int_0^\infty d\tau \left[I(\tau) - \frac{I(0)}{2} \right] \cos(\omega\tau).$$

If we record $I(\tau) = I(d/c)$ for a range of path length differences d we can calculate the spectral distribution $\tilde{I}(\omega)$ by taking the inverse cosine transform. This is the basis of Fourier transform spectroscopy (FTS). FTS is often more efficient than using a scanning spectrometer since a larger intensity is incident on the detector which reduces the influence of photon shot noise and detector noise. Figure 3.13 shows an example of the output signal for a two frequency input.

An important practical question is how large a scan range is needed to resolve a frequency interval $\delta\omega$. From the usual Fourier uncertainty relations we know that $\delta\omega \sim 1/\delta\tau$. Thus we should set $\delta\tau > 1/\delta\omega_{\min}$ where $\delta\omega_{\min}$ is the minimal spectral resolution we are interested in. Say we wish to resolve a spectral interval of $\delta\lambda = 0.01$ nm centered at $\lambda = 500$ nm. Then $\delta\omega = 2\pi\delta\nu = \frac{2\pi c}{\lambda^2}\delta\lambda$. The corresponding frequency resolution is $\delta\nu = \delta\omega/2\pi = 12$ GHz. To resolve this we need $\delta\tau > \frac{\lambda^2}{2\pi c \delta\lambda} = 13$ ps. The physical distance the path length difference needs to be scanned is

$$\delta d = c\delta\tau = \frac{\lambda^2}{2\pi \delta\lambda} \simeq 4 \text{ mm}.$$

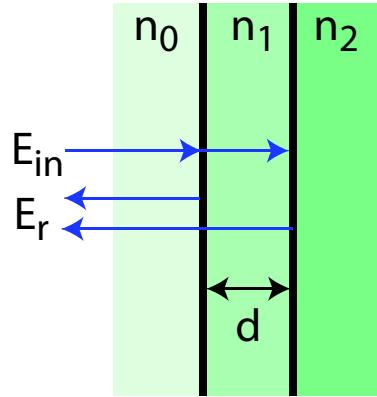


Figure 3.14: Reflection from a thin film of thickness d . The external medium has index n_0 , the film has index n_1 and the substrate has index n_2 .

3.6 Multilayer coatings

The utility of the Fabry-Pérot resonator, and many other optical instruments, is dependent on the availability of mirrors with very high reflectivity. The technology for making high reflectivity mirrors has advanced to the point where more than five nines of reflectivity are commercially available. Such mirrors are fabricated using multilayer dielectric coatings. As is the case with lens systems for high performance imaging the design of multilayer coatings is based on computerized numerical calculations. We can nevertheless understand some of the basic ideas from analytical considerations.

3.6.1 Quarterwave anti-reflection coating

We will first consider the task of producing a low reflectivity or “anti-reflection” coating on a high index substrate. Such coatings are needed for lenses used in imaging systems to avoid ghosting from undesired multiple reflections. The geometry is shown in Fig. 3.14. We know from our analysis of the Fabry-Pérot interferometer that the reflected field should be found by adding up an infinite number of partial reflections. We can nonetheless gain a qualitative understanding of the behavior by considering the first two reflected waves. There is the external reflection from the incident beam, $E_{r1} = E_{in}r_{01}$ and the first reflection from the substrate, $E_{r2} = E_{in}t_{01}r_{12}t_{10}e^{i2dkn_1}$. The notation for the Fresnel coefficients uses r_{ij} for the reflection inside medium i with the second medium labeled j and t_{ij} the transmission from medium i to medium j . This notation will be useful below in the treatment of multilayer films.

Qualitatively we expect the reflection to be minimized if E_{r1} and E_{r2} interfere destructively. Assume a phase convention where t_{01}, t_{10} are both positive. We then require that $\text{Re} \left[\frac{r_{12}}{r_{01}} e^{i2dkn_1} \right]$ is negative. At normal incidence the Fresnel reflection amplitudes are negative when the second medium has a higher index. If we set $n_2 > n_1 > n_0$ then we will get destructive interference when $2dkn_1 = \pi$ or $d = \lambda/(4n_1)$. In order to match the magnitude of E_{r1} and E_{r2} we need to choose the film index n_1 judiciously.

At normal incidence $r_{ij} = \frac{n_i - n_j}{n_i + n_j}$. If n_i, n_j are similar in magnitude then $|r_{ij}|$ will be small

and $|t_{ij}|$ will be close to unity. This suggests the condition $r_{01} = r_{12}$ or

$$n_1 = \sqrt{n_0 n_2}.$$

We see that a quarter wave film ($d = \lambda/4n_1$) with n_1 the geometric mean of the external and substrate indices can be expected to minimize the reflection coefficient. To use this idea for air the external medium and a glass substrate we need $n_1 \sim \sqrt{1.5} = 1.22$. This condition is approximately met by MgF₂ which has an index of 1.38.

A quantitative evaluation of the reflected field can be found using Eq.(3.7) which becomes in the notation of this section

$$E_r = E_{\text{in}} \frac{r_{01}}{R_{01}} \frac{R_{01} - g}{1 - g}$$

with

$$g = \sqrt{R_{01} R_{12}} e^{i\phi_0} e^{i\phi_1(\lambda)} \frac{r_{10}}{|r_{10}|} \frac{r_{12}}{|r_{12}|}$$

and $\phi_1 = 2dkn_1/\lambda$. Setting $\phi_0 = 0, \phi_1 = \pi$ we get $g = \sqrt{R_{01} R_{12}}$ and

$$E_r = E_{\text{in}} \frac{n_0 n_2 - n_1^2}{n_0 n_2 + n_1^2}. \quad (3.27)$$

Perhaps surprisingly the condition for canceling the reflection, $n_1 = \sqrt{n_0 n_2}$, which we found from a simplified consideration of only the first two reflected waves, is borne out by a full analysis. The reason why this condition works perfectly is that it corresponds to an impedance matched cavity. Recall the Fabry-Pérot impedance matching condition when there are no internal losses is simply $R_{\text{in}} = R_{\text{out}}$ which coincides with $n_1 = \sqrt{n_0 n_2}$.

Using a MgF₂ layer $n_1 = 1.38$ on a glass substrate $n_2 = 1.52$ we find an intensity reflectivity of

$$\frac{I_r}{I_{\text{in}}} = 0.014$$

which is considerably less than that of uncoated glass which has $\frac{I_r}{I_{\text{in}}} = \left(\frac{1.5-1}{1.5+1}\right)^2 = 0.04$.

The attentive reader may notice one peculiar point. In our analysis of the Fabry-Pérot resonator we found that transmission resonances occur for $2dkn = 2\pi N$ with N an integer. Thus the first resonance is at $d = \lambda/2$. Here the first resonance is at $d = \lambda/4$. Why do we have an apparent discrepancy?

3.6.2 Quarterwave high-reflection coating

We now wish to enhance the reflectivity of the substrate. From our study of the Fabry-Pérot resonator we know that the transmitted intensity never goes exactly to zero. We therefore do not expect to be able to make a perfect reflector. Nevertheless extremely high performance can be achieved with a multilayer coating.

The most basic idea uses the quarter-wave antireflection coating but with the low index film replaced by a high index film. With $n_1 > n_2 > n_0$ the reflection coefficient r_{12} becomes

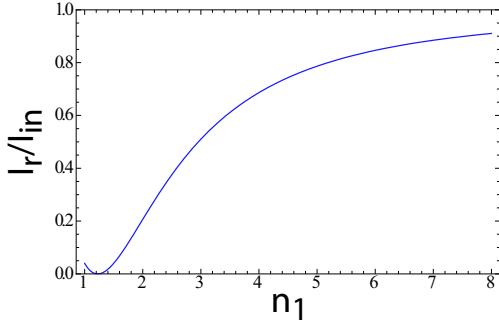


Figure 3.15: Reflectance of quarter wave film with index n_1 in between air ($n_0 = 1$) and glass ($n_2 = 1.5$).

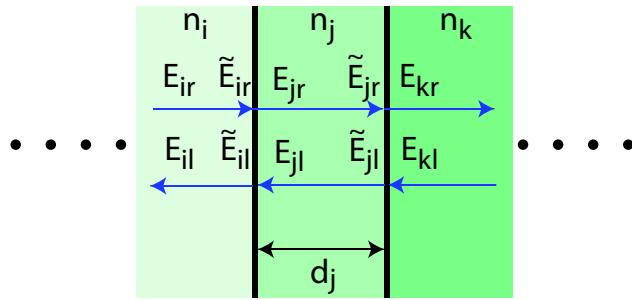


Figure 3.16: Wave propagation through a multilayer film.

positive and instead of destructive interference we get constructive interference and enhanced reflectivity. We now have $g = -\sqrt{R_{01}R_{12}}$ but Eq. (3.27) remains unchanged giving

$$E_r = E_{in} \frac{n_0 n_2 - n_1^2}{n_0 n_2 + n_1^2}.$$

For example using TiO_2 which has $n = 2.30$ we get $I_r/I_{in} = 0.31$. As can be seen in Fig. 3.15 very high reflectance requires large n_1 . Unfortunately there are no convenient transparent materials available which motivates the use of multilayer films as will be discussed in the next section.

3.6.3 Multilayer coatings

High performance antireflection or highreflection films rely on multilayer coatings. We can analyze these cases using a matrix formalism. Consider the geometry in Fig. 3.16. Each layer has index n_j and thickness d_j . The fields propagating to the right and the left at the left hand side of layer j are labeled E_{jr}, E_{jl} . The fields at the right hand side of the same layer are $\tilde{E}_{jr}, \tilde{E}_{jl}$ and they are related by

$$\begin{pmatrix} E_{jr} \\ E_{jl} \end{pmatrix} = \begin{pmatrix} e^{-i\phi_j} & 0 \\ 0 & e^{i\phi_j} \end{pmatrix} \begin{pmatrix} \tilde{E}_{jr} \\ \tilde{E}_{jl} \end{pmatrix},$$

with $\phi_j = kn_j d_j$ and $k = 2\pi/\lambda_{\text{vac}}$. We can also write this in matrix notation as $\mathbf{E}_j = \Phi_j \tilde{\mathbf{E}}_j$ where $\mathbf{E}_j = \begin{pmatrix} E_{jr} \\ E_{jl} \end{pmatrix}$. Concentrating on the interface between layers j and k we have

$$\begin{aligned} E_{kr} &= t_{jk} \tilde{E}_{jr} + r_{kj} E_{kl}, \\ \tilde{E}_{jl} &= r_{jk} \tilde{E}_{jr} + t_{kj} E_{kl}. \end{aligned}$$

Solving for the \tilde{E} we find

$$\begin{aligned} \tilde{E}_{jr} &= \frac{1}{t_{jk}} E_{kr} - \frac{r_{kj}}{t_{jk}} E_{kl}, \\ \tilde{E}_{jl} &= \frac{r_{jk}}{t_{jk}} E_{kr} + \frac{t_{kj} t_{jk} - r_{kj} r_{jk}}{t_{jk}} E_{kl}. \end{aligned}$$

Using the Stokes relation (1.34) $t_{kj} t_{jk} - r_{kj} r_{jk} = 1$ we can write the solution as $\tilde{\mathbf{E}}_j = \mathbf{G}_{jk} \mathbf{E}_k$ where

$$\mathbf{G}_{jk} = \frac{1}{t_{jk}} \begin{pmatrix} 1 & r_{jk} \\ r_{jk} & 1 \end{pmatrix}.$$

Combining with the Φ_j matrix we have for the transformation between two layers

$$\mathbf{E}_j = \Phi_j \mathbf{G}_{jk} \mathbf{E}_k = \mathbf{M}_{jk} \mathbf{E}_k \quad (3.28)$$

with

$$\mathbf{M}_{jk} = \frac{1}{t_{jk}} \begin{pmatrix} e^{-i\phi_j} & r_{jk} e^{-i\phi_j} \\ r_{jk} e^{i\phi_j} & e^{i\phi_j} \end{pmatrix}. \quad (3.29)$$

This can be straightforwardly extended to an arbitrary number of layers simply by multiplying together the matrices for each layer.

Let's consider the most general situation where there is an external beam incident from the left ($E_{0r} \neq 0$), and also a beam incident from the right at the last layer labeled N ($E_{Nl} \neq 0$). If the total transformation matrix is $\mathbf{M}_{0N} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$ then we can solve for the output fields to find

$$E_{0l} = \frac{M_{21}}{M_{11}} E_{0r} + \frac{M_{11} M_{22} - M_{12} M_{21}}{M_{11}} E_{Nl}, \quad (3.30a)$$

$$E_{Nr} = \frac{1}{M_{11}} E_{0r} - \frac{M_{12}}{M_{11}} E_{Nl}. \quad (3.30b)$$

Equations (3.30) provide a compact solution to any multilayer interference problem. When there is only a beam incident from the left the effective reflection coefficient is $r = \frac{E_{0l}}{E_{0r}} = \frac{M_{21}}{M_{11}}$ and when there is only a beam incident from the right $r = \frac{E_{Nr}}{E_{Nl}} = -\frac{M_{12}}{M_{11}}$.

As an example let's check the previous results for a single $\lambda/4$ film with incident medium 0, quarter wave layer 1, and substrate 2. The transformation between layers 0 and 2 is $\mathbf{E}_0 = \mathbf{M} \mathbf{E}_2$ with

$$\mathbf{M} = M_{01} M_{12} = \frac{1}{t_{01} t_{12}} \begin{pmatrix} e^{-i(\phi_0 + \phi_1)} (1 + e^{i2\phi_1} r_{01} r_{12}) & e^{-i(\phi_0 + \phi_1)} (e^{i2\phi_1} r_{01} + r_{12}) \\ e^{i(\phi_0 - \phi_1)} (r_{01} + e^{i2\phi_1} r_{12}) & e^{i(\phi_0 - \phi_1)} (e^{i2\phi_1} + r_{01} r_{12}) \end{pmatrix}.$$

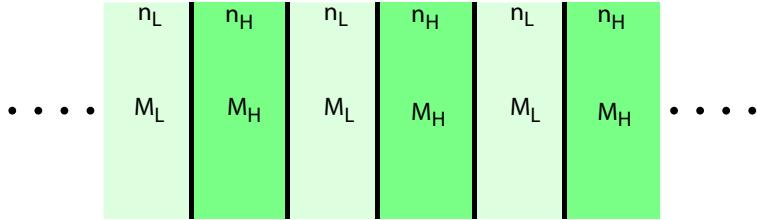


Figure 3.17: High reflectivity coating made of alternating quarterwave layers of high and low index.

We take $\phi_0 = 0$ since we define the incident fields as those at the surface of the film, $\phi_1 = \pi/2$, $r_{01} = \frac{n_0 - n_1}{n_0 + n_1}$, $r_{12} = \frac{n_1 - n_2}{n_1 + n_2}$, $t_{01} = \frac{2n_0}{n_0 + n_1}$, and $t_{12} = \frac{2n_1}{n_1 + n_2}$. Plugging in we get

$$\mathbf{M} = \frac{i}{2n_0 n_1} \begin{pmatrix} -n_0 n_2 - n_1^2 & n_0 n_2 - n_1^2 \\ -n_0 n_2 + n_1^2 & n_0 n_2 + n_1^2 \end{pmatrix}.$$

With the boundary condition on the right hand side that $E_{2l} = 0$ we get the solutions

$$\begin{aligned} r &\equiv \frac{E_{0l}}{E_{0r}} = \frac{C}{A} = \frac{n_0 n_2 - n_1^2}{n_0 n_2 + n_1^2} \\ t &\equiv \frac{E_{2r}}{E_{0r}} = \frac{1}{A} = i \frac{2n_0 n_1}{n_0 n_2 + n_1^2}. \end{aligned}$$

The expression for r agrees with (3.27) and $|r|^2 + \frac{n_2}{n_0}|t|^2 = 1$ agrees with (1.31).

3.6.4 Multilayer high reflectivity coating

As an example of the use of Eqs. (3.30) let's calculate the reflectivity for a stack of alternating high and low index $\lambda/4$ layers as shown in Fig. 3.17. A pair of $\lambda/4$ layers with indices n_L, n_H give

$$\mathbf{M}_L = \frac{i}{2n_L} \begin{pmatrix} -n_H - n_L & n_H - n_L \\ n_L - n_H & n_H + n_L \end{pmatrix}, \quad \mathbf{M}_H = \frac{i}{2n_H} \begin{pmatrix} -n_H - n_L & n_L - n_H \\ n_H - n_L & n_H + n_L \end{pmatrix},$$

and the two-layer segment $\mathbf{M}_L \mathbf{M}_H$ gives

$$\mathbf{M}_s = \mathbf{M}_L \mathbf{M}_H = \frac{-1}{2n_H n_L} \begin{pmatrix} n_H^2 + n_L^2 & n_H^2 - n_L^2 \\ n_H^2 - n_L^2 & n_H^2 + n_L^2 \end{pmatrix}.$$

Note that $\det(\mathbf{M}_s) = 1$. The matrix for N such segments is $\mathbf{M}_N = \mathbf{M}_s^N$. To calculate this we can use a result known as Sylvester's theorem. This says that for matrices with $\det(\mathbf{M}) = 1$ the matrix $\mathbf{M}_N = \mathbf{M}^N$ is given by

$$\mathbf{M}_N = \frac{1}{\sin(\theta)} \begin{pmatrix} M_{11} \sin(N\theta) - \sin[(N-1)\theta] & M_{12} \sin(N\theta) \\ M_{21} \sin(N\theta) & M_{22} \sin(N\theta) - \sin[(N-1)\theta] \end{pmatrix}$$

with $\cos(\theta) = (M_{11} + M_{22})/2$.

For a large number of layers we can neglect edge effects due to air on the left and a glass substrate on the right. In this limit the intensity reflectance is

$$R_N = \left| \frac{M_{N,21}}{M_{N,11}} \right|^2 = \frac{M_{21}^2 \sin^2(N\theta)}{(M_{11} \sin(N\theta) - \sin[(N-1)\theta])^2}.$$

Let us use MgF₂ with $n_L = 1.38$ and TiO₂ with $n_H = 2.30$. For $N = 2, 4, 6, 8, 10$ corresponding to 4, 8, 12, 16, 20 layers we find $R = 0.5937, 0.9350, 0.9913, 0.9989, 0.9999$.

We can also derive a closed form expression for R_N in terms of n_H, n_L as follows. The segment matrix $\mathbf{M}_s = \mathbf{M}_L \mathbf{M}_H$ is a real symmetric matrix. It can be diagonalized by an orthogonal matrix \mathbf{P} constructed from the eigenvectors of \mathbf{M}_s . To do this we calculate the eigenvectors and eigenvalues of \mathbf{M}_s which are $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $-n_H/n_L, -n_L/n_H$. We then use the eigenvectors to form the matrix

$$\mathbf{P} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

and

$$\mathbf{P}^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

We can then diagonalize \mathbf{M}_s as $\mathbf{D} = \mathbf{P} \mathbf{M}_s \mathbf{P}^{-1} = \begin{pmatrix} -n_H/n_L & 0 \\ 0 & -n_L/n_H \end{pmatrix}$ which has the eigenvalues of \mathbf{M}_s on the diagonal. Doing this we see that $\mathbf{M}_s = \mathbf{P}^{-1} \mathbf{D} \mathbf{P}$ and

$$\begin{aligned} \mathbf{M}_N &= (\mathbf{M}_s)^N \\ &= \mathbf{P}^{-1} \mathbf{D}^N \mathbf{P} \\ &= \frac{1}{2} \begin{pmatrix} (n_H/n_L)^{2N} + 1 & (n_H/n_L)^{2N} - 1 \\ (n_H/n_L)^{2N} - 1 & (n_H/n_L)^{2N} + 1 \end{pmatrix}. \end{aligned}$$

The reflectivity is

$$R_N = \left| \frac{M_{N,21}}{M_{N,11}} \right|^2 = \left[\frac{(n_H/n_L)^{2N} - 1}{(n_H/n_L)^{2N} + 1} \right]^2. \quad (3.31)$$

We see that for $n_H > n_L$, $R_N \rightarrow 1$ as $N \rightarrow \infty$.

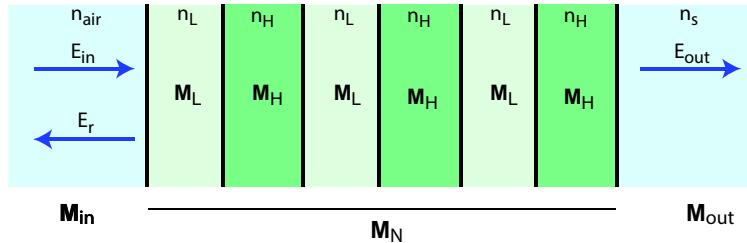


Figure 3.18: High reflectivity coating made of alternating quarterwave layers of high and low index with air on the left and a substrate with index n_s on the right.

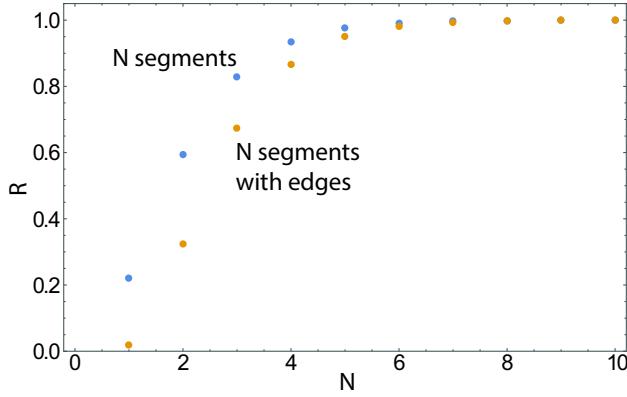


Figure 3.19: High reflectivity coating without (blue dots) and with (yellow dots) inclusion of edge layers for $n_L = 1.38$, $n_H = 2.30$ and $n_{\text{air}} = 1$, and $n_s = 1.5$.

Let's check the validity of this result when we include the edge layers. The full geometry is shown in Fig. 3.18. The matrix connecting air and substrate is

$$\begin{aligned} \mathbf{M} &= \mathbf{M}_{\text{air,L}} \mathbf{M}_N \mathbf{M}_{\text{H,s}} \\ &= \frac{1}{t_{\text{air,L}}} \begin{pmatrix} e^{-i\phi_{\text{air}}} & r_{\text{air,L}} e^{-i\phi_{\text{air}}} \\ r_{\text{air,L}} e^{i\phi_{\text{air}}} & e^{i\phi_{\text{air}}} \end{pmatrix} \mathbf{M}_N \frac{1}{t_{\text{H,s}}} \begin{pmatrix} e^{-i\phi_s} & r_{\text{H,s}} e^{-i\phi_s} \\ r_{\text{H,s}} e^{i\phi_s} & e^{i\phi_s} \end{pmatrix}. \end{aligned}$$

To evaluate this expression use $n_{\text{air}} = 1$, $r_{\text{air,L}} = \frac{n_{\text{air}} - n_L}{n_{\text{air}} + n_L}$, $t_{\text{air,L}} = \frac{2n_{\text{air}}}{n_{\text{air}} + n_L}$, $r_{\text{H,s}} = \frac{n_H - n_s}{n_H + n_s}$, $t_{\text{H,s}} = \frac{2n_H}{n_H + n_s}$, $\phi_{\text{air}} = 0$, $\phi_s = \pi/2$. The coating matrix is then

$$\mathbf{M} = \frac{in_s}{2n_H} \begin{pmatrix} -\left(\frac{n_H}{n_L}\right)^{2N} - \frac{n_L n_H}{n_{\text{air}} n_s} & \left(\frac{n_L}{n_H}\right)^{2N} - \frac{n_L n_H}{n_{\text{air}} n_s} \\ -\left(\frac{n_H}{n_L}\right)^{2N} + \frac{n_L n_H}{n_{\text{air}} n_s} & \left(\frac{n_L}{n_H}\right)^{2N} + \frac{n_L n_H}{n_{\text{air}} n_s} \end{pmatrix}.$$

The intensity reflectivity is

$$R = \left| \frac{M_{21}}{M_{11}} \right|^2 = \left[\frac{\left(\frac{n_H}{n_L}\right)^{2N} - \frac{n_L n_H}{n_{\text{air}} n_s}}{\left(\frac{n_H}{n_L}\right)^{2N} + \frac{n_L n_H}{n_{\text{air}} n_s}} \right]^2. \quad (3.32)$$

Comparing (3.31) with (3.32) we see that inclusion of the edge layers has no effect if $\frac{n_L n_H}{n_{\text{air}} n_s} = 1$ which simply corresponds to the air and substrate having the same refractive indices as the coating layers. For the coating indices used above $n_L = 1.38$, $n_H = 2.30$ and $n_{\text{air}} = 1$, $n_s = 1.5$ corresponding to glass we find $\frac{n_L n_H}{n_{\text{air}} n_s} = 2.12$. Figure 3.19 shows the reflectivity including the edge contributions for different numbers of layers. We see that the edge layers reduce the reflectivity when there are only a few layers but have a minimal effect when N is large.

Various filter morphologies can be constructed with similar methods. Coatings that have bandpass characteristics, specific wavelength or angular responses, etc. can be synthesized using multiple layers with different thicknesses and indices. It is generally more difficult to

design a broadband anti-reflection coating than a high reflector. If light is reflected from any of the layers it will exit back towards the source and be lost whereas for a high reflector if the initial layers do not reflect the light, subsequent layers can still be effective. As in the case of lens design this subject is an art, but there is much accumulated knowledge available in specialized books on the topic[19].

Chapter 4

Geometrical optics

The short wavelength limit of the Maxwell equations describes rays that propagate in straight lines in homogeneous media. In this chapter we will explore the effect of planar and curved surfaces on light rays which propagate along straight lines between surfaces. We will justify this statement using the eikonal in Sec. 4.5.

4.1 Reflection and refraction

Light rays are reflected and refracted at interfaces as seen in Fig. 4.1. On reflection the incident and reflected angles are equal, $\theta_r = \theta_i$. On refraction we have Snell's law

$$n_1 \sin \theta_i = n_2 \sin \theta_t$$

where the angles are measured from the interface normal. Snell's law can be derived from the Maxwell equations or from Fermat's principle of least time.

4.1.1 Paraxial Ray Matrices

Following the path of a ray through multiple surfaces becomes cumbersome due to the sin that appears in Snell's law. For small angles $\sin \theta \approx \theta$ and the algebra is greatly simplified.

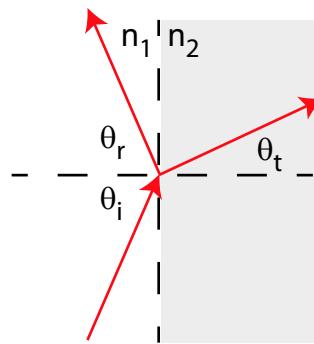


Figure 4.1: Reflection and transmission at an interface.

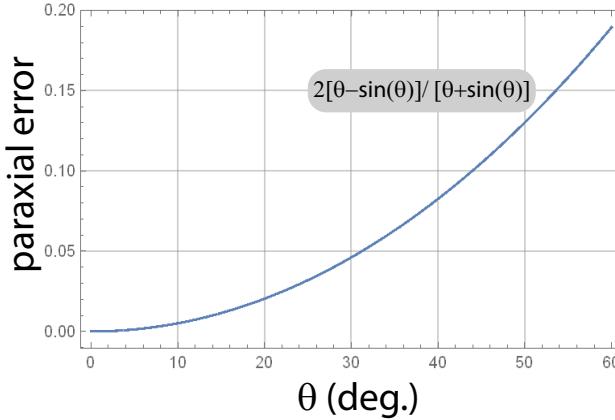


Figure 4.2: Fractional error in approximating $\sin(\theta)$ by θ .

This approximation corresponds to a near axial or *paraxial* description. Using a paraxial approximation the propagation of rays is efficiently described using ray matrices.

A useful abstraction is the concept of an optical ray that propagates in a straight line, without diffracting. It is an idealized construction that corresponds to the mathematical limit of $\lambda \rightarrow 0$. In a plane perpendicular to the optical axis z a ray is described fully by its transverse coordinate x and angle with the z axis θ . Propagation through a length of homogeneous space l gives $x(l) = x(0) + l \tan(\theta(0))$ and $\theta(l) = \theta(0)$. If we limit ourselves to rays that make a small angle with respect to the propagation axis then $\tan \theta \approx \theta$ (this is referred to as paraxial propagation) and the transformation of the ray parameters is

$$\begin{pmatrix} x \\ \theta \end{pmatrix} \rightarrow \begin{pmatrix} x' \\ \theta' \end{pmatrix} = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \theta \end{pmatrix}. \quad (4.1)$$

If we define a vector $\mathbf{s} = (x, \theta)$ propagation can be written as $\mathbf{s}(l) = \mathbf{M}_l \mathbf{s}(0)$ with

$$\mathbf{M}_l = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix}.$$

As indicated, the elements of ray matrices are conventionally referred to by the letters A, B, C, D . The error incurred by the paraxial approximation is shown in Fig. 4.2. For angles of not more than 30 deg. from the optical axis the fractional error is less than 5%.

Paraxial propagation through any optical element can be characterized in terms of the $ABCD$ matrix for that element. For example a thin lens of focal length f transforms rays as shown in Fig. 4.3. The matrix connecting the ray parameters directly before and after the lens leaves the position x unchanged so $A = 1$ and $B = 0$. When $x = 0$ the input and output angles are equal so $D = 1$. Finally, we see that when $x = f\theta$, $\theta' = Cx + \theta = 0$ so $C = -1/f$. The thin lens transformation is therefore

$$\begin{pmatrix} x' \\ \theta' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} x \\ \theta \end{pmatrix}. \quad (4.2)$$

More complex systems are described by a transformation matrix \mathbf{M} found from the matrix product of the individual matrices: $\mathbf{M} = \prod_i \mathbf{M}_i$. For future reference the ray matrices for

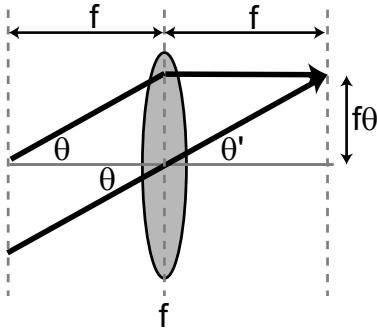


Figure 4.3: Ray propagation through a thin lens.

some common optical elements are given in Fig. 4.4. Note that if the optical element is tilted with respect to the optical axis we need to distinguish between tangential and sagittal rays. Tangential rays propagate in the plane that is normal to the axis about which the optical element is rotated. Sagittal rays lie in the plane containing the axis of rotation. Attention should be drawn to the fact that some authors use a different convention for the ray vector that replaces θ by $n\theta$ where n is the local refractive index. See the book, A. E. Siegman, Lasers, for details. Our definitions coincide with those in the original paper of Kogelnik and Li from 1966[20].

The elemental ray matrices have unit determinant, and hence any extended system is described by a matrix of unit determinant¹. There are thus only three independent components of the ray matrix. It is an exceedingly useful result that the ray matrices found from consideration of the propagation of rays, without account of diffraction, can be used to describe the diffractive propagation of light through a wide class of systems. This is discussed in Sec. 5.1.8.

4.1.2 Sign conventions

The sign conventions we use for ABCD matrices are as follows.

- 1) Rays travel from left to right.
- 2) Positions above (below) the optical axis have $x > 0$ ($x < 0$).
- 3) Rays pointing towards positive x (negative x) have $\theta > 0$ ($\theta < 0$).
- 4) Curved surfaces with center of curvature to the right (left) of the surface have $R > 0$ ($R < 0$).
- 5) Object distances are positive (negative) when the object is to the left (right) of the optical element.
- 6) Image distances are positive (negative) when the image is to the right (left) of the optical element.

Note that these definitions are not universal and different sign conventions can be found in different texts.

¹This is only true when the input and output planes are in media with the same refractive index. Otherwise the determinant is given by $AD - BC = n_1/n_2$.

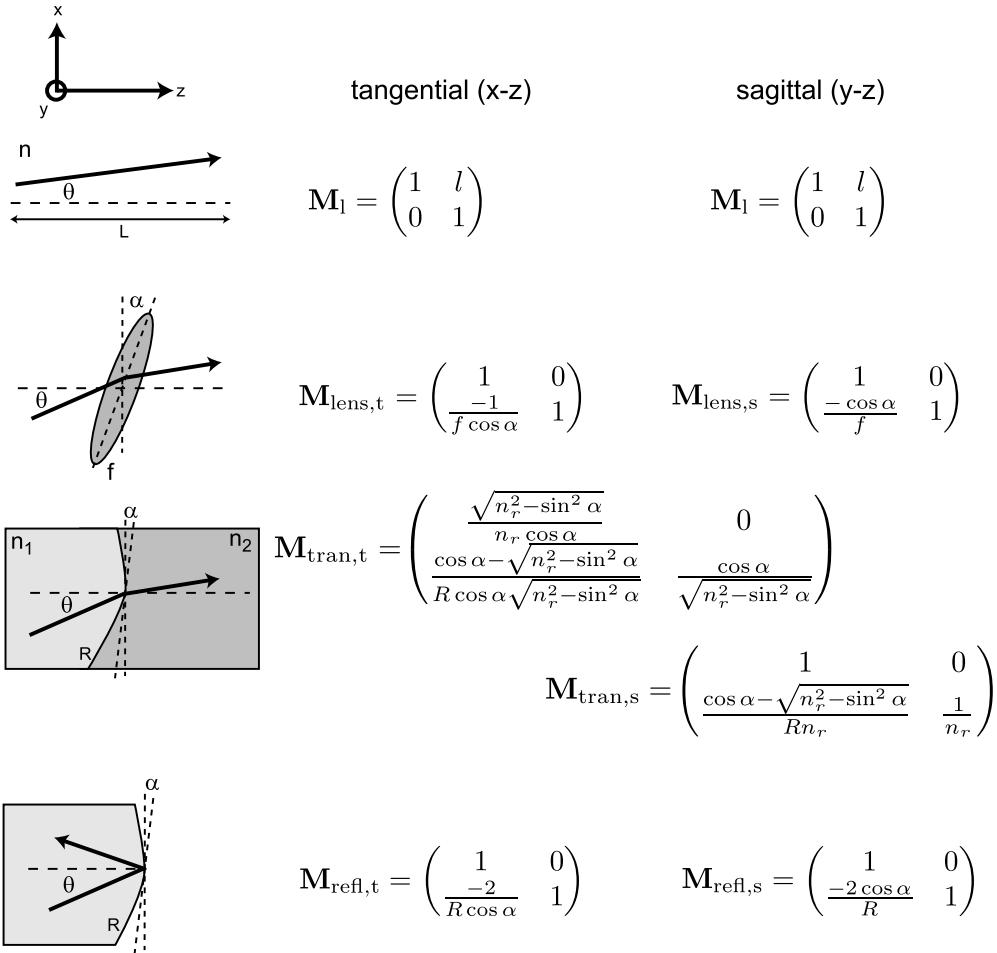


Figure 4.4: Ray matrices in the tangential and sagittal planes. The relative refractive index is $n_r = n_2/n_1$. In the fourth and fifth lines $R > 0$ for the interface concave to the right. The matrices for transmission through the tilted interface can be found in [21].

4.1.3 Ray matrix for a concave mirror

As an example of how to derive the ray matrix for a more complicated geometry consider the concave mirror in Fig. 4.5. The center of curvature of the mirror is at C , the object is at P and the image is formed at Q . The object distance is $\overline{OP} = d_o$ and the image distance is $\overline{OQ} = d_i$. Using the law of sines we have $R/\sin(\theta_1) = (d_o - R)/\sin(\theta)$ and $(R - d_i)/\sin(\theta') = R/\sin(\pi - \theta_2) = R/\sin(\theta_2)$. Reflection on the mirror gives $\theta' = \theta$ from which it follows that

$$\sin(\theta) = \frac{d_o - R}{R} \sin(\theta_1) = \frac{R - d_i}{R} \sin(\theta_2)$$

or

$$\frac{\sin(\theta_1)}{\sin(\theta_2)} = \frac{R - d_i}{d_o - R}.$$

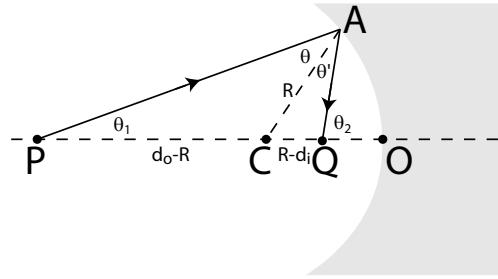


Figure 4.5: Geometry for imaging with a concave mirror.

We then make the paraxial approximation and write this as or

$$\frac{\theta_1}{\theta_2} = \frac{R - d_i}{d_o - R}.$$

Then use $d_o\theta_1 = d_i\theta_2$ so

$$\frac{d_i}{d_o} = \frac{R - d_i}{d_o - R}.$$

Rearranging gives $2/R = 1/d_o + 1/d_i$ and the focal length of the mirror is $f = R/2$. The ray matrix is thus

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix}.$$

A derivation of the more complex case where the mirror is tilted with respect to the optical axis so the matrix is different in the tangential and sagittal planes can be found in [21].

4.1.4 Imaging with a thin lens

We can use the matrices for linear propagation and a thin lens to derive the formula for imaging with a lens. An imaging system takes rays from an object plane at $z = 0$ to an image plane at $z = L$. The geometry is shown in Fig. 4.6. If the lens is at $z = d_1$ and $d_2 = L - d_1$ the transformation matrix is

$$\begin{pmatrix} x' \\ \theta' \end{pmatrix} = \mathbf{M}_l(d_2) \mathbf{M}_{\text{lens}}(f) \mathbf{M}_l(d_1) \begin{pmatrix} x \\ \theta \end{pmatrix}.$$

Note that the matrices should be multiplied with the matrix of the first element on the right, progressing to the matrix of the last element on the left. Multiplying out gives a composite matrix

$$\mathbf{M} = \begin{pmatrix} 1 - \frac{d_2}{f} & d_1 + d_2 - \frac{d_1 d_2}{f} \\ -\frac{1}{f} & 1 - \frac{d_1}{f} \end{pmatrix}. \quad (4.3)$$

Image formation means that all rays that start at x end at the same position x' irrespective of the value of θ . This occurs when $B = 0$ or

$$\frac{1}{f} = \frac{1}{d_1} + \frac{1}{d_2}. \quad (4.4)$$

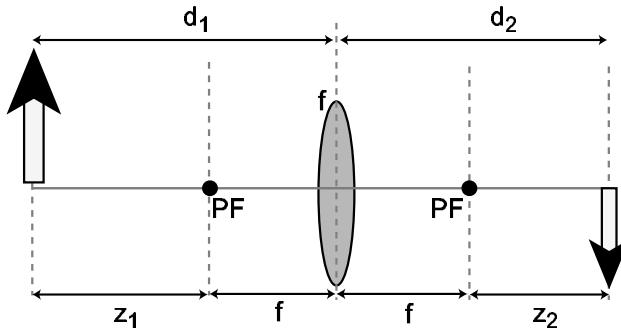


Figure 4.6: Imaging with a thin lens. The object and image distances are d_1 and d_2 , respectively. The distances measured from the principal foci are z_1 and z_2 .

The magnification is given by $x'/x = 1 - d_2/f = -d_2/d_1$.

The imaging condition can be written in a different form by introducing the principal foci shown in Fig. 4.6. The principal focus is the point which a parallel bundle of rays propagating along the optical axis are focused to. For a thin lens the principal foci are a distance f before and after the lens. The distances measured from the principal foci are $z_1 = d_1 - f$ and $z_2 = d_2 - f$. Introducing these quantities into (4.4) and rearranging we find

$$z_1 z_2 = f^2$$

which is often referred to as the Newtonian form of the lens equation.

4.1.5 Negative lens

A lens with index greater than 1 and concave surfaces has a negative focal length. An incident bundle of rays parallel to the optical axis will become divergent after passage through the lens so no real image is formed to the right. Nonetheless there is an image plane containing a virtual image and located to the left of the lens as shown in Fig. 4.7.

To analyze this situation we put $f = -|f|$ so that Eq. (4.3) becomes

$$\mathbf{M} = \begin{pmatrix} 1 + \frac{d_2}{|f|} & d_1 + d_2 + \frac{d_1 d_2}{|f|} \\ \frac{1}{|f|} & 1 + \frac{d_1}{|f|} \end{pmatrix}. \quad (4.5)$$

The image plane is at

$$d_2 = -\frac{d_1 |f|}{d_1 + |f|}$$

which is negative for $d_1 > 0$. In other words when the object is in front of the lens an image is formed to the left of the image. This is now a virtual image with positive magnification given by

$$A = 1 + \frac{d_2}{|f|} = \frac{|f|}{d_1 + |f|}$$

which is positive for $d_1 > -|f|$.

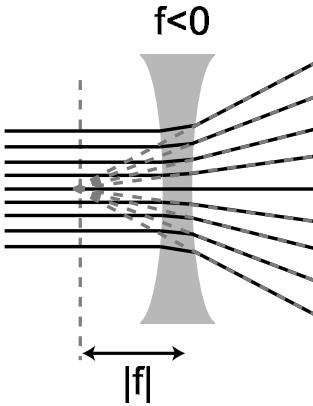


Figure 4.7: A negative lens forms a virtual image to the left of the lens.

4.1.6 Compound lens

We can form a compound lens by placing two lenses with focal lengths f_1, f_2 next to each other with separation d as shown in Fig. 4.8. Propagation from object to image planes is described by a matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

The matrix elements are

$$A = \frac{d(d_2 - f_2) + f_1 f_2 - d_2(f_1 + f_2)}{f_1 f_2}, \quad (4.6a)$$

$$B = \frac{d(d_1 - f_1)(d_2 - f_2) + d_2 f_1 f_2 - d_1[d_2(f_1 + f_2) - f_1 f_2]}{f_1 f_2}, \quad (4.6b)$$

$$C = \frac{d - f_1 - f_2}{f_1 f_2}, \quad (4.6c)$$

$$D = \frac{d(d_1 - f_1) + f_1 f_2 - d_1(f_1 + f_2)}{f_1 f_2}. \quad (4.6d)$$

Setting $C = -1/f$ we find for the effective focal length of the compound lens

$$f = \frac{f_1 f_2}{f_1 + f_2 - d}. \quad (4.7)$$

Note that when $d = 0$ then $f = f_1 f_2 / (f_1 + f_2)$ or $1/f = 1/f_1 + 1/f_2$. Thus adjacent lenses combine like parallel resistors as regards the composite focal length. The focusing power proportional to $1/f$ combines like resistors in series.

The image plane is located at the value of d_2 for which the B element of the propagation matrix is zero. Solving $B = 0$ gives

$$d_{2,\text{image}} = \frac{(dd_1 - df_1 - d_1 f_1) f_2}{dd_1 - df_1 - d_1 f_1 - d_1 f_2 + f_1 f_2}.$$

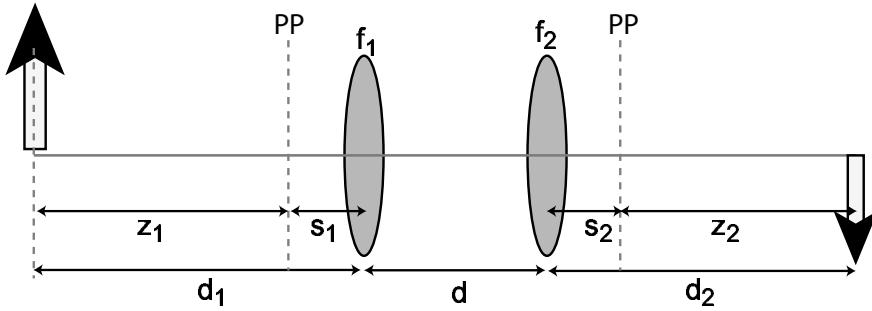


Figure 4.8: Compound lens composed of two thin lenses. The principal planes are labeled PP.

When $d_1 = f_1$ then $d_{2,\text{image}} = f_2$, independent of the value of d . This is to be expected since when $d_1 = f_1$ the first lens creates a parallel bundle of rays from the object. The parallel bundle is then focused to an image at the principal focus located at $d_2 = f_2$.

We can also find the image magnification, which is given by the A matrix element. Substituting $d_2 \rightarrow d_{2,\text{image}}$ we find

$$A = \frac{f_1 f_2}{d(d_1 - f_1) + f_1 f_2 - d_1(f_1 + f_2)}.$$

Note that $A \neq d_{2,\text{image}}/d_1$ which is given by

$$\frac{d_{2,\text{image}}}{d_1} = \frac{[dd_1 - (d + 2d_1)f_1]f_2}{d_1[d(d_1 - f_1) + f_1 f_2 - d_1(f_1 + f_2)]}.$$

Although the compound lens acts as though it has an effective focal length given by (4.7) the simple relations between focal length, image and object distances, and magnification that described a thin lens no longer apply.

4.1.7 Principal Planes

The concept of principal planes allows us to describe a compound lens as an effective thin lens, provided we measure the object and image distances from the principal planes. Let the distances from the principal planes be z_1, z_2 as shown in Fig. 4.8, with the principal planes located at distances s_1, s_2 from the lenses. We wish to choose the position of the principal planes such that the transformation matrix for the compound lens takes the same form as the transformation matrix for a thin lens, as given by Eq. (4.3), except that d_1, d_2 are replaced by z_1, z_2 . Explicitly we solve

$$\begin{pmatrix} A & B \\ -\frac{1}{f} & D \end{pmatrix} = \begin{pmatrix} 1 - \frac{z_2}{f} & z_1 + z_2 - \frac{z_1 z_2}{f} \\ -\frac{1}{f} & 1 - \frac{z_1}{f} \end{pmatrix}$$

with $z_1 = d_1 - s_1$, $z_2 = d_2 - s_2$. The matrix entries are given above and solving we find

$$s_1 = \frac{df_1}{d - f_1 - f_2}, \quad (4.8a)$$

$$s_2 = \frac{df_2}{d - f_1 - f_2}. \quad (4.8b)$$

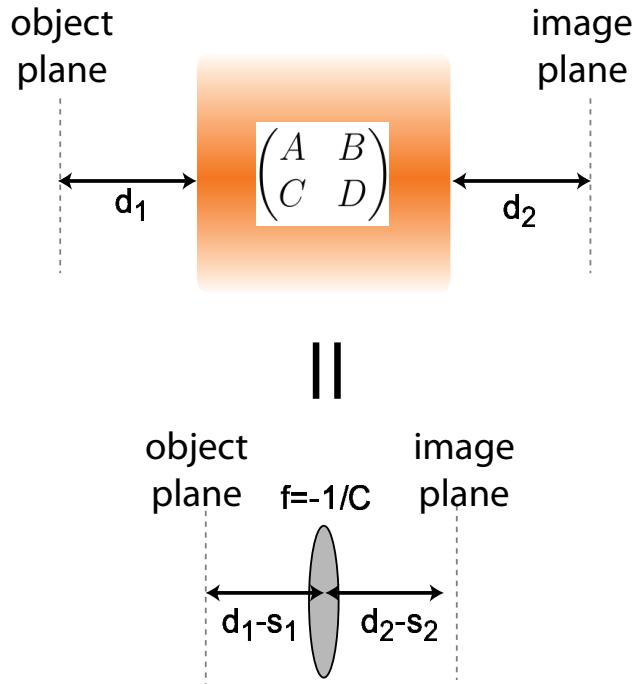


Figure 4.9: Principal plane analysis. A composite optical system (top) is equivalent to an effective thin lens with the propagation distances measured to and from the principal planes (bottom).

Thus the compound lens acts on optical rays as an effective thin lens with the image and object distances measured from the principal planes s_1, s_2 .

The real utility of this result arises from the fact that it is not limited to a compound lens, but is true for an arbitrary optical transformation described by an ABCD matrix. Surprisingly, any optical system can be described as an effective thin lens, as regards propagation measured from the principal planes. Another way of thinking about the principal planes is to recognize that an arbitrary ABCD ray transformation acts as though all of the ray refraction takes place at the principal planes. This result greatly simplifies the analysis of complex, multi-element optical systems.

To prove this consider the situation shown in Fig. 4.9. The matrix for propagation between the principal planes is

$$\begin{aligned} \mathbf{M}_{pp-pp} &= \begin{pmatrix} 1 & s_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 1 & s_1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} A + s_2C & s_1A + B + s_1s_2C + s_2D \\ C & s_1C + D \end{pmatrix}. \end{aligned}$$

Setting this equal to $M_{\text{thin lens}} = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}$ we find a solution for

$$\begin{aligned} f &= -\frac{1}{C}, \\ s_1 &= \frac{1-D}{C}, \\ s_2 &= \frac{1-A}{C}. \end{aligned}$$

Let's verify that we reproduce the principal plane locations found for the compound lens. The ABCD matrix for the two lenses is

$$\begin{pmatrix} 1 & 0 \\ -1/f_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_1 & 1 \end{pmatrix} = \begin{pmatrix} 1-d/f_1 & d \\ \frac{d-f_1-f_2}{f_1f_2} & 1-d/f_2 \end{pmatrix}.$$

Thus the principal planes are at

$$\begin{aligned} s_1 &= \frac{df_1}{d-f_1-f_2}, \\ s_2 &= \frac{df_2}{d-f_1-f_2}, \end{aligned}$$

which agrees with Eqs. (4.8). The distances s_1, s_2 can be positive or negative. A positive(negative) s_1 means that the principal plane is to the left(right) of the edge of the optical element in the object space. A positive(negative) s_2 means that the principal plane is to the right(left) of the edge of the optical element in the image space.

4.1.8 Lens Maker's formula

As another example of the use of the ray matrices consider the thick lens shown in Fig. 4.10. The first surface has radius of curvature $R_1 > 0$, the lens material has index n , and the second surface has radius of curvature $R_2 < 0$. Using the appropriate matrices from Fig. 4.4 the matrix for the lens is

$$\begin{aligned} \mathbf{M} &= \begin{pmatrix} 1 & 0 \\ (n-1)/R_2 & n \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ ((1-n)/nR_1) & 1/n \end{pmatrix} \\ &= \begin{pmatrix} 1 + \frac{d(1-n)}{nR_1} & \frac{d}{n} \\ \frac{(n-1)[d(1-n)+n(R_1-R_2)]}{nR_1R_2} & 1 + \frac{d(n-1)}{nR_2} \end{pmatrix}. \end{aligned}$$

The focal length is given by the lens maker's formula

$$\frac{1}{f} = (n-1) \left[\frac{1}{R_1} - \frac{1}{R_2} + \frac{(n-1)d}{nR_1R_2} \right]. \quad (4.9)$$

The principal planes are located at

$$\begin{aligned} s_1 &= \frac{dR_1}{(n-1)d - n(R_1 - R_2)}, \\ s_2 &= \frac{dR_2}{d(1-n) + n(R_1 - R_2)}. \end{aligned}$$

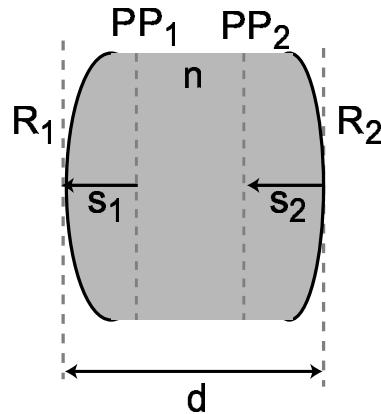


Figure 4.10: Thick lens.

Let's consider some numerical examples. For a symmetric bi-convex lens take $R_1 = 10$ cm, $R_2 = -10$ cm, $d = 0.2$ cm, and $n = 1.5$. These values give

$$f = 10.03 \text{ cm}, \quad s_1 = -0.067 \text{ cm}, \quad s_2 = -0.067 \text{ cm}.$$

We see that the principal planes are at negative distances, i.e. they are inside the lens, approximately 1/3 of the thickness from each surface. For a negative bi-convex lens take the same values except $R_1 = -10$ cm, $R_2 = 10$ cm. We find

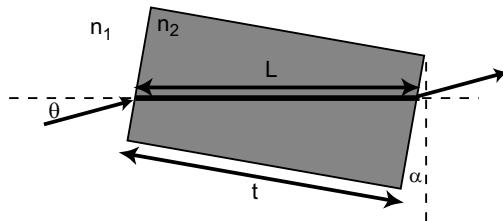
$$f = -9.97 \text{ cm}, \quad s_1 = -0.066 \text{ cm}, \quad s_2 = -0.066 \text{ cm}.$$

For a plano-convex lens we could use $R_1 = \infty$ cm, $R_2 = -10$ cm, $d = 0.2$ cm, $n = 1.5$ giving

$$f = 20. \text{ cm}, \quad s_1 = -0.13 \text{ cm}, \quad s_2 = 0. \text{ cm}.$$

4.1.9 Optical plates

As an example of the use of the ray matrices to analyze other optical elements consider the slab of index n_2 shown in Fig. 4.11. The thickness of the slab is t , and it is rotated so that the axis of the slab is at an angle α with respect to the propagation axis. The internal ray length is then $L = t/\sqrt{1 - n_1^2 \sin^2(\theta + \alpha)/n_2^2}$. Using the matrices given in Fig. 4.4 we

Figure 4.11: A tilted slab of index n_2 and thickness t .

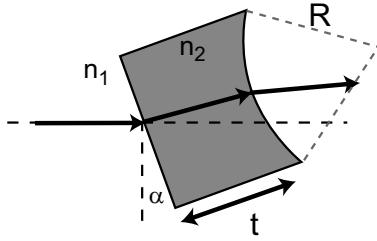


Figure 4.12: A tilted mirror with radius of curvature R , thickness t , and substrate index n_2 .

can find the matrix describing propagation through the slab. The general expression is not compact. Of most interest is the situation where the beam is incident at Brewster's angle for which $\tan(\theta + \alpha) = n_2/n_1$ and $L = t\sqrt{1 + n_2^2/n_1^2}/(n_2/n_1)$. The matrix of the slab for tangential rays is then

$$\mathbf{M}_t = \begin{pmatrix} \frac{n_1}{n_2} & 0 \\ 0 & \frac{n_2^2}{n_1^2} \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{n_2}{n_1} & 0 \\ 0 & \frac{n_1^2}{n_2^2} \end{pmatrix} = \begin{pmatrix} 1 & \frac{\sqrt{1+(\frac{n_2}{n_1})^2}}{(\frac{n_2}{n_1})^2} t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{L}{(\frac{n_2}{n_1})^3} \\ 0 & 1 \end{pmatrix} \quad (4.10)$$

and for sagittal rays²

$$\mathbf{M}_s = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n_2}{n_1} \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{pmatrix} = \begin{pmatrix} 1 & \frac{\sqrt{1+(\frac{n_2}{n_1})^2}}{(\frac{n_2}{n_1})^2} t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{L}{(\frac{n_2}{n_1})} \\ 0 & 1 \end{pmatrix}. \quad (4.11)$$

An optical element that is often used for optical pumping of laser or nonlinear optical cavities is a tilted, partially transmitting mirror as shown in Fig. 4.12. Application of the ray matrices from Fig. 4.4 leads to

$$\mathbf{M}_t = \begin{pmatrix} \frac{e_1 e_2}{\cos^2(\alpha)} & \frac{e_1 t}{e_2 n_r^2 \cos \alpha} \\ \frac{e_2 n_r}{e_1 R \cos \alpha} \left(\frac{e_1}{n_r \cos \alpha} - 1 \right) & 1 + \frac{t}{e_2 n_r^2 R} \left(\sec \alpha - \frac{n_r}{e_1} \right) \end{pmatrix} \quad (4.12a)$$

$$\mathbf{M}_s = \begin{pmatrix} 1 & \frac{t}{e_2 n_r} \\ \frac{1}{R} (e_1 - n_r \cos \alpha) & 1 + \frac{t}{e_2 n_r R} (e_1 - n_r \cos \alpha) \end{pmatrix} \quad (4.12b)$$

where $n_r = n_2/n_1$, $e_1 = \sqrt{1 - n_r^2 \sin^2 \alpha}$, and $e_2 = \sqrt{1 - \sin^2 \alpha / n_r^2}$. An initially round beam entering from the left is transformed into an elliptically shaped astigmatic beam. Optimization of a nonlinear cavity where the pump beam passes through this type of tilted curved mirror requires careful matching to the cavity mode or precompensation of the pump beam to balance the mirror induced astigmatism.

²These results agree with Appendix A in [22] and with Eqs. (A.7) and (A.10) in [23], after accounting for their use of a different convention for the ray vector.

4.1.10 Brightness

An important question is the possibility of increasing the brightness of an optical beam using a system of optical elements described by $ABCD$ matrices. We can think of an optical beam as being a bundle of rays with a characteristic width Δx and angular divergence $\Delta\theta$. In one transverse dimension the brightness of a beam is defined as $B = P/(\Delta x)(\Delta\theta)$, where P is the power.

Transformation by an optical system described by matrix \mathbf{M} results in $x' = Ax + B\theta$ and $\theta' = Cx + D\theta$. Therefore

$$\begin{aligned} (\Delta x')^2 &\equiv \langle (x' - \langle x' \rangle)^2 \rangle \\ &= \langle x'^2 \rangle - \langle x' \rangle^2 \\ &= A^2(\Delta x)^2 + B^2(\Delta\theta)^2 + 2AB(\langle x\theta \rangle - \langle x \rangle \langle \theta \rangle) \end{aligned} \quad (4.13)$$

and

$$(\Delta\theta')^2 = C^2(\Delta x)^2 + D^2(\Delta\theta)^2 + 2CD(\langle x\theta \rangle - \langle x \rangle \langle \theta \rangle). \quad (4.14)$$

We assume that the input beam is centered on the x axis so $\langle x \rangle = 0$ and that the x and θ distributions are uncorrelated so that $\langle x\theta \rangle = 0$. We then have, using $AD - BC = 1$,

$$\begin{aligned} (\Delta x')^2(\Delta\theta')^2 &= (A^2D^2 + B^2C^2)(\Delta x)^2(\Delta\theta)^2 + A^2C^2(\Delta x)^2 + B^2D^2(\Delta\theta)^2 \\ &= (1 + 2ABCD)(\Delta x)^2(\Delta\theta)^2 + A^2C^2(\Delta x)^2 + B^2D^2(\Delta\theta)^2 \\ &= (\Delta x)^2(\Delta\theta)^2 + (AC\Delta x + BD\Delta\theta)^2 \\ &\geq (\Delta x)^2(\Delta\theta)^2. \end{aligned} \quad (4.15)$$

Thus $\Delta x'\Delta\theta' \geq \Delta x\Delta\theta$ so $B' \leq B$. The brightness cannot be increased by a linear optical system. The brightness is maximized when $AC = 0$ and $BD = 0$. Since $AD - BC = 1$ it is impossible for $A \& B$ or $C \& D$ to simultaneously vanish so the condition for maximal brightness is $A = D = 0$ or $B = C = 0$. We can identify the product $\Delta x\Delta\theta$ as the phase space density of the beam. The impossibility of reducing the phase space density is an example of a general result from classical mechanics known as Liouville's Theorem (see for example Landau & Lifshitz, Mechanics).

4.2 Aberrations

The paraxial approximation is just that, an approximation. A more careful treatment that accurately follows the refraction at each interface without any approximation to $\sin(\theta)$ shows that even when the imaging condition is fulfilled rays originating from a single object point may not converge to a single point in the image plane. In addition the separation of object points may be distorted in the image plane. These aberrations become more pronounced as the ray angles with respect to the optical axis increase. The variation of refractive index with wavelength leads to chromatic aberrations, even for small ray angles.

Aberrations can be corrected for using multiple lenses with different types of glass designed to compensate both chromatic and other errors. The study of aberrations was originally based on analytical methods and there exist extensive classical results. With the advent of optical design software for ray tracing the study of aberrations for even very complex optical systems can now be quickly performed on computers.

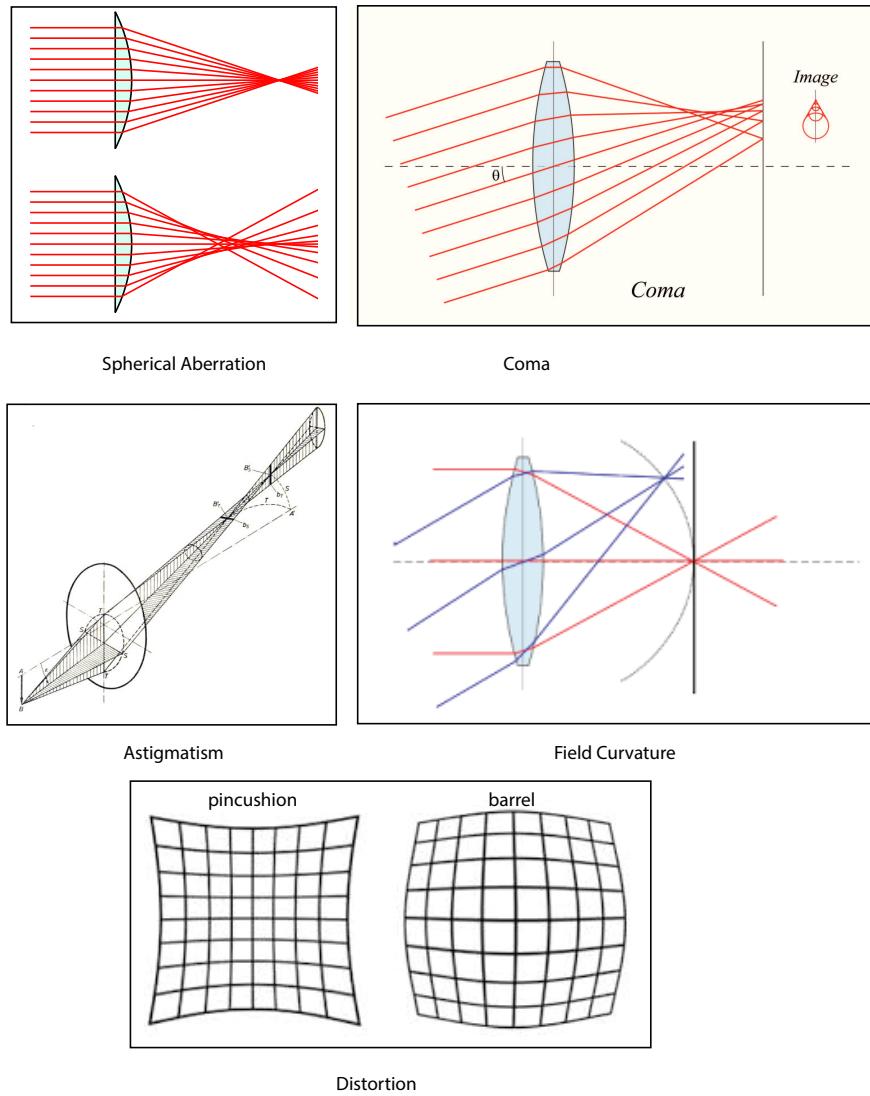


Figure 4.13: Seidel aberrations.

4.2.1 Primary monochromatic aberrations

Despite the availability of computer programs for ray tracing it is still useful to understand the basic types of aberration that occur. Aberrations that depend on the wavelength of the light due to dispersion of the optical refractive index, $n = n(\lambda)$ are called chromatic aberrations. Aberrations that occur for a single wavelength are called monochromatic aberrations. The most important monochromatic aberrations can be categorized analytically by approximating

$$\sin(\theta) \simeq \theta - \theta^3/6, \quad \cos(\theta) \simeq 1 - \theta^2/2, \quad \tan(\theta) \simeq \theta + \theta^3/3.$$

This approximation leads to five types of third order aberration. These were categorized in a series of 1856 papers by Seidel[24, 25, 26] who considered a centered system of spherical surfaces and are known as the Seidel aberrations shown in Fig. 4.13. They are

- Spherical aberration. Rays from a point on the optical axis do not converge to a single point on axis.
- Coma. Rays from an off-axis image point do not converge to a single point. The distribution of the rays looks like a comet, hence the name coma.
- Astigmatism. Rays in the tangential and sagittal planes are focused at different axial locations.
- Field curvature. The surface on which rays are imaged is not planar but lies on a curved surface.
- Distortion. Rays from an object point converge to a single image point but the distance of the image point from the optical axis is not proportional to the distance of the object point.

Spherical aberration calculation

Consider the geometry of Fig. 4.14. A ray with angle θ_1 refracts at a surface of curvature R at height $x_1 = h$. The normal to the surface makes an angle α with the optical axis. The local normal to the surface can be written as $\hat{\mathbf{n}} = -\cos(\alpha)\hat{z} + \sin(\alpha)\hat{x}$ with $\sin(\alpha) = h/R$. The ray starts in a medium with index n_1 and refracts into a medium with index n_2 . After refraction the ray is at height $x_2 = h$ and has angle with respect to the surface normal of

$$\theta' = \sin^{-1} \left[\frac{n_1}{n_2} \sin(\theta_1 + \alpha) \right]. \quad (4.16)$$

The refracted ray meets the planar back surface of the lens at height h' with angle from the normal of $\theta'' = \alpha - \theta'$. The height h' is given by

$$h' = h - t' \tan(\theta'') \quad (4.17)$$

with $t' = t - R(1 - \cos(\alpha))$. The ray then refracts giving a ray heading towards the optical axis with angle

$$\begin{aligned} \theta_2 &= \sin^{-1} \left[\frac{n_2}{n_1} \sin(\theta'') \right] \\ &= \sin^{-1} \left[\frac{n_2}{n_1} \sin \left(\alpha - \sin^{-1} \left[\frac{n_1}{n_2} \sin(\theta_1 + \alpha) \right] \right) \right]. \end{aligned} \quad (4.18)$$

The ray crosses the optical axis at a distance after the front surface of the lens given by

$$d = t + \frac{h'}{\tan(\theta_2)} \quad (4.19)$$

where t is the lens thickness.

Combining Eqs. (4.16 - 4.19) we obtain an expression for the distance d_o as a function of the ray height h . The paraxial approximation implies that the ratio $\eta = h/R$ is a small parameter. After some algebra we find for the simplest case of $\theta_1 = 0$ and $t = 0$

$$d = \frac{R}{n_r - 1} + \frac{n_r^2(2 - n_r) - 2}{2n_r(n_r - 1)} R\eta^2 + \mathcal{O}(\eta^4)$$

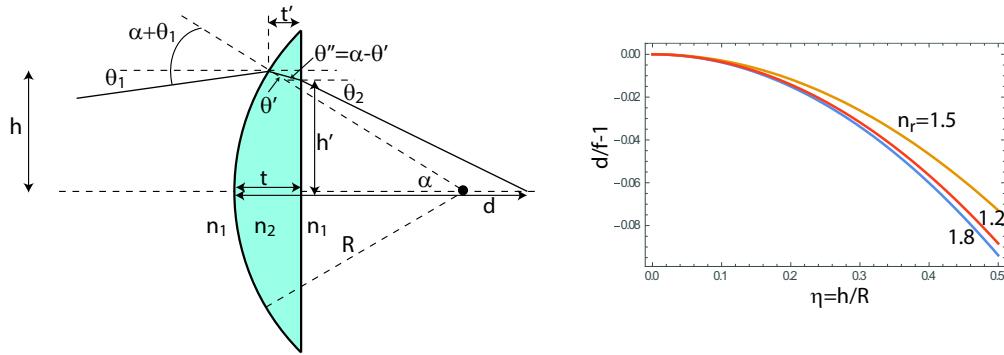


Figure 4.14: Spherical aberration of a plano-convex lens.

where the relative refractive index is $n_r = n_2/n_1$. From the lens maker's formula Eq. (4.9) the focal length is $f = R/(n_r - 1)$ so

$$\frac{d}{f} = 1 + \frac{n_r^2(2 - n_r) - 2}{2n_r} \eta^2 + \mathcal{O}(\eta^4). \quad (4.20)$$

The coefficient in front of η^2 determines the amount of spherical aberration of the lens. The fractional shift of the focal position as a function of η is shown in Fig. 4.14. It is a happy coincidence that for the plano-convex lens the spherical aberration is minimized for $n_r = 1.466$ which is close to the index of common glass.

The other Seidel aberrations can be analyzed by ray tracing in the same manner as we have done for the spherical aberration. The geometrical details are complicated and the results can be found in many optics books.

4.2.2 Chromatic aberration

In most glasses the refractive index increases towards shorter wavelengths. This can be seen for two common glass types in Fig. 4.15. Thus blue light experiences greater bending than red

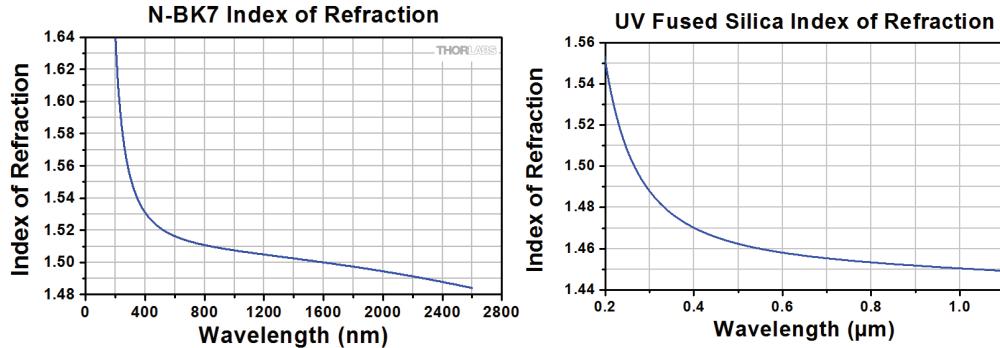


Figure 4.15: Refractive index of BK-7 (left) and fused silica (right) as a function of wavelength. Plots taken from Thorlabs catalog.

glass type	BK7	fused silica	SF11	
B_1	1.03961212	0.6961663	1.73759695	
B_2	0.231792344	0.4079426	0.313747346	
B_3	1.01046945	0.8974794	1.89878101	
C_1	0.0774642^2	0.0684043^2	0.114842^2	μm^2
C_2	0.141485^2	0.1162414^2	0.249613^2	μm^2
C_3	10.1765^2	9.896161^2	12.4594^2	μm^2
Abbe number	64.17	67.82	25.68	

Table 4.1: Sellmeier coefficients for some common glasses. Data from refractive index database at <https://refractiveindex.info>.

light. This leads to the image location varying with the wavelength which results in colored fringes. The chromatic aberration can lead to axial variations in image position (referred to as axial color) or transverse variations (referred to as lateral color). To correct for chromatic aberration the lens designer combines glasses with different dispersion characteristics to compensate for this effect. A lens that is designed to have no chromatic aberration at two distinct wavelengths is called an achromat. Such lenses will have relatively good performance over a broad range of wavelengths. Lenses that are corrected at three distinct wavelengths require more complex designs and are called apochromats.

The dispersion of transparent materials is conveniently described by the Sellmeier equation[27]

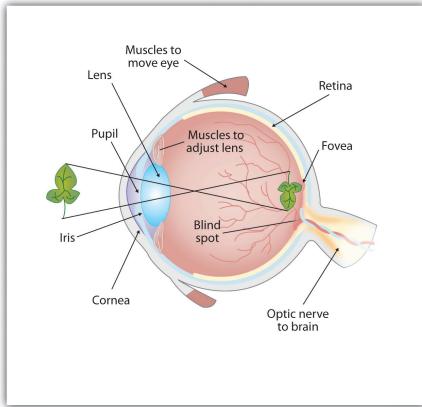
$$n^2(\lambda) = 1 + \sum_j \frac{B_j}{\lambda^2 - C_j} \lambda^2. \quad (4.21)$$

Here λ is the vacuum wavelength and the index depends on a sum over absorption resonances labeled with index j of strength B_j at wavelengths $\sqrt{C_j}$. For typical glasses the Sellmeier coefficients B_j, C_j are measured by fitting to experimental data over a range of wavelengths and a three term summation yields accurate results with a fractional error of $\sim 10^{-6}$. Some sample values are given in Table 4.1. The coefficients for common glasses correspond to two resonances in the deep ultraviolet and an infrared resonance near $10 \mu\text{m}$. This type of description is widely used for design of imaging instruments and for calculation of dispersion in crystals to determine phase matching conditions in nonlinear wave mixing processes. This latter application is discussed in more detail in Ch. 11.

The Sellmeier equation gives a full description of the wavelength dependent refractive index. It is also useful to have a single number that characterizes the amount of dispersion of an optical material. The most commonly used quantity is the Abbe number defined as

$$V_d = \frac{n_D - 1}{n_F - n_C}.$$

Here V is the Abbe number and n_D, n_F, n_C are the indices of refraction of the material at the wavelengths of the D, F and C Fraunhofer lines which correspond to atomic transitions at wavelengths of 589.3 nm (Sodium), 486.13 nm (Hydrogen), and 656.27 nm (Hydrogen) respectively. A large Abbe number indicates low dispersion whereas a small Abbe number is large dispersion.



<http://open.lib.umn.edu/intropsyc/chapter/4-2-seeing/>

Figure 4.16: Optics of the human eye.

4.3 Image forming instruments

4.3.1 The eye

The eye is a precision optical instrument that focuses light onto the retina. Focusing is done by the combination of the cornea and the eye lens. The cornea has a fixed focal length while the eye lens responds to muscle tension by changing shape and focusing power in order to bring objects at different distance into focus on the retina. The image on the retina is inverted and the wiring from the retina to the brain corrects this.

The image distance from the eye lens to the retina is fixed and is typically $d_i \simeq 1.7$ cm. The power of the lens is measured in diopters $D = 1/f$ which are defined as the inverse of the focal length in m. A typical person has an eye with a relaxed focusing power of $D_0 \simeq 58.8$ diopters of which 43 come from the cornea and the rest come from the eye lens which is variable. Muscle tension on the eye lens is used to increase the focusing power, an ability which decreases with age. A young person can typically achieve an additional 20 diopters, by age 25 about 10 diopters, and by age 50 about 1 diopter. This limits the ability to form an image of objects at short distance, hence the need for reading glasses.

In terms of the focusing power the object distance is³

$$d_o = \frac{1}{D - 1/d_i}.$$

Using $d_i = 0.017$ m the relaxed eye will focus a distant object onto the retina. Assuming an accommodation of $D_a = 6$ in an adult we get a total power of $D = D_0 + D_a = 64.8$ giving $d_0 = 16.7$ cm. A nominal value of 15 cm is typically assumed as the shortest distance an eye can focus to.

The size of an object viewed by the eye depends on the magnification $M = |d_i/d_o|$. At the nominal closest viewing distance of $d_o = 15$ cm the eye has a magnification of $M = 1.7/15 = 0.11$. This can be improved on with a simple magnifying glass which is a lens with a power D_{mg} . If the magnifying glass is brought close to the eye the focusing power becomes $D = D_0 + D_{mg}$.

³To be precise the imaging equation should be modified to account for the index of the fluid inside the eye. We will sidestep this complication.

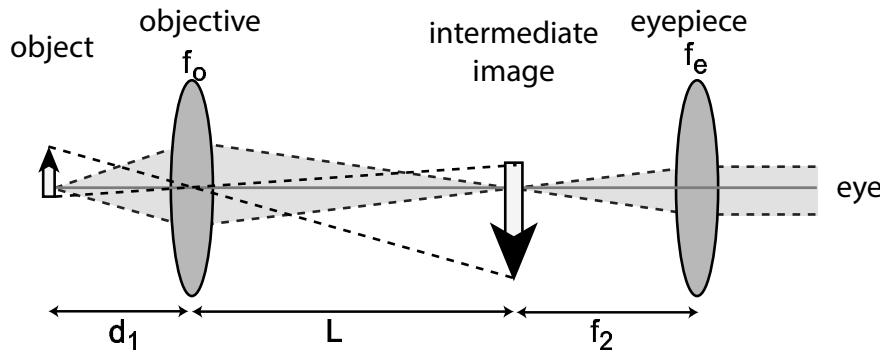


Figure 4.17: Microscope with objective lens and eyepiece.

Suppose the magnifying glass has power $D_{mg} = 40$ (this corresponds to $f = 2.5$ cm). Then a relaxed eye can image an object at distance $d = 1/(D - 1/d_i) = 2.5$ cm giving a magnification of $d_i/d = 1.7/2.5 = 0.68$ which is an increase of $\times 6$ relative to the unaided eye. Note that the magnifying glass forms a virtual image in front of the lens at a distance which can be comfortably viewed by the eye.

Another way to express the improvement is that we get a magnification by a factor of d_o/f where d_o is the shortest object distance of the unaided eye. Since shorter f requires thicker and more strongly curved lenses there is a practical limit of about $\times 25$ for a magnifying glass. This can be improved on using a microscope.

4.3.2 Microscope

The microscope as shown in Fig. 4.17 has two lenses, the objective lens which forms a magnified real image, followed by an eyepiece which gives an additional magnifying factor in the same way as the magnifying glass. Most modern microscopes are standardized to have a distance $L = 16$ cm between the objective and the intermediate real image. The objective is often a multi-element lens that has been designed to minimize aberrations. Focal lengths as short as $f_o \simeq 1$ mm are common. This give a magnification of 160 at the intermediate focus. With an eyepiece that gives an additional magnification of $\times 20$ total magnification factors of several thousand are possible.

The ultimate limit for the smallest object that can be viewed is not set by the magnification but by the wave nature of light and diffraction. As we will see later on the resolution limit of the objective lens is approximately $\lambda/(2 NA)$ where λ is the wavelength and NA is the numerical aperture of the lens. The numerical aperture is $NA = n_o \sin(\theta)$ where n_o is the refractive index surrounding the object and θ is the half opening angle of the lens. High resolution microscope objectives can have a numerical aperture $NA = 0.9$, or even greater than 1 for oil immersion devices. Thus objects smaller than the wavelength can be viewed.

4.3.3 Telescope

Telescopes are similar to microscopes, but designed to view very distant objects. A basic refracting telescope is shown in Fig. 4.18. In typical usage the object distance d_o is very large compared to the focal length of the objective lens f_o . Thus the intermediate image is formed close to the back focal plane of the lens. The eyepiece and the objective are separated by the sum of their focal lengths $L = f_o + f_e$ and in this configuration the ray matrix from the front of the objective lens to

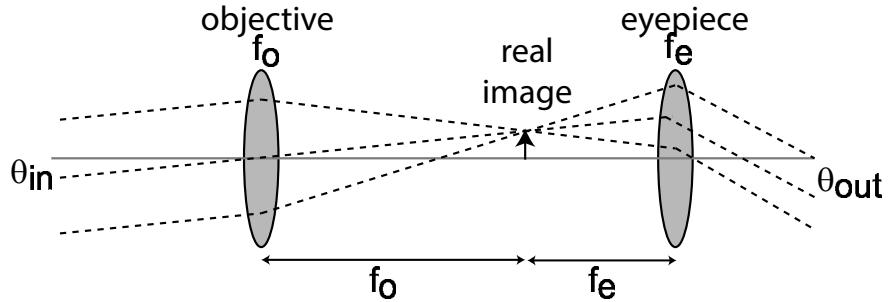


Figure 4.18: Refracting telescope with objective lens of focal length f_o and eyepiece of focal length f_e .

the back of the eyepiece is

$$M_{\text{tel}} = \begin{pmatrix} 1 & 0 \\ -1/f_e & 1 \end{pmatrix} \begin{pmatrix} 1 & f_o + f_e \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_o & 1 \end{pmatrix} = \begin{pmatrix} -f_e/f_o & f_e + f_o \\ 0 & -f_o/f_e \end{pmatrix}.$$

We see that the effective focal length is infinite so the focusing power is zero. Such an instrument is called afocal. Parallel rays are transformed into parallel rays, but at a larger angle to the optical axis. The telescope provides angular magnification of

$$M_\theta = \frac{\theta_{\text{out}}}{\theta_{\text{in}}} = -\frac{f_o}{f_e}.$$

The image at infinity in the image space is inverted. This can be corrected using the Galilean version of the telescope which uses a negative lens for the eyepiece and forms an upright image. The image magnification seen after focusing by the eye lens onto the retina is M_θ which is limited by the length of the telescope tube since

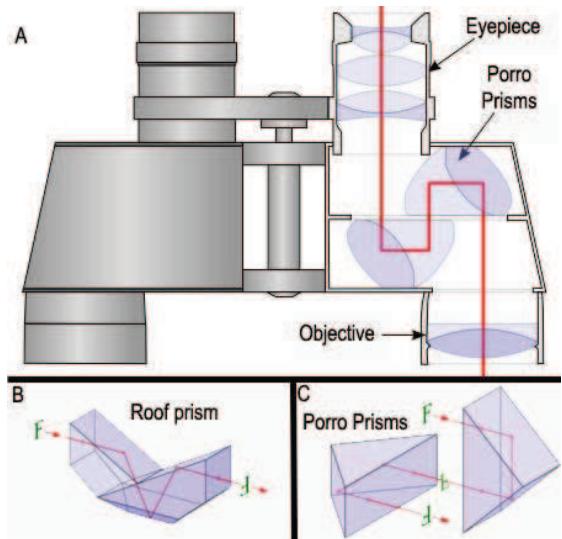
$$M_\theta = -\frac{L - f_e}{f_e} \simeq -\frac{L}{f_e}$$

at high magnification.

The limit of the achievable magnification is governed by two considerations. It becomes impractical to make the lens tube more than a few meters long and still be able to point the telescope. In order to collect sufficient light the objective lens diameter should be increased as the magnification is increased but it is difficult to fabricate very large lenses without defects in the body of the glass.

Both of these limitations can be sidestepped with the reflecting telescope which is the design used in all large, modern astronomical telescopes. In the reflecting telescope the objective lens is replaced by a large primary mirror so the problem of glass defects is avoided. A flat or spherical secondary mirror diverts the focused rays away from the optical axis for viewing by an observer. In this design the focal length of the objective can be made very long, giving a large magnification, without the length of the lens tube increasing correspondingly giving a much more compact construction than a refracting telescope of the same magnification. The first reflecting telescope was constructed by Newton.

The primary mirror should be figured as a paraboloid to avoid astigmatism, but is then subject to strong off-axis coma which limits the field of view. This can be corrected using the Schmidt design[28] which uses a spherical primary mirror to reduce the off-axis coma, and a refractive corrector plate to correct the axial astigmatism. The Schmidt design is popular for small to moderate sized telescopes that take large field images for sky surveys.



http://www.montaukobservatory.com/observer/theobserver1_5.htm

Figure 4.19: Binocular optics.

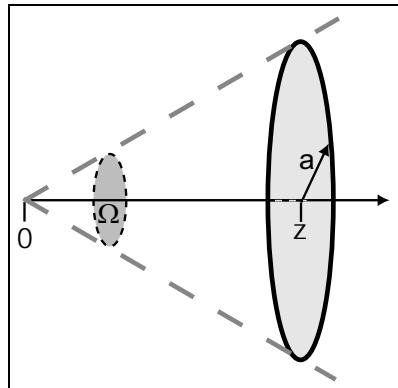


Figure 4.20: Calculation of solid angle of window or lens.

4.3.4 Binoculars

Binoculars are essentially two telescopes placed next to each other, one for each eye. The image inversion is corrected using a pair of prisms as in Fig. 4.19.

4.4 Solid Angle

The amount of light collected by an imaging instrument is important for obtaining detailed images in as short a time as possible. The collection efficiency is determined by the solid angle of the first lens of the instrument. Consider the geometry of Fig. 4.20 where a lens or window of radius a has its center a distance z from the origin. The solid angle subtended by the lens at the origin is given

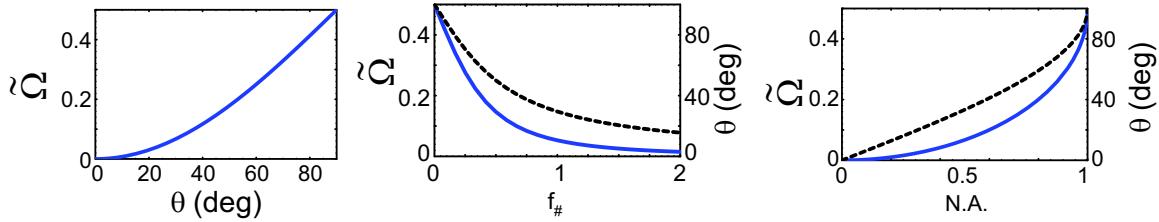


Figure 4.21: Fractional solid angle as a function of cone angle θ , $f_{\#}$, and NA . The dashed line is the cone angle as a function of the $f_{\#}$ or NA .

by

$$\begin{aligned}
 \Omega &= \int_A \frac{\hat{R} \cdot \hat{n}}{R^2} da \\
 &= \int_0^{2\pi} d\phi \int_0^a d\rho \frac{\cos(\tan^{-1} \frac{\rho}{z})}{z^2 + \rho^2} \rho \\
 &= \frac{2\pi}{z^2} \int_0^a d\rho \frac{\rho}{\left(1 + \frac{\rho^2}{z^2}\right)^{3/2}} \rho \\
 &= 2\pi \left(1 - \frac{1}{\sqrt{1 + a^2/z^2}}\right). \tag{4.22}
 \end{aligned}$$

The f-number of the lens is⁴ $f_{\#} = z/(2a)$ so the solid angle can be written as

$$\Omega = 2\pi \left(1 - \frac{1}{\sqrt{1 + 1/(2f_{\#})^2}}\right) \simeq \frac{\pi}{4f_{\#}^2}$$

where the last equality holds for large $f_{\#}$. Camera lenses often specify the $f_{\#}$. The speed of a lens is proportional to the light gathering power which is proportional to the inverse of the $f_{\#}$.

It is also customary to specify optical lenses in terms of the numerical aperture defined as $NA = n \sin(\theta)$ where n is the index of refraction of the medium the lens is embedded in, and $\theta = \tan^{-1}(a/z)$ is the cone angle of the lens. When $n = 1$ we have

$$\Omega = 2\pi \left(1 - \sqrt{1 - NA^2}\right) \simeq \pi NA^2,$$

where the last equality holds for small NA .

It is convenient to compare the fractional solid angle as a function of the cone angle θ , the $f_{\#}$, and the NA . We have for $\tilde{\Omega} = \Omega/4\pi$,

$$\begin{aligned}
 \tilde{\Omega}_{\theta} &= \frac{1}{2} \left[1 - \frac{1}{\sqrt{1 + \tan^2(\theta)}}\right] \\
 \tilde{\Omega}_{f_{\#}} &= \frac{1}{2} \left[1 - \frac{1}{\sqrt{1 + \frac{1}{(2f_{\#})^2}}}\right] \\
 \tilde{\Omega}_{NA} &= \frac{1}{2} \left(1 - \sqrt{1 - NA^2}\right). \tag{4.23}
 \end{aligned}$$

⁴It is common notation to write the f-number as $f/\#$.

Note that for a cone angle of 45 deg. $\theta = \pi/4$ and $f_{\#} = 0.5$, and $NA = .707$. Finally, for convenience in converting from $f_{\#}$ to NA we note that

$$NA = \frac{1}{\sqrt{1 + 4f_{\#}^2}}, \quad \text{and} \quad f_{\#} = \frac{\sqrt{1 - NA^2}}{2NA}.$$

4.5 Eikonal equation

So far we have considered optical rays that propagate in straight lines but are reflected or refracted at surfaces separating media with different indices of refraction. In many cases there can be a continuous variation of the optical properties of a medium. For example in gases the refractive index is proportional to the density of atoms or molecules. In an ideal gas the density is in turn inversely proportional to the temperature. Temperature gradients therefore lead to index gradients which change the ray path in a continuous fashion. Another example is an optical fiber, or a lens, with a gradient index profile. The index profile is imprinted into the glass in the manufacturing process and is frozen in place.

We can describe the ray path in a continuously varying medium with the geometry of Fig. 4.22. Let the ray path be given by the vector $\mathbf{r}(s)$ with s a scalar quantity which parameterizes distance along the ray. The local tangent vector to the ray is $\hat{\epsilon}_t = \frac{d\mathbf{r}}{ds}$. The time it takes for light to follow a ray path \mathcal{C} is

$$t = \int_{\mathcal{C}} \frac{ds}{v(\mathbf{r})} = \frac{1}{c} \int_{\mathcal{C}} ds n(\mathbf{r}).$$

The optical path length \mathcal{L} is defined as

$$\mathcal{L} = ct = \int_{\mathcal{C}} ds n(\mathbf{r}). \quad (4.24)$$

The differential optical path can be expressed in two different ways. From (4.24) we have $d\mathcal{L} = n(\mathbf{r})ds = n(\mathbf{r})\hat{\epsilon}_t \cdot d\mathbf{r}$ and formally considering $\mathcal{L} = \mathcal{L}(\mathbf{r})$ we can write $d\mathcal{L} = \frac{\partial \mathcal{L}}{\partial \mathbf{r}} \cdot d\mathbf{r} = \nabla \mathcal{L} \cdot d\mathbf{r}$. Equating the two expressions gives

$$\nabla \mathcal{L} = n(\mathbf{r})\hat{\epsilon}_t.$$

Squaring this relation we arrive at

$$(\nabla \mathcal{L})^2 = n^2(\mathbf{r}),$$

which is known as the eikonal equation.

It is instructive to derive this equation in a different way. Assume time harmonic fields at frequency ω , $\vec{\mathcal{E}}(\mathbf{r}, t) = \vec{\mathcal{E}}(\mathbf{r})e^{-i\omega t}$, $\vec{\mathcal{H}}(\mathbf{r}, t) = \vec{\mathcal{H}}(\mathbf{r})e^{-i\omega t}$. Without loss of generality we factor the spatial dependence of the fields into a slowly varying part and a phase depending on the optical path length as

$$\begin{aligned} \vec{\mathcal{E}}(\mathbf{r}) &= \vec{\mathcal{E}}_0(\mathbf{r})e^{ik_0\mathcal{L}(\mathbf{r})}, \\ \vec{\mathcal{H}}(\mathbf{r}) &= \vec{\mathcal{H}}_0(\mathbf{r})e^{ik_0\mathcal{L}(\mathbf{r})}. \end{aligned}$$

Here $k_0 = \omega/c$ is the wavenumber of the light in vacuum. Taking the curl we get

$$\begin{aligned} \nabla \times \vec{\mathcal{E}} &= [\nabla \times \vec{\mathcal{E}}_0 - ik_0(\nabla \mathcal{L} \times \vec{\mathcal{E}}_0)]e^{ik_0\mathcal{L}(\mathbf{r})} = i\omega\mu\vec{\mathcal{H}}, \\ \nabla \times \vec{\mathcal{H}} &= [\nabla \times \vec{\mathcal{H}}_0 - ik_0(\nabla \mathcal{L} \times \vec{\mathcal{H}}_0)]e^{ik_0\mathcal{L}(\mathbf{r})} = -i\omega\epsilon\vec{\mathcal{E}}. \end{aligned}$$

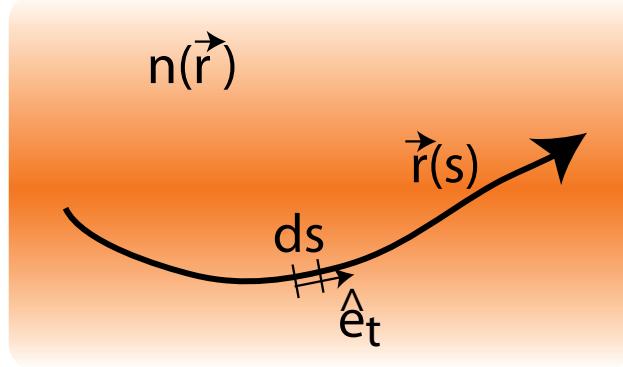


Figure 4.22: Ray propagation in a medium with continuously varying index $n(\mathbf{r})$.

The last equality on each line follows from the Maxwell equations. Thus

$$\nabla \mathcal{L} \times \vec{\mathcal{H}}_0 = -\frac{i}{k_0} \nabla \times \vec{\mathcal{H}}_0 + \frac{\omega}{k_0} \epsilon \vec{\mathcal{E}}_0$$

or

$$\nabla \mathcal{L} \times \vec{\mathcal{H}}_0 - c\epsilon \vec{\mathcal{E}}_0 = -\frac{i}{k_0} \nabla \times \vec{\mathcal{H}}_0.$$

Similarly we find

$$\nabla \mathcal{L} \times \vec{\mathcal{E}}_0 + c\mu \vec{\mathcal{H}}_0 = -\frac{i}{k_0} \nabla \times \vec{\mathcal{E}}_0.$$

The ray approximation is that $\lambda_0 \rightarrow 0$ or $k_0 \rightarrow \infty$ so the right hand sides can be neglected leaving

$$\begin{aligned} \nabla \mathcal{L} \times \vec{\mathcal{H}}_0 - c\epsilon \vec{\mathcal{E}}_0 &= 0, \\ \nabla \mathcal{L} \times \vec{\mathcal{E}}_0 + c\mu \vec{\mathcal{H}}_0 &= 0. \end{aligned}$$

Eliminating \mathcal{H}_0 this becomes

$$\nabla \mathcal{L} \times (\nabla \mathcal{L} \times \vec{\mathcal{E}}_0) + n^2 \vec{\mathcal{E}}_0 = 0,$$

where we have used $c^2 \epsilon \mu = \frac{\epsilon \mu}{\epsilon_0 \mu_0} = n^2$. Expanding the triple product gives

$$(\nabla \mathcal{L} \cdot \vec{\mathcal{E}}_0) \nabla \mathcal{L} - (\nabla \mathcal{L})^2 \vec{\mathcal{E}}_0 + n^2 \vec{\mathcal{E}}_0 = 0.$$

The first factor can be shown to vanish in the ray limit⁵ which leaves $-(\nabla \mathcal{L})^2 \vec{\mathcal{E}}_0 + n^2 \vec{\mathcal{E}}_0 = 0$ or

$$(\nabla \mathcal{L})^2 = n^2$$

which is the eikonal equation.

⁵Use $0 = \nabla \cdot (\epsilon \vec{\mathcal{E}}) = (\epsilon \nabla \cdot \vec{\mathcal{E}}_0 + \nabla \epsilon \cdot \vec{\mathcal{E}}_0 + ik_0 \epsilon \vec{\mathcal{E}}_0 \cdot \nabla \mathcal{L}) e^{ik_0 \mathcal{L}}$. In the ray limit $k_0 \rightarrow \infty$ so we must have $\vec{\mathcal{E}}_0 \cdot \nabla \mathcal{L} \rightarrow 0$.

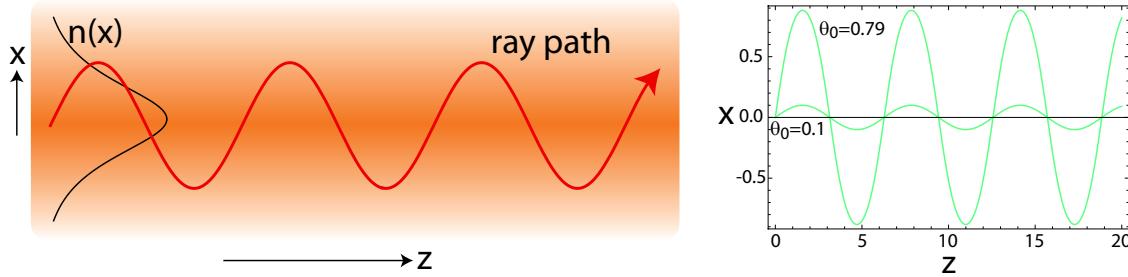


Figure 4.23: Periodic ray trajectories in a gradient index fiber. The numerical solutions are for $n(x) = 1.2 \operatorname{sech}(x)$, and $\theta(0) = 0.1, 0.79$.

4.5.1 Ray equation

Using the eikonal equation we can derive an equation for the ray path $\mathbf{r}(s)$. We have

$$\begin{aligned} \frac{d}{ds} \nabla \mathcal{L} &= \nabla(\nabla \mathcal{L}) \cdot \frac{d\mathbf{r}}{ds} = \frac{d\mathbf{r}}{ds} \cdot \nabla(\nabla \mathcal{L}) = \hat{\epsilon}_t \cdot \nabla(\nabla \mathcal{L}) = \frac{\nabla \mathcal{L}}{n(\mathbf{r})} \cdot \nabla(\nabla \mathcal{L}) = \frac{\nabla [(\nabla \mathcal{L})^2]}{2n(\mathbf{r})} \\ &= \frac{\nabla [n^2(\mathbf{r})]}{2n(\mathbf{r})} = \nabla n(\mathbf{r}). \end{aligned}$$

We can also write

$$\frac{d}{ds} \nabla \mathcal{L} = \frac{d}{ds} [n(\mathbf{r}) \hat{\epsilon}_t] = \frac{d}{ds} \left[n(\mathbf{r}) \frac{d\mathbf{r}}{ds} \right].$$

Equating the expressions for $\frac{d}{ds} \nabla \mathcal{L}$ gives

$$\frac{d}{ds} \left[n(\mathbf{r}) \frac{d\mathbf{r}}{ds} \right] = \nabla n(\mathbf{r}). \quad (4.25)$$

This is the equation for the optical ray path in an inhomogeneous medium which can be thought of as a generalization of Snell's law. When $n(\mathbf{r}) = n_0$ is a constant we get immediately

$$\frac{d}{ds} \left(n_0 \frac{d\mathbf{r}}{ds} \right) = 0 \quad \rightarrow n_0 \frac{d\mathbf{r}}{ds} = \text{constant}.$$

The solution is $\mathbf{r} = \mathbf{c}_1 + s\mathbf{c}_2$, with $\mathbf{c}_1, \mathbf{c}_2$ constant vectors. We see that rays in a homogeneous optical medium indeed travel in straight lines.

Let's use the ray equation to calculate the path of rays in a medium with a continuously varying index. A common and important example of this is a graded index optical fiber as shown in Fig. 4.23. The index is a function of the radial coordinate, $n = n(x)$ and is a maximum on the axis of the fiber. This geometry supports bound rays that propagate along the fiber with a maximum value of $x = x_{\max}$ for all z . There are two types of rays: meridional rays that pass through the fiber axis, and skew rays that start off-axis and follow a spiral path.

Let's consider a meridional ray propagating mainly along \hat{z} for which the motion can be described in terms of a 2D geometry, $\mathbf{r} = x(z)\hat{x} + z\hat{z}$. We parameterize the transverse coordinate $x = x(z)$ by the axial coordinate z . At a distance s along the ray the unit tangent vector makes an angle $\theta(x)$ with the \hat{z} axis, i.e. $\hat{\epsilon}_t = \sin \theta \hat{x} + \cos \theta \hat{z}$. Using $\hat{\epsilon}_t = \frac{d\mathbf{r}}{ds}$ we see that $\frac{dx}{ds} = \sin \theta$ and $\frac{dz}{ds} = \cos \theta$. The two components of (4.25) are then

$$\frac{d}{ds} [n(x) \sin \theta(x)] = \frac{dn(x)}{dx}, \quad \frac{d}{ds} [n(x) \cos \theta(x)] = 0.$$

The second equation can be integrated to give

$$n(x) \cos \theta(x) = \bar{n} = n(0) \cos \theta(0), \quad (4.26)$$

with \bar{n} a constant. If n has a maximum at $x = 0$ and decreases monotonically to both sides away from the axis then the ray is confined and has a maximum value x_{\max} which occurs when $\cos \theta(x_{\max}) = 1$. The index of refraction at x_{\max} is

$$\bar{n} = n(x_{\max}) \cos \theta(x_{\max}) = n(0) \cos \theta(0).$$

The actual value of x_{\max} depends on the initial angle and the shape of the index profile and is found from solving $n(x_{\max}) = n(0) \cos \theta(0)$.

To find the ray path rewrite we can rewrite (4.26) as $n \frac{dz}{ds} = \bar{n}$ so $\frac{d}{ds} = \frac{\bar{n}}{n} \frac{d}{dz}$. The equation for the x component of the ray path can then be written as

$$\frac{d}{ds} [n(x) \sin \theta(x)] = \frac{\bar{n}}{n} \frac{d}{dz} [n(x) \sin \theta(x)] = \frac{\bar{n}}{n} \frac{d}{dz} \left[n(x) \frac{dx}{ds} \right] = \frac{\bar{n}}{n} \frac{d}{dz} \left[\bar{n} \frac{dx}{dz} \right] = \frac{dn(x)}{dx}$$

or

$$\bar{n}^2 \frac{d^2 x}{dz^2} = \frac{1}{2} \frac{d}{dx} n^2.$$

This has the same form as for the motion of a particle in a potential. As in mechanics problems we can integrate by multiplying with $\frac{dx}{dz}$ to get

$$\bar{n}^2 \frac{d^2 x}{dz^2} \frac{dx}{dz} = \frac{1}{2} \frac{dx}{dz} \frac{d}{dx} n^2 = \frac{1}{2} \frac{d}{dz} n^2$$

which integrates to

$$\bar{n}^2 \left(\frac{dx}{dz} \right)^2 = n^2 + c_1.$$

At $x = 0$ the slope is $\frac{dx}{dz}|_{x=0} = \frac{n(0)}{\bar{n}} \frac{dx}{ds}|_{x=0} = \frac{n(0)}{\bar{n}} \sin \theta(0)$ and

$$c_1 = \bar{n}^2 \frac{n^2(0)}{\bar{n}^2} \sin^2 \theta(0) - n^2(0) = -n^2(0) \cos^2 \theta(0) = -\bar{n}^2.$$

Thus the ray path is given by

$$\bar{n}^2 \left(\frac{dx}{dz} \right)^2 = n^2 - \bar{n}^2.$$

To solve for the trajectory $z(x)$ we write

$$\bar{n} \frac{dx}{dz} = \sqrt{n^2 - \bar{n}^2}$$

and invert to get

$$z(x) = \bar{n} \int_0^x \frac{dx'}{\sqrt{n^2(x') - \bar{n}^2}},$$

assuming the ray begins at $x = 0$ with angle $\theta(0)$.

For some choices of the profile $n(x)$ this can be solved analytically. For example using $n(x) = n(0) \operatorname{sech}(\alpha x)$ with α a constant the solution is

$$x = \frac{1}{\alpha} \sinh^{-1} [\sinh(\alpha x_{\max}) \sin(\alpha z)]$$

where $\bar{n} = n(x_{\max}) = n(0) \cos \theta(0)$. A few trajectories for different values of $\theta(0)$ are shown in Fig. 4.23. As the initial angle is increased the ray reaches larger values of x_{\max} while the axial period of the trajectory is unchanged.

Chapter 5

Fourier optics

Fourier optics relies on spectral decomposition of optical fields as a route to calculating field changes under propagation. The spatial amplitude $A(x, y)$ can be represented by a Fourier transform as $\tilde{A}(k_x, k_y)$ where k_x, k_y are the transverse components of the wavevector. Consider rays propagating along the \hat{z} axis and at an angle θ to the axis as shown in Fig. 5.1. After a distance z the axial ray acquires a propagation phase $\phi_0 = kz$. The tilted ray travels a longer distance and acquires a phase $\phi_\theta = kz / \cos(\theta)$. Using $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ and $\cos(\theta) = k_z/k$ we find

$$\delta\phi = \phi_\theta - \phi_0 \simeq \frac{k_x^2 + k_y^2}{2k} z + \mathcal{O}((k_x^2 + k_y^2)^2).$$

Propagation of the component $\tilde{A}(k_x, k_y)$ directed along $\hat{\theta}$ an axial distance z results in a quadratic phase factor $\delta\phi \sim k_x^2 + k_y^2$. Since different transverse wavevectors pick up different propagation phases the spatial distribution found after transforming back to $A(x, y)$ changes. In this way, even though rays propagate in a straight line in a homogeneous medium, optical fields diffract. Diffractive spreading of localized fields is an important concept in optics and leads to resolution limits in imaging and spectroscopy.

A physical picture of diffraction can be based on the Huygens' construction[29] as depicted in Fig. 5.2. Each point on the surface of a wavefront acts as a source for expanding spherical waves that combine to give a new downstream wavefront. A mathematical formulation of Huygens' construction suitable for quantitative calculations was provided by Fresnel in 1816[30]. Subsequent work showed that modifications to the Huygens-Fresnel theory were necessary to make it fully

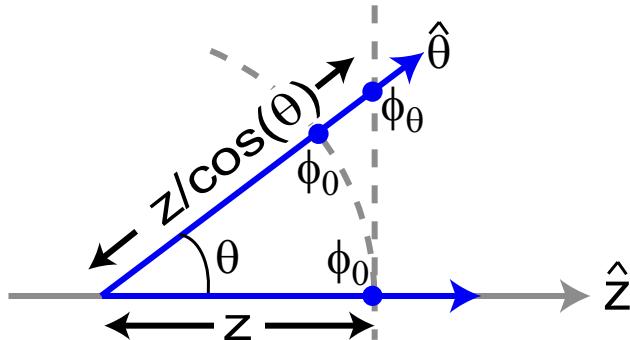


Figure 5.1: Phase delay for off-axis ray propagation.

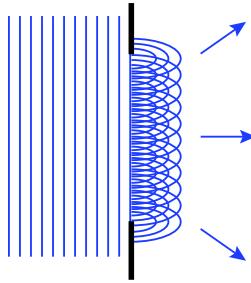


Figure 5.2: Huygens' construction for describing propagation of wavefronts.

mathematically self-consistent. In this chapter we will not follow the historical development but will start in Sec. 5.1 with a paraxial theory of diffraction which follows from the Maxwell equations. After developing the Fresnel and Fraunhofer approximations we will show the connection to the more accurate Huygens-Fresnel theory, in a form established by Rayleigh and Sommerfeld, in Sec. 5.1.4.

5.1 Paraxial propagation and diffraction

In a homogeneous medium with linear electric and magnetic properties and no free charges each component U of the electric and magnetic fields satisfies the scalar Helmholtz equation (1.4)

$$\frac{\partial^2 U}{\partial t^2} - v^2 \nabla^2 U = 0.$$

Here $v = c/n$ is the phase velocity in the medium with refractive index n . Assuming a time dependence $e^{-i\omega t}$ we find

$$\nabla^2 U + k^2 n^2 U = 0$$

where $k = 2\pi/\lambda_{\text{vac}}$ is the vacuum wavenumber. The effective wavenumber inside the medium is $kn = \omega n/c$ where ω is the angular frequency of the field, c is the speed of light, and n is the index of refraction. In the following we will mostly assume $n = 1$. Although the Helmholtz equation is an exact description of the field evolution in a homogeneous medium it is only an approximation when the refractive index is spatially varying. Provided the length scales of the variation are large compared to the wavelength of light the Helmholtz equation provides a good description. In the opposite regime of nano-optics other methods such as finite difference time domain (FDTD) numerical solutions of the Maxwell equations are required. With the availability of fast computers FDTD simulations have become widely used[31].

The scalar Helmholtz equation admits exact solutions including plane waves $U = Ae^{i\mathbf{k}\cdot\mathbf{r}}$ and spherical waves $U = (A/r)e^{ikr}$ with A constant. The Helmholtz equation is an elliptic partial differential equation so the general solution requires specifying boundary condition on the surface of an enclosed region.

In propagation problems we are interested in describing the transverse distribution of a light beam that propagates in a definite direction as shown in Fig. 5.3. Choosing \hat{z} as the propagation direction, and seeking a solution of the form $U = A(x, y, z)e^{ikz}$ gives

$$\left(\nabla^2 A + i2k \frac{\partial A}{\partial z} - k^2 A + k^2 A \right) e^{ikz} = 0$$

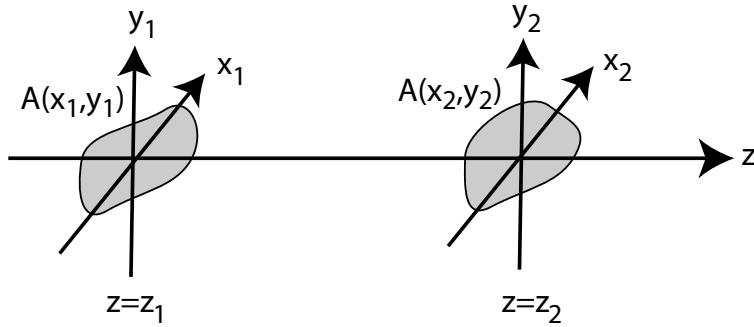


Figure 5.3: Coordinate system for optical propagation.

which simplifies to

$$\nabla_{\perp}^2 A + \frac{\partial}{\partial z} \left(\frac{\partial A}{\partial z} + i2kA \right) = 0 \quad (5.1)$$

where $\nabla_{\perp}^2 A = \partial^2 A / \partial x^2 + \partial^2 A / \partial y^2$. Solutions of Eq. (5.1) are exact solutions of the Helmholtz equation. We now introduce a slowly varying envelope approximation by assuming $(\partial A / \partial z) \ll 2kA$ which results in

$$\frac{\partial A}{\partial z} - \frac{i}{2k} \nabla_{\perp}^2 A = 0. \quad (5.2)$$

All of Fresnel and Fraunhofer diffraction theory for scalar fields are described by the solutions of Eq. (5.2) which is called the paraxial wave equation. It is a partial differential equation of parabolic type. Given an appropriate set of initial conditions such as $A(x, y, 0)$ the solution is uniquely determined for all $A(x, y, z > 0)$.

We can develop a physical picture of diffractive propagation by decomposing the field $A(x, y)$ into a superposition of plane waves with transverse components $\mathbf{k}_{\perp} = k_x \hat{x} + k_y \hat{y}$ and amplitudes $\tilde{A}(k_x, k_y)$. A plane wave propagating in the direction $\hat{\theta}$ with $\mathbf{k}_{\perp} = (k_x, k_y)$ picks up a propagation

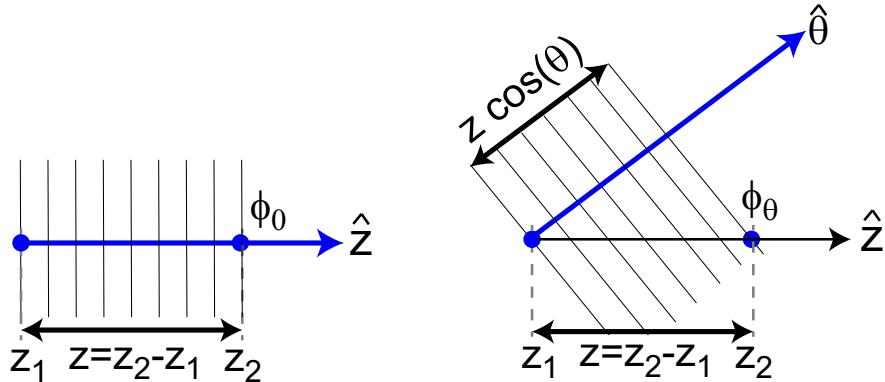


Figure 5.4: Phase delay for off-axis ray propagation. The effective phase between points separated by z along the \hat{z} axis is larger for axial propagation with $\hat{k} = \hat{z}$ (left) than for off-axis propagation with $\hat{k} = \hat{\theta}$ (right).

phase from $\mathbf{r}_1 = (0, 0, z_1)$ to $\mathbf{r}_2 = (0, 0, z_2)$ of

$$e^{i\phi_\theta} = e^{i\mathbf{k}\cdot(\mathbf{r}_2 - \mathbf{r}_1)} = e^{ik_z(z_2 - z_1)}.$$

The phase difference compared to a plane wave propagating along \hat{z} is

$$\delta\phi = \phi_\theta - \phi_0 = (k_z - k)(z_2 - z_1).$$

Making a paraxial approximation we write $k_z = \sqrt{k^2 - k_\perp^2} \simeq k - \frac{k_\perp^2}{2k}$ so

$$\delta\phi = -\frac{k_\perp^2}{2k}(z_2 - z_1).$$

Therefore off-axis waves pick up a differential propagation phase compared to an axial wave given by

$$\tilde{A}(\mathbf{k}_\perp, z_2) = \tilde{A}(\mathbf{k}_\perp, z_1)e^{-i\frac{k_\perp^2}{2k}(z_2 - z_1)}.$$

The effective “diffractive” wavenumber is $k_d = \frac{k_x^2 + k_y^2}{2k}$. Note that there is a phase *advance* for the off-axis plane wave components, even though the propagation length for a ray to reach the plane $z = z_2$ is lengthened as shown in Fig. 5.1. The relative phase of the spectral components proportional to k_\perp^2 results in a new field distribution $A(x, y, z_2)$ which is found from combining the plane wave components with the new phases. This results in the physical phenomena of diffraction and spreading of localized wave packets, even though each individual plane wave propagates along a straight line.

5.1.1 Fresnel Diffraction

The above discussion can be put in a formal framework using the mathematics of Fourier transforms. The spatial spectrum of $A(x, y) = A(\boldsymbol{\rho})$ is given by

$$\tilde{A}(\mathbf{k}_\perp) = \frac{1}{2\pi} \iint_{-\infty}^{\infty} d\boldsymbol{\rho} A(\boldsymbol{\rho}) e^{-i\boldsymbol{\rho}\cdot\mathbf{k}_\perp}, \quad (5.3)$$

and the field is found from the inverse transform

$$A(\boldsymbol{\rho}) = \frac{1}{2\pi} \iint_{-\infty}^{\infty} d\mathbf{k}_\perp \tilde{A}(\mathbf{k}_\perp) e^{i\boldsymbol{\rho}\cdot\mathbf{k}_\perp}. \quad (5.4)$$

Using the shorthand notation $\mathcal{F}[\dots] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\boldsymbol{\rho} \dots e^{-i\boldsymbol{\rho}\cdot\boldsymbol{\rho}}$ we have $\mathcal{F}[\partial A/\partial x] = -ik_x \mathcal{F}[A]$

and we get $\frac{\partial \tilde{A}}{\partial z} = -(i/2k)k_\perp^2 \tilde{A}$. This can be solved as $\tilde{A}(\mathbf{k}_\perp, z_2) = \tilde{A}(\mathbf{k}_\perp, z_1)e^{-i(1/2k)k_\perp^2 z}$ where $k_\perp^2 = \mathbf{k}_\perp \cdot \mathbf{k}_\perp = k_x^2 + k_y^2$ and $z = z_2 - z_1$. Transforming back to the spatial domain we find

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \mathcal{F}^{-1} [\mathcal{F}[A_1(\boldsymbol{\rho}_1)] H(k_\perp, z)] \\ &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} d\mathbf{k}_\perp e^{i\mathbf{k}_\perp \cdot \boldsymbol{\rho}_2} e^{-i\frac{k_\perp^2}{2k}(z_2 - z_1)} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) e^{-i\mathbf{k}_\perp \cdot \boldsymbol{\rho}_1} \end{aligned} \quad (5.5)$$

where the transfer function of free space is

$$H(k_\perp, z) = e^{-i(k_\perp^2/2k)z}. \quad (5.6)$$

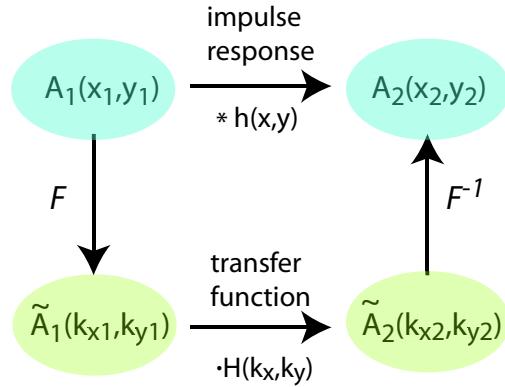


Figure 5.5: Paraxial evolution can be solved by convolution with the impulse response in the spatial domain or multiplication by the transfer function in the Fourier domain.

As discussed above the transfer function H is just the phase shift picked up by a wave propagating with a transverse wavenumber k_{\perp} . While the integral in (5.5) extends to infinite values of k_x, k_y it is important to recognize that contributions from wavenumbers with $k_{\perp}^2 > k^2$ represent nonpropagating or evanescent waves. The propagation factor for an oblique wave is

$$e^{i\sqrt{k^2 - k_{\perp}^2}z} = e^{-\sqrt{k_{\perp}^2 - k^2}z}.$$

When $k_{\perp}^2 > k^2$ this is a decaying exponential. Although such waves do not propagate, or carry energy along \hat{z} , they must be included in the Fourier integral to correctly describe the small scale features of the diffraction pattern. We can estimate the minimum spatial feature size as the interference pattern period created by two waves with $k_{\perp} = \pm k\hat{x}$. The included angle is π and the interference period is $\Lambda = \lambda/2$. This simple estimate underpins the notion that the resolution limit of optics is of order the wavelength. Resolution below a wavelength is possible using nonlinear material response or using super-resolution techniques which trade off better resolution against reduced energy throughput.

As an alternative to the Fourier space transfer function we can use the Fourier convolution theorem to write the solution in the form of a convolution with the impulse response of free space or Green function. The Fourier convolution theorem says that if $\tilde{A} = \tilde{B}\tilde{C}$ then $A = \mathcal{F}^{-1}[\tilde{A}] = \frac{1}{2\pi}\mathcal{F}^{-1}[\tilde{B}]\ast\mathcal{F}^{-1}[\tilde{C}]$ so¹

$$\begin{aligned} A_2(\rho_2) &= \frac{1}{2\pi}A_1 * h \\ &= \frac{-ik}{2\pi z} \iint_{-\infty}^{\infty} d\rho_1 A_1(\rho_1) e^{i(k/2z)|\rho_2 - \rho_1|^2} \end{aligned} \quad (5.7)$$

where the impulse response, or Green function, for free space propagation is

$$h(\rho) = \mathcal{F}^{-1}[H] = \frac{-ik}{z} e^{i\frac{k}{2z}\rho^2}. \quad (5.8)$$

It can be verified that Eqs. (5.5,5.7) are mathematically equivalent. The two methods for solving Eq. (5.2) are shown pictorially in Fig. 5.5. Either the real space or Fourier space approach can be

¹The factor of 2π in Eq. (5.7) is more generally $(2\pi)^{D/2}$ where D is the dimensionality of the space.

used as they give identical results. The choice of how to perform a particular calculation is a matter of convenience and numerical efficiency. The availability of fast discrete transforms using the FFT algorithm often makes the Fourier space approach preferable. It is worth noting however that the impulse response form of Fresnel diffraction can be cast as a Fourier transform by expanding the exponential in (5.7) to get

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \frac{-ik}{2\pi z} e^{i(k/2z)\rho_2^2} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 \left[A_1(\boldsymbol{\rho}_1) e^{i(k/2z)\rho_1^2} \right] e^{-i(k/z)\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}_1} \\ &= \frac{-ik}{z} e^{i(k/2z)\rho_2^2} \mathcal{F} \left[A_1(\boldsymbol{\rho}_1) e^{i(k/2z)\rho_1^2} \right]_{\mathbf{k}_{\perp,1}=(k/z)\boldsymbol{\rho}_2}. \end{aligned} \quad (5.9)$$

Thus, at the cost of multiplication by a quadratic phase, Fresnel diffraction can be calculated using a Fourier transform.

The solution given by Eqs. (5.5,5.7) is known as the Fresnel approximation or Fresnel diffraction. We have found an exact solution to the paraxial wave equation (5.2) which was derived using the approximation $\partial A/\partial z \ll 2kA$. The Fresnel impulse response $h(\rho)$ is a paraxial approximation to an expanding spherical wave since

$$\begin{aligned} \left(e^{-ikz} \right) \frac{1}{r} e^{ikr} &= e^{-ikz} \frac{1}{\sqrt{x^2 + y^2 + z^2}} e^{ik\sqrt{x^2 + y^2 + z^2}} \\ &\simeq \frac{1}{z} \left(1 - \frac{x^2 + y^2}{2z^2} + \frac{3(x^2 + y^2)^2}{8z^4} \right) e^{ik\frac{x^2 + y^2}{2z}} e^{-ik\frac{(x^2 + y^2)^2}{8z^3}}. \end{aligned} \quad (5.10)$$

The prefactor of e^{-ikz} accounts for the fact that we are solving for A and $A = e^{-ikz}U$ where U is the function that solves the Helmholtz equation. The paraxial approximation we have made above in arriving at Eq. (5.8) amounts to neglecting $(x^2 + y^2)/2z^2$ in the amplitude and $k(x^2 + y^2)^2/8z^3$ in the phase of a true spherical wave. This is valid provided $k(x^2 + y^2)^2/8z^3 \ll \pi$ or

$$z \gg z_{\text{Fresnel}} = \left[\frac{k}{8\pi} [(x_2 - x_1)^2 + (y_2 - y_1)^2]_{\max}^2 \right]^{1/3}. \quad (5.11)$$

This limit is referred to as the Fresnel approximation. It is a sufficient condition for the use of Fresnel diffraction theory, but it may be overly restrictive in particular cases. For $\lambda = 1 \mu\text{m}$ and $|x_2 - x_1|_{\max} = |y_2 - y_1|_{\max} = 1 \text{ mm}$ we get $z_{\text{Fresnel}} \gg 10 \text{ mm}$.

5.1.2 Fraunhofer Diffraction

If we are willing to limit ourselves to even longer distances than z_{Fresnel} we can further simplify Eq. (5.7). For large z we approximate the exponential factor under the integral by

$$\begin{aligned} e^{i(k/2z)[(x_2 - x_1)^2 + (y_2 - y_1)^2]} &= e^{i(k/2z)(x_1^2 + x_2^2 + y_1^2 + y_2^2)} e^{-i(k/z)(x_1 x_2 + y_1 y_2)} \\ &\simeq e^{i(k/2z)(x_2^2 + y_2^2)} e^{-i(k/z)(x_1 x_2 + y_1 y_2)}. \end{aligned}$$

Here we have discarded the quadratic phase in plane 1. This implies that the size of the source aperture is not too large. With this approximation we find

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \frac{-ik}{2\pi z} e^{i\frac{k}{2z}\rho_2^2} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) e^{-i(k/z)\boldsymbol{\rho}_1 \cdot \boldsymbol{\rho}_2} \\ &= -\frac{ik}{z} e^{i\frac{k}{2z}\rho_2^2} \mathcal{F} [A_1(\boldsymbol{\rho}_1)]_{\mathbf{k}_{\perp,1}=(k/z)\boldsymbol{\rho}_2}. \end{aligned} \quad (5.12)$$

This is known as the Fraunhofer approximation. In this limit we see that apart from a quadratic phase factor, $A_2(x_2, y_2)$ is proportional to the Fourier transform of the input field. The ease with which optical systems calculate Fourier transforms has wide ranging consequences for optical signal processing, and we will return to this topic later on.

The Fraunhofer approximation is valid for

$$z \gg z_{\text{Fraunhofer}} = (k/2\pi) (x_1^2 + y_1^2)_{\max}. \quad (5.13)$$

For $\lambda = 1 \mu\text{m}$ and $|x_1|_{\max} = |y_1|_{\max} = 1 \text{ mm}$ we get $z_{\text{Fraunhofer}} \gg 2 \text{ m}$. This approximation is two orders of magnitude more restrictive in the range of z allowed than the Fresnel approximation. It will often turn out that both Fresnel and Fraunhofer approximations work quite well outside their strict validity limits. However, it is always necessary to check the results obtained in such a case. For the Fraunhofer approximation we can check by comparing the results with a Fresnel diffraction calculation. For Fresnel diffraction we would have to check against a more accurate approach, such as the Rayleigh-Sommerfeld formulation of diffraction discussed in Sec. 5.1.4 below.

5.1.3 Diffraction from an aperture

Let's use these results to calculate the diffracted field after a rectangular aperture. For simplicity consider a one-dimensional geometry where the aperture has unit transmission for $-L/2 < x < L/2$ and is infinitely extended in y . let the incident field be a uniform plane wave $A(\rho_1) = A_0$. The Fresnel diffraction pattern is

$$\begin{aligned} A(x_2, y_2) &= \frac{-ikA_0}{2\pi z} \int_{-\infty}^{\infty} dy e^{i\frac{k}{2z}(y_2-y)^2} \int_{-L/2}^{L/2} dx e^{i\frac{k}{2z}(x_2-x)^2} \\ &= -iA_0 \left(\frac{ik}{2\pi z} \right)^{1/2} \int_{-L/2}^{L/2} dx e^{i\frac{k}{2z}(x_2-x)^2} \\ &= -iA_0 \left(\frac{i}{2} \right)^{1/2} \int_{u_1}^{u_2} du e^{i\frac{\pi}{2}u^2} \end{aligned}$$

where we have made the change of variable to $u^2 = \frac{k}{\pi z}(x-x_2)^2$ and $u_1 = \sqrt{k/(\pi z)}(-L/2-x_2)$, $u_2 = \sqrt{k/(\pi z)}(L/2-x_2)$. The remaining integral over u can be written in terms of what are known as Fresnel integrals

$$C(s) = \int_0^s du \cos \left(\frac{\pi}{2} u^2 \right), \quad S(s) = \int_0^s du \sin \left(\frac{\pi}{2} u^2 \right).$$

These functions are plotted in Fig. 5.6. A parametric plot of $C(s)$ vs. $S(s)$ gives a spiral, known as the Cornu spiral, with accumulation points at $(0.5, 0.5)$, $(-0.5, -0.5)$ for $s \rightarrow \pm\infty$. The Cornu spiral was introduced to provide a graphical means of calculating the integrals, although with computers this is no longer needed.

In terms of Fresnel integrals the diffracted field is

$$A(x_2) = -iA_0 \left(\frac{i}{2} \right)^{1/2} \{ [C(u_2) - C(u_1)] + i[S(u_2) - S(u_1)] \}, \quad (5.14)$$

and the intensity is

$$I(x_2) = \frac{\epsilon_0 c}{2} \frac{|A_0|^2}{2} \{ [C(u_2) - C(u_1)]^2 + [S(u_2) - S(u_1)]^2 \}.$$

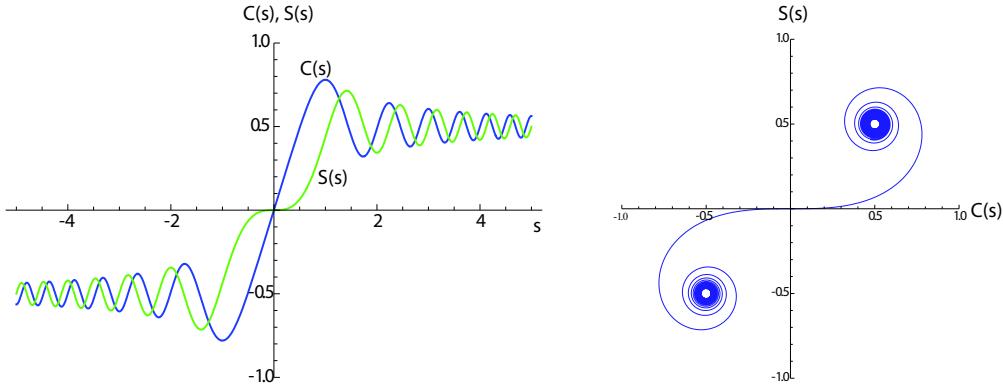


Figure 5.6: Fresnel integrals $C(s)$, $S(s)$ and Cornu spiral.

Figure 5.7 shows the intensity at different z values ranging from $5.4 - 540 \mu\text{m}$. These values can be compared with the Fresnel diffraction distance of (5.11) which we estimate, using the parameters in the figure, to be

$$z_{\text{Fresnel}} = \left[\frac{k}{8\pi} [(x_2 - x_1)^2 + (y_2 - y_1)^2]_{\max}^2 \right]^{1/3} \simeq \left[\frac{k}{8\pi} [L^2 + L^2]_{\max}^2 \right]^{1/3} = 54. \mu\text{m}.$$

We see that all plots in the figure are safely inside the region of validity of the Fresnel approximation. The Fraunhofer approximation is valid for z larger than

$$z_{\text{Fraunhofer}} = (k/2\pi) (x_1^2 + y_1^2)_{\max} \simeq (k/2\pi)L^2 = 400. \mu\text{m}.$$

The last plot in the figure is in the Fraunhofer limit and indeed we see a smooth intensity profile which is close to the Fourier transform of the aperture function squared. Let's check that the Fresnel and Fraunhofer expressions agree in this limit. The field in the Fraunhofer approximation is from (5.12)

$$\begin{aligned} A(x_2, y_2) &= \frac{-ik}{2\pi z} A_0 e^{i\frac{k}{2z}(x_2^2 + y_2^2)} \int_{-\infty}^{\infty} dy \int_{-L/2}^{L/2} dx e^{-i(k/z)(xx_2 + yy_2)} \\ &= \frac{-ik}{2\pi z} A_0 e^{i\frac{k}{2z}(x_2^2 + y_2^2)} \int_{-\infty}^{\infty} dy e^{-i(k/z)yy_2} \int_{-L/2}^{L/2} dx e^{-i(k/z)xx_2}. \end{aligned}$$

The y integral gives a δ function which is singular. To bypass this complication we write the field as

$$A(x_2, y_2) = A_0 \left[\left(\frac{-ik}{2\pi z} \right)^{1/2} e^{i\frac{k}{2z}y_2^2} \int_{-\infty}^{\infty} dy e^{-i(k/z)yy_2} \right] \left[\left(\frac{-ik}{2\pi z} \right)^{1/2} e^{i\frac{k}{2z}x_2^2} \int_{-L/2}^{L/2} dx e^{-i(k/z)xx_2} \right].$$

The first square bracket has all the y, y_2 dependence and for an aperture that is infinitely extended

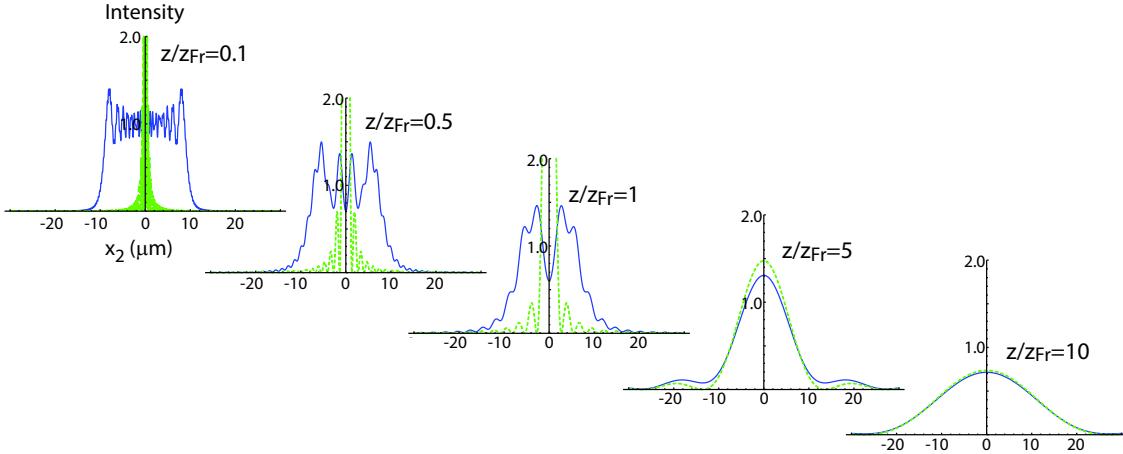


Figure 5.7: Diffracted intensity using Fresnel diffraction (blue solid line) and Fraunhofer diffraction (green dashed line) from a slit of width $L = 20.$ μm , with $\lambda = 1.$ μm at different distances $z/z_{\text{Fresnel}} = 0.1, 0.5, 1, 5, 10.$ The intensity has been normalized to the incident intensity at the aperture plane. The aperture parameters give $z_{\text{Fresnel}} = 54.$ μm and $z_{\text{Fraunhofer}} = 400.$ $\mu\text{m}.$

along y this dependence vanishes. We will therefore write the effective one-dimensional solution as

$$\begin{aligned} A(x_2) &= A_0 \left(\frac{-ik}{2\pi z} \right)^{1/2} e^{i\frac{k}{2z}x_2^2} \int_{-L/2}^{L/2} dx e^{-i(k/z)xx_2} \\ &= A_0 e^{-i\pi/4} \left(\frac{2z}{\pi k} \right)^{1/2} e^{i\frac{k}{2z}x_2^2} \frac{\sin(kLx_2/2z)}{x_2}. \end{aligned} \quad (5.15)$$

We can verify that this is correct by comparing with (5.14) in the limit of $z \rightarrow \infty$ where Fresnel and Fraunhofer diffraction calculations must agree. In this limit $u_1, u_2 \rightarrow 0$ and we can use the expansions $C(u) \simeq u$, $S(u) \sim \frac{\pi u^3}{6}$ to get

$$\begin{aligned} A_{\text{Fresnel}}(x_2) &\simeq -iA_0 \left(\frac{i}{2} \right)^{1/2} \left[(u_2 - u_1) + i \left(\frac{\pi u_2^3}{6} - \frac{\pi u_1^3}{6} \right) \right], \\ &\simeq A_0 e^{-i\pi/4} \left(\frac{k}{2\pi z} \right)^{1/2} L, \end{aligned}$$

where we have kept only the leading term in inverse powers of z . Expanding (5.15) we get for the Fraunhofer calculation

$$A_{\text{Fraunhofer}}(x_2) \simeq A_0 e^{-i\pi/4} \left(\frac{k}{2\pi z} \right)^{1/2} L$$

which agrees with the Fresnel result. Comparing the solid and dashed lines in Fig. 5.7 we see that for $z \lesssim z_{\text{Fraunhofer}}/2$ the Fresnel and Fraunhofer calculations give very different results, while for large z we get the same result from both approximations.

5.1.4 Beyond Fresnel diffraction

Figure 5.7 clearly shows that the results of Fresnel and Fraunhofer diffraction theory differ for $z \ll z_{\text{Fraunhofer}}$. On the other hand for $z \ll z_{\text{Fresnel}}$ we have no way of knowing how accurate the calculated intensity pattern is. To understand the limits of Fresnel diffraction theory we need a more accurate theory to compare it with. A more accurate theory of diffraction, although still limited to scalar fields, can be based on the Huygens-Fresnel theory whereby each point on the source wavefront acts as a source of spherical waves which interfere to create the observed diffracted field. A self-consistent and mathematically correct description of this basic idea turns out to be a remarkably difficult and subtle problem. Details can be found in [5, 32].

Instead of pursuing a formal, more accurate solution of the Helmholtz equation let's try and modify what we have so far to achieve better accuracy. Within the framework of the paraxial wave equation we arrived at an impulse response given by Eq. (5.8), $h(\rho) = \frac{-ik}{z} e^{ik\rho^2/2z}$. This impulse response differs from a true spherical wave by neglect of higher order terms in ρ^2/z^2 as was shown in (5.10). A first guess at an improved theory would be to simply replace our paraxial approximation to h by

$$h_r = -ik \frac{e^{ikr}}{r}$$

where $r = |\mathbf{r}_2 - \mathbf{r}_1|$ is the three-dimensional distance from object point \mathbf{r}_1 to image point \mathbf{r}_2 . Using this in (5.7) we get

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \frac{1}{2\pi} A_1 * h_r \\ &= \frac{-ik}{2\pi} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) \frac{e^{ik|\mathbf{r}_2 - \mathbf{r}_1|}}{|\mathbf{r}_2 - \mathbf{r}_1|} \end{aligned} \quad (5.16)$$

We are still assuming $A_1(\boldsymbol{\rho}_1), A_2(\boldsymbol{\rho}_2)$ to be planar fields but include the full three-dimensional distance r in the integration. It turns out that (5.16) agrees with more accurate theories of diffraction apart from a missing obliquity factor χ . In the Kirchhoff theory[33] $\chi = \frac{1+\cos(\hat{\mathbf{r}}_{12} \cdot \hat{z})}{2}$ where \hat{z} defines the optical axis, and $\hat{\mathbf{r}}_{12}$ is a unit vector from object to image points. For small angles $\hat{\mathbf{r}}_{12} \cdot \hat{z} \simeq 1$ so this factor is only important for large angular spreads. The Kirchhoff theory is not mathematically self-consistent and an improved approach due to Rayleigh and Sommerfeld[34, 35, 36, 37] yields $\chi = \cos(\hat{\mathbf{r}}_{12} \cdot \hat{z})$. In practice it turns out that these two forms of the obliquity factor yield very similar results[38].

With the Rayleigh and Sommerfeld obliquity factor the diffraction integral reads

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \frac{1}{2\pi} A_1 * h_r \\ &= \frac{-ik}{2\pi} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) \frac{e^{ik|\mathbf{r}_2 - \mathbf{r}_1|}}{|\mathbf{r}_2 - \mathbf{r}_1|} \cos(\hat{\mathbf{r}}_{12} \cdot \hat{z}). \end{aligned} \quad (5.17)$$

Figure 5.8 compares Fresnel diffraction with Eq. (5.17) for a square aperture. We see that at $z = z_{\text{Fresnel}}$ there is perfect agreement in the intensity profile, at $z = 0.5z_{\text{Fresnel}}$ small deviations are apparent, and at $z = 0.1z_{\text{Fresnel}}$ large deviations appear. The Fresnel calculation over emphasizes the off-axis peaks at small z compared to the more accurate approach. This can be attributed to the missing obliquity factor which suppresses the large angle diffraction.

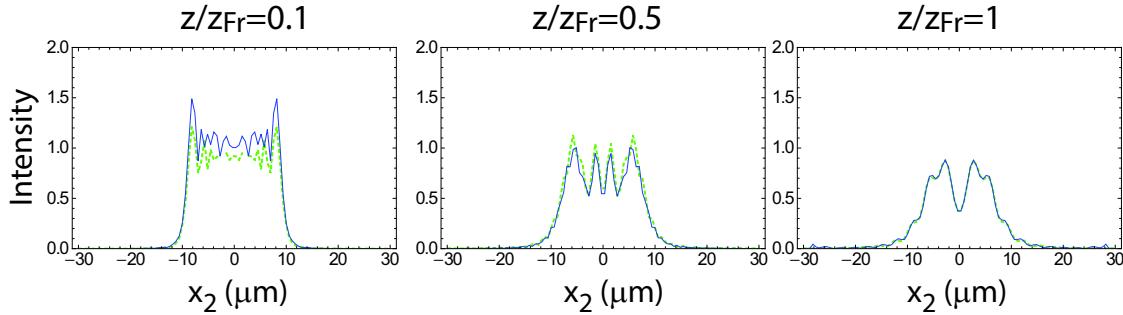


Figure 5.8: Diffracted intensity from Fresnel diffraction theory (blue solid line) and Rayleigh-Sommerfeld theory Eq. (5.17) (green dashed line) from a square aperture of width $L = 20. \mu\text{m}$, with $\lambda = 1. \mu\text{m}$ at different distances $z/z_{\text{Fresnel}} = 0.1, 0.5, 1$. The intensity is shown as a function of x_2 for $y_2 = 0$ and has been normalized to the incident intensity at the aperture plane. The aperture parameters give $z_{\text{Fresnel}} = 54. \mu\text{m}$.

5.1.5 Vectorial diffraction theory

If we wish to improve the accuracy of the calculation even further it is necessary to include vectorial effects. This can be done rigorously using finite difference time domain (FDTD) numerical simulations of Maxwell equations or approximately with analytical approaches due to Wolf and Richards[39, 40] who extended the Debye approximation[41], which is valid for large Fresnel number, to include vectorial effects.

The Debye approximation is based on ...♣?? to be added

5.1.6 Fourier transforming properties of lenses

A lens of focal length f can be described as a phase plate with transmission function $t = e^{-i(k/2f)(x_f^2+y_f^2)}$ where x_f, y_f are the coordinates in the plane of the lens. Let us place a lens at a distance L_1 after the input plane and calculate the output field a distance L_2 after the lens as shown in Fig. 5.9. Using Fresnel diffraction theory we find

$$\begin{aligned} A_2(\rho_2) &= \left(\frac{-ik}{2\pi L_2} \right) \int d\rho_f A(\rho_f) e^{-i(k/2f)\rho_f^2} e^{i(k/2L_2)|\rho_2 - \rho_f|^2} \\ &= \left(\frac{-ik}{2\pi L_2} \right) \int d\rho_f e^{-i(k/2f)\rho_f^2} e^{i(k/2L_2)|\rho_2 - \rho_f|^2} \left(\frac{-ik}{2\pi L_1} \right) \int d\rho_1 A(\rho_1) e^{i(k/2L_1)|\rho_f - \rho_1|^2}, \end{aligned}$$

with $\rho = (x, y)$. Reversing the order of integration, we have for the integral over the lens coordinates

$$\begin{aligned} &\int d\rho_f e^{-i(k/2f)\rho_f^2} e^{i(k/2L_2)|\rho_2 - \rho_f|^2} e^{i(k/2L_1)|\rho_f - \rho_1|^2} \\ &= e^{i(k/2L_2)\rho_2^2} e^{i(k/2L_1)\rho_1^2} \int d\rho_f e^{i\frac{k}{2}\rho_f^2\left(\frac{1}{L_1} + \frac{1}{L_2} - \frac{1}{f}\right)} e^{-ik\rho_f \cdot \left(\frac{\rho_1}{L_1} + \frac{\rho_2}{L_2}\right)} \\ &= i \frac{2\pi}{k\left(\frac{1}{L_1} + \frac{1}{L_2} - \frac{1}{f}\right)} \exp\left[-\frac{ik}{2} \frac{\rho_1^2/L_1 + \rho_2^2/L_2 - f|\rho_1 - \rho_2|^2/(L_1L_2)}{f/L_1 + f/L_2 - 1}\right]. \end{aligned}$$

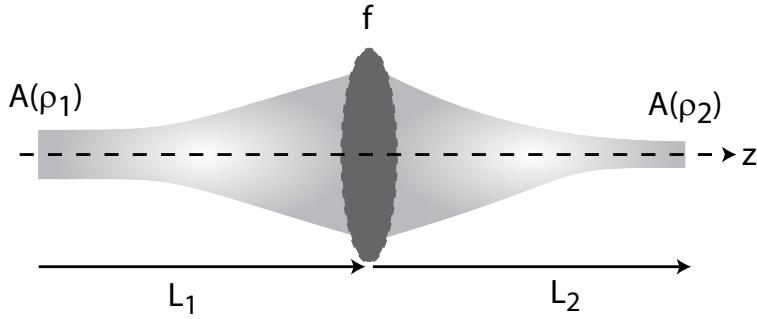


Figure 5.9: Field transformation with a lens.

The output field is thus

$$A(\rho_2) = -i \frac{k}{2\pi(L_1 + L_2 - L_1 L_2/f)} \int d\rho_1 A(\rho_1) \exp \left[-\frac{ik}{2} \frac{\rho_1^2/L_1 + \rho_2^2/L_2 - f|\rho_1 - \rho_2|^2/(L_1 L_2)}{f/L_1 + f/L_2 - 1} \right].$$

This expression simplifies in the back focal plane at a distance $L_2 = f$ after the lens where

$$\begin{aligned} A_2(\rho_2) &= -\frac{ik}{2\pi f} e^{i\frac{k\rho_2^2}{2f}(1-L_1/f)} \int d\rho_1 A(\rho_1) e^{-\frac{ik}{f}\rho_1 \cdot \rho_2} \\ &= -\frac{ik}{f} e^{i\frac{k\rho_2^2}{2f}(1-L_1/f)} \mathcal{F}[A(\rho_1)]_{\mathbf{k}_\perp = \frac{k\rho_2}{f}}. \end{aligned}$$

Hence the output intensity is proportional to the spatial spectrum of the Fourier transform of the input field. When the input field is in the front focal plane of the lens ($L_1 = f$) the quadratic phase factor vanishes and

$$\begin{aligned} A_2(\rho_2) &= -\frac{ik}{f} \mathcal{F}[A_1(\rho_1)]_{\mathbf{k}_\perp = \frac{k\rho_2}{f}} \\ &= -\frac{ik}{f} \frac{1}{2\pi} \iint d\rho_1 A_1(\rho_1) e^{i\frac{k}{f}\rho_1 \cdot \rho_2}. \end{aligned} \quad (5.18)$$

We see that a single lens can be used to create an exact Fourier transform of an arbitrary paraxial field. This result is valid for both the Fresnel and Fraunhofer approximations.

5.1.7 Fields with radial symmetry

When the field has radial symmetry the Fourier transform relations (5.3,5.4) can be simplified to one-dimensional integrals. Put $x = \rho \cos \theta$, $y = \rho \sin \theta$, $dxdy = \rho d\rho d\theta$ and $k_x = k_\perp \cos \phi$, $k_y = k_\perp \sin \phi$, $dk_x dk_y = k_\perp dk_\perp d\phi$ to get

$$\begin{aligned} \tilde{A}(k_\perp, \phi) &= \frac{1}{2\pi} \int_0^\infty d\rho \rho A(\rho) \int_0^{2\pi} d\theta e^{-i\rho k_\perp (\cos \phi \cos \theta + \sin \phi \sin \theta)} \\ &= \frac{1}{2\pi} \int_0^\infty d\rho \rho A(\rho) \int_0^{2\pi} d\theta e^{-i\rho k_\perp \cos(\phi - \theta)}. \end{aligned}$$

The angular integral can be expressed as

$$\int_0^{2\pi} d\theta e^{-i\rho k_{\perp} \cos(\phi-\theta)} = \int_0^{2\pi} d\theta e^{-i\rho k_{\perp} \cos(\theta)} = 2\pi J_0(\rho k_{\perp}).$$

We see that \tilde{A} depends only on k_{\perp} and can be written as

$$\tilde{A}(k_{\perp}) = \int_0^{\infty} d\rho \rho A(\rho) J_0(\rho k_{\perp}) \quad (5.19)$$

with the inverse transform

$$A(\rho) = \int_0^{\infty} dk_{\perp} k_{\perp} \tilde{A}(k_{\perp}) J_0(\rho k_{\perp}). \quad (5.20)$$

Thus the Fourier transform for problems with radial symmetry has a Bessel function kernel. This is referred to as a Hankel transform.

The analog of (5.18) for the field transformation by a lens is

$$\begin{aligned} A_2(\rho_2) &= -\frac{ik}{f} \mathcal{F}[A_1(\rho_1)]_{k_{\perp}=\frac{k\rho_2}{f}} \\ &= -\frac{ik}{f} \int d\rho_1 \rho_1 A_1(\rho_1) J_0(\rho_1 \rho_2 k/f). \end{aligned} \quad (5.21)$$

5.1.8 More general transformations

We can generalize Fresnel diffraction theory to the propagation of a paraxial field through an arbitrary optical system described by an ABCD ray matrix. The output field can be written in terms of a Green function as

$$A_2(\rho_2) = \iint d\rho_1 \mathcal{G}(\rho_2; \rho_1) A_1(\rho_1), \quad (5.22)$$

where the Green function is given by[42, 43].

$$\mathcal{G}(\rho_2; \rho_1) = \frac{-ik}{2\pi B} \exp \left[i \frac{k}{2B} (A\rho_1^2 + D\rho_2^2 - 2\rho_1 \cdot \rho_2) \right] \quad (5.23)$$

and the ray matrix describing propagation from planes 1 → 2 is

$$\mathbf{M} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \quad (5.24)$$

In imaging systems the matrix element B is zero in which case Eq. (5.23) cannot be used directly. To proceed start by assuming $B \neq 0$, and using $AD - BC = 1$ rewrite (5.23) as

$$\mathcal{G}(\rho_2; \rho_1) = \frac{-ik}{2\pi B} e^{i\frac{kC}{2A}\rho_2^2} \exp \left[i \frac{kA}{2B} \left[\left(x_1 - \frac{x_2}{A} \right)^2 + \left(y_1 - \frac{y_2}{A} \right)^2 \right] \right]. \quad (5.25)$$

Taking the limit as $B \rightarrow 0$ and using $\delta(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\sqrt{i\pi\varepsilon}} e^{i\frac{x^2}{\varepsilon}}$ we find

$$\mathcal{G}(\rho_2; \rho_1) = D e^{i\frac{kCD}{2}\rho_2^2} \delta(x_1 - x_2/A) \delta(y_1 - y_2/A). \quad (5.26)$$

The δ functions in Eq. (5.26) demonstrate that $B = 0$ results in imaging with magnification A in accordance with the ray optics treatment of Ch. 4. Note, however, that there is a quadratic phase in the output plane $\frac{kCD}{2}\rho_2^2$ that is not accounted for in a ray optics calculation. Furthermore, as we will see in the following sections, point sources in the object plane do not lead to points in the image plane when we account for the finite size of apertures. The standard ABCD treatment does not account for finite aperture size and therefore predicts perfect paraxial imaging. The ABCD formalism can be extended to account for finite aperture size by allowing for complex matrices[44].

5.1.9 Summary of diffraction formulae

In this section we list the most important formulae for diffraction calculations together with the equation numbers where they were defined.

Fresnel diffraction, transfer function, 2D Cartesian coordinates:

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \mathcal{F}^{-1}[\mathcal{F}[A_1(\boldsymbol{\rho}_1)]H(k_\perp, z)] \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} d\mathbf{k}_\perp e^{i\mathbf{k}_\perp \cdot \boldsymbol{\rho}_2} e^{-i\frac{k^2}{2k}(z_z - z_1)} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) e^{-i\mathbf{k}_\perp \cdot \boldsymbol{\rho}_1} \quad (5.5) \\ H(k_\perp, z) &= e^{-i(k_\perp^2/2k)z}. \end{aligned} \quad (5.6)$$

Fresnel diffraction, impulse response, 2D Cartesian coordinates:

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \frac{1}{2\pi} A_1 * h = \frac{-ik}{2\pi z} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) e^{i(k/2z)|\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1|^2} \quad (5.7) \\ &= \frac{-ik}{z} e^{i(k/2z)\rho_2^2} \mathcal{F} \left[A_1(\boldsymbol{\rho}_1) e^{i(k/2z)\rho_1^2} \right]_{\mathbf{k}_{\perp,1}=(k/z)\boldsymbol{\rho}_2} \quad (5.9) \\ h(\rho) &= \mathcal{F}^{-1}[H] = \frac{-ik}{z} e^{i\frac{k}{2z}\rho^2}. \end{aligned} \quad (5.8)$$

Fresnel diffraction, transfer function, radial symmetry:

$$\begin{aligned} A_2(\rho_2) &= \mathcal{F}^{-1}[\mathcal{F}[A_1(\boldsymbol{\rho}_1)]H(k_\perp, z)] \\ &= \int_0^{\infty} dk_\perp e^{-i\frac{k^2}{2k}(z_2 - z_1)} k_\perp J_0(k_\perp \rho_2) \int_0^{\infty} d\rho_1 A_1(\rho_1) \rho_1 J_0(\rho_1 k_\perp). \end{aligned} \quad (5.30)$$

Fresnel diffraction, impulse response, radial symmetry:

$$A_2(\rho_2) = \frac{-ik}{z} e^{i\frac{k\rho_2^2}{2z}} \int_0^{\infty} d\rho_1 A_1(\rho_1) \rho_1 e^{i\frac{k\rho_1^2}{2z}} J_0\left(\frac{k\rho_2 \rho_1}{z}\right). \quad (5.30)$$

Fraunhofer diffraction, 2D Cartesian coordinates:

$$\begin{aligned} A_2(\boldsymbol{\rho}_2) &= \frac{-ik}{2\pi z} e^{i\frac{k}{2z}\rho_2^2} \iint_{-\infty}^{\infty} d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) e^{-i(k/z)\boldsymbol{\rho}_1 \cdot \boldsymbol{\rho}_2} \\ &= -\frac{ik}{z} e^{i\frac{k}{2z}\rho_2^2} \mathcal{F} [A_1(\boldsymbol{\rho}_1)]_{\mathbf{k}_{\perp,1}=(k/z)\boldsymbol{\rho}_2}. \end{aligned} \quad (5.12)$$

Fraunhofer diffraction, 1D:

$$A_2(x_2) = \left(\frac{-ik}{2\pi z}\right)^{1/2} e^{i\frac{k}{2z}x_2^2} \int_{-\infty}^{\infty} dx_1 A_1(x_1) e^{-i(k/z)x_1 x_2} \quad (5.15)$$

Fourier transformation by lens:

$$A_2(\rho_2) = -\frac{ik}{f} \mathcal{F}[A_1(\rho_1)]_{\mathbf{k}_\perp=\frac{k\rho_2}{f}}. \quad (5.18)$$

5.2 Applications

In the following sections we explore a few examples of Fourier optical phenomena.

5.2.1 Airy disk

A first example of the use of the Hankel transform arises in computing the focal plane pattern formed by uniform illumination of a circular aperture of diameter d . Using (5.21) for a circularly symmetric field and assuming uniform illumination in the front focal plane of the lens we find from Eq. (5.18)

$$\begin{aligned} A(\rho_2) &= -\frac{ik}{f} \int_0^{d/2} d\rho_1 \rho_1 J_0\left(\frac{k\rho_1 \rho_2}{f}\right) \\ &= -id \frac{J_1\left(\frac{dk\rho_2}{2f}\right)}{2\rho_2}. \end{aligned} \quad (5.27)$$

The $J_1(x)$ Bessel function has a first zero away from the origin at $x_1 = 3.83$ so the field in the back focal plane vanishes at

$$\rho_2 = 7.66 \frac{f}{dk} = 1.22 \frac{\lambda f}{d}.$$

This result was first obtained by Airy in 1835[45]. The intensity which is proportional to $|A(\rho_2)|^2$ is plotted in Fig. 5.10. The total power in the Airy diffraction pattern is

$$P = \frac{\epsilon_0 c}{2} \int_0^\infty d\rho_2 2\pi \rho_2 A(\rho_2)^2 = \frac{\epsilon_0 c \pi d^2}{2} \frac{1}{4}$$

which is just the intensity intercepted by the lens. The power inside the first zero is

$$P_0 = \frac{\epsilon_0 c}{2} \int_0^{1.22 \frac{\lambda f}{d}} d\rho_2 2\pi \rho_2 A(\rho_2)^2 = \frac{\epsilon_0 c}{2} 0.66 d^2.$$

A fraction of $0.66/(\pi/4) = 0.84$ of the energy is focused in the central spot.

For an annulus of inner and outer radii a_1, a_2 we get

$$\begin{aligned} A(\rho_2) &= -ia_2 \frac{J_1\left(\frac{a_2 k \rho_2}{f}\right)}{\rho_2} + ia_1 \frac{J_1\left(\frac{a_1 k \rho_2}{f}\right)}{\rho_2} \\ &= i \frac{a_1 J_1\left(\frac{a_1 k \rho_2}{f}\right) - a_2 J_1\left(\frac{a_2 k \rho_2}{f}\right)}{\rho_2}. \end{aligned}$$

Referring to Fig. 5.10 we see that the annulus has the effect of suppressing the central intensity and putting relatively more energy into the rings.

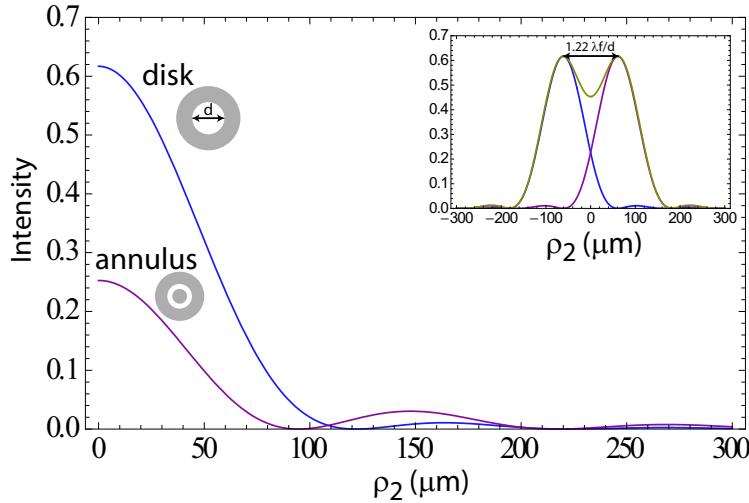


Figure 5.10: Airy diffraction rings formed by uniform illumination of a lens. The circular lens parameters are diameter $d = 100 \mu\text{m}$, focal length $f = 1 \text{ cm}$, $\lambda = 1 \mu\text{m}$. The annulus has inner diameter $0.6d$ and outer diameter d . The intensity is normalized to the incident intensity at the lens.

The ring structure sets the diffraction limited angular resolution of a lens. Suppose we use a lens to view two distant stars with an angular separation of θ . The images in the back focal plane will be separated by $\delta\rho_2 = \sin(\theta)f$. If we say that the first zero of one image coincides with the peak of the adjacent image then there will be a visible dip of the total intensity in between the images. This is shown in the inset to the figure and is referred to as the Rayleigh resolution limit. This limit occurs when $\sin(\theta)f = 1.22\lambda f/d$ or

$$\sin(\theta) = 1.22 \frac{\lambda}{d}. \quad (5.28)$$

This expression gives the angular resolution limit of a telescope according to the Rayleigh criterion. The same expression can be used to estimate the spatial resolution of a microscope. A lens with numerical aperture $NA = \sin(\theta)$ can resolve two point sources separated by lateral distance d provided

$$d > 1.22 \frac{\lambda}{NA}. \quad (5.29)$$

Equations (5.28,5.29) do not give the ultimate resolution limit of a telescope since the dip is clearly visible at the Rayleigh separation. Furthermore optimized detection of the collected light can be used to further enhance the resolution with an ultimate limit set by the intrinsic quantum noise of the light. Techniques for improving the resolution of an imaging system go under the name of “superresolution”[46, 47].

5.2.2 Spot of Arago

An interesting phenomenon appears if we consider the diffracted field without a focusing lens behind a uniformly illuminated circular plate which blocks the light for $0 \leq \rho \leq a$. With circular symmetry

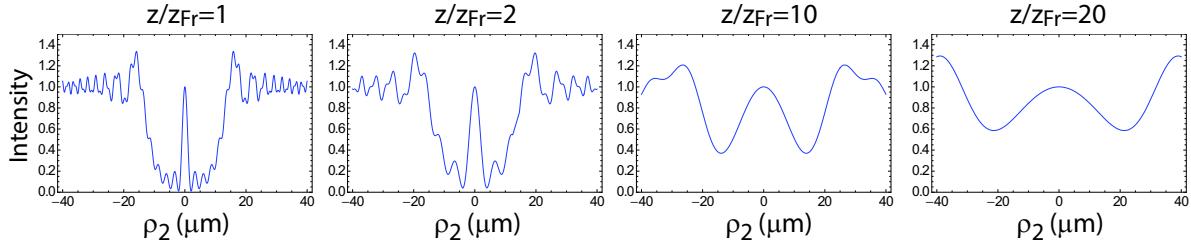


Figure 5.11: Spot of Arago calculated for $\lambda = 1.$ μm , $a = 10.$ μm , giving $z_{\text{Fresnel}} = 54$ μm . The diffracted intensity normalized to the incident intensity is shown for $z/z_{\text{Fresnel}} = 1, 2, 10, 20$.

Eq. (5.7) for Fresnel diffraction becomes

$$\begin{aligned} A(\rho_2) &= \frac{-ik}{2\pi z} \int_{-\infty}^{\infty} d\rho A(\rho) e^{i(k/2z)|\rho_2 - \rho|^2} \\ &= \frac{-ik}{z} e^{i\frac{k\rho_2^2}{2z}} \int_0^{\infty} d\rho A(\rho) \rho e^{i\frac{k\rho^2}{2z}} J_0\left(\frac{k\rho_2 \rho}{z}\right). \end{aligned} \quad (5.30)$$

For the uniformly illuminated circular plate we get

$$\begin{aligned} A(\rho_2) &= \frac{-ik}{z} e^{i\frac{k\rho_2^2}{2z}} \int_a^{\infty} d\rho \rho e^{i\frac{k\rho^2}{2z}} J_0\left(\frac{k\rho_2 \rho}{z}\right) \\ &= \frac{-ik}{z} e^{i\frac{k\rho_2^2}{2z}} \int_0^{\infty} d\rho \rho e^{i\frac{k\rho^2}{2z}} J_0\left(\frac{k\rho_2 \rho}{z}\right) + \frac{ik}{z} e^{i\frac{k\rho_2^2}{2z}} \int_0^a d\rho \rho e^{i\frac{k\rho^2}{2z}} J_0\left(\frac{k\rho_2 \rho}{z}\right). \end{aligned}$$

The infinite range integral is evaluated using $\int_0^{\infty} dx x e^{i a x^2} J_0(bx) = \frac{i}{2a} e^{-ib^2/4a}$ to give

$$A(\rho_2) = 1 + \frac{ik}{z} e^{i\frac{k\rho_2^2}{2z}} \int_0^a d\rho \rho e^{i\frac{k\rho^2}{2z}} J_0\left(\frac{k\rho_2 \rho}{z}\right).$$

When there is no blocking plate, $a = 0$ and $A(\rho_2) = 1$ as expected. Surprisingly when there is a plate the intensity is always nonzero at the origin. We can see this by calculating

$$A(0) = 1 + \frac{ik}{z} \int_0^a d\rho \rho e^{i\frac{k\rho^2}{2z}} = e^{i\frac{ka^2}{2z}}.$$

For all values of z there is unit intensity on axis behind the blocking disk. This is known as the spot of Arago and is shown in Fig. 5.11. Due to diffractive spreading the spot is only clearly visible for z up to about $10z_{\text{Fresnel}}$.

5.2.3 Fresnel zones

A phenomenon related to that of the previous section is the possibility of focusing without using a lens. Consider a uniformly illuminated source plane with radial coordinate ρ and an image plane a distance z after it with radial coordinate ρ_2 . On axis in the image plane $\rho_2 = 0$ and Eq. (5.30) predicts a field

$$A(0, z) = \frac{ik}{z} A_0 \frac{1}{2\pi} \int_0^{\infty} d\rho 2\pi \rho e^{i\frac{k\rho^2}{2z}}$$

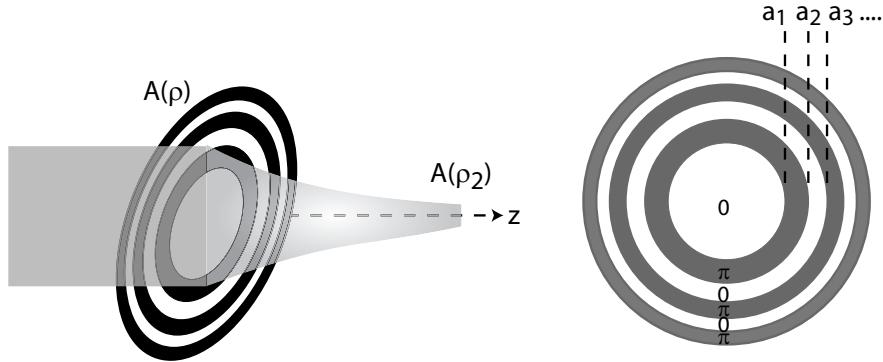


Figure 5.12: Fresnel zone plate for lensless focusing.

where A_0 is the field at the source plane. The integrand has a phase which starts at 0 and grows proportional to ρ^2 . The phase reaches the value $p\pi$ for $a_p = \sqrt{p2\pi z/k} = \sqrt{p\lambda z}$. The source area between radii a_{p-1} and a_p is

$$\sigma = \pi a_p^2 - \pi a_{p-1}^2 = \pi \lambda z [p - (p - 1)] = \pi \lambda z,$$

which is constant. We see that the source plane can be divided into zones of equal area, with the diffracted phase at the image plane varying by π from zone to zone. These are known as Fresnel zones.

The contributions to the diffracted field on axis from neighboring zones tend to cancel out since they have opposite signs. However, if we block every other zone, as in Fig. 5.12, the zones which are transmitted all add constructively and a large intensity will be seen on-axis. This construction is known as a Fresnel zone plate.

We can estimate the on-axis intensity as follows. The p^{th} zone contributes a power $\frac{\epsilon_0 c}{2} |A_0|^2 \sigma |\chi_p|^2$ where

$$\begin{aligned} \chi_p &= \frac{1}{\sigma} \int_{a_{p-1}}^{a_p} d\rho 2\pi\rho e^{i\frac{k\rho^2}{2z}} \\ &= -\frac{2i}{\pi} e^{i\pi p}, \end{aligned}$$

so $|\chi_p|^2 = 4/\pi^2 \simeq 0.405$. Due to the phase averaging within each zone the lens is only about 40% efficient. The integrated intensity from N zones is thus

$$I_N = \frac{\epsilon_0 c}{2} |A_0|^2 \frac{4N}{\pi^2} \frac{\sigma}{\sigma_f},$$

where σ_f is the area of the focused spot. Relative to the incident intensity $I_0 = \frac{\epsilon_0 c}{2} |A_0|^2$ we find

$$\frac{I_N}{I_0} = N \frac{4\lambda z}{\pi \sigma_f}.$$

A calculation of σ_f can be made using Fresnel integrals as in the spot of Arago calculation. Assuming $k\rho\rho_2/z \ll 1$ we expand the Bessel function as $J_0(x) \simeq 1 - x^2/4$ and find that the field from the p^{th} zone vanishes at $\rho_2 \simeq \sqrt{\lambda z/(\pi^2 p)}$ for $p \gg 1$. The focal spot size decreases at large p . The

limit on the focusing is set by $p = 1$ for which case the field vanishes at $\rho_2 = \sqrt{2\lambda z/\pi^2}$ and we can estimate the focal spot area as $\sigma_f \sim \pi(2\lambda z/\pi^2) = (2/\pi)(\lambda z)$ and

$$\frac{I_N}{I_0} \sim 2N$$

for N alternating zones added together. While the zone plate does not produce better focusing than a refractive lens it is useful for wavelengths where it is difficult to produce materials with desired refractive properties. This is the case in x-ray optics where Fresnel zone plates are commonly used for focusing. Another application, and the original motivation for the development of the Fresnel lens, is for lighthouses. Compared to a refractive lens the Fresnel design is preferable due to the possibility of fabricating a large mobile lens with reduced mass.

For a given aperture of radius a the Fresnel number N_{Fresnel} is defined as the number of Fresnel zones seen in the image plane. This is given by $\sqrt{N_{\text{Fresnel}}\lambda z} = a$ or

$$N_{\text{Fresnel}} = \frac{a^2}{\lambda z}.$$

Note that $N_{\text{Fresnel}} = 1$ corresponds to $z = a^2/\lambda \sim z_{\text{Fraunhofer}}$ (see Eq. (5.13)). Put another way if the source aperture has $N_{\text{Fresnel}} \leq 1$ then Fraunhofer diffraction is adequate to calculate the image field. Conversely if $N_{\text{Fresnel}} > 1$ we must use Fresnel diffraction theory.

5.2.4 Talbot effect

Another interesting piece of Fourier optics is the Talbot effect which results in the phenomenon of lensless imaging of periodic objects. Consider an optical field $A(x, z)$ that is periodic along x with period L_x . This could be a field with periodic amplitude, phase, or both. Periodicity implies that the field can be written as

$$A(x, z_0) \sim \cos(2\pi x/L_x) \sim e^{i2\pi x/L_x} + e^{-i2\pi x/L_x}. \quad (5.31)$$

From Eq. (5.31) we identify the field as being due to the interference of plane waves with transverse wave numbers $k_x = \pm 2\pi/L_x$. Transverse wave numbers that are a factor m larger, with m an integer, will also synthesize a field with period L_x so a general periodic field can be written as

$$A(x, z_0) = \sum_{m=1}^{\infty} c_m e^{imk_x x} + c_m^* e^{-imk_x x} \quad (5.32)$$

with the c_m complex amplitudes.

Propagation of the field (5.32) an axial distance L_z will add quadratic phases to the plane wave components giving a new field

$$A(x, z_0 + L_z) = \sum_{m=1}^{\infty} \left(c_m e^{imk_x x} + c_m^* e^{-imk_x x} \right) e^{-i\frac{m^2 k_x^2}{2k} L_z}. \quad (5.33)$$

The new field will be identical to the original field when $\frac{m^2 k_x^2}{2k} L_z = m^2 2\pi$ or

$$L_z = \frac{4\pi k}{k_x^2} = \frac{2L_x^2}{\lambda} \equiv L_{\text{Talbot}}.$$

We see that the periodic field distribution exactly reproduces itself at multiples of the Talbot length L_{Talbot} . This phenomenon was discovered by Talbot[48], and has been studied with both optical and matter waves.

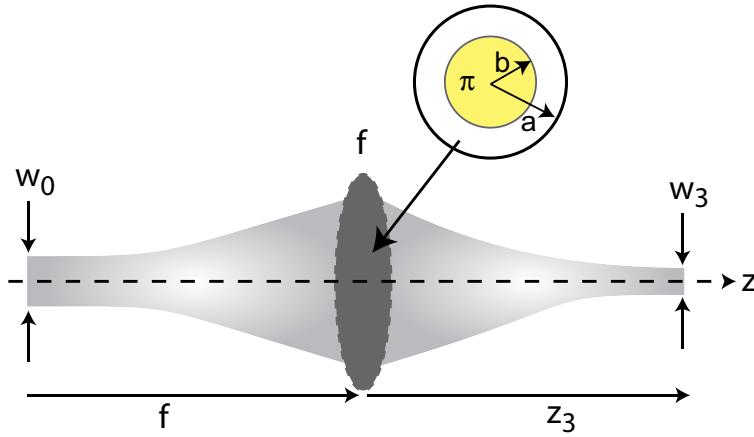


Figure 5.13: Gaussian beam illuminating a lens with a π phase in the inner region of radius b .

It is instructive to rewrite the field with $z_0 = 0$ as

$$A(x, z) = \sum_{m=1}^{\infty} \left(c_m e^{imk_x x} + c_m^* e^{-imk_x x} \right) e^{-i2\pi m^2 \frac{z}{L_{\text{Talbot}}}} \quad (5.34)$$

which makes explicit the periodicity at the Talbot length. At subharmonics of the Talbot length there are other effects. At $z = L_{\text{Talbot}}/2$ the pattern is shifted half a period along x . In addition if the field is weakly amplitude modulated and the modulation is an even function of x then at $z = L_{\text{Talbot}}/4$ the amplitude modulation is converted into a pure phase modulation. At other subharmonic distances higher frequency copies of the original field are observed. A more extensive discussion can be found in review articles[49, 50].

5.2.5 Bottle beam lenses

As another example of the use of the Hankel transform let's consider a thin lens of focal length f with radius a and an inner region of radius $b < a$ which has a π phase shift relative to the outer as shown in Fig. 5.13. We will illuminate the lens with a Gaussian beam with waist w_0 in the front focal plane. This configuration can be used to create a Bottle Beam optical trap (BBT). We can develop an analytical description of the field in the focal region using Fresnel diffraction theory.

Call the field in the front focal plane A_0 then at the lens the field is

$$A_1 = \mathcal{F}^{-1}[\mathcal{F}[A_0(\rho)]H(k_{\perp}, f)]$$

with $H(k_{\perp}, z) = e^{-i\frac{k_{\perp}^2 z}{2k}}$. Passage through the thin lens with two annular regions is accounted for by multiplying with the transmission function

$$t(\rho) = e^{-i\frac{k\rho^2}{2f}} [1 - 2 \text{circ}(\rho/b)]$$

where we have introduced the radial step function $\text{circ}(\rho) = 1$ for $\rho \leq 1$ and $\text{circ}(\rho) = 0$ for $\rho > 1$. We will assume the field does not extend to the outer boundary of the lens and ignore the radius

a. We thus get $A_2(\rho) = tA_1(\rho)$ and the field in the output plane a distance z after the lens is given by

$$\begin{aligned} A_3(\rho_3, z) &= \mathcal{F}_{32}^{-1}[\mathcal{F}_{22}[A_2(\rho_2)]H(k_{\perp 2}, z)] \\ &= \mathcal{F}_{32}^{-1}[\mathcal{F}_{22}[t(\rho_1)\mathcal{F}_{10}^{-1}[\mathcal{F}_{00}[A_0(\rho_0)]H(k_{\perp 0}, f)]]H(k_{\perp 2}, z)]. \end{aligned} \quad (5.35)$$

Subscripts on the variables and operators have been introduced to indicate different transverse coordinates corresponding to the different planes in Fig. 5.13. Writing out (5.35) explicitly the output field can be expressed as

$$\begin{aligned} A_3(\rho_3, z_3) &= \int dk_{\perp 1} k_{\perp 1} J_0(\rho_3 k_{\perp 1}) e^{-i\frac{k_{\perp 1}^2 z_3}{2k}} \int d\rho_1 \rho_1 J_0(\rho_1 k_{\perp 1}) e^{-i\frac{k\rho_1^2}{2f}} [1 - 2 \text{circ}(\rho_1/b)] \\ &\quad \times \int dk_{\perp 0} k_{\perp 0} J_0(\rho_1 k_{\perp 0}) e^{-i\frac{k_{\perp 0}^2 f}{2k}} \int d\rho_0 \rho_0 J_0(\rho_0 k_{\perp 0}) A_0(\rho_0). \end{aligned} \quad (5.36)$$

The factor $[1 - 2 \text{circ}(\rho_1/b)]$ splits the result into two terms. The first one, which is independent of b , results in a Gaussian beam with waist $w_3 = \lambda f / (\pi w_0)$ at $z_3 = f$. The second term is more complicated and can be written as

$$A_{3b} = -\frac{2A_0}{1 + iw_3/w_0} \int_0^\infty dk_{\perp 1} k_{\perp 1} J_0(\rho_3 k_{\perp 1}) e^{-i\frac{w_0 w_3 s k_{\perp 1}^2}{4}} \int_0^b d\rho_1 \rho_1 J_0(k_{\perp 1} \rho_1) e^{-h\rho_1^2}$$

with $h = k^2 w_0^2 / (4f^2 - i2fk w_0^2) = 1/(w_3^2 - iw_0 w_3)$. We have also introduced a normalized axial coordinate $s = z_3/f$, so that $s = 1$ corresponds to the back focal plane. Reversing the order of integration gives

$$\begin{aligned} &\int_0^\infty dk_{\perp 1} k_{\perp 1} J_0(\rho_3 k_{\perp 1}) e^{-i\frac{w_0 w_3 s k_{\perp 1}^2}{4}} \int_0^b d\rho_1 \rho_1 J_0(k_{\perp 1} \rho_1) e^{-h\rho_1^2} \\ &= \int_0^b d\rho_1 \rho_1 e^{-h\rho_1^2} \int_0^\infty dk_{\perp 1} k_{\perp 1} J_0(\rho_3 k_{\perp 1}) J_0(k_{\perp 1} \rho_1) e^{-i\frac{w_0 w_3 s k_{\perp 1}^2}{4}}. \end{aligned}$$

To evaluate the integral over $dk_{\perp 1}$ use (Watson, p.395)

$$\int_0^\infty dz z J_\nu(az) J_\nu(bz) e^{-pz^2} = \frac{1}{2p} e^{-\frac{a^2+b^2}{4p}} I_\nu\left(\frac{ab}{2p}\right). \quad (5.37)$$

Here I_ν is a modified Bessel function and the result is valid provided $\text{Re}(\nu) > -1/2$ and $\text{Re}(p) > 0$. In our case $p = iw_0 w_3 s / 4$ and $\text{Re}(p) = 0$ which violates the condition on p but (5.37) is still valid since we have² $\text{Re}(a) > 0$, $\text{Re}(b) > 0$. Proceeding with $a = \rho_1$, $b = \rho_3$, and $p = iw_0 w_3 s / 4$ we find

$$A_{3b} = 4i \frac{A_0}{w_3 s (w_0 + iw_3)} e^{i\frac{\rho_3^2}{w_0 w_3 s}} \int_0^b d\rho_1 \rho_1 e^{\left(-h + \frac{i}{w_0 w_3 s}\right) \rho_1^2} J_0\left(\frac{2\rho_1 \rho_3}{w_0 w_3 s}\right).$$

²To show this use (G & R 6.729, 1,2)

$$\begin{aligned} \int_0^\infty dz z J_\nu(az) J_\nu(bz) \cos(p' z^2) &= \frac{1}{2p'} \sin\left(\frac{a^2 + b^2}{4p'} - \frac{\nu\pi}{2}\right) J_\nu\left(\frac{ab}{2p'}\right), \\ \int_0^\infty dz z J_\nu(az) J_\nu(bz) \sin(p' z^2) &= \frac{1}{2p'} \cos\left(\frac{a^2 + b^2}{4p'} - \frac{\nu\pi}{2}\right) J_\nu\left(\frac{ab}{2p'}\right), \end{aligned}$$

provided $\text{Re}(\nu) > -2$, $\text{Re}(a) > 0$, $\text{Re}(b) > 0$, $\text{Re}(p') > 0$. These can be combined to give

$$\int_0^\infty dz z J_\nu(az) J_\nu(bz) e^{-ip' z^2} = \frac{-i}{2p'} e^{i\left(\frac{a^2+b^2}{4p'} - \frac{\nu\pi}{2}\right)} J_\nu\left(\frac{ab}{2p'}\right),$$

This last integral arises in the problem of diffraction from a finite circular aperture and can be expressed in terms of Lommel functions of two variables ([51], Watson p.540)

$$U_1(u, v) = u \int_0^1 dt t J_0(vt) \cos \left[\frac{u}{2}(1-t^2) \right] \quad (5.38a)$$

$$U_2(u, v) = u \int_0^1 dt t J_0(vt) \sin \left[\frac{u}{2}(1-t^2) \right]. \quad (5.38b)$$

The field is then

$$A_{3b} = 4i \frac{A_0 b^2 e^{-iu/2}}{w_3 s(w_0 + iw_3)u} e^{i\frac{\rho_3^2}{w_0 w_3 s}} [U_1(u, v) + iU_2(u, v)]$$

with

$$u = -2b^2 \frac{w_3^2 - w_0^2(s-1) + iw_0 w_3 s}{w_0 w_3 s(w_0^2 + w_3^2)}, \quad v = \frac{2b\rho_3}{w_0 w_3 s}.$$

This completes the calculation of the field of the bottle beam. To evaluate the field and intensity we must numerically evaluate the Lommel functions. This can be done either by numerical integration of (5.38) or by evaluation of the functions expressed as infinite sums of Bessel functions (Watson p. 537)

$$\begin{aligned} U_1(u, v) &= \sum_{m=0}^{\infty} (-1)^m \left(\frac{u}{v} \right)^{1+2m} J_{1+2m}(v), \\ U_2(u, v) &= \sum_{m=0}^{\infty} (-1)^m \left(\frac{u}{v} \right)^{2+2m} J_{2+2m}(v). \end{aligned}$$

Let us calculate the intensity of the bottle beam as a function of ρ_3 at the focal plane $z_3 = f$ or $s = 1$. We find

$$I(\rho_3, s = 1) = \frac{\epsilon_0 c}{2} \left| -i A_0 \frac{w_0}{w_3} e^{-\rho_3^2/w_3^2} + A_{3b} \right|^2.$$

The factor of $-i$ in front of the Gaussian is a propagation phase due to the Gouy term in the expression for a Gaussian beam. The intensity is plotted in Fig. 5.14 for some representative parameters. Note that the on-axis intensity does not go to zero for any value of b due to the small diffractive phase shift which prevents perfect field cancellation. This could be corrected for by using a phase shift that is slightly different from π (analysis to be added). As the axial plane is moved away from the focus the on-axis intensity grows. We see that the minimum of the potential is about 2/3 of the radial peak in the focal plane. Figure 5.15 shows that the intensity profile is not symmetric about the focal plane as z is varied.

The on-axis intensity at $\rho_3 = 0$ should be as small as possible for a bottle beam. On axis

$$\begin{aligned} A_{3b}(0) &= 4i \frac{A_0 b^2 e^{-iu/2}}{w_3 s(w_0 + iw_3)u} [U_1(u, 0) + iU_2(u, 0)] \\ &= 4 \frac{A_0 b^2}{w_3 s(w_0 + iw_3)u} \left(1 - e^{-iu/2} \right), \end{aligned}$$

and putting $p = ip'$ we get

$$\int_0^\infty dz z J_\nu(az) J_\nu(bz) e^{-pz^2} = \frac{1}{2p} e^{-\frac{a^2+b^2}{4p}} e^{-i\frac{\nu\pi}{2}} J_\nu\left(\frac{iab}{2p}\right) = \frac{1}{2p} e^{-\frac{a^2+b^2}{4p}} I_\nu\left(\frac{ab}{2p}\right),$$

which is the same as (5.37) even though $\text{Re}(p) = 0$. Thus (5.37) can be extended to the case of $\text{Re}(p) = 0$ provided $\text{Re}(a) > 0$, $\text{Re}(b) > 0$.

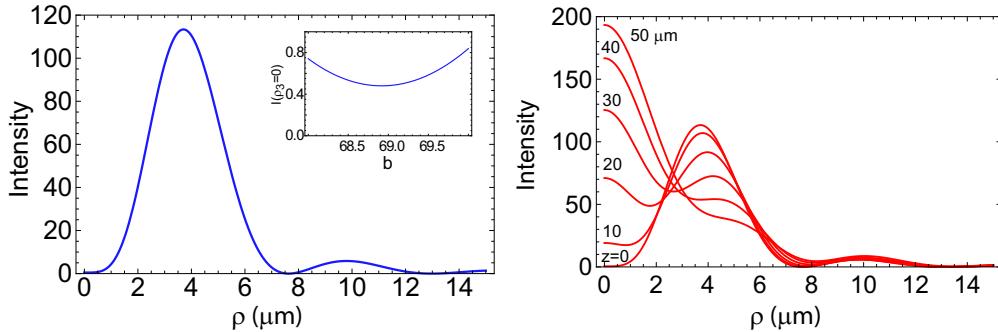


Figure 5.14: BoB profiles for $w_0 = 82.76 \mu\text{m}$, $w_3 = 3 \mu\text{m}$, $b = 69 \mu\text{m}$ and $s = 1$ (left). These parameters correspond to $\lambda = 0.78 \mu\text{m}$ and $f = 1 \text{ mm}$. The plotted intensity is normalized to the peak of the input Gaussian with waist w_0 . The inset shows that the on-axis intensity does not vanish for any b . On the right the transverse profiles are shown at axial displacements up to $50 \mu\text{m}$ from the focal plane.

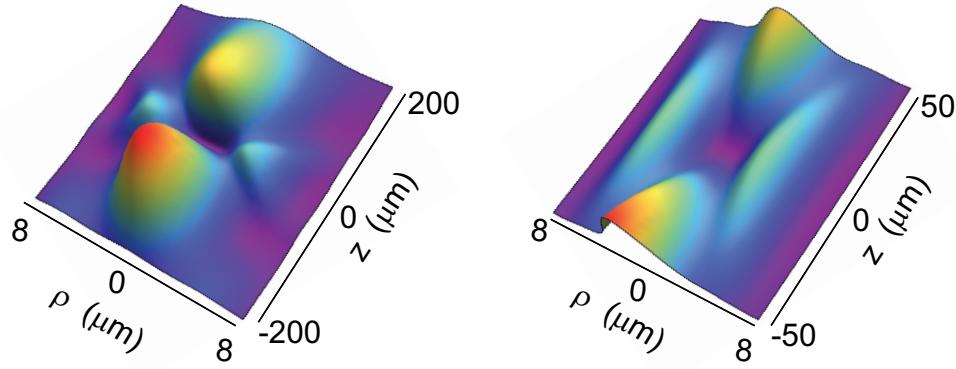


Figure 5.15: Three dimensional intensity distribution about the focus. Same parameters as in Fig. 5.14.

where we have used $U_1(u, 0) = \sin(u/2)$, $U_2(u, 0) = 2 \sin^2(u/4)$. The intensity is thus zero when

$$e^{-iu/2} = 1 + i \frac{w_3^2 - w_0^2(s-1) + iw_0w_3s}{2w_3(w_0 - iw_3)}. \quad (5.40)$$

Solving for u and then b we verify that there are no solutions for $s = 1$ and real b . We can get an approximate solution for b in the limit of $w_3 \ll w_0$ and $s = 1$ which gives for the example in Fig. 5.14

$$\text{Re}(b) \approx 71.3 \mu\text{m} + \mathcal{O}((w_3/w_0)^2)$$

provided we choose the appropriate branch of the solution to (5.40). This is close to the numerically found optimum of $b = 69 \mu\text{m}$.

5.3 Fourier optical image processing

Sophisticated image processing tasks can be performed rapidly using optical transformations. A few examples are given in this section.

5.3.1 Repeated objects

The Fourier transformation properties of lenses can be used for image processing. Let's start by considering the Fourier transform of a repeated object. An important example of a repeated object is a diffraction grating which is useful for determining the amount of energy as a function of frequency or wavelength in an optical beam.

Consider an object $A(\rho)$ which is some arbitrary optical field that depends on the two-dimensional coordinate $\rho = x\hat{x} + y\hat{y}$. A displaced version of the object centered at ρ_j can be written as

$$\begin{aligned} A(\rho - \rho_j) &= A(\rho) * \delta(\rho - \rho_j) \\ &= \int d\rho' A(\rho') \delta(\rho' - (\rho - \rho_j)) \\ &= A(\rho - \rho_j). \end{aligned}$$

Convolving any well behaved function with a delta function centered at ρ_j gives a copy of the function displaced by ρ_j .

Suppose that we have N such displaced copies of the object centered at coordinates ρ_j , $j = 1, N$. Using the convolution theorem for Fourier transforms the transform of the object field is

$$\begin{aligned} \mathcal{F} \left[\sum_j A(\rho - \rho_j) \right] &= 2\pi \sum_j \mathcal{F}[A(\rho) * \delta(\rho - \rho_j)] \\ &= 2\pi \mathcal{F}[A(\rho)] \sum_j \mathcal{F}[\delta(\rho - \rho_j)] \\ &= 2\pi \tilde{A}(\mathbf{k}_\perp) \sum_j \frac{1}{2\pi} e^{-i\rho_j \cdot \mathbf{k}_\perp} \\ &= \tilde{A}(\mathbf{k}_\perp) \sum_j e^{-i\rho_j \cdot \mathbf{k}_\perp}. \end{aligned}$$

If the copies of the object are uniformly spaced on a regular grid separated by a displacement vector \mathbf{h} we can put $\rho_j = j\mathbf{h}$ and the sum becomes

$$\begin{aligned} \sum_{j=1}^N e^{-i\rho_j \cdot \mathbf{k}_\perp} &= \sum_{j=1}^N e^{-ij\mathbf{h} \cdot \mathbf{k}_\perp} \\ &= e^{-i\mathbf{h} \cdot \mathbf{k}_\perp N} \frac{e^{i\mathbf{h} \cdot \mathbf{k}_\perp N} - 1}{e^{i\mathbf{h} \cdot \mathbf{k}_\perp} - 1}. \end{aligned}$$

The intensity of the transformed object will thus be

$$\begin{aligned} I &= |\tilde{A}(\mathbf{k}_\perp)|^2 \left| \frac{e^{i\mathbf{h} \cdot \mathbf{k}_\perp N} - 1}{e^{i\mathbf{h} \cdot \mathbf{k}_\perp} - 1} \right|^2 \\ &= |\tilde{A}(\mathbf{k}_\perp)|^2 \frac{\sin^2(\mathbf{h} \cdot \mathbf{k}_\perp N/2)}{\sin^2(\mathbf{h} \cdot \mathbf{k}_\perp/2)}. \end{aligned} \tag{5.41}$$

Equation (5.41) can be understood as follows. The Fourier transform of N copies of an object, uniformly spaced by multiples of \mathbf{h} is the Fourier transform of the object multiplied by a function $\sin^2(\alpha N)/\sin^2(\alpha)$ with $\alpha = \mathbf{h} \cdot \mathbf{k}_\perp/2$. If $\tilde{A}(\mathbf{k}_\perp)$ has a maximum at the origin then N copies of the object will give a maximum intensity that is increased by a factor of N^2 . This is to be expected since we have N times the field added together and squaring to get the intensity gives a factor N^2 .

The width of the transform will be reduced by the $\sin^2(\alpha N)$ factor which has a zero at $\alpha_0 N = \pi$ or

$$\alpha_0 = \mathbf{h} \cdot \mathbf{k}_{\perp,0}/2 = \pi/N.$$

Thus the width of the central maximum will be reduced by a factor of N . In addition there will be secondary maxima at zeroes of the denominator when $\mathbf{h} \cdot \mathbf{k}_\perp = 2m\pi$ or

$$k_{\perp,m} = m \frac{2\pi}{h}$$

when \mathbf{k}_\perp is parallel to \mathbf{h} .

These features of an intensity increase by N^2 , a reduction in width of N , and periodic maxima of the transform are true for generic objects $A(\rho)$. They are particularly useful for achieving enhanced spectral sensitivity.

5.3.2 Periodic apertures

Let's check an explicit example of a repeated 1D aperture. Let the aperture function $A(x)$ be a periodic array of N slits separated by h and each of width δx . The Fraunhofer diffraction pattern in the focal plane of a lens of focal length f is

$$A(x_2) = \left(\frac{-ik}{2\pi f} \right)^{1/2} \int_{-\infty}^{\infty} dx A(x) e^{-i \frac{kxx_2}{f}}.$$

The aperture function is

$$\begin{aligned} A(x) &= \sum_{j=1}^N \text{rect} \left(\frac{x - (j - \frac{N+1}{2}) h}{\delta x/2} \right) \\ &= \sum_{j=1}^N \text{rect} \left(\frac{x}{\delta x/2} - \frac{x_j}{\delta x/2} \right) \end{aligned}$$

with $x_j = [j - (N + 1)/2]h$ and $\text{rect}(x) = 1$ for $|x| \leq 1$. The Fourier transform then takes the form

$$A(x_2) = \left(\frac{-ik}{2\pi f} \right)^{1/2} \sum_{j=1}^N e^{i \frac{kx_2 x_j}{f}} \int_{-\infty}^{\infty} dx \text{rect}[x/(\delta x/2)] e^{-i \frac{kxx_2}{f}}.$$

The integral no longer depends on j and we get

$$A(x_2) = \left(\frac{-ik(\delta x)^2}{2\pi f} \right)^{1/2} \text{sinc} \left[\frac{k\delta x x_2}{2f} \right] \sum_{j=1}^N e^{i \frac{kxx_j}{f}}.$$

The geometric sum is evaluated using $\sum_{j=1}^N e^{-iaj} = \frac{e^{-iaN} - 1}{1 - e^{ia}}$ so

$$A(x_2) = \left(\frac{-ik(\delta x)^2}{2\pi f} \right)^{1/2} \text{sinc} \left[\frac{k\delta x x_2}{2f} \right] \frac{\sin \left(\frac{Nkhx_2}{2f} \right)}{\sin \left(\frac{khx_2}{2f} \right)}. \quad (5.42)$$

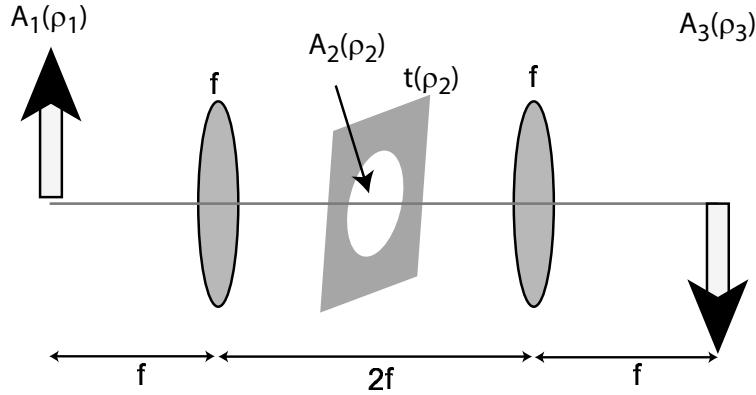


Figure 5.16: Fourier filtering setup. The filter function $t(\rho_2)$ is positioned in the Fourier plane between the lenses.

This formula will reappear when we discuss the spectral resolution of a diffraction grating. The sinc function is just the Fourier transform of one unit cell of the grating, while the factor of $\frac{\sin\left(\frac{Nkhx_2}{2f}\right)}{\sin\left(\frac{khx_2}{2f}\right)}$ gives N times narrower lobes which are repeated every $\Delta x_2 = 2\pi f/(kh)$. This analysis implies that the spectroscopic resolution of a diffraction grating increases with N , the number of grating periods.

Binary amplitude holography

Let's consider a specific example of Eq. (5.42) with the following parameters: $\lambda = 1 \mu\text{m}$, $h = 15 \mu\text{m}$, $\delta x = 13 \mu\text{m}$, $f = 2 \text{ cm}$, and $N = 100$. We then find

$$A(x_2) = (-i)^{1/2} 0.0919 \operatorname{sinc}(0.00204 x_2) \frac{\sin(0.2356 x_2)}{\sin(0.002356 x_2)} \quad (5.43)$$

with x_2 measured in microns.

5.3.3 Fourier filtering

Spectral filtering of images can be performed using the $4 - f$ optical processor shown in Fig. 5.16. An input image $A_1(\rho_1)$ is Fourier transformed by a lens with focal length f which gives the Fourier transform $A_2(\rho_2) \sim \mathcal{F}[A_1(\rho_1)]_{\mathbf{k}_{\perp 1}=k\rho_2/f}$ in the back focal plane of the first lens. This is then transformed again with the second lens. When $L_1 = L_2 = f$ and $L = 2f$ the field in the back focal plane of the second lens $A_3(\rho_3)$ is an inverted image of the input field. We can modify this image by inserting amplitude or phase filters $t(\rho_2)$ into the Fourier plane. Using the Fourier convolution theorem we find

$$A_3(\rho_3) = -\frac{ik}{f} \mathcal{F}[t(\rho_2)A_2(\rho_2)]_{\mathbf{k}_{\perp 2}=k\rho_3/f} \quad (5.44a)$$

$$= -\frac{ik}{f} \frac{1}{2\pi} \mathcal{F}[t(\rho_2)]_{\mathbf{k}_{\perp 2}=k\rho_3/f} * \mathcal{F}[A_2(\rho_2)]_{\mathbf{k}_{\perp 2}=k\rho_3/f}. \quad (5.44b)$$

We see that the Fourier filtering operation can be expressed in two equivalent ways. As a filtering operation described by Eq. (5.44a) where we multiply the transform of the input field by the

Fourier plane filter $t(\boldsymbol{\rho}_2)$ and then transform again to get the output. Alternatively we convolve the transform of the input field with the transform of the filter function. This is the content of Eq. (5.44b).

Although the convolution theorem provides a compact expression for the output field it is easy to run into trouble in the evaluation of the convolution. Let us evaluate the output field explicitly using

$$A_3(\boldsymbol{\rho}_3) = -\frac{ik}{f} \frac{1}{2\pi} \iint d\boldsymbol{\rho}_2 t(\boldsymbol{\rho}_2) A_2(\boldsymbol{\rho}_2) e^{-i\frac{k\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}_3}{f}}.$$

Then use the Fourier representations

$$t(\boldsymbol{\rho}_2) = \frac{1}{2\pi} \iint d\boldsymbol{\rho}' \tilde{t}(\boldsymbol{\rho}') e^{i\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}'}$$

$$A_2(\boldsymbol{\rho}_2) = \frac{1}{2\pi} \iint d\boldsymbol{\rho}'' \tilde{A}_2(\boldsymbol{\rho}'') e^{i\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}''}$$

to get

$$\begin{aligned} A_3(\boldsymbol{\rho}_3) &= -\frac{ik}{f} \frac{1}{2\pi} \iint d\boldsymbol{\rho}'' \tilde{A}_2(\boldsymbol{\rho}'') \iint d\boldsymbol{\rho}' \tilde{t}(\boldsymbol{\rho}') \\ &\times \frac{1}{4\pi^2} \iint d\boldsymbol{\rho}_2 e^{-i\frac{k\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}_3}{f}} e^{i\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}''} e^{i\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}'} \\ &= -\frac{ik}{f} \frac{1}{2\pi} \iint d\boldsymbol{\rho}'' \tilde{A}_2(\boldsymbol{\rho}'') \iint d\boldsymbol{\rho}' \tilde{t}(\boldsymbol{\rho}') \delta(\boldsymbol{\rho}' + \boldsymbol{\rho}'' - \frac{k}{f}\boldsymbol{\rho}_3) \\ &= -\frac{ik}{f} \frac{1}{2\pi} \iint d\boldsymbol{\rho}' \tilde{t}(\boldsymbol{\rho}') \tilde{A}_2(\frac{k}{f}\boldsymbol{\rho}_3 - \boldsymbol{\rho}') \\ &= -\frac{ik}{f} \frac{1}{2\pi} \iint d\boldsymbol{\rho}' \tilde{t}(\frac{k}{f}\boldsymbol{\rho}_3 - \boldsymbol{\rho}') \tilde{A}_2(\boldsymbol{\rho}'), \end{aligned} \quad (5.45)$$

which makes explicit (5.44b).

As a check let $t(\boldsymbol{\rho}_2) = 1$, then $\tilde{t}(\boldsymbol{\rho}') = 2\pi\delta(\boldsymbol{\rho}')$ and

$$\begin{aligned} A_3(\boldsymbol{\rho}_3) &= -\frac{ik}{f} \iint d\boldsymbol{\rho}' \tilde{A}_2(\boldsymbol{\rho}') \delta(\frac{k}{f}\boldsymbol{\rho}_3 - \boldsymbol{\rho}') \\ &= -\frac{ik}{f} \tilde{A}_2(\frac{k}{f}\boldsymbol{\rho}_3) \\ &= -\frac{ik}{f} \frac{1}{2\pi} \iint d\boldsymbol{\rho}_2 A_2(\boldsymbol{\rho}_2) e^{-i\frac{k}{f}\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}_3} \\ &= \left(-\frac{ik}{f}\right)^2 \iint d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) \frac{1}{4\pi^2} \iint d\boldsymbol{\rho}_2 e^{-i\frac{k}{f}\boldsymbol{\rho}_2 \cdot \boldsymbol{\rho}_3} e^{-i\frac{k}{f}\boldsymbol{\rho}_1 \cdot \boldsymbol{\rho}_2} \\ &= \left(-\frac{ik}{f}\right)^2 \iint d\boldsymbol{\rho}_1 A_1(\boldsymbol{\rho}_1) \delta(\frac{k}{f}(\boldsymbol{\rho}_3 + \boldsymbol{\rho}_1)) \\ &= -A_1(-\boldsymbol{\rho}_3). \end{aligned} \quad (5.46)$$

We obtain an inverted image as expected with a phase factor of $-1 = (-i)^2$ which results from diffraction. With other filter functions we can implement low pass, high pass, or bandpass operations. We can also implement directional filtering, e.g. to suppress or enhance edges parallel to a desired transverse direction.

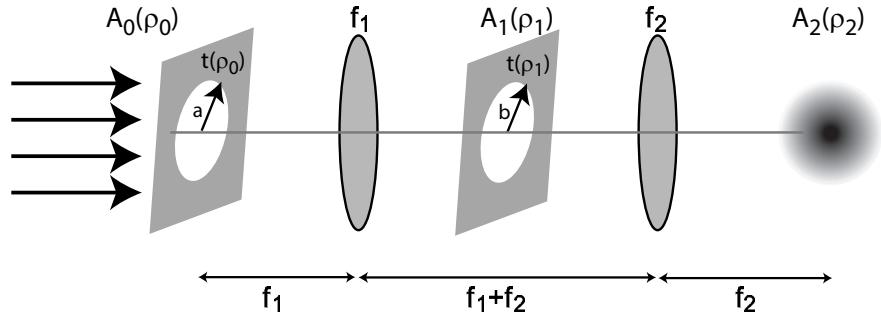


Figure 5.17: Fourier filtering setup for beam shaping starting with a uniform intensity input.

This type of imaging arrangement can be used to establish the diffraction limited resolution of an optical imaging system. Suppose the input field is a sinusoidal pattern with period Λ . We can think of this as being created by two plane waves propagating at angles $\pm\theta/2$ from the optical axis where $\Lambda = \lambda/(2 \sin \theta/2)$. These waves have transverse wavevector components $\pm k_{\perp}$ where $\sin \theta/2 = k_{\perp}/k$. Upon Fourier transformation the field A_2 will have strong components at $\rho_2 = fk_{\perp}/k$. Due to the finite diameter D of the first lens the maximum value of ρ_2 that passes through the optical system is $\rho_{2,\max} = D/2$. The maximum transverse wavenumber is thus $k_{\perp,\max} = \frac{k}{f}\rho_{2,\max} = \frac{kD}{2f}$. This corresponds to $\sin(\theta_{\max}/2) = D/2f$ and

$$\Lambda_{\min} = \frac{\lambda}{2 \sin(\theta_{\max}/2)} = \lambda \frac{f}{D}.$$

We see that the minimum spatial scale that propagates through the optical system is proportional to λ . Recalling that the numerical aperture of the lens is defined as $NA = \sin(\theta/2)$ we can write this simply as

$$\Lambda_{\min} = \frac{\lambda}{2(NA)}$$

which shows that the spatial resolution is inversely proportional to the numerical aperture.

5.3.4 Beam shaping

One application of Fourier filtering is in the preparation of amplitude modulated beam profiles. Consider the $4f$ arrangement of Fig. 5.17. An aperture of radius a is uniformly illuminated in the input plane. A lens of focal length f_1 produces an Airy pattern in the back focal plane that is filtered by an aperture of radius b . The field is then Fourier transformed to the output plane with a lens of focal length f_2 . Using Eqs. (5.21,5.27) the output field is given by

$$A_2(\rho_2) = -A_0 \frac{ak}{f_2} \int_0^b d\rho_1 J_0\left(\frac{\rho_2 k}{f_2} \rho_1\right) J_1\left(\frac{ak}{f_1} \rho_1\right) \quad (5.47)$$

with A_0 the input plane amplitude. The finite integral of Bessel functions (5.47) can be expressed as a power series in b using[52]

$$\int_0^b dz J_0(cz) J_1(dz) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!(j+1)!(2j+2)} {}_2F_1(-j, -1-j; 1; c^2/d^2) b^{2+2j} (d/2)^{1+2j}. \quad (5.48)$$

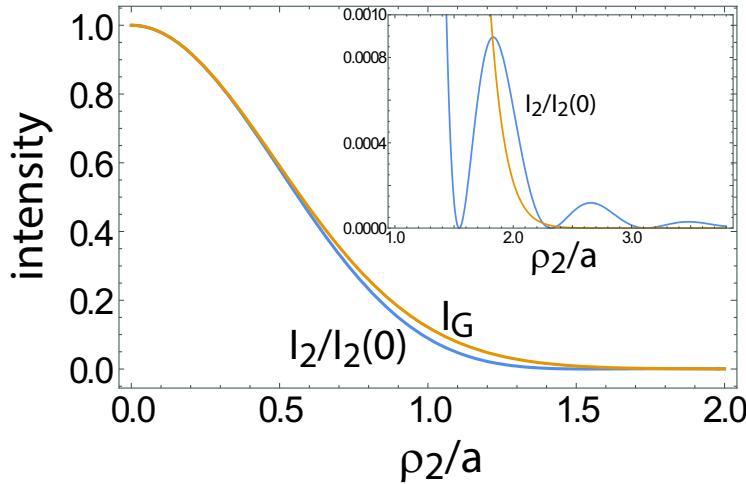


Figure 5.18: Near Gaussian intensity profile I_2 obtained by Fourier filtering of the Airy pattern resulting from uniform illumination of a circular aperture.

Here ${}_2F_1$ is the hypergeometric function. We can simplify the result by taking $f_1 = f_2 = f$ and setting $b = \frac{f}{ak}x_1$ where $x_1 = 3.8317$ is the first zero of J_1 . This corresponds to setting the aperture size in the filter plane to block the Airy rings outside of the central lobe. Doing so incurs only a small power loss since the integrated power in the central lobe is 0.84 of the total power $I_0\pi a^2$, with I_0 the input intensity.

With these choices the output field can be expressed as a power series in ρ_2/a . The leading terms are

$$\frac{I_2(\rho_2)}{I_0} = 1.978 - 4.147 \left(\frac{\rho_2}{a}\right)^2 + 3.918 \left(\frac{\rho_2}{a}\right)^4 - \dots$$

We will refer to this beam profile as an Airy-Gauss (AG) beam since it is created by filtering an Airy pattern and the intensity has a near Gaussian form, which near the origin is a quadratic function of ρ_2 , as is seen in Fig. 5.18. Matching the quadratic term with that of a Gaussian intensity profile³ $I_G = e^{-2\rho_2^2/w^2}$ we find $w = 0.974a$. Thus, to a good approximation, Fourier filtering of a uniformly illuminated circular aperture produces a Gaussian profile with waist parameter slightly less than the aperture radius a . Although the filtered Airy beam is not a pure Gaussian, and has secondary lobes as seen in the inset of Fig. 5.18, the lobes are sufficiently weak that the profile after diffractive propagation remains close to that of a Gaussian. It is interesting to note that time reversal symmetry implies that by starting with a Gaussian beam and propagating through a similar double aperture setup it is possible to efficiently prepare a near uniform top hat type beam.

This Fourier filtering approach to beam shaping can be readily extended to create an array of Gaussian like beams. Consider Fig. 5.19 where the input aperture is replaced by an array of apertures on a two-dimensional grid with spacing d . The field transmitted through each aperture will be of the form given by (5.47) but will appear at position $-\rho_{ij}$ in the output plane where ρ_{ij} is the input position of the corresponding aperture. Provided the spacing satisfies $d > \sim 3a$ the interference between adjacent beams will be negligible.

The efficiency of the array creation can be defined as $\epsilon = I_t/I_d$ where I_t is the peak intensity of an output beam and $I_d = P/d^2$ is the input intensity with P the power per $d \times d$ unit cell. We

³Gaussian beam profiles are analyzed in detail in Sec. 6.1.

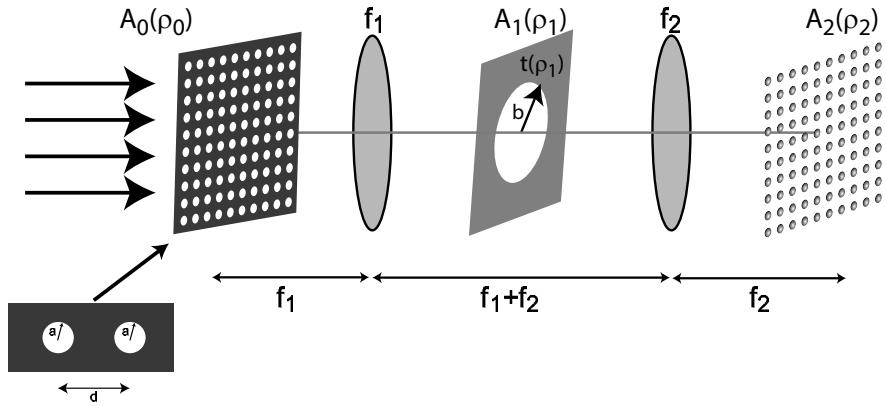


Figure 5.19: Fourier filtering setup for creating an array of Airy-Gauss beams. Each aperture in the input plane has radius a as in Fig. 5.17.

can write

$$I_t = .84 \frac{P \frac{\pi a^2}{d^2}}{\pi a^2} \times 1.978 = 1.66 I_d$$

so $\epsilon = 1.66$, independent of the value of a .

It is also possible to create an array of dark spots with Gaussian profiles, suitable for trapping atoms at a local minimum of the optical intensity. Let the input plane aperture be an array of reflecting spots of radius a and otherwise fully transmitting. The field to the right of the input aperture can be written as

$$E = E_d - r \sum_{ij} E_{ij}$$

where E_d is the amplitude of the plane wave covering the entire aperture, E_{ij} is the field transmitted by aperture ij in Fig. 5.19, and r is the reflectivity of each spot. The plane wave, which is much broader than the field of a single aperture, will be fully transmitted through the Fourier filter. Therefore the field in the output plane will be

$$E_2 = -E_d - r \sum_{ij} E_{2,ij}$$

where $E_{2,ij}$ is the field of Eq. (5.47) centered at position $-\rho_{ij}$ in the output plane. Choosing $r = 1/\sqrt{1.66} = 0.78$ there will be a zero in the field at $-\rho_{ij}$ surrounded by an intensity pattern with a Gaussian profile.

The efficiency is given by

$$\epsilon = \frac{I_t}{I_d} = \frac{I_d}{I_d} = 1.$$

This is lower than for the bright spot array but compares favorably with dark spots created with a Gaussian beam array using diffractive optical elements which has $\epsilon \leq 0.51$ [53] or a line array which has $\epsilon \leq 0.97$.[54] In practice the Fourier filtering approach is expected to have substantially better efficiency than a line array since the diffractive multi-spot gratings used to prepare such an array have efficiencies ~ 0.75 whereas a top hat beam shaper needed for input to the optical system can have near 100% efficiency.

Localization

When used for particle or atom trapping the most important parameters are the depth of the trap, which is proportional to I_t , and the spatial localization. When the trapped objects have motional energy that is small compared to the depth of the trapping potential the degree of localization is governed by the quadratic variation of the intensity near the trap center. For bright traps, which localize the particle near the intensity maximum we can write $U = U_0(1 - \alpha_{\perp}\rho^2 - \alpha_{\parallel}z^2 + \dots)$. Here ρ is the radial coordinate and z is the axial coordinate along the trap axis. It follows from the virial theorem that for a particle with motional temperature T we have

$$\begin{aligned} 2U_0\alpha_{\perp}\langle\rho^2\rangle &= 2k_B T \\ 2U_0\alpha_{\parallel}\langle z^2\rangle &= k_B T \end{aligned}$$

with k_B the Boltzmann constant. The standard deviations of the particle position are therefore

$$\begin{aligned} \sigma_{\rho} &= \sqrt{\langle\rho^2\rangle} = \frac{1}{\alpha_{\perp}^{1/2}} \left(\frac{k_B T}{U_0} \right)^{1/2}, \\ \sigma_z &= \sqrt{\langle z^2\rangle} = \frac{1}{(2\alpha_{\parallel})^{1/2}} \left(\frac{k_B T}{U_0} \right)^{1/2}. \end{aligned} \quad (5.49)$$

For an ideal Gaussian beam with waist parameter w_G we have (see Sec. 6.1) $\alpha_{\perp} = 2/w_G^2$, $\alpha_{\parallel} = \frac{\lambda^2}{\pi^2 w_G^4}$ with λ the optical wavelength so

$$\begin{aligned} \frac{\sigma_{\rho}}{\left(\frac{k_B T}{U_0}\right)^{1/2}} &\equiv \tilde{\sigma}_{\rho} = \frac{w_G}{\sqrt{2}}, \\ \frac{\sigma_z}{\left(\frac{k_B T}{U_0}\right)^{1/2}} &\equiv \tilde{\sigma}_z = \frac{\pi w_G^2}{\sqrt{2}\lambda}. \end{aligned}$$

For the AG beam we have $w_{AG} = 0.974a$ giving position deviations

$$\tilde{\sigma}_{\rho} = \frac{0.97a}{\sqrt{2}}, \quad (5.51a)$$

$$\tilde{\sigma}_z = \frac{\pi 0.95a^2}{\sqrt{2}\lambda}. \quad (5.51b)$$

Equations (5.50,5.51) give the position spreads for bright optical traps. For a dark optical trap created by interfering a Gaussian beam with a plane wave the axial profile far from the origin is different than that of a bright trap due to the variation of the field phase with z given by $\phi(z) = \tan^{-1}[z/(\pi w_G^2/\lambda)]$. This is illustrated in Fig. 5.20. Note that the axial profiles are somewhat different for Airy-Gauss and Gaussian beams. Nevertheless the leading quadratic terms are unchanged so the localization parameters are still given by Eqs. (5.50,5.51). With the parameters in Fig. 5.20 we find $\tilde{\sigma}_{\rho} = 0.69 \mu\text{m}$ and $\tilde{\sigma}_z = 2.6 \mu\text{m}$. With a temperature to trap depth ratio of less than a factor of 9, which is standard for atoms in optical traps, this implies sub micron localization in all dimensions.

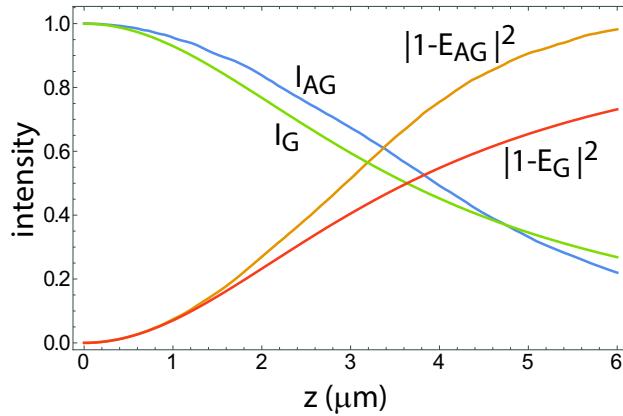


Figure 5.20: On axis intensity as a function of axial coordinate z computed by Fresnel diffraction. The profiles are an Airy-Gauss beam (I_{AG}), a dark Airy-Gauss beam ($|1-E_{AG}|^2$), a Gaussian beam (I_G), and a dark Gaussian beam ($|1-E_G|^2$). Parameters for numerical calculations were $a = b = 1.0 \text{ } \mu\text{m}$, $\lambda = 0.825 \text{ } \mu\text{m}$, $f = 2 \text{ } \mu\text{m}$ for the Airy-Gauss beam and $w_G = 0.974a$ for the Gaussian beam.

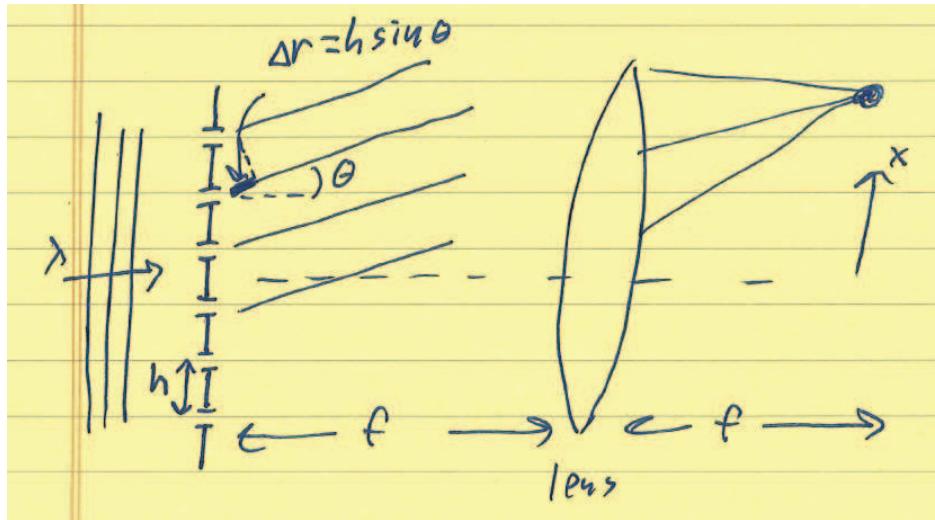


Figure 5.21: Interference of N plane waves emitted from apertures separated by h .

5.4 Diffraction gratings

Interference of optical waves leads to periodic intensity patterns, with the period a function of the optical wavelength. As the number of interfering waves is increased the resulting pattern acquires sharper features that can be used to accurately determine the wavelength of the light. This is the basic idea underlying the diffraction grating.

Consider the interference of plane waves emitted from two apertures separated by h and observed on a screen a distance L away. The intensity pattern is periodic with period $\Lambda = \frac{\lambda}{2\sin(\theta/2)}$ where $\tan(\theta/2) = h/L$. For small angles $\theta \simeq h/L$ and $\Lambda \simeq \lambda/\theta \simeq \lambda L/h$.

Let's now generalize to the case of N interfering plane waves. Consider the geometry of Fig.

5.21 which shows N source apertures with separation h in the front focal plane of a lens with focal length f . Rays propagating at angle θ from the optical axis will be imaged to a point $x = f \tan(\theta)$ in the back focal plane of the lens. If the path length from aperture j to the observation point is r_j the optical field due to the N beams can be written as

$$A \sim A_0 \sum_{j=1}^N e^{ikr_j} = A_0 \sum_{j=1}^N e^{ikr_0} e^{ikh \sin(\theta) j} = A_0 e^{ikr_0} \sum_{j=1}^N e^{ikh \sin(\theta) j}$$

where A_0 is the amplitude of each beam and r_0 is the common path length independent of j . Using the geometric sum $\sum_{j=1}^N e^{iaj} = e^{ia} \frac{e^{iaN} - 1}{e^{ia} - 1}$ we find for the intensity

$$I \sim |A|^2 = |A_0|^2 \frac{\sin^2 [kh \sin(\theta) N/2]}{\sin^2 [kh \sin(\theta)/2]}.$$

This is the same type of expression as that for the Fourier transform of a repeated object in Eq. (5.41). We see that the peak intensity at $\theta = 0$ is $N^2 |A_0|^2$. The angular width to the first zero is found from $kh \sin(\theta) N/2 = \pi$ giving

$$\theta_{\text{width}} = \sin^{-1} \left(\frac{2\pi}{Nkh} \right) = \sin^{-1} \left(\frac{\lambda}{Nh} \right).$$

There are secondary maxima when the denominator vanishes at angles

$$\theta_m = \sin^{-1} \left(\frac{2\pi m}{kh} \right) = \sin^{-1} \left(\frac{m\lambda}{h} \right)$$

with m an integer. These are the Bragg angles for diffraction at order m . The angular position of the diffraction order is sensitive to wavelength according to

$$\frac{d\theta_m}{d\lambda} = \frac{m/h}{\left(1 - \frac{m^2\lambda^2}{h^2}\right)^{1/2}}.$$

The uncertainty in θ can be related to the uncertainty in λ according to

$$\theta_m + \theta_{\text{width}} = \sin^{-1} \left(\frac{m(\lambda + \lambda_{\text{width}})}{h} \right) \simeq \sin^{-1} \left(\frac{m\lambda}{h} \right) + \frac{\frac{m\lambda_{\text{width}}}{h}}{\left(1 - \frac{m^2\lambda^2}{h^2}\right)^{1/2}}.$$

It follows that

$$\theta_{\text{width}} \simeq \frac{m/h}{\left(1 - \frac{m^2\lambda^2}{h^2}\right)^{1/2}} \lambda_{\text{width}}.$$

More details to be added....

Chapter 6

Gaussian beams

6.1 Lowest order Gaussian mode

The amplitude of a Gaussian laser beam in a plane where the wavefront is flat can be written as

$$A = A_0 e^{-\frac{r^2}{w^2}}, \quad (6.1)$$

where w is the width. The diameter of the beam between points where the intensity has fallen to $1/e^2$ of the maximum is the “Gaussian” diameter and is given by $d = 2w$. It is sometimes useful to work with the full width at half-maximum (FWHM) of the intensity. This is given by $d_{\text{FWHM}} = \sqrt{2 \ln 2} w \simeq 1.177w$.

The power in the beam (6.1) is

$$P = \left(\frac{\epsilon_0}{2} nc\right) 2\pi \int_0^\infty dr r |A|^2 = \left(\frac{\epsilon_0}{2} nc\right) \frac{\pi w^2}{2} |A_0|^2. \quad (6.2)$$

A useful expression for the spatial variation of the intensity is $I(r) = (2P/(\pi w^2))e^{-2r^2/w^2}$.

As we saw in the previous section propagation from the front to the back focal plane of a lens results in Fourier transformation of the spatial structure of the field. In general a Fourier transformed field will have a different structure than the original field. For example a fringe pattern $A_1 \sim \cos(qx_1)$ transforms to $A_2 \sim \delta(x_2 - \lambda f q) + \delta(x_2 + \lambda f q)$ and a “top-hat” beam $A_1(x_1) \sim \text{rect}(x_1 - a)$ transforms to a sinc function $A_2 \sim \sin(x_2 a / \lambda f) (x_2 a / \lambda f)$.

A natural question arises as to whether or not there exist fields that retain the same spatial structure apart from an overall scale change under Fourier transformation, or under the more general transformation given by Eqs. (5.22,5.23). Indeed there are families of functions that are Fourier self-similar. For a comprehensive discussion of this topic see the book on Fourier integrals by Titchmarsh[55]. Perhaps the simplest example of a Fourier self-similar field is the Gaussian, $A_1 \sim e^{-x_1^2/w^2}$ which transforms into $A_2 \sim e^{-x_2^2 w^2/\lambda^2 f^2}$. The Gaussian function together with higher order generalizations in terms of Hermite-Gauss functions play a very important role in optics as they are the natural modes in terms of which we can decompose localized beams. The lowest order field emitted by a well-aligned laser is also a Gaussian.

It is therefore natural to look for Gaussian type solutions of Eq. (5.2). Let’s find the lowest order solution, the fundamental Gaussian beam. To do so we guess a solution of the form

$$A = A_0 e^{-i(P(z) - \frac{k}{2q(z)} r^2)}, \quad (6.3)$$

assume cylindrical symmetry so that $\nabla_{\perp}^2 = \partial_{rr} + (1/r)\partial_r$ and plug in to Eq. (5.2) to get

$$\left(\frac{k}{q(z)}\right)^2 \left(1 - \frac{dq}{dz}\right) r^2 - i2k \left(\frac{1}{q(z)} - i\frac{dP}{dz}\right) = 0. \quad (6.4)$$

This equation must hold for all r so we get

$$\frac{dq}{dz} = 1 \quad \text{and} \quad \frac{dP}{dz} = -\frac{i}{q}. \quad (6.5)$$

The solution for q is

$$q(z) = q(0) + z \quad (6.6)$$

from which we find

$$P(z) = -i \ln \left(1 + \frac{z}{q(0)}\right). \quad (6.7)$$

We have put the integration constant to zero since it merely gives an overall phase shift.

At this point it is convenient to introduce two real beam parameters: the radius of curvature R and the radius of the width w . Put

$$\frac{1}{q(z)} = \frac{1}{R(z)} + i\frac{\lambda}{\pi nw^2(z)}. \quad (6.8)$$

Now define the plane $z = 0$ to be where q is purely imaginary, this will lead to localized solutions with an energy that decays exponentially as $r \rightarrow \infty$. Then

$$\frac{1}{q(0)} = i\frac{\lambda}{\pi nw_0^2} \quad (6.9)$$

where $w_0 = w(0)$. After some algebra we find that (6.3) can be written as

$$A = A_0 \frac{w_0}{w(z)} e^{-i\eta(z)} e^{i\frac{k}{2R(z)}r^2} e^{-\frac{r^2}{w^2(z)}}, \quad (6.10)$$

where

$$w^2(z) = w_0^2 \left(1 + \frac{z^2}{L_R^2}\right), \quad (6.11a)$$

$$R(z) = z \left(1 + \frac{L_R^2}{z^2}\right), \quad (6.11b)$$

$$\eta(z) = \tan^{-1} \left(\frac{z}{L_R}\right), \quad (6.11c)$$

$$L_R = \frac{\pi nw_0^2}{\lambda} = \frac{kw_0^2}{2}. \quad (6.11d)$$

A fundamental beam profile is shown in Fig. 6.1. The parameters are the beam waist w_0 , which is the minimum value of the width $w(z)$, the radius of curvature $R(z)$, the Gouy phase $\eta(z)$ and the Rayleigh length L_R . The Rayleigh length is the characteristic distance over which a Gaussian type beam remains collimated. When $z = L_R$, w increases from w_0 to $\sqrt{2}w_0$. As an example for $\lambda = 0.63 \mu\text{m}$ (HeNe laser) and $w_0 = 1 \text{ mm}$ we get $L_R = 5.0 \text{ m}$. Some authors also define a confocal length $b = 2L_R$.

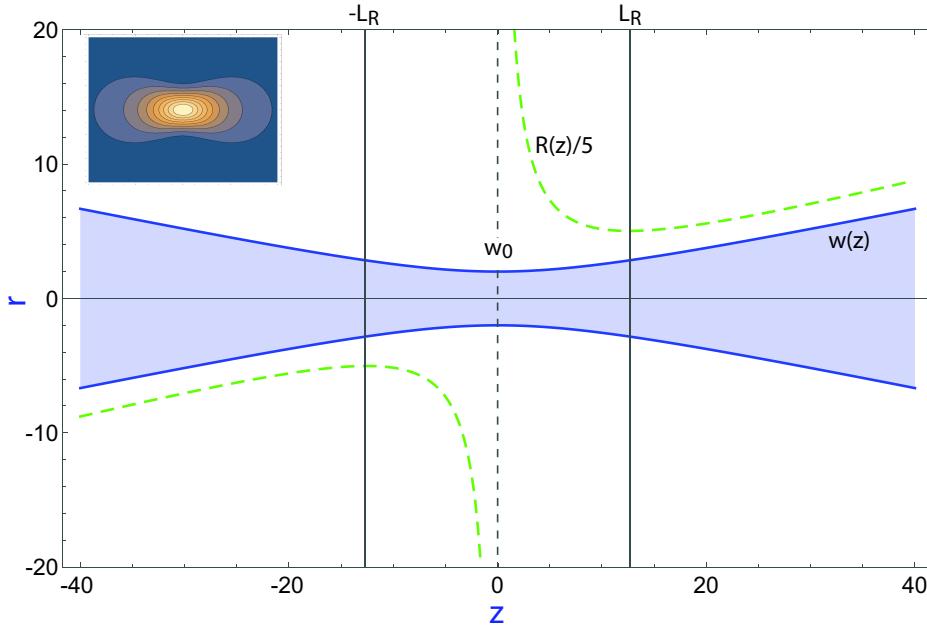


Figure 6.1: Gaussian beam of wavelength $\lambda = 1$, and waist $w_0 = 2\lambda$. The inset shows intensity contours.

The Gouy phase adds a phase shift of π when a Gaussian beam goes through a focus compared to the phase of a plane wave[56, 57]. The physical origin of the Gouy phase is related to the transverse structure of the field[58]. Higher order beams have larger Gouy phase shifts.

We motivated the Gaussian beam solution by seeking a field that is an eigensolution of paraxial propagation. If we start with a Gaussian beam and send it through any paraxial optical system characterized by an ABCD matrix then the beam coming out will also be Gaussian. To calculate the output beam we can use Eq. (5.22). Given a Gaussian beam characterized by $q_1 = q(z_1)$ at plane $z = z_1$ it will be transformed into a new Gaussian beam with $q_2 = q(z_2)$ at plane $z = z_2$. Evaluation of Eq. (5.22) shows that

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D} \quad (6.12)$$

where $ABCD$ are the elements of the ray matrix describing the propagation from planes 1 to 2. Note that the transformation law is particularly simple for propagation in a homogeneous medium of length L : $q_2 = q_1 + L$.

This perhaps curious looking result can be thought of as a generalization of the behavior of rays to beams that diffract. Suppose we let $w \rightarrow \infty$ then Eq. (6.8) becomes $q = R$, where R is the radius of curvature. A ray with parameters (x, θ) crossed the axis at $z = -x/\theta$. Therefore the ray can be thought of as being part of a beam with radius of curvature $R = x/\theta$. We can then ask how the radius of curvature changes under propagation. We have

$$R = \frac{x}{\theta} \rightarrow R' = \frac{x'}{\theta'} = \frac{Ax + B\theta}{Cx + D\theta} = \frac{Ax/\theta + B}{Cx/\theta + D} = \frac{AR + B}{CR + D}.$$

If we label the input and output radii of curvature as R_1 and R_2 we have

$$R_2 = \frac{AR_1 + B}{CR_1 + D}$$

which is the same transformation as Eq. (6.12) for the q parameter. Thus we have the remarkable result that the complex q parameter transforms in exactly the same way as the real radius of curvature of a ray.

As a first example using Eq. (6.12) let's calculate the width of a Gaussian beam that has an initial waist w_0 in a medium with index n_1 . Initially the q parameter is $q_1 = -i\pi n_1 w_0^2/\lambda = -iL_R$. The new q parameter is

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D} = \frac{-iL_R A + B}{-iL_R C + D} = \frac{B^2 + A^2 L_R^2}{BD + ACL_R^2 + i(AD - BC)L_R} \quad (6.13)$$

so

$$\text{Im} \left[\frac{1}{q_2} \right] = \frac{(AD - BC)L_R}{B^2 + A^2 L_R^2} = \frac{\lambda}{\pi n_2 w_2^2}.$$

The output width is therefore

$$w_2^2(z) = \frac{n_1}{n_2} w_0^2 \frac{A^2 + \frac{B^2}{L_R^2}}{AD - BC}.$$

The determinant is $AD - BC = n_1/n_2$ so that

$$w_2(z) = w_0 \sqrt{A^2 + \frac{B^2}{L_R^2}}, \quad (6.14)$$

which reduces to Eq. (6.11a) for a straight segment of length z . Note that $w_2(z)$ is in general not a beam *waist* but is the *width* of the beam at plane z_2 . We will usually denote the beam waist by an extra subscript of 0 as in w_{20} . In the particular case where the waist occurs at plane z_2 the width is equal to the waist. To find the position of the new waist we require $R_2 = \infty$ or $\text{Re} \left[\frac{1}{q_2} \right] = 0$ which gives the condition

$$BD + ACL_R^2 = 0. \quad (6.15)$$

As we will discuss below this condition, which can be thought of as the requirement for imaging of a beam waist to a beam waist, does not in general coincide with the condition for imaging in a ray optics approximation which is simply $B = 0$. The question arises as to whether or not the geometrical imaging condition $B = 0$ can also coincide with imaging of the beam waist. If $B = 0$ the condition for imaging the waist is $AC = 0$. Since the ray matrices satisfy $AD - BC = n_1/n_2$, if $B = 0$ we must have $AD = n_1/n_2$ which implies $A \neq 0$. The requirement for simultaneous geometric imaging together with reimaging of the beam waist is therefore $B = 0$ and $C = 0$. For propagation through a single lens $C = -1/f$ so this is not possible, but it may be satisfied in more complicated multi-lens situations as discussed in Sec. 6.3.4. Of particular interest are zoom lenses which provide geometrical and beam waist imaging between fixed planes while allowing the magnification to be varied.

6.1.1 Higher order Hermite-Gaussian beams

The circularly symmetric solution (6.3) is only the lowest order mode among an infinite set of successively higher order beams. To find the higher order solutions we start with the separable ansatz

$$A_{mn}(x, y) = A_0 \frac{w_0}{w(z)} e^{-i\eta(z)} e^{i\frac{k}{2R(z)}r^2} e^{-\frac{r^2}{w^2(z)}} f_m(x, z) f_n(y, z), \quad (6.16)$$

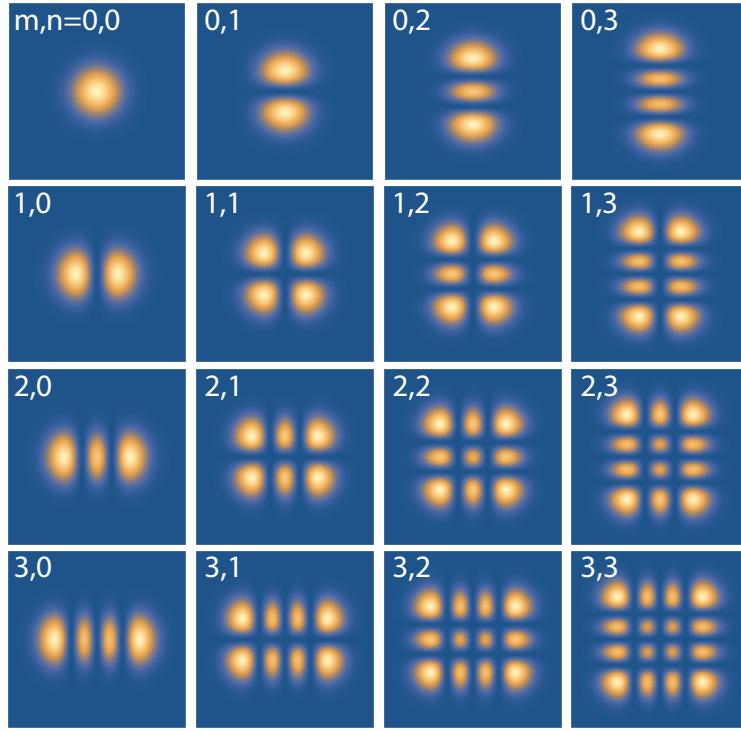


Figure 6.2: Intensity of Hermite-Gaussian beams in the focal plane $z = 0$ with $\lambda = 0.5$, $w_0 = 1$ in regions of size 6×6 .

with f_m, f_n unknown functions to be determined. This form of solution is substituted into the paraxial wave equation (5.2), we separate variables, and note that the solution of

$$\frac{d^2u}{dx^2} - 2x\frac{du}{dx} + 2nu = 0$$

is the n^{th} order Hermite polynomial $u = H_n(x)$, to find

$$A_{mn}(x, y) = A_0 \frac{w_0}{w(z)} H_m \left(\sqrt{2} \frac{x}{w(z)} \right) H_n \left(\sqrt{2} \frac{y}{w(z)} \right) e^{-i\eta_{m,n}(z)} e^{i\frac{k}{2R(z)}r^2} e^{-\frac{x^2+y^2}{w^2(z)}}, \quad (6.17)$$

where $\eta_{m,n}(z) = (1 + m + n)\eta(z)$ and η, R, w are given by Eqs. (6.11). The normalization integral for Hermite polynomials is

$$\int_{-\infty}^{\infty} d\xi e^{-\xi^2} H_m(\xi) H_n(\xi) = \sqrt{\pi} 2^m m! \delta_{mn}.$$

We can therefore write the normalized transverse mode which has integral of the field squared set to unity as

$$A_{mn}(x, y, z) = \left[\frac{2}{\pi 2^{m+n} m! n!} \right]^{1/2} \frac{1}{w(z)} H_m \left(\sqrt{2} \frac{x}{w(z)} \right) H_n \left(\sqrt{2} \frac{y}{w(z)} \right) e^{-i\eta_{m,n}(z)} e^{i\frac{k}{2R(z)}r^2} e^{-\frac{x^2+y^2}{w^2(z)}}. \quad (6.18)$$

These are called Hermite-Gaussian beams. The intensity of the first few beams is shown in Fig. 6.2.

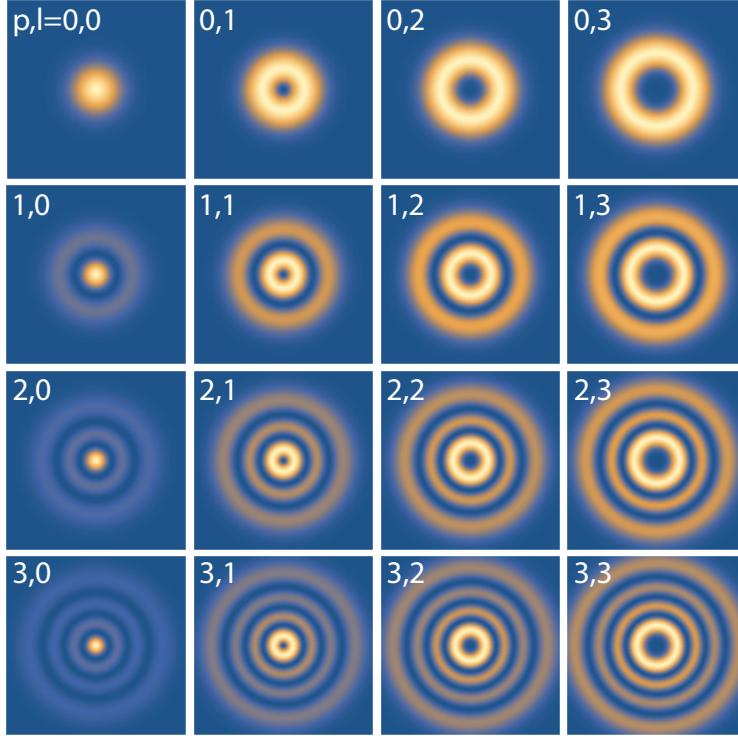


Figure 6.3: Intensity of Laguerre-Gaussian beams in the focal plane $z = 0$ with $\lambda = 0.5$, $w_0 = 1$ in regions of size 6×6 .

6.1.2 Laguerre-Gaussian beams

Beyond the Hermite-Gaussian beams there exist several families of Gaussian beam solutions of the paraxial wave equation associated with orthonormal polynomials. Cartesian coordinates were used to describe the Hermite-Gaussian beams in the previous section. Circular cylindrical coordinates $\rho = \sqrt{x^2 + y^2}$ and z are used for Laguerre-Gaussian beams with radial symmetry.

The Laguerre-Gaussian beams are of particular interest in connection with orbital angular momentum. They are given by

$$A_{pl}(\rho, \phi, z) = \left[\frac{2p!}{\pi(p+|l|)!} \right]^{1/2} \frac{1}{w(z)} \left[\frac{\sqrt{2}\rho}{w(z)} \right]^{|l|} L_p^{|l|} \left(\frac{2\rho^2}{w^2(z)} \right) e^{-i\eta_{2p,|l|}(z)} e^{i\frac{k}{2R(z)}r^2} e^{-\frac{\rho^2}{w^2(z)}} e^{-il\phi}.$$

Here p, l are integers and $p \geq 0$. The normalization has been chosen so that $2\pi \int_0^\infty d\rho \rho |A_{pl}|^2 = 1$. The Gouy phase factor is $\eta_{2p,|l|}(z) = (2p + |l| + 1)\eta(z)$.

The intensity of a Laguerre-Gaussian beam is proportional to

$$|A_{pl}|^2 = \frac{2p!}{\pi(p+|l|)!} \frac{1}{w^2(z)} \left[\frac{\sqrt{2}\rho}{w(z)} \right]^{2|l|} \left[L_p^{|l|} \left(\frac{2\rho^2}{w^2(z)} \right) \right]^2 e^{-\frac{2\rho^2}{w^2(z)}}.$$

The intensity of the first few Laguerre-Gaussian beams is shown in Fig. 6.3. When $p = 0$, the Laguerre polynomial simplifies to $L_0^l = 1$ and the field amplitude near the origin is $A_{0l} \sim \rho^l$. Any Laguerre-Gaussian mode can be written in terms of Hermite-Gaussian functions and vice versa[59]. Beams p, l and $p, -l$ have the same intensity distribution but the opposite sign of azimuthal phase.

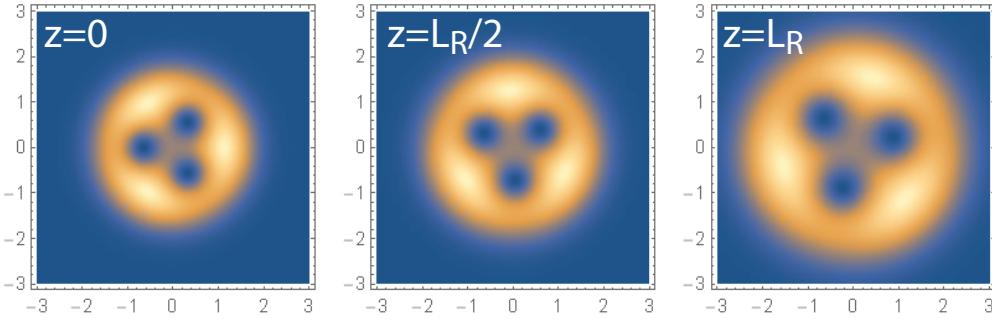


Figure 6.4: Intensity of $LG_{0,3} + 0.75LG_{0,0}$ for (left to right) $z = 0, L_R/2, L_R$ with $\lambda = 0.5$, $w_0 = 1$ in regions of size 6×6 .

A photon in spatial mode A_{pl} carries $\hbar l$ of orbital angular momentum. This is in addition to the spin angular momentum that depends on the polarization state.

Superpositions of Laguerre-Gaussian beams exhibit curious intensity patterns that rotate as they propagate. Figure 6.4 shows an example of the rotating intensity pattern from the field $A = LG_{0,3} + 0.75LG_{0,0}$.

6.1.3 Ince-Gaussian beams

In addition to the Hermite-Gaussian and Laguerre-Gaussian families there are additional complete sets of eigenmodes. The Ince-Gaussian beams which are defined in terms of Ince polynomials are solutions of the paraxial wave equation in elliptical coordinates and are intermediate between Hermite-Gaussian and Laguerre-Gaussian modes[60].

6.2 Nondiffracting beams

The Helmholtz equation in three dimensions is separable in eleven different orthogonal coordinate systems. Of these there are four coordinate systems that separate into a longitudinal and transverse part. These are Cartesian coordinates, circular cylindrical, elliptic cylindrical, and parabolic cylindrical. Separable solutions can be written in the form

$$U(\boldsymbol{\rho}, z) = A(\boldsymbol{\rho})Z(z)$$

where $\boldsymbol{\rho}$ represents the transverse coordinates. Such solutions are nondiffracting since $A(\boldsymbol{\rho})$ does not depend on the propagation coordinate z . In Cartesian coordinates $\boldsymbol{\rho} = x\hat{x} + y\hat{y}$ and the solutions are plane waves $U = A_0 e^{i(k_x x + k_y y)} e^{ik_z z}$ with amplitude A_0 . These are trivially nondiffracting waves but do not provide a localized beam profile.

A more interesting case is the Bessel function solution in circular cylindrical coordinates \mathbf{r}, \mathbf{z} where $\mathbf{r} = \rho\hat{\rho} + \phi\hat{\phi}$. The Cartesian coordinates are related by $x = \rho \cos(\phi), y = \rho \sin(\phi), z = \mathbf{z} \cdot \hat{z}$ and separable solutions take the form

$$U = A_0 J_n(k_\rho \rho) e^{ik_z z} e^{\pm i n \phi}.$$

Here J_n is the Bessel function of integer order n and $k_\rho^2 + k_z^2 = k^2 = \omega^2/c^2$. These are nondiffracting beams with a peak intensity at the origin for the J_0 solution and near the origin for $n > 0$ as shown

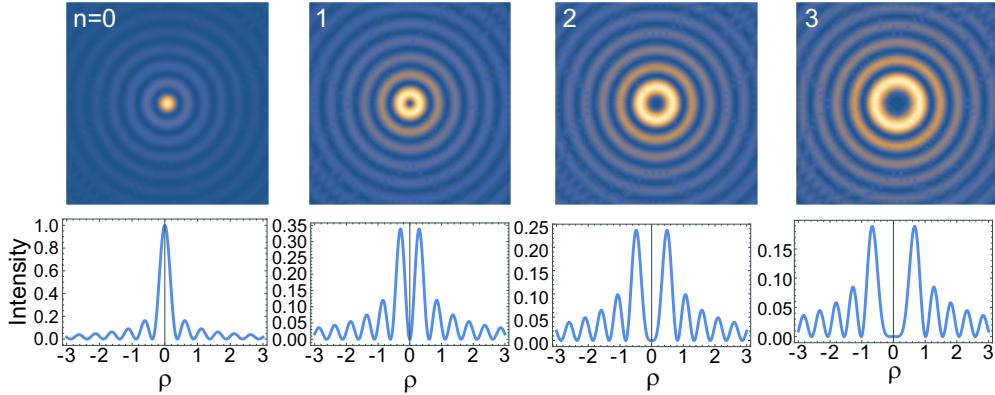


Figure 6.5: Bessel beam intensity profiles for $n = 0, 1, 2, 3$ and $k_\rho = 2\pi$.

in Fig. 6.5. Strictly speaking these are not physically realizable beams since they carry infinite energy, nevertheless apertured versions propagate long distances with minimal spreading[61].

Nondiffracting solutions in parabolic and elliptic cylindrical coordinates in the form of Weber beams[62] and Mathieu beams[63] respectively are also known.

6.3 Gaussian beam transformation by a lens

Consider transformation of a Gaussian beam by a thin lens as shown in Fig. 6.6. For the geometry of Fig. 6.6 the transformation matrix is

$$\mathbf{M} = \begin{pmatrix} 1 - \frac{z_2}{f} & z_1 + z_2 - \frac{z_1 z_2}{f} \\ -\frac{1}{f} & 1 - \frac{z_1}{f} \end{pmatrix}$$

and Eqs. (6.8,6.12) lead to

$$\begin{aligned} R_2 &= \frac{(L_1^2 + z_1^2)z_2^2 - 2fz_2L_1^2 - 2fz_1z_1(z_1 + z_2) + f^2L_1^2 + f^2(z_1 + z_2)^2}{(L_1^2 + z_1^2)z_2 + (z_1 + z_2)f^2 - fL_1^2 - fz_1(z_1 + 2z_2)}, \\ w_2^2 &= \frac{\lambda}{\pi L_1 f^2} [(L_1^2 + z_1^2)z_2^2 - 2fz_2(L_1^2 + z_1^2 + z_1z_2) + f^2(L_1^2 + (z_1 + z_2)^2)]. \end{aligned}$$

Here we have introduced the Rayleigh length of the input beam $L_1 = \pi w_{10}^2/\lambda$. The output waist occurs at a value of z_2 such that $R_2(z_2) = \infty$. To find z_2 we note that the input waist has q parameter $q_1 = -iL_1$ and at the output waist q_2 is pure imaginary. This gives the condition $BD + ACL_1^2 = 0$ where A, B, C, D are the elements of the ray matrix connecting planes 1 and 2. Solving for this condition gives

$$z_2 = f \frac{z_1(z_1 - f) + L_1^2}{(z_1 - f)^2 + L_1^2}$$

and

$$w_{20}^2 = w_{10}^2 \frac{f^2}{(z_1 - f)^2 + L_1^2}.$$

The simplest situation occurs when $z_1 = f$, i.e. the incident waist is in the front focal plane of the lens. In this case we find the output waist is at $z_2 = f$ and $w_{20}^2 = \lambda f^2/\pi L_1$. When we have a

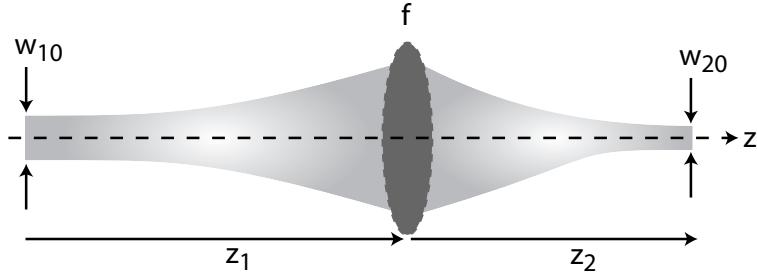


Figure 6.6: Gaussian beam transformation by a lens of focal length f . The incident waist w_{10} is positioned a distance z_1 in front of the lens. The output waist w_{20} occurs a distance z_2 after the lens.

positive lens with $f > 0$ this can be written as

$$w_{10}w_{20} = \frac{\lambda f}{\pi}. \quad (6.19)$$

This very useful result is exact for paraxial Gaussian beams. We will refer to this condition as a confocal arrangement since the beam waists lie in the front and back focal planes of the lens. Introducing the “confocal” waist $w_c = \sqrt{\lambda f / \pi}$ which is invariant under the lens transformation and the dimensionless waist parameters $\omega_1 = w_{10}/w_c, \omega_2 = w_{20}/w_c$ Eq. (6.19) takes on the simplified form

$$\omega_1\omega_2 = 1. \quad (6.20)$$

When $z_1 \neq f$ it is useful to describe the lens transformation in terms of dimensionless lengths $\zeta_1 = z_1/f, \zeta_2 = z_2/f$. In terms of these new variables we find

$$\zeta_2 = \frac{\zeta_1(\zeta_1 - 1) + \omega_1^4}{(\zeta_1 - 1)^2 + \omega_1^4} \quad (6.21)$$

and

$$\omega_1^2\omega_2^2 = \frac{1}{1 + \frac{(\zeta_1 - 1)^2}{\omega_1^4}}. \quad (6.22)$$

Equation (6.22) reduces to (6.20) when $z_1 = f$ so $\zeta_1 = 1$. Figure 6.7 shows representative curves of ζ_2 and ω_2 for different input beam waists.

Some useful limits can be extracted from Eqs. (6.21,6.22). The maximum and minimum values of ζ_2 occur when $\zeta_1 = 1 \pm \omega_1^2$ which give $\zeta_2 = 1 \pm \frac{1}{2\omega_1^2}$. Thus for small ω_1 the distance to the waist approaches $\frac{1}{2\omega_1^2}$ or $f \times \frac{w_c^2}{2w_{10}^2}$. The corresponding output waist values satisfy $\omega_1^2\omega_2^2 = 1/2$. This can also be expressed as $\omega_2^2 = \frac{1}{2|\zeta_1 - 1|}$. The maximum possible value of ω_2 occurs when $\zeta_1 = 1$ which gives $\omega_1\omega_{2,\max} = 1$. We see that for a given input beam waist the largest possible output waist occurs when the input waist is placed in the front focal plane.

Note that when $\zeta_1 > 1$ we always get $\zeta_2 > 1$. Stated in words: when the input waist is more than a focal length away from the lens, the output waist is also more than a focal length away. The slope of the ζ_2/ζ_1 mapping increases inversely with ω_1 . Thus when the input waist is small compared to the confocal waist w_c , i.e. $\omega_1 \ll 1$ the slope is large and a small shift of the input waist position causes a large shift of the output waist position. Conversely when $w_{10} \gg w_c$, i.e. $\omega_1 \gg 1$ the output waist position is pinned at $\zeta_2 = 1$, or $z_2 = f$. Qualitatively this regime occurs

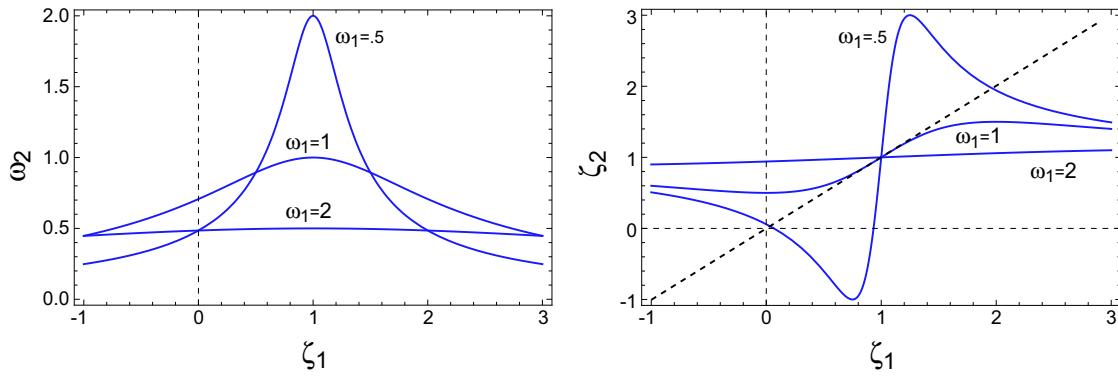


Figure 6.7: Output waist size (left) and output waist position (right) for transformation by a single lens.

when the input waist is large enough that $L_1 \gg f$ so the output waist appears at $z_2 = f$ after the lens.

When $\zeta_1 < 1$ we get $\zeta_2 < 1$. In other words when the input waist is closer than a focal length from the lens the output waist is also closer than a focal length, provided ω_1 is not too small. When $\omega_1 \ll 1$ ζ_2 can be negative which means the waist is in front of the lens, and no real waist occurs after the lens.

The somewhat counterintuitive behavior of the waist transformation can be contrasted with the requirement for imaging in a geometrical optics approximation $B = 0$, which gives the lens makers formula for a single lens $1/f = 1/z_1 + 1/z_2$. In geometrical imaging bringing the object closer to the lens pushes the image further away. The explanation of the lack of “waist imaging” when there is geometrical imaging is due to phase errors. The geometrical imaging condition implies that all rays leaving a single point in the object plane also meet at a single point in the image plane. However this condition says nothing about the path lengths of the rays or about the relative path lengths for different pairs of object and image plane points. Thus the geometrical imaging condition does not preclude quadratic phase errors which lead to non imaging of a Gaussian beam waist.

In many applications we wish to focus a Gaussian beam to a waist that is small compared to w_c . Thus $\omega_1 \gg 1$ and $\omega_2 \ll 1$. In this situation the relative motion of the output waist as the input waist position changes is most usefully quantified by normalizing the waist positions z_1, z_2 to the input and output Rayleigh lengths L_1, L_2 . We thus introduce normalized deviations from the focal planes as

$$\begin{aligned}\zeta'_1 &= \frac{z_1 - f}{L_1} \\ \zeta'_2 &= \frac{z_2 - f}{L_2}.\end{aligned}$$

When $\zeta'_1 \ll 1$ it is not hard to show that $\zeta'_2 \simeq \zeta'_1$. In other words the relative position error normalized to the Rayleigh length is a constant under transformation between front and back focal planes.

6.3.1 Gaussian beam focusing

Another useful formula is the numerical aperture needed to focus a Gaussian beam to a desired spot size. Assume a waist of w_0 which expands to $w(z) = w_0 \sqrt{1 + z^2/L_R^2}$. An aperture with a

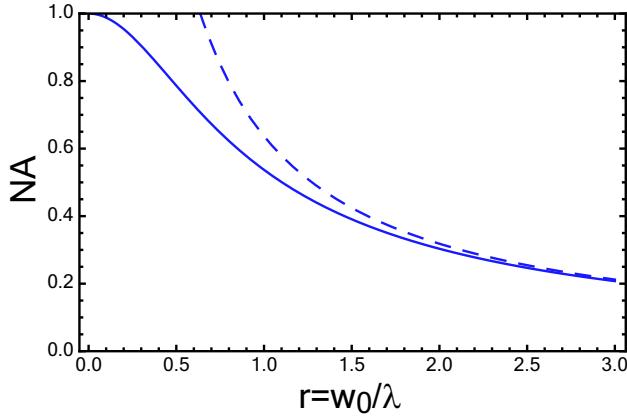


Figure 6.8: Numerical aperture for a given focusing ratio at aperture parameter $p = 4$. The dashed line shows a large r approximation.

diameter that is p times larger than $w(z)$ fills a cone of half angle θ . The numerical aperture of the aperture is

$$\begin{aligned} NA &= \sin \theta \\ &= \frac{\frac{pw_0}{2} \sqrt{1 + z^2/L_R^2}}{\sqrt{z^2 + \frac{p^2 w_0^2}{4} (1 + z^2/L_R^2)}} \\ &= \frac{pw_0}{2} \frac{\left(\frac{1}{z^2} + \frac{1}{L_R^2}\right)^{1/2}}{\sqrt{1 + \frac{p^2 w_0^2}{4} \left(\frac{1}{z^2} + \frac{1}{L_R^2}\right)}}. \end{aligned}$$

At large distances $z \gg L_R$ this tends to

$$NA \simeq \frac{pw_0}{2L_R} \frac{1}{\sqrt{1 + \frac{p^2 w_0^2}{4z_R^2}}}.$$

The transmission of a Gaussian beam through an aperture that is p times larger in diameter than the beam waist is

$$\begin{aligned} T &= \frac{2}{\pi w^2} \int_0^{pw/2} dr 2\pi r e^{-2r^2/w^2} \\ &= 1 - e^{-p^2/2}. \end{aligned}$$

An aperture diameter of four times the waist is a good rule of thumb since $p = 4$ gives $T = 0.9997$. Using $p = 4$ we find

$$NA \simeq \frac{2w_0}{L_R} \frac{1}{\sqrt{1 + \frac{4w_0^2}{L_R^2}}} = \frac{1}{\sqrt{1 + \frac{\pi^2 w_0^2}{4\lambda^2}}}. \quad (6.23)$$

Figure 6.8 shows the required numerical aperture from Eq. (6.23) as a function of the ratio $r = w_0/\lambda$. In the small aperture limit of $r \gg 1$ we find

$$NA \simeq \frac{2}{\pi r} = \frac{2\lambda}{\pi w_0}$$

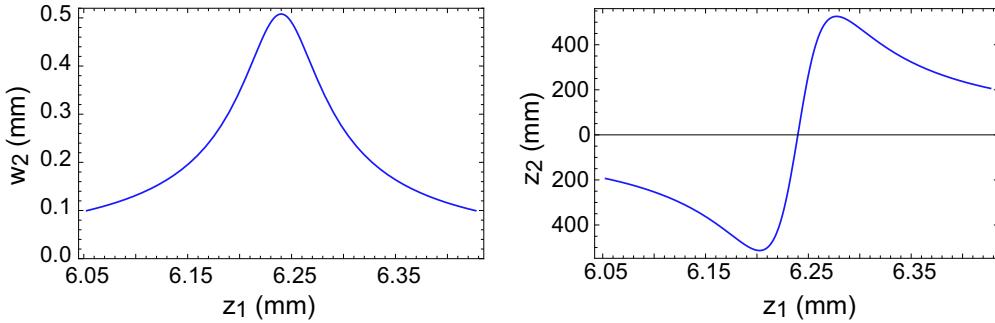


Figure 6.9: Output waist (left) and waist position (right) for collimation of the light emitted by a single mode fiber.

which is shown as a dashed line in the figure.

6.3.2 Fiber collimation

A specific case of the waist transformations occurs in the common optical task of collimating a beam from a single mode fiber. The fiber output will be taken to be a beam of waist w_1 . The beam propagates a distance z_1 , and is then focused by a lens of focal length f which forms a new waist z_2 after the lens, which is exactly the geometry of Fig. 6.6.

Rewriting Eqs. (6.21,6.22) the new waist occurs at

$$z_2 = f \frac{(z_1/f)(z_1/f - 1) + (w_1/w_c)^4}{(z_1/f - 1)^2 + (w_1/w_c)^4}$$

and has the value

$$w_2^2 = w_1^2 \frac{w_c^4}{w_1^4 + w_c^4(z_1/f - 1)^2}.$$

Here $w_c = \sqrt{\lambda f / \pi}$ is the confocal waist. Let us take $\lambda = .78 \mu\text{m}$, $w_1 = 3.05 \mu\text{m}$, and $f = 6.24 \text{ mm}$ which give $w_c = 39.4 \mu\text{m}$. The output waist, and waist position for values of z_1 close to f are shown in Fig. 6.9.

A commonly made mistake is to place the lens further than f away from the fiber end face so that the waist is smaller than the maximum. This occurs if we try to collimate the fiber output by minimizing the spot size on a distant screen a distance z_s from the lens. We see from the figure that for the parameters used $z_2 < 0.5 \text{ m}$. If z_s is larger than this it is not possible to adjust the lens so the waist is on the screen. In this situation the beam radius on the screen is given by

$$w_s(z_1) = w_2(z_1) \sqrt{1 + (z_s - z_2(z_1))^2 / L_2^2}$$

where $L_2 = \pi w_2(z_1)^2 / \lambda$.

Figure 6.10 shows the shift of the apparent best collimation position z_1 as a function of z_s . Comparison with Fig. 6.9 shows that to produce the maximum waist with an error less than 5% it is necessary to have the screen at least 5 m away. Such a large distance is inconvenient in most laboratory settings.

It might appear that a possible procedure is as follows. We minimize the waist on a screen that is say 1 m away. Then monitor the beam width a distance f after the lens with a beam scanner

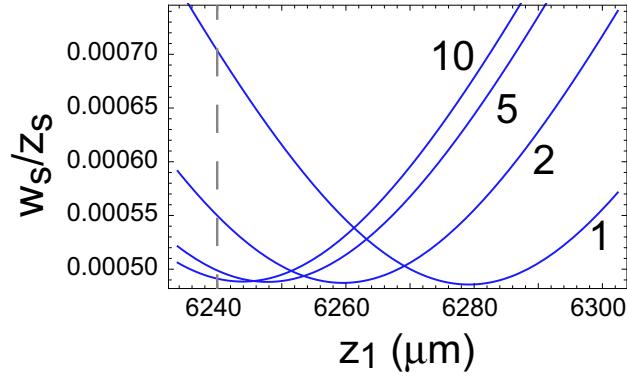


Figure 6.10: Width on a screen at distance z_s vs. lens position z_1 . The curves are labeled with $z_s = 1, 2, 5, 10$ m.

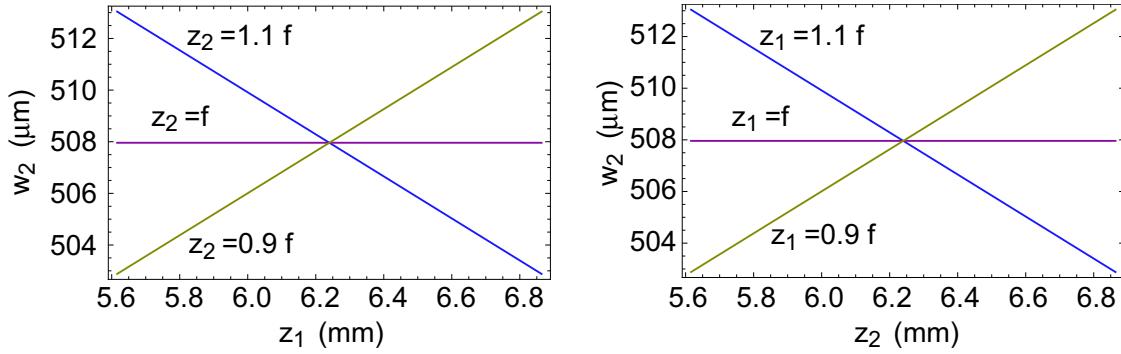


Figure 6.11: Beam width as z_1 is varied (left) and as z_2 is varied (right). The flat line in the left hand graph is an exact result for all z_1 . The flat line in the right hand graph is an approximate result for z_2 similar to f .

and move the fiber a small distance towards the lens to maximize the measured width. However, the beam width a distance $z_2 = f$ after the lens is independent of z_1 ! It can easily be shown that the width at $z_2 = f$ is simply $w_2 = \lambda f / \pi w_1$ which is a constant. Thus moving the fiber will not change the measured beam width.

The behavior of the beam width for different values of z_1, z_2 is shown in Fig. 6.11. The correct position z_1 can be set by measuring the beam width as z_2 is changed, and adjusting z_1 so that the beam width stays almost constant, despite changing z_2 . Setting $z_1 = f$ we find

$$w_2 = w_1 \left(1 + \frac{z_2^2}{f^2} - 2 \frac{z_2}{f} + \frac{f^2}{L_1^2} \right)^{1/2}.$$

For $z_2 \approx f$ we have $w_2 \approx w_1 f / L_1 = \lambda f / \pi w_1$. Large values of z_2 result in the usual quadratic behavior $w_2 \simeq w_1 z_2 / f$, but for z_2 similar to f the beam width will be closest to constant when $z_1 = f$ as shown in the right hand side of Fig. 6.11.

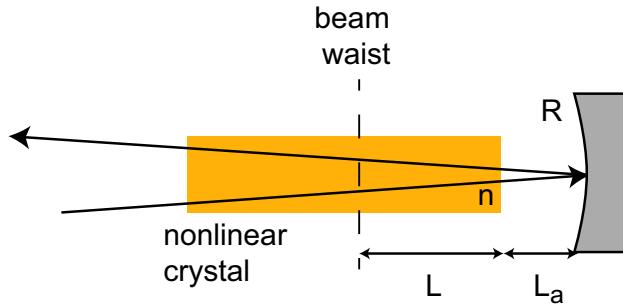


Figure 6.12: Refocusing of light in a crystal with a curved mirror

6.3.3 Refocusing with a mirror

Another example arises in the context of optical frequency conversion in a nonlinear crystal. As shown in Fig. 6.12 we wish to refocus a beam waist in a second pass through a nonlinear crystal. The ABCD matrix from the beam waist plane to the mirror and back is

$$\begin{aligned} \mathbf{M} &= \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/n \end{pmatrix} \begin{pmatrix} 1 & L_a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix} \begin{pmatrix} 1 & L_a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \frac{2}{R}(L_a + L/n) & -\frac{2}{R}(L_a + L/n)(L_a - R + L/n) \\ \frac{-2}{nR} & 1 - \frac{2}{R}(L_a + L/n) \end{pmatrix}. \end{aligned} \quad (6.24)$$

Setting $BD + ACL_1^2 = 0$ gives the condition $R = 2(L/n + L_a)$ for reimaging the waist. This result could also have been deduced from the behavior of a thin lens without multiplying all the ABCD matrices together by recalling that the effective diffraction length inside a medium of index n is L/n so the effective distance to the mirror is $L/n + L_a$ and the focal length is $f = R/2$. Thus the waist is in the front focal plane when $R = 2f = 2(L/n + L_a)$.

6.3.4 Zoom lenses for Gaussian beams

The question arises as to whether or not the geometrical imaging condition $B = 0$ can also coincide with reimaging of the beam waist which requires $BD + ACL_1^2 = 0$ as discussed above. We see that if $B = 0$ the condition for a new waist is $AC = 0$. Since the ray matrices are unimodular $AD - BC = 1$, if $B = 0$ we must have $AD = 1$ which implies $A \neq 0$. The requirement for simultaneous geometric imaging together with reimaging of the beam waist is therefore $B = 0$ and $C = 0$. For a single lens $C = -1/f$ so this is not possible, but it may be satisfied in more complicated multi-lens situations as discussed in this section. Of particular interest are zoom lenses which provide geometrical and beam waist imaging between fixed planes while allowing the magnification to be varied.

6.3.5 Two-lens zoom

A two-lens configuration does not function as a true zoom lens, but it can provide variable magnification with relatively small axial displacement of the beam waist[64]. This is called a “varifocal” zoom. The analysis proceeds as follows. The ray matrix for the two-lens zoom shown in Fig. 6.13

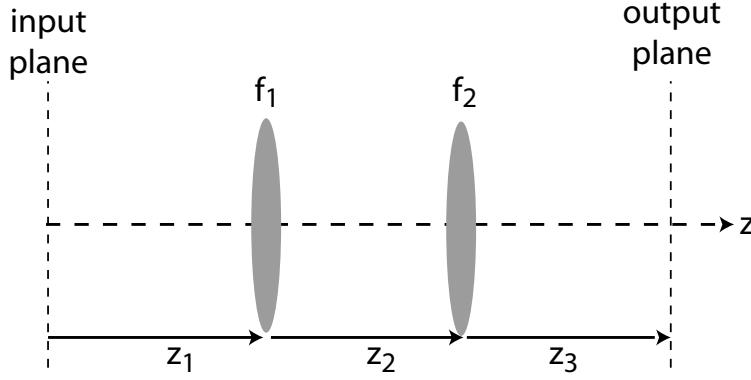


Figure 6.13: Layout of a two lens zoom.

is

$$\mathbf{M} = \frac{1}{f_1 f_2} \times \begin{pmatrix} f_1(f_2 - z_3) + z_2 z_3 - f_2(z_2 + z_3) & f_1 [(z_1 + z_2 + z_3)f_2 - (z_1 + z_2)z_3] - z_1 [(z_2 + z_3)f_2 - z_2 z_3] \\ z_2 - f_1 - f_2 & (f_2 - z_1 - z_2)f_1 + z_1(z_2 - f_2). \end{pmatrix}$$

The geometrical imaging condition is $B = 0$ or

$$z_2 = f_1 + \frac{f_1^2}{z_1 - f_1} + \frac{f_2 z_3}{z_3 - f_2}. \quad (6.25)$$

For $B = 0$ the waist imaging condition is $C = 0$ or

$$z_2 = f_1 + f_2. \quad (6.26)$$

Simultaneous solution of (6.25,6.26) gives

$$z_3 = \frac{f_2}{f_1} (f_1 + f_2) - \frac{f_2^2}{f_1^2} z_1.$$

The geometrical magnification as well as the ratio of output to input waists is then given by

$$A = \frac{f_1(f_2 - z_3)}{f_2(f_1 - z_1)} = -\frac{f_2}{f_1}.$$

Although the two-lens configuration can simultaneously provide geometrical imaging and waist imaging it cannot do so while providing zoom functionality since the magnification is fixed at $-f_2/f_1$.

If we keep $B = 0$ but drop the requirement that the waist appears at the geometrical image plane, i.e. $z_2 \neq f_1 + f_2$ then the geometrical image appears at

$$z_{3g} = \frac{f_2 z_1 z_2 - f_1 f_2 (z_1 + z_2)}{f_1 (f_2 - z_1 - z_2) + z_1 (z_2 - f_2)} \quad (6.27)$$

and the magnification is

$$A = \frac{f_1 f_2}{f_1 (f_2 - z_1 - z_2) + z_1 (z_2 - f_2)}.$$

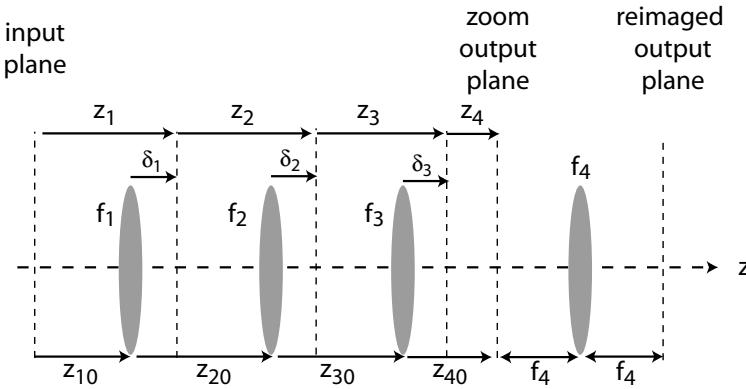


Figure 6.14: Layout of a three lens zoom. The three lens zoom is afocal so we include a fixed f_4 to transform the output. The nominal distances are $z_{10}, z_{20}, z_{30}, z_{40}$. The lenses move by amounts $\delta_1, \delta_2, \delta_3$ during zoom operation giving actual distances $z_1 = z_{10} + \delta_1$, $z_2 = z_{20} + \delta_2 - \delta_1$, $z_3 = z_{30} + \delta_3 - \delta_2$, and $z_4 = z_{40} - \delta_3$.

A parfocal zoom lens provides variable magnification without axial image displacement, i.e. $z_1 + z_2 + z_{3g} = Z = \text{constant}$. Inserting into (6.27) we can write the object to image distance as

$$Z = z_1 + z_2 + \frac{f_2 z_1 z_2 - f_1 f_2 (z_1 + z_2)}{f_1 (f_2 - z_1 - z_2) + z_1 (z_2 - f_2)}. \quad (6.28)$$

It is not hard to show that there are no choices of f_1, f_2 that allow the magnification to be varied while keeping Z fixed. We conclude that a parfocal or true zoom lens is not possible with only two lenses.

However variable magnification accompanied by some motion of the image plane, or varifocal zoom behavior, is possible. The relation between magnification and image shift is found by rewriting (6.28) as

$$Z = z_1 + z_2 + A \left(\frac{z_1 z_2}{f_1} - z_1 - z_2 \right). \quad (6.29)$$

If we put $z_2 = z_1 f_1 / (z_1 - f_1)$ the term in parentheses vanishes and $Z = \frac{z_1^2}{z_1 - f_1}$ while $A = \frac{f_1}{f_1 - z_1}$ so $Z = -A z_1^2 / f_1$. This is not a useful solution since the smallest image shift is at $z_1 \approx 0$ which means $z_2 \approx 0$ which is not physical, as the lenses would be on top of each other. Let us instead choose $z_1 = f_1$ which gives ♠?? (to be added, but this is not very interesting).

6.3.6 Three-lens zoom

To provide geometrical and waist imaging with variable magnification we turn to a three lens configuration as shown in Fig. 6.14. This can provide variable magnification between fixed input and output planes and is referred to as a “parfocal” zoom. The distance from input plane to zoom output plane is $Z = z_1 + z_2 + z_3 + z_4 = z_{10} + z_{20} + z_{30} + z_{40}$. If we start with nominal positions $z_{10} = f_1$, $z_{20} = f_1 + f_2$, $z_{30} = f_2 + f_3$, $z_{40} = f_3$ then we have geometrical and Gaussian waist imaging to the reimaged output plane with static magnification $M_0 = f_2 f_4 / (f_1 f_3)$. The lens values can be chosen as desired, also either positive or negative, to give a desired static magnification. Negative lenses (concave) lead to a shorter overall zoom lens and may be preferred for compactness.

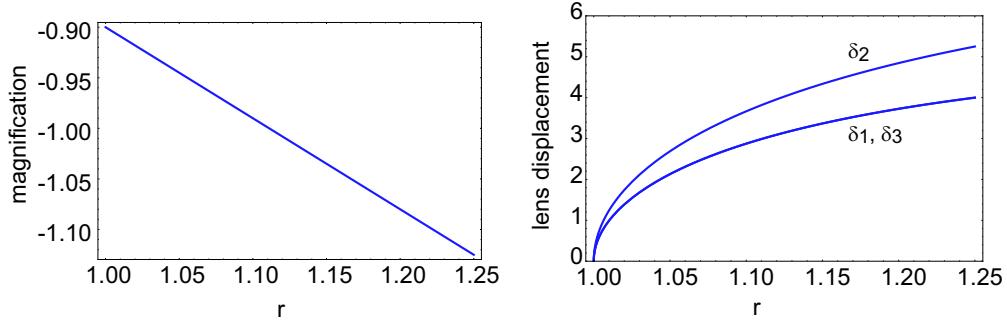


Figure 6.15: Three lens zoom as in Fig. 6.14 with $f_1 = 5.$, $f_2 = -2.5$, $f_3 = 5.$, $f_4 = 9..$ The actual magnification is shown on the left and the lens displacements on the right.

By moving the lenses, i.e. changing $\delta_1, \delta_2, \delta_3$ away from zero we get a new magnification $M = rM_0$. ABCD analysis shows that this can be done while keeping $B = C = 0$ provided

$$\begin{aligned}\delta_1 &= \sqrt{r-1} \frac{f_1^2[(f_1^2 - rf_2^2)(f_3^2 - rf_2^2)]^{1/2}}{r^2 f_2^3 - rf_1^2 f_2} \\ \delta_2 &= \sqrt{r-1} \frac{(r^2 f_2^4 - f_1^2 f_3^2)}{rf_2[(f_1^2 - rf_2^2)(f_3^2 - rf_2^2)]^{1/2}} \\ \delta_3 &= \sqrt{r-1} \frac{(rf_2^2 - f_1^2)f_3^2}{rf_2[(f_1^2 - rf_2^2)(f_3^2 - rf_2^2)]^{1/2}}.\end{aligned}$$

These solutions are written in a form valid for $r > 1$ and $(f_1^2 - rf_2^2)(f_3^2 - rf_2^2) > 1$. If we wish to vary the zoom magnification both smaller and larger than a nominal value we can simply start with M_0 at the small end and use these solutions for $r > 1$. An example is shown in Fig. 6.15. The initial magnification with the lenses confocally spaced is $M_0 = -0.9$ so varying r between 1 and 1.25 changes the magnification to $M = -1.125$. The total length of the zoom is $Z = 15$.

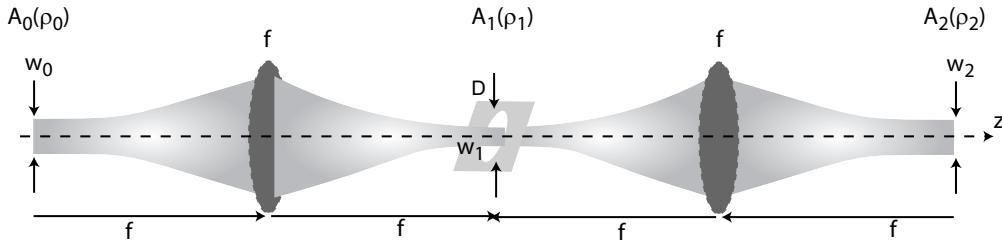


Figure 6.16: Gaussian beam filtered with a circular aperture.

6.4 Apertured Gaussian Beams

The power transmission of a Gaussian beam with waist w through a circular aperture of diameter D is

$$T = \frac{2}{\pi w^2} \int_0^{D/2} dr 2\pi r e^{-2r^2/w^2} = 1 - e^{-\frac{D^2}{2w^2}}. \quad (6.30)$$

Choosing $D = 2w, 3w, 4w$ gives $T = 0.8647, 0.98889, 0.99966$.

Aperturing the beam is useful for removing the intensity tails and for smoothing the spatial profile. This is typically done in a Fourier filtering configuration shown in Fig. 6.16. Say we have a Gaussian beam with waist w_0 but the beam has some additional spatial structure we wish to suppress. We Fourier transform to a new waist w_1 , pass through an aperture of diameter D and transform back to the original waist w_0 . If the aperture were not present the output beam would be the same as the input beam. With an aperture spatial structure gets blocked in the Fourier plane. If the aperture is too big unwanted spatial frequencies giving beam structure will still be transmitted. If the aperture is too small, we lose power, and we add “ringing” to the output beam due to the discontinuity in intensity at the aperture edges.

It is not obvious what the optimal aperture diameter should be. We can estimate the ringing effect as follows. The discontinuity in the beam amplitude at the aperture edge is $\delta A = e^{-D^2/4w_1^2}$ for a beam with unit peak amplitude. If we think about the fringe due to this discontinuity on the background of a unit amplitude Gaussian we get a peak intensity of $\delta I = (1 + \delta A)^2 - 1 \simeq 2\delta A$. Taking $D = 4w_1$ to get more than 0.999 transmission gives $\delta I = 0.037$, or 3.7 %. This is a rough estimate, but it highlights the problem that very small intensity changes can lead to relatively large interference.

We can calculate more precisely with a Fourier model. Let's first consider the ideal case of a centered circular aperture with diameter D and an on axis Gaussian beam of waist w_1 , $A_1(\rho_1) = A_{10}e^{-\rho_1^2/w_1^2}$. The field in the back focal plane has radial symmetry and can be expressed as

$$A_2(\rho_2) = -i \frac{k}{f} A_{10} \int_0^{D/2} d\rho_1 \rho_1 e^{-\rho_1^2/w_1^2} J_0\left(\frac{k}{f}\rho_1 \rho_2\right). \quad (6.31)$$

Put $t = \frac{2}{D}\rho_1$, $v = \frac{kD}{2f}\rho_2$, $u = -iD^2/2w_1^2$, to get

$$A_2(\rho_2) = \left(\frac{kw_1^2}{2f} A_{10}\right) u \int_0^1 dt t e^{-i\frac{u}{2}t^2} J_0(vt).$$

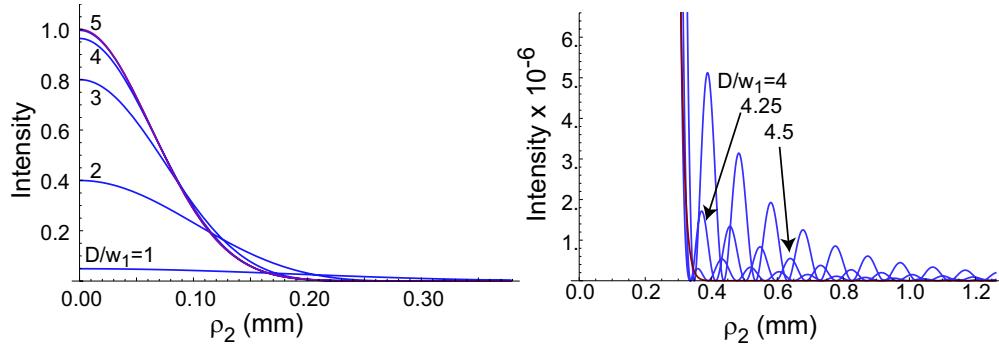


Figure 6.17: Intensity profiles in back focal plane for $w_1 = w_2 = 0.126$ mm, $\lambda = 0.5$ μm , $f = 100$. mm and $D/w_1 = 1, 2, 3, 4, 5$ (left). The profile for $D/w_1 = 5$ is indistinguishable from that with $D = \infty$ shown in purple. The right hand plot shows the weak oscillating tails for $D/w_1 = 4, 4.25, 4.5$.

The integral can be expressed in terms of the Lommel functions (5.38)

$$U_1(u, v) = u \int_0^1 dt t J_0(vt) \cos \left[\frac{u}{2}(1 - t^2) \right] \quad (6.32a)$$

$$U_2(u, v) = u \int_0^1 dt t J_0(vt) \sin \left[\frac{u}{2}(1 - t^2) \right] \quad (6.32b)$$

whence

$$A_2(\rho_2) = \left(\frac{kw_1^2}{2f} A_{10} \right) e^{-vu/2} (U_1 + iU_2). \quad (6.33)$$

Let's check the amplitude at the origin $\rho_2 = 0$. We find

$$A_2(0) = \left(\frac{kw_1^2}{2f} A_{10} \right) e^{-vu/2} (U_1(u, 0) + iU_2(u, 0)).$$

From the function definitions we see that

$$U_1(u, 0) = \sin(u/2), \quad U_2(u, 0) = 2 \sin^2(u/4)$$

so

$$A_2(0) = -i \left(\frac{kw_1^2}{2f} A_{10} \right) \left[1 - e^{-D^2/4w_1^2} \right]. \quad (6.34)$$

When there is no aperture the output field is a Gaussian with waist $w_2 = 2f/kw_1$ and amplitude

$$A_2(\rho_2) = -i A_{10} \frac{w_1}{w_2} e^{-\rho_2^2/w_2^2} = -i A_{10} \frac{kw_1^2}{2f} e^{-k^2 w_1^2 \rho_2^2 / 4f^2}$$

which agrees with the first term of (6.34). The peak intensity falls off faster than the power transmission given by Eq. (6.30) which implies that the beam effectively is broadened.

The intensity profiles for several values of aperture diameter are shown in Fig. 6.17. As the aperture diameter is decreased the transmitted power is reduced and the beam becomes broader but there is no obvious modulation imparted to the main beam. There is weak modulation in the

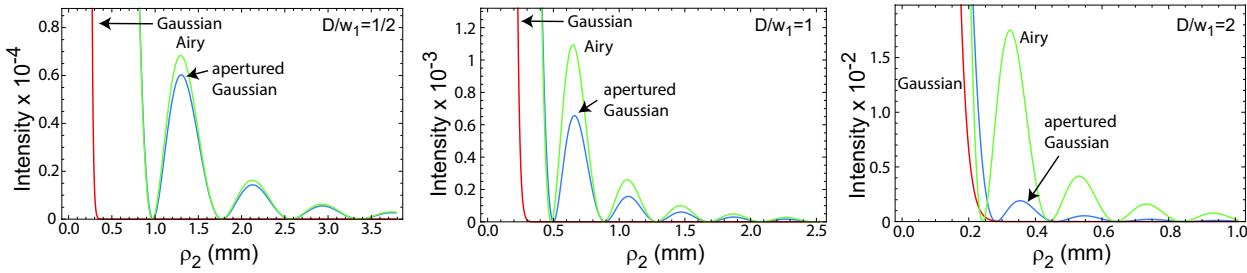


Figure 6.18: Airy ring structure of the intensity profiles in the back focal plane for $D/w_1 = 0.5, 1, 2$, and all other parameters the same as in Fig. 6.17. The red curves are for a Gaussian beam with no aperture, the blue curves are the apertured Gaussian and the green curves are from the Airy formula (5.27).

far wings for $D/w_1 \sim 4$, at $< 10^{-6}$ in relative intensity. For $D \sim w_1$ the output is essentially the transform of a uniform circular disk and Airy rings can be seen, as shown in Fig. 6.18. When $D < w_1$ the apertured intensity is close to uniform and the intensity profile matches closely that of the Airy expression (5.27). As D/w_1 increases the intensity inside the aperture becomes more and more nonuniform and the antinodes of the diffraction are in approximately the same positions as predicted by the Airy formula, but are substantially suppressed.

We conclude from the above plots that aperturing at $D \sim 4w_1$ does not impart significant extra modulation to the central lobe of the beam. This calculation leaves open the question of why the estimate based on the edge discontinuity overestimates the beam modulation. This may be due to the fact that a Gaussian beam picks up a diffractive phase of $\pi/2$ when propagating from Fourier to image planes. By Babinet's principle we can think of the interference as arising from superposition of the transformed whole Gaussian minus the transform of the excised wings of the Gaussian. If the wings do not acquire the same $\pi/2$ phase they interfere with the main beam in quadrature which suppresses the modulation amplitude.

6.4.1 Apertured beam in intermediate plane

The question arises as to the structure of the apertured beam if we look in a plane a distance z away from the Fourier plane. We can calculate this using Fresnel diffraction as

$$\begin{aligned} A_3(\rho_3) &= -i\frac{k}{z}e^{i\frac{k}{2z}\rho_3^2} \int_0^\infty d\rho_2 A_2(\rho_2)\rho_2 e^{i\frac{k}{2z}\rho_2^2} J_0(k\rho_3\rho_2/z) \\ &= -A_{10}\frac{k^2}{fz}e^{i\frac{k}{2z}\rho_3^2} \int_0^\infty d\rho_2 \rho_2 e^{i\frac{k}{2z}\rho_2^2} J_0\left(\frac{k}{z}\rho_3\rho_2\right) \int_0^{D/2} d\rho_1 \rho_1 e^{-\rho_1^2/w_1^2} J_0\left(\frac{k}{f}\rho_1\rho_2\right) \\ &= -A_{10}\frac{k^2}{fz}e^{i\frac{k}{2z}\rho_3^2} \int_0^{D/2} d\rho_1 \rho_1 e^{-\rho_1^2/w_1^2} \int_0^\infty d\rho_2 \rho_2 e^{i\frac{k}{2z}\rho_2^2} J_0\left(\frac{k}{z}\rho_3\rho_2\right) J_0\left(\frac{k}{f}\rho_1\rho_2\right). \end{aligned}$$

Gradshteyn and Ryzhik 6.633 give

$$\int_0^\infty dx x J_0(ax) J_0(bx) e^{-c^2 x^2} = \frac{1}{2c^2} e^{-\frac{a^2+b^2}{4c^2}} I_0\left(\frac{ab}{2c^2}\right) \quad a > 0, \quad b > 0, \quad |\arg c| < \pi/4.$$

We have $a = k\rho_3/z, b = k\rho_1/f$ which are both positive but $c^2 = -ik/2z$ so $\arg c = -\pi/4$. We can

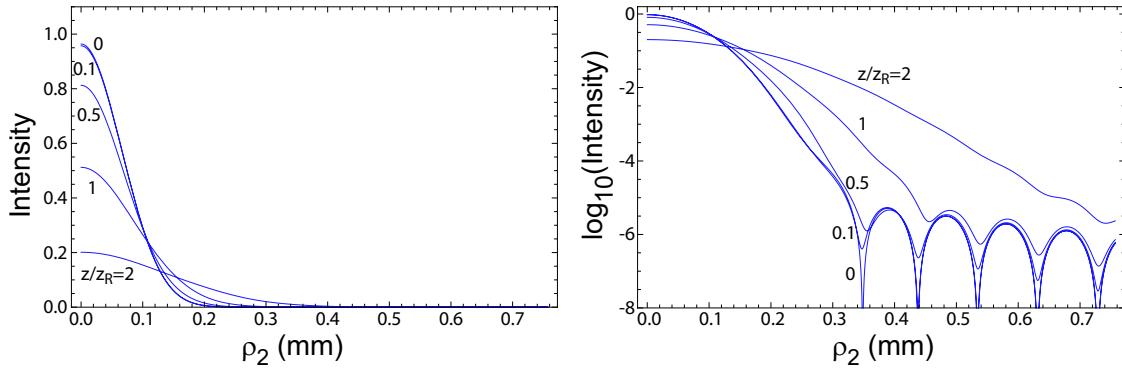


Figure 6.19: Intensity profiles for $D/w_1 = 4$, and all other parameters the same as in Fig. 6.17 for $z = (0, 0.1, 0.5, 1, 2)L_R$ with $L_R = \pi w_2^2/\lambda = 200$. mm.

add a small convergence factor $\epsilon > 0$ to the integral so $c^2 = \epsilon - ik/2z$ and use

$$\int_0^\infty d\rho_2 \rho_2 e^{(i\frac{k}{2z}-\epsilon)\rho_2^2} J_0\left(\frac{k}{z}\rho_3\rho_2\right) J_0\left(\frac{k}{f}\rho_1\rho_2\right) = \frac{1}{2\epsilon - ik/z} e^{-\frac{k^2(\rho_3^2 + \rho_1^2)}{4\epsilon - 2ik/z}} I_0\left(\frac{k^2}{2\epsilon - ik/z}\rho_1\rho_3\right).$$

Letting $\epsilon \rightarrow 0$ we get a well behaved result which we use to express the output field as

$$A_3(\rho_3) = -iA_{10}\frac{k}{f} \int_0^{D/2} d\rho_1 \rho_1 e^{-\rho_1^2/w_1^2} e^{-i\frac{kz}{2f^2}\rho_1^2} J_0\left(\frac{k}{f}\rho_1\rho_3\right).$$

This last expression can again be written in terms of Lommel functions as

$$A_3(\rho_3) = -i \frac{A_{10}}{z/f - i2f/kw_1^2} e^{-iu/2} [U_1(u, v) + iU_2(u, v)] \quad (6.35)$$

with

$$u = \frac{kzD^2}{4f^2} - i\frac{D^2}{2w_1^2}, \quad v = \frac{kD}{2f}\rho_3.$$

We can check that letting $z \rightarrow 0$ and $\rho_3 \rightarrow \rho_2$ we recover (6.33) as expected. Intensity profiles for several values of z are shown in Fig. 6.19. We see the expected smooth diffractive spreading of the beam with only weak intensity modulation in the wings.

Finally we can ask what the beam looks like in the near field a distance z from the aperture plane. Using Fresnel diffraction formulae we have

$$A_3(\rho_3) = -i\frac{k}{z} e^{i\frac{k}{2z}\rho_3^2} \int_0^{D/2} d\rho_1 e^{-\rho_1^2/w_1^2} \rho_1 e^{i\frac{k}{2z}\rho_1^2} J_0(k\rho_3\rho_1/z).$$

This is the same as (6.31) with some changes of variables so the solution is

$$A_3(\rho_3) = i \frac{A_{10}}{1 + i2z/kw_1^2} e^{i\frac{k}{2z}\rho_3^2} e^{-iu/2} (U_1(u, v) + iU_2(u, v)) \quad (6.36)$$

with

$$u = -i\frac{D^2}{2} \left(\frac{1}{w_1^2} - i\frac{k}{2z} \right), \quad v = \frac{kD}{2z}\rho_3.$$

♣?? add plot

Chapter 7

Optical resonators

An important application of ray matrices arises in the analysis of resonant optical cavities. In Sec. 3.2 the Fabry-Pérot interferometer was analyzed assuming excitation with infinitely wide plane waves. As pointed out in Eq. (3.22) wide input beams are needed to achieve high finesse in devices with planar mirrors. This limit can be removed by using spherical mirror cavities that support a spectrum of stable eigenmodes. We can analyze the stability and resonance properties of these cavities using the formalism of ray matrices. Implicit in this approach is the assumption that each optical element has a transverse size that is large compared to the width of the resonating eigenmode. When this assumption does not hold a more sophisticated analysis is required that accounts for scattering at the edges of physical elements[65, 66, 67].

7.1 Resonator eigenmodes

An eigenmode satisfies $q_1 = q_2 = q$ which gives

$$\frac{1}{q} = \frac{D - A}{2B} \pm \frac{1}{2B} \sqrt{(A - D)^2 + 4BC}$$

where A, B, C, D are the elements of the ray matrix describing one resonator round trip. Since the ray matrices are unimodular we can write this as

$$\frac{1}{q} = \frac{D - A}{2B} \pm \frac{i}{2B} \sqrt{4 - (A + D)^2}.$$

The q parameter is related to the waist and radius of curvature of the cavity mode by

$$\frac{1}{q} = \frac{1}{R} + i \frac{\lambda}{\pi n w^2}.$$

The condition for a confined mode is that the square of the waist radius is positive, which requires $|A + D| \leq 2$. The mode parameters are then

$$R = \frac{2B}{D - A}, \quad (7.1a)$$

$$w^2 = \frac{2\lambda}{\pi n} \frac{|B|}{[4 - (A + D)^2]^{1/2}}. \quad (7.1b)$$

A waist occurs when $R = \infty$ which requires $A = D$ so that

$$w_0^2 = \pm \frac{\lambda}{\pi n} \frac{B}{(1 - A^2)^{1/2}}. \quad (7.2)$$

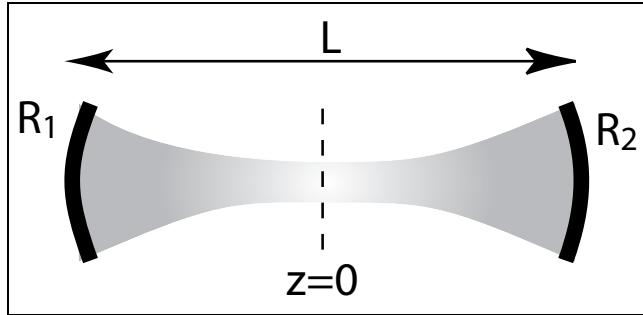


Figure 7.1: Linear resonator with spherical mirrors.

Note that the elements of the round trip ray matrix, and therefore the solutions for R and w , depend on the choice of reference plane. Different choices will lead to different solutions. Alternatively we can choose a desired reference plane and then use the ray matrix for propagation to find the mode parameters at a different resonator plane. It is often convenient to choose a reference plane with some obvious symmetry where e.g. a waist occurs.

7.2 Linear resonator

An important example is provided by the two-mirror linear cavity shown in Fig. 7.1. Two mirrors with radii of curvature R_1 and R_2 are separated by a distance L . We start with a reference plane at the center of the cavity. The beam at this plane has an initial q parameter given by $q(z = 0) = q_1$. After one round trip through the cavity we have

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D}$$

where A, B, C, D are the elements of the round trip ray matrix which is given by

$$\mathbf{M} = \mathbf{M}_l(L/2)\mathbf{M}_r(R_2)\mathbf{M}_l(L)\mathbf{M}_r(R_1)\mathbf{M}_l(L/2). \quad (7.3)$$

Here $\mathbf{M}_l, \mathbf{M}_r$ are the ray matrices for linear propagation and mirror reflections from Fig. 4.4. We find

$$A = 1 + \frac{2L^2 - L(R_1 + 3R_2)}{R_1 R_2} \quad (7.4a)$$

$$B = 2L + \frac{2L^3 - 3L^2(R_1 + R_2)}{2R_1 R_2} \quad (7.4b)$$

$$C = \frac{4L - R_1 - R_2}{R_1 R_2} \quad (7.4c)$$

$$D = 1 + \frac{2L^2 - L(3R_1 + R_2)}{R_1 R_2} \quad (7.4d)$$

We see that when $R_1 = R_2$ then $A = D$ and the waist is at the center of the cavity. The sign convention for R_1, R_2 is the same as for the focal length $f = R/2$ of the equivalent lens in an unfolded version of the resonator. Thus, for the cavity in Fig. 7.1, $R_{1,2}$ are both positive.

The stability condition $|A + D| < 2$ takes the form

$$-1 < \frac{L(L - R_1 - R_2)}{R_1 R_2} < 0.$$

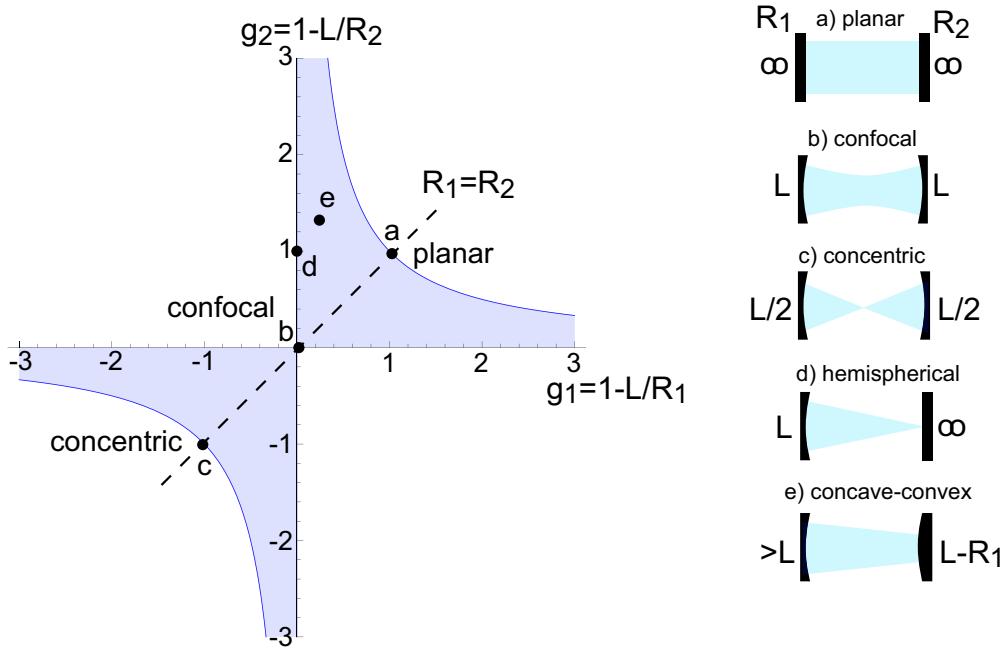


Figure 7.2: Stability diagram for a linear resonator. The shaded regions are stable, and everywhere else is unstable. The points marked with dots are symmetric plane-plane ($g_1 = g_2 = 1, R_1 = R_2 = \infty$), confocal ($g_1 = g_2 = 0, L = R_1 = R_2$), and concentric ($g_1 = g_2 = -1, L = 2R_1 = 2R_2$). The figure also shows asymmetric hemispherical and concave-convex resonators.

It is customary to define parameters $g_1 = 1 - L/R_1$ and $g_2 = 1 - L/R_2$. The stability condition can then be written as

$$0 < g_1 g_2 < 1.$$

The stability diagram in the g_1, g_2 plane is shown in Fig. 7.2.

To calculate the waist size we first determine the position of the waist, which could be inside or outside the cavity. The matrix of Eq. (7.3) cannot be used when $R_1 \neq R_2$ since the plane where the waist occurs is not in the center of the cavity. To find the waist position assume a starting reference plane that is offset by a distance δL from the cavity center. Calculating the round trip matrix and setting $A = D$ gives

$$\delta L = \frac{L(R_2 - R_1)}{2(2L - R_1 - R_2)}.$$

We have defined the position as $\delta L > 0$ when the waist is to the left of the cavity center, i.e. closer to mirror 1. We can check that when $R_1 = \infty$ then $\delta L = L/2$ and the waist is at mirror 1. Although the radii of curvature of the self-consistent modes at the mirrors match the mirror radii of curvature in the linear cavity this is not generally true. For example in the ring cavity of Sec. 7.6 the radius of curvature of the stable mode does not match the cavity mirror radii.

Calculating with a reference plane offset by δL we find the round trip matrix and then using Eq. (7.2) we find

$$w_0^2 = \frac{\lambda L}{\pi} \frac{\sqrt{g_1 g_2 (1 - g_1 g_2)}}{g_1 + g_2 - 2g_1 g_2}. \quad (7.5)$$

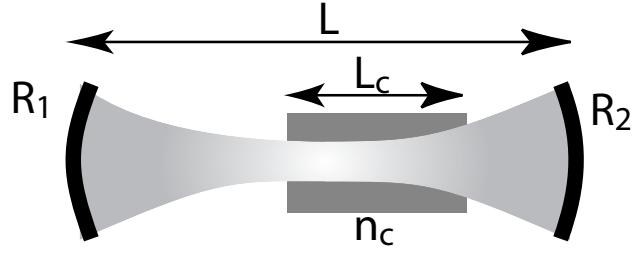


Figure 7.3: Linear resonator containing an optical crystal of length L_c and refractive index n_c .

For a symmetric resonator $R_1 = R_2 = R$ so $g_1 = g_2 = g$ we get

$$w_0 = \left(\frac{\lambda L}{2\pi} \right)^{1/2} \left[\frac{1+g}{1-g} \right]^{1/4} = \left(\frac{\lambda}{2\pi} \right)^{1/2} [L(2R-L)]^{1/4}.$$

Let's evaluate this for the symmetric resonators at the stability boundaries. For the confocal resonator $L = R$, $g = 0$, and taking the limit of (7.5) with L'Hospital's rule we find¹ $w_0 = (\frac{\lambda L}{2\pi})^{1/2}$. Note that this result could have been easily derived using the expression for transformation of a Gaussian beam by a lens $w_1 w_2 = \lambda f / \pi$ with $w_1 = w_2 = w_0$ and $f = R/2 = L/2$. The Rayleigh length of the beam is $L_R = L/2$ so the beam width on the mirrors is $\sqrt{2}w_0$. The confocal resonator has the special property that the mode is intrinsically stable against small misalignments, even though it is marginally stable (it sits on the stability boundary in Fig. 7.2). The mode stability derives from the fact that the center of curvature of each mirror is at the position of the other mirror.

For the concentric resonator $L = 2R$ and $g = -1$ so $w_0 = 0$. The ABCD analysis breaks down in this case. The actual mode size is determined by the size of the end mirrors which are filled by the cavity mode and a more sophisticated analysis is needed to predict the resonator mode.

For the planar resonator $g = 1$ so $w_0 = \infty$. Again the ABCD analysis breaks down. The actual mode size is determined by diffraction from the edges of the cavity mirrors and requires a more sophisticated analysis.

7.3 Linear cavity with internal optical element

Another important example is provided by the two-mirror linear cavity with an internal optical element shown in Fig. 7.3. Two mirrors with radii of curvature R_1 and R_2 are separated by a distance L and there is an element of length L_c with index n_c . This type of cavity is commonly used for second harmonic generation and optical parametric oscillation and as a laser, where the element represents the gain medium.

We could repeat the analysis of Sec. 7.2 to find the stability condition but this is not necessary. When there is no internal crystal the matrix for propagation between the mirrors is $\mathbf{M}_l(L) = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}$. When the crystal is present this becomes $\mathbf{M}_l(L_{\text{eff}})$ where the effective diffractive length

¹ Although Eq. (7.4b) gives $B = 0$ for the confocal resonator $A^2 = 1$ so Eq. (7.2) is indeterminate and it is necessary to take a limit to find the finite waist value.

is

$$L_{\text{eff}} = L - L_c + \frac{L_c}{n} = L - L_c \left(1 - \frac{1}{n}\right).$$

It can be checked that L_{eff} is independent of the exact position of the crystal inside the resonator. All results from Sec. 7.2 are still valid with the substitution $L \rightarrow L_{\text{eff}}$ and, when the waist occurs inside the optical element, $\lambda \rightarrow \lambda/n_c$. Although L_{eff} is independent of the position of the internal element the position of the beam waist does depend on the position of the element as can be found from an ABCD analysis.

It is notable that for $n > 1$, $L_{\text{eff}} < L$ so the diffractive length corresponding to one cavity round trip is reduced. On the other hand the phase accumulation due to propagation is proportional to

$$L_\phi = (L - L_c) + nL_c = L + (n - 1)L_c.$$

We see that $L_\phi > L$ for $n > 1$.

7.4 Resonant frequencies

The resonant frequencies of plane wave fields inside a planar cavity were found in Sec. 3.2 from the condition $2Lkn = 2\pi q$ where q is the longitudinal mode index. Resonant modes in spherical cavities have a transverse field profile which can correspond to any of the higher order modes detailed in Sec. 6.1.1. The resonant frequency is different for different transverse modes due to the Gouy phase factor $\eta_{mn}(z) = (1 + m + n)\eta(z)$. Here m, n label the transverse modes with $m = n = 0$ the lowest order Gaussian beam mode. Any of the transverse modes can be excited with proper mode matching in a stable cavity. Because the mode frequency depends on q, m , and n an optical resonator can be used for combined spectral and spatial filtering of an optical field.

The resonant frequencies are found by starting at the reference plane where the waist occurs and enforcing the condition

$$2Lkn(z) - (1 + m + n)\eta_{\text{rt}} = 2\pi q \quad (7.6)$$

where η_{rt} is the accumulated Gouy phase in one cavity round trip.

It can be shown, after considerable algebra, (?? details to be added) that the resonant frequencies are

$$\nu_{qmn} = \nu_{\text{FSR}} \left[q + \frac{1 + m + n}{\pi} \cos^{-1}(\pm\sqrt{g_1 g_2}) \right]. \quad (7.7)$$

Here ν_{FSR} is the cavity free spectral range calculated for a plane wave and the sign of $\sqrt{g_1 g_2}$ is given by the sign of g_1 (g_1 and g_2 have the same sign for stable modes). The part proportional to $(m + n) \cos^{-1}(\pm\sqrt{g_1 g_2})$ arises from the higher order Gouy phase.

Let's calculate the mode frequencies for the three basic linear resonators, planar, confocal, and concentric. For the planar resonator $g_1 = g_2 = 1$ and

$$\nu_{qmn} = q\nu_{\text{FSR}}.$$

All transverse modes have the same frequency since there are no focusing elements, and no Gouy phase shift.

For the confocal resonator with $g_1 = g_2 = 0$ and index of refraction $n(z) = 1$ we find

$$\nu_{qmn} = \left[\left(q + \frac{1}{2} \right) + \left(\frac{m + n}{2} \right) \right] \nu_{\text{FSR}}. \quad (7.8)$$

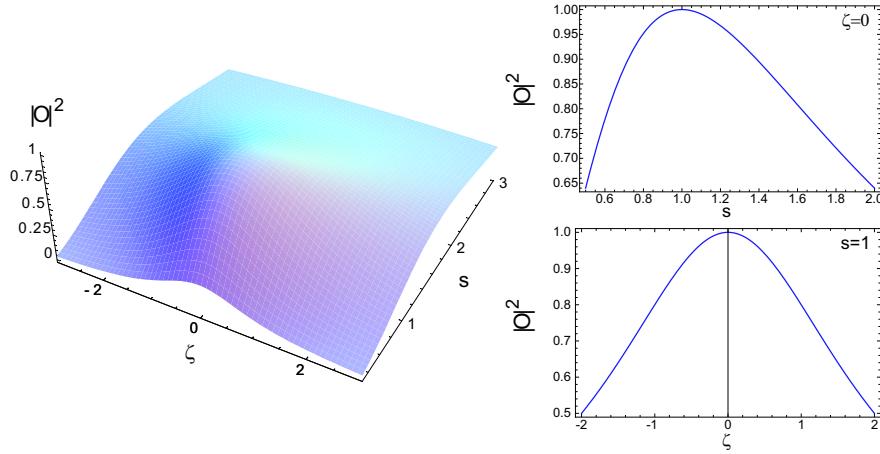


Figure 7.4: Power overlap between a Gaussian beam and a Gaussian cavity mode. The line plots show the variation with s and ζ .

When $m+n$ is an even integer transverse modes associated with longitudinal index q are degenerate with longitudinal mode $q' = q + (m+n)/2$. In other words all modes with $m+n$ even, including $m+n = 0$ will be simultaneously resonant. When $m+n$ is an odd integer the transverse modes are spaced halfway in between axial modes. By tuning the resonator length we can select simultaneous resonance for all even or all odd transverse modes.

For the concentric resonator $g_1 = g_2 = -1$, and

$$\nu_{qmn} = (q + 1 + m + n) \nu_{\text{FSR}}.$$

Different transverse modes are all simultaneously degenerate with axial modes at multiples of the free spectral range.

We see that the planar and concentric geometries allow simultaneous resonance of all transverse modes. Nevertheless these geometries are not preferable in practice since they are very sensitive to small cavity misalignments. The confocal geometry is relatively insensitive to alignment errors while allowing simultaneous resonance of all even or odd transverse modes.

7.5 Mode matching

An important practical aspect when working with optical cavities is the necessity of mode matching an input beam to the cavity mode. We can define an overlap function between an incident beam A_1 and a cavity mode amplitude A_2 by the integral

$$O = \int dx dy A_1^* A_2. \quad (7.9)$$

The profiles A_1 and A_2 are functions of z , however it is straightforward to show that provided A_1 and A_2 are solutions of the paraxial equation (5.2) the overlap O is independent of z . We can therefore simplify the analysis by assuming the cavity mode has a waist at $z = 0$ and calculating the overlap in this plane. Parameterizing the ratio of the waists of the amplitudes by $s = w_{01}/w_{02}$ and the axial displacement of the incident beam waist by $\zeta = z_1/z_{R2}$ we find that the power overlap is given by

$$|O|^2 = \frac{4s^2}{(1+s^2)^2 + \zeta^2}. \quad (7.10)$$

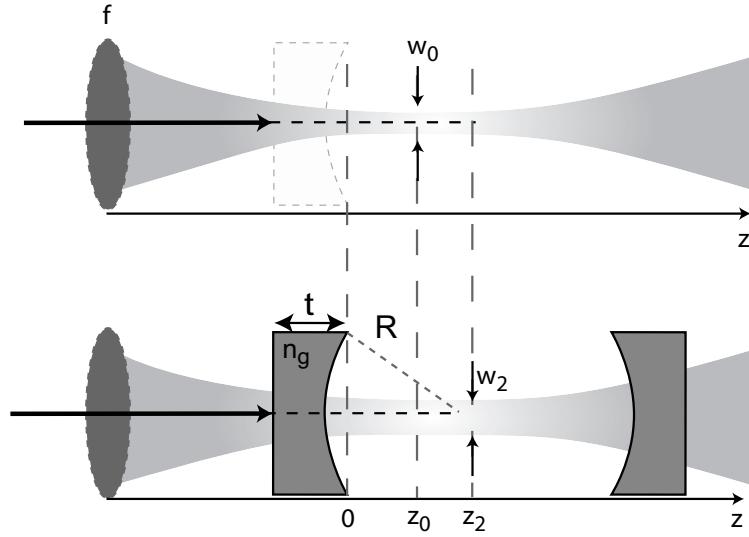


Figure 7.5: Mode matching to a linear resonator. Without the input coupler (top) the focusing lens produces a waist w_0 at position z_0 . With the input coupler in place (bottom) the waist is transformed to w_2 at position z_2 .

A plot of the overlap is shown in Fig. 7.4. In order to exceed 90% power overlap the beam waist ratio must not deviate by more than $-20\% - +30\%$ and the waist position must not be displaced by more than about 1/2 of a Rayleigh length. These requirements also apply to the problem of maximizing coupling efficiency into a single mode optical fiber. It should be apparent that mode matching to higher order transverse modes requires preparing the input beam to also have a higher order mode structure.

In order to achieve good mode matching the input beam parameters must be carefully controlled. Consider the geometry shown in Fig. 7.5. A lens with focal length f is positioned to create a beam waist w_0 at $z = z_0$. When the beam is transmitted through the front mirror (thickness t and index n_g) of the linear cavity (the input coupler) it acquires a radius of curvature at $z = 0$ given by

$$R_1 = R \frac{L_R^2 + z'^2}{\tilde{n}(L_R^2 + z'^2) - Rz'}$$

where $L_R = \pi w_0^2 / \lambda$, $\tilde{n} = n_g - 1$, and $z' = z_0 + \frac{n_g - 1}{n_g}t$. The input coupler is concave to the right so $R > 0$. As long as $R_1 < 0$ a new waist will be formed to the right of the mirror. If $R_1 > 0$ the new waist is to the left of the mirror, i.e. the beam is diverging in the resonator and will certainly be poorly mode matched. The transition occurs at a cavity mirror radius of

$$R_m = \tilde{n} \frac{L_R^2 + z'^2}{z'}.$$

For $R < R_m$ the new waist will be to the left of the front mirror. Assuming $R > R_m$ the waist is formed at

$$z_2 = R \frac{(R - \tilde{n}z')z' - \tilde{n}L_R^2}{(R - \tilde{n}z')^2 + \tilde{n}^2 L_R^2}. \quad (7.11)$$

The value of the new waist is

$$w_2^2 = w_0^2 \frac{R^2}{\tilde{n}^2 L_R^2 + (R - \tilde{n}z')^2}. \quad (7.12)$$

Equations (7.11,7.12) determine z_2, w_2 as functions of z_0, w_0 . For practical use they must be inverted to find the corresponding values of z_0, w_0 . As they are nonlinear equations it is convenient to solve them graphically.

Let's look at a representative case of modematching to a confocal cavity with $R = 0.1$ m mirrors and a wavelength of $\lambda = 1 \mu\text{m}$. The confocal cavity waist is $w_2 = \sqrt{\lambda R/2\pi} = 126 \mu\text{m}$ at $z_2 = R/2 = 5 \text{ cm}$. To modematch to this waist we should generate a beam with

$$w_0 = w_2 \frac{\sqrt{\tilde{n}^2 L_R^2 + (z'')^2}}{R}.$$

Solving (7.11) for z_0 we find

$$z_0 = \frac{R^2 + 2\tilde{n}R(z_2 - \frac{\tilde{n}}{n_g}t) - 2z_2\frac{\tilde{n}^3}{n_g}t \pm \sqrt{R^4 - 4L_R^2\tilde{n}^2(R + \tilde{n}z_2)^2}}{2\tilde{n}(R + \tilde{n}z_2)}$$

7.6 Ring cavity

Ring cavities are used frequently for optically pumped lasers such as the Ti:Sa and for nonlinear interactions. They have several distinguishing features relative to linear cavities that are advantageous. First they support unidirectional oscillation which prevents spatial hole burning, and second the round trip losses for parametric interactions are reduced relative to linear cavities without requiring that particular phase conditions be satisfied by the cavity mirrors.

A prototypical ring cavity in a bow-tie geometry containing a crystal used at normal incidence is shown in Fig. 7.6. The geometrical round trip cavity length is L while the curved mirrors with radius of curvature R are separated by a distance L_2 . The optical crystal has length L_c and index of refraction n_c . The distance from the crystal end surface to a curved mirror is then $(L_2 - L_c)/2$ and the distance between the curved mirrors for a ray traversing the long side of the cavity is $L - L_2$. With these definitions we find that the round trip ray matrix, starting from the center of the crystal as the reference plane, has elements

$$A = \frac{2(L - L_2)(L_2 - L'_c) - 2R(L - L'_c) + R^2}{R^2} \quad (7.13)$$

$$B = \frac{[L_2 - R - L'_c][(L - L_2)(L_2 - L'_c) - R(L - L'_c)]}{R^2} \quad (7.14)$$

$$C = \frac{4(L - R - L_2)}{n_c R^2} \quad (7.15)$$

$$D = A \quad (7.16)$$

where $L'_c = L_c \frac{n_c - 1}{n_c}$.

Since $A = D$ the waist is located at the reference plane in the center of the crystal as expected. To see how the cavity mode waist depends on parameters we have plotted w found from Eq. (7.1) in Fig. 7.7 as a function of the mirror separation L_2 for several values of the cavity length. The condition for a confined mode ($|A + D| < 2$) can be used to find limits on the possible values of L_2 . A little algebra shows that a stable mode exists for

$$\frac{L + L'_c}{2} - \frac{1}{2}\sqrt{(L - L'_c)^2 - 4R(L - L'_c) + 4R^2} < L_2 < \frac{L + L'_c}{2} - \frac{1}{2}\sqrt{(L - L'_c)^2 - 4R(L - L'_c)}.$$

This analysis is incomplete due to astigmatism which results in different stability conditions in the tangential and saggital planes.

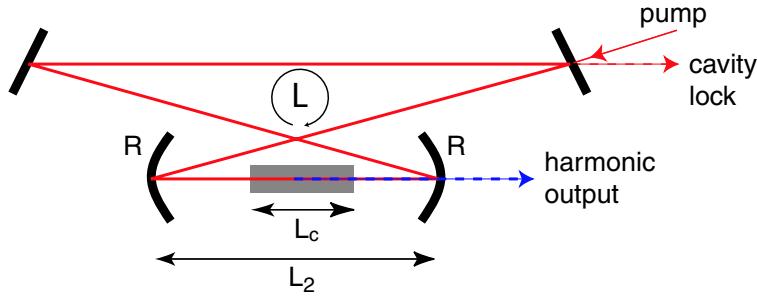


Figure 7.6: Ring bowtie cavity. The fields are drawn for a harmonic generation configuration.

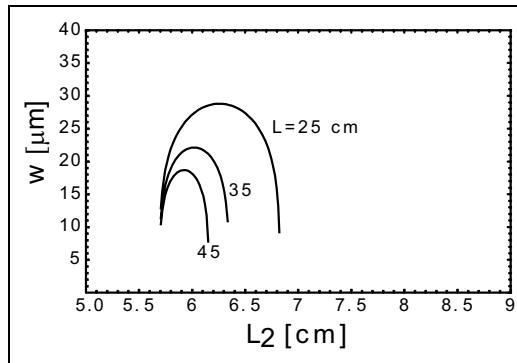


Figure 7.7: Fundamental mode waist for a bowtie cavity with $R = 4$ cm, $L_c = 3$ cm, $n_c = 2.3$, and $\lambda = 0.96 \mu\text{m}$.

7.6.1 Ring cavity with angled crystal

Ring cavities are often used for optically pumped lasers such as dye, or Ti:Sapphire lasers. To avoid reflection losses at the crystal surface it is common to orient the crystal at Brewster's angle as shown in Fig. 7.8. In this case the astigmatism of the angled crystal can be used to compensate the mirror astigmatism, as was first shown by Kogelnik, et al.[68].

Accurate modeling of the cavity requires the optical propagation lengths between the curved mirrors to be accounted for carefully. We assume the cavity beams propagate in a medium with index n_1 and the crystal has index n_c . Brewster's angle is $\theta_B = \tan^{-1}(n_r)$ where $n_r = n_c/n_1$. The refracted angle inside the crystal is $\theta'_B = \tan^{-1}(1/n_r) = \cos^{-1}(n_r/\sqrt{1+n_r^2})$. The refracted beam has a geometrical path length inside the crystal of L_c . Formulae for the x and y displacements of the beam in the crystal are given in Fig. 7.8. The total round trip length is $L_T = L_1 + L_2 + L_{d1} + L_{d2}$. The distance L_1 is the optical path length between the curved mirrors (the actual x distance between the mirrors is slightly smaller). One way to analyze the mode structure is to fix L_T and look at what happens as L_1 is varied. We assume L_c, n_c, θ are fixed constants. The remaining cavity distances are then found from

$$\begin{aligned} L_2 &= (L_{d1} + L_{d2}) \cos(2\theta) - L_1 + (L_c - t_x) \\ L_{d1} &= h / \sin(2\theta) \\ L_{d2} &= (h + t_y) / \sin(2\theta) \end{aligned}$$

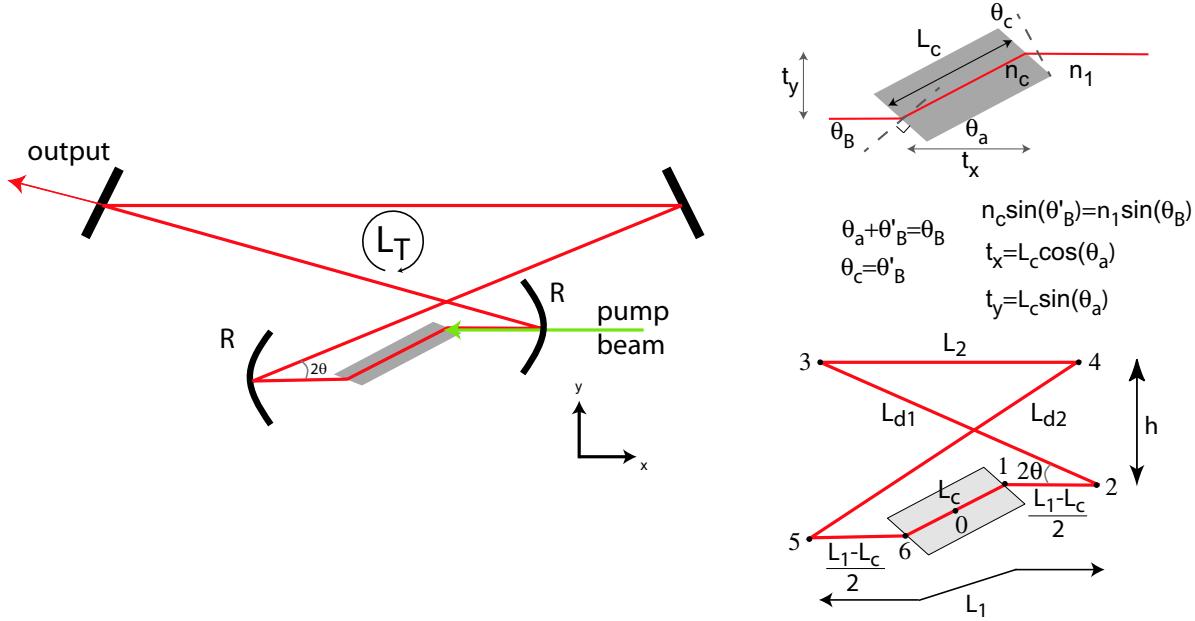


Figure 7.8: Ring cavity for an optically pumped laser with angled crystal. The geometrical round trip distance L_T is $L_T = L_2 + L_{d1} + L_{d2} + L_u$. The path length inside the crystal is L_c . The detail in the upper right shows the geometry of the Brewster cut crystal. The beam path in the lower right with positions 1 – 6 is used for generating a detailed optical layout. Position 0 is at the center of the crystal.

and

$$L_T - L_1 = L_{d1} + L_{d2} + L_2 = \frac{1 + \cos(2\theta)}{\sin(2\theta)}(t_y + 2h) - L_1 + L_c - t_x.$$

The last equation can be solved for $h = h(L_T, L_1)$ which then determines the other lengths L_2, L_{d1}, L_{d2} .

Using the above definitions the round trip ray matrix, starting from the center of the crystal as the reference plane, is given by

$$\begin{aligned} \mathbf{M}_t &= \mathbf{M}_l(L_c/2)\mathbf{M}_{i,t}(n_1, n_c, \theta_B)\mathbf{M}_l((L_1 - L_c)/2)\mathbf{M}_{r,t}(R, \theta)\mathbf{M}_l(L_T - L_1) \\ &\quad \times \mathbf{M}_{r,t}(R, \theta)\mathbf{M}_l((L_1 - L_c)/2)\mathbf{M}_{i,t}(n_c, n_1, \theta'_B)\mathbf{M}_l(L_c/2) \end{aligned}$$

where the matrices are defined in Fig. 4.4. Subscript t denotes tangential matrices and an analogous expression holds for matrices in the sagittal plane.

A sample calculation for a compact Ti:Sa ring laser is shown in Fig. 7.9. Although the astigmatism is well compensated, the tangential and sagittal waists inside the cavity differ by a factor of about 2 which complicates mode matching.

The free spectral range in frequency units is

$$\nu_{FSR} = \frac{c}{(L_T - L_c)n_1 + L_c n_c}.$$

For the parameters of Fig. 7.9 we find $\nu_{FSR} = 1.90$ GHz.

For the purposes of checking the mechanical layout it is useful to find explicit values for the coordinates of the cavity beam vertices. These are given in Table 7.1.

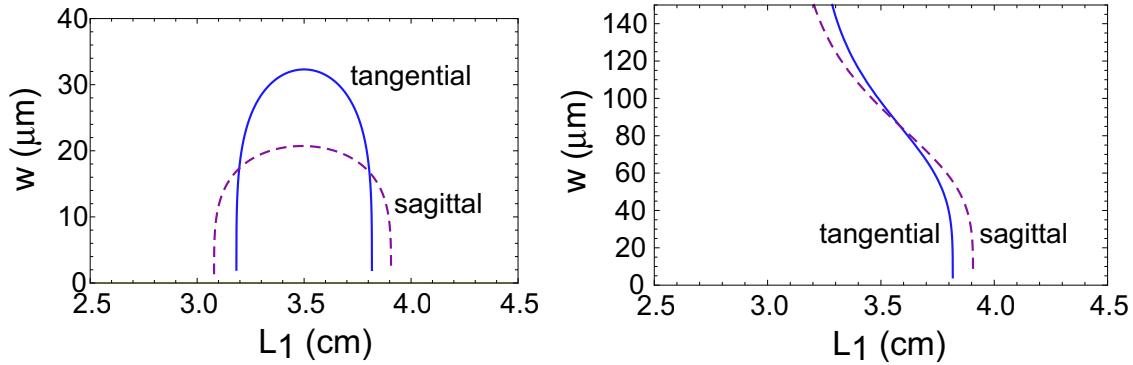


Figure 7.9: Calculated waists at the center of the Ti:Sa crystal (left) and between the flat mirrors (right). Parameters: $\lambda = 0.685 \mu\text{m}$, $L_c = 1 \text{ cm}$, $n_c = 1.77$, $R = 2.5 \text{ cm}$, $\theta = 19 \text{ deg.}$, $L_T = 15 \text{ cm}$, $h = 2.3 \text{ cm}$.

Position	x	y
1	$t_x/2$	$t_y/2$
2	$x_1 + (L_1 - L_c)/2$	y_1
3	$x_2 - L_{d1} \cos(2\theta)$	$y_1 + h$
4	$x_5 + L_{d2} \cos(2\theta)$	$y_1 + h$
5	$-x_2$	$-y_1$
6	$-x_1$	$-y_1$

Table 7.1: Coordinates of beam vertices 1 – 6 in the Ti:Sa ring cavity.

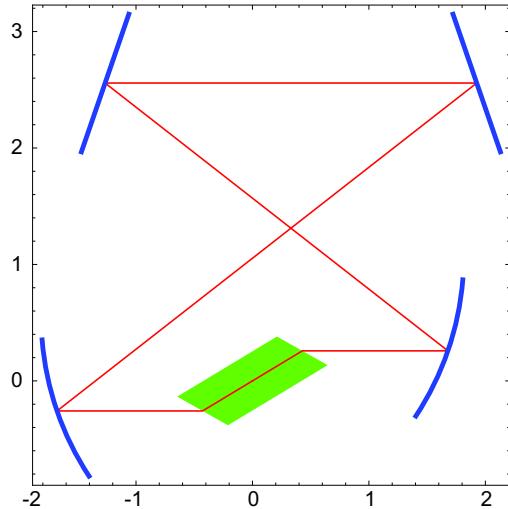


Figure 7.10: Layout of the Ti:Sa cavity for the parameters of Fig. 7.9.

Chapter 8

Light and Matter

In this chapter we will develop some basic aspects of the interaction of light and matter.

8.1 Einstein's 1917 theory

Einstein's 1917 paper[69], written four years after Bohr's theory of the atom and eight years before the invention of modern quantum mechanics, established a theoretical framework that is still valuable today. The theory predates quantum mechanics, yet it uses essentially quantum mechanical notions of transitions between stationary states, and is fully consistent with later results of quantum theory. Einstein considered a two-level atom with lower level 1 and upper level 2 interacting with a radiation field with energy density ρ_ω per unit frequency interval. Einstein assumed the density of atoms in the two levels was N_1 and N_2 and postulated the following rate equations

$$\frac{dN_1}{dt} = AN_2 - B_{2\leftarrow 1}\rho_\omega N_1 + B_{1\leftarrow 2}\rho_\omega N_2, \quad (8.1a)$$

$$\frac{dN_2}{dt} = -AN_2 + B_{2\leftarrow 1}\rho_\omega N_1 - B_{1\leftarrow 2}\rho_\omega N_2. \quad (8.1b)$$

As shown in Fig. 8.1 A is the rate for spontaneous decay from level 2 \rightarrow 1, $B_{2\leftarrow 1}$ is the coefficient for induced absorption of a photon resulting in a transition from 1 \rightarrow 2, and $B_{1\leftarrow 2}$ is the coefficient for induced emission of a photon resulting in a transition from 2 \rightarrow 1. Note that the total atomic density is conserved since $\dot{N}_1 + \dot{N}_2 = 0$. The units of the quantities appearing in the above equations are $[N_1] = [N_2] = \text{m}^{-3}$, $[A] = \text{s}^{-1}$, $[B] = \text{m}^3 \text{J}^{-1} \text{s}^{-2}$, and $[\rho_\omega] = \text{Jsm}^{-3}$.

In thermal equilibrium the populations are governed by Boltzmann factors,

$$N_1 = \frac{e^{-U_1/k_B T}}{Z}, \quad N_2 = \frac{e^{-U_2/k_B T}}{Z},$$

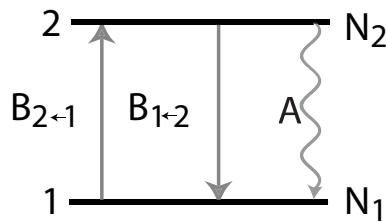


Figure 8.1: Two level atom and transitions due to Einstein A, B coefficients.

with $Z = e^{-U_1/k_B T} + e^{-U_2/k_B T}$. Since the populations are stationary ($\dot{N}_i = 0$) we can also write

$$A = \left(B_{2 \leftarrow 1} \frac{N_1}{N_2} - B_{1 \leftarrow 2} \right) \rho_\omega = \left(B_{2 \leftarrow 1} e^{U_{21}/k_B T} - B_{1 \leftarrow 2} \right) \rho_\omega$$

where $U_{21} = U_2 - U_1$ is the energy difference between the two states.

The equilibrium radiation density is also a function of temperature. It can be written as $\rho_\omega = \hbar\omega n_\omega f(T)$ where $\hbar\omega$ is the energy of a photon with frequency ω , n_ω is the density of states in frequency space, and $f(T)$ gives the excitation number of each state at temperature T . From a classical counting of modes, including a factor of two for two orthogonal polarization states, $n_\omega = \omega^2/\pi^2 c^3$. Since photons are spin 1 bosonic particles they satisfy Bose-Einstein statistics and $f(T) = \frac{1}{e^{\hbar\omega/k_B T} - 1}$ which corresponds to the temperature dependence of Planck's blackbody spectrum.

Einstein used the assumption of thermal equilibrium between radiation and matter to give a very simple derivation of $f(T)$ and the Planck spectrum. We note that the A coefficient satisfies

$$A = \frac{\hbar\omega^3}{\pi^2 c^3} \left(B_{2 \leftarrow 1} e^{U_{21}/k_B T} - B_{1 \leftarrow 2} \right) f(T).$$

We now require that all three unknowns A , $B_{2 \leftarrow 1}$, and $B_{1 \leftarrow 2}$ are constants that do not depend on temperature. These constants depend on the microscopic details of the radiation - matter interaction and quantum mechanics is needed to calculate them. With the additional assumption that the stimulated rates are the same for absorption and emission, $B_{2 \leftarrow 1} = B_{1 \leftarrow 2}$, we get

$$A = \frac{\hbar\omega^3}{\pi^2 c^3} B_{2 \leftarrow 1} \left(e^{U_{21}/k_B T} - 1 \right) f(T).$$

In order for A and $B_{2 \leftarrow 1}$ to be independent of T we require

$$f(T) = \frac{1}{e^{U_{21}/k_B T} - 1}$$

so that¹

$$A = \frac{\hbar\omega^3}{\pi^2 c^3} B_{2 \leftarrow 1}. \quad (8.2)$$

It follows that the radiation energy density per unit angular frequency is

$$\rho_\omega = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{U_{21}/k_B T} - 1} = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{\hbar\omega/k_B T} - 1},$$

which is indeed the Planck spectrum. Note that if we had assumed the Planck spectrum instead of deriving it, the requirement that A be constant would also have led to the equality of $B_{1 \leftarrow 2}$ and $B_{2 \leftarrow 1}$.

Thus Einstein provided a simple derivation of the Planck law based on a few reasonable assumptions: 1) thermal equilibrium between matter and radiation, 2) the rates of atomic excitation and de-excitation are due to elementary processes of spontaneous emission, stimulated emission, and stimulated absorption, 3) the coefficients governing these rates are constants independent of temperature, and 4) the coefficients for stimulated emission and absorption are the same. An important aspect of the theory that is not needed to derive the Planck spectrum but is important for

¹Equation (8.2) will contain different numerical factors if B is defined per unit Hz instead of per unit angular frequency as we have done. In that case using $\rho_\nu = \rho_\omega \frac{d\omega}{d\nu} = 2\pi\rho_\omega$ we would have found $A = 2\pi \frac{\hbar\omega^3}{\pi^2 c^3} B_{2 \leftarrow 1} = \frac{8\pi\hbar\nu^3}{c^3} B_{2 \leftarrow 1}$.

understanding the interaction of light and matter is that spontaneously emitted photons, which are due to the A coefficient, are emitted in random directions. Stimulated emission, due to the $B_{1\leftarrow 2}$ coefficient, results in adding photons to the same mode of the field that stimulated the emission. The added photons are fully coherent in amplitude and phase with the existing field. This provides for the possibility of gain and laser action as we discuss briefly in Sec. 8.1.1.

Degeneracy

When level 1 has degeneracy factor g_1 and level 2 has degeneracy g_2 Eq. (8.2) is modified to

$$g_2 A = \frac{\hbar\omega^3}{\pi^2 c^3} g_1 B_{2\leftarrow 1}$$

$$A = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{g_1}{g_2} B_{2\leftarrow 1} = \frac{\hbar\omega^3}{\pi^2 c^3} B_{1\leftarrow 2}.$$

There are no degeneracy factors in the second equality since both sides refer to a transition from $2 \rightarrow 1$ and the factors cancel. We therefore see that

$$B_{2\leftarrow 1} = \frac{g_2}{g_1} B_{1\leftarrow 2}.$$

The A and B coefficients, as well as the relations between them, have been derived from an empirical model of the radiation-matter interaction and by appealing to concepts of classical statistical mechanics. The actual value of the $B_{1\leftarrow 2}$ coefficient can be calculated from classical radiation theory. The Einstein relations can then be used to find A . The results so obtained are confirmed by fully quantum mechanical calculations, but these lie outside the scope of this course.

8.1.1 Optical gain

Einstein's 1917 theory implies that matter can amplify a light beam. Say we have a sample of two-level atoms and we prepare them so that $N_2 > N_1$. This does not happen in equilibrium as it would imply $T < 0$ but is possible if there is an external source of energy that prepares atoms in state 2.

Such a situation with $N_2 > N_1$ is referred to as population inversion. Light that is resonant with the $1 - 2$ transition experiences loss due to stimulated absorption and gain due to stimulated emission. When $N_2 > N_1$ the rate for stimulated emission is larger than the rate for stimulated absorption and resonant radiation is amplified. The amplification process redistributes the level populations and tends to reduce the inversion. To have a steady state amplification an external source of energy is required to maintain inversion.

If the amplifying medium is placed inside an optical resonator and the round trip gain is greater than the round trip loss any small resonant light will be amplified. As the light grows in intensity the gain will be reduced due to saturation effects until a steady state condition is reached. This corresponds to an oscillator and the steady state output is that of a LASER - Light Amplification by Stimulated Emission of Radiation².

Although the physics of light amplification is embodied in Einstein's 1917 paper it took more than 30 years until the first demonstration of this effect in the 1950s at microwave frequencies[70, 71]. In 1960 oscillation and the first laser were demonstrated at optical wavelengths[72]. The laser is now one of the important devices of our modern information age.

²A more accurate name would perhaps be Light Oscillator by Stimulated Emission of Radiation, but the corresponding acronym is not good.

8.2 Absorption area law

A useful result for the absorption cross section of a weak optical beam can be derived as follows. In (8.1) we assume the radiation field is weak so $N_2 \ll N_1$. We then make the replacement $\rho_\omega \rightarrow I_\omega/c$, where I_ω is the optical intensity per unit frequency, with units of $[I_\omega] = \text{J/m}^2$. The rate at which atoms are excited by photon absorption is then

$$R_{2 \leftarrow 1} = B_{2 \leftarrow 1} \frac{I_\omega(\omega)}{c} N_1 = N_1 \frac{\pi^2 c^2}{\hbar \omega^3} A I_\omega(\omega).$$

Let us now assume that the intensity has a finite width frequency spectrum. In other words the radiation is not monochromatic and the spectrum may be broad compared to the width of the absorption cross section, as seen in Fig. 8.2. The atomic absorption cross section $\sigma(\omega)$ is typically a strongly peaked function of frequency, with the peak absorption occurring when $\omega = \omega_{21}$ the frequency of the atomic transition. The units of the cross section are $[\sigma] = \text{m}^2$. The rate at which photons are absorbed per unit volume is found by integrating the product of the photon flux times the absorption cross section times the density of atoms to give

$$R_{\text{abs}} = \int d\omega \frac{I_\omega(\omega)}{\hbar \omega} \sigma(\omega) N_1.$$

Setting $R_{2 \leftarrow 1} = R_{\text{abs}}$ gives

$$\int d\omega \frac{I_\omega(\omega)}{\hbar \omega} \sigma(\omega) = \frac{\pi^2 c^2}{\hbar \omega^3} A I_\omega(\omega). \quad (8.3)$$

Let us now assume that $\sigma(\omega)$ is strongly peaked at the transition frequency $\omega_{21} = U_{21}/\hbar$, such that the spectral width of the absorption cross section $\delta\omega$ satisfies $\delta\omega \ll \omega_{21}$. We also assume that the intensity spectrum can be approximated as a constant with value $I_\omega(\omega) = I_\omega(\omega_{21})$. Then

$$\int d\omega \frac{I_\omega(\omega)}{\hbar \omega} \sigma(\omega) \simeq \frac{I_\omega(\omega_{21})}{\hbar \omega_{21}} \int d\omega \sigma(\omega) \quad (8.4)$$

and comparison with (8.3) gives

$$\int d\omega \sigma(\omega) = \frac{\pi^2 c^2}{\omega_{21}^2} A. \quad (8.5)$$

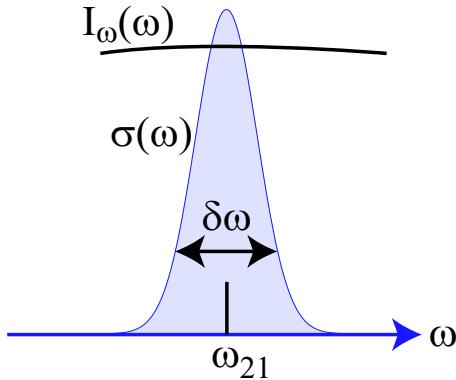


Figure 8.2: Broad band light spectrum interacting with a narrowband resonance.

We thus arrive at the area law relating the cross section to the Einstein A coefficient

$$\int d\omega \sigma(\omega) = \frac{\lambda_{21}^2}{4} A.$$

The calculation only considered radiation propagating in a particular direction. If we assume that the absorption cross section is independent of the orientation of the absorber and allow for three spatial dimensions we find

$$\int d\omega \sigma(\omega) = \frac{3\lambda_{21}^2}{4} A. \quad (8.6)$$

This is known as the absorption area theorem which can be used to estimate A based on absorption measurements. It is particularly useful since it is independent of the type of transition involved and is valid not only for electric dipole allowed transitions, but also higher multipoles. We will show in what follows that $\delta\omega \sim A$ so that $\int d\omega \sigma(\omega) \sim \delta\omega\sigma(\omega_{21}) \sim A\sigma(\omega_{21})$ and $\sigma(\omega_{21}) \sim \lambda_{21}^2$. The resonant photon absorption cross section of an atom is order λ^2 which can be orders of magnitude larger than the geometrical size of the atom.

8.3 Resonant absorption due to narrowband radiation

In experiments with lasers the radiation often has a frequency spectrum that is narrow compared to the width of the atomic resonance. Let's calculate the absorption rate per atom in this case. The situation is shown in Fig. 8.3. Radiation with field amplitude $E(\omega)$ is absorbed by an atom in lower level $|1\rangle$. After excitation to the upper level $|2\rangle$ the atom decays and emits a photon. The lifetime of the upper level is $\tau = 1/A = 1/\gamma$, where we will use the symbol $\gamma = A$ to denote the linewidth of the transition. The energy-time uncertainty condition $\Delta U \Delta t \sim \hbar$ then implies $\Delta U \sim \hbar/\tau \sim \hbar\gamma$.

In order to make the expression for the energy uncertainty of the excited state precise we proceed as follows. When the atom is in the upper level it decays at rate γ . Another way of saying this is that the rate of energy loss satisfies $\frac{dU_2}{dt} = -\gamma U_2$ which has the solution $U_2(t) = U_2(0)e^{-\gamma t}$. From a quantum mechanical perspective the state of the atom is $|\psi(t)\rangle = c_1(t)|1\rangle + c_2(t)e^{-i\omega_{21}t}|2\rangle$ with $\omega_{21} = (U_2 - U_1)/\hbar$. The energy in the excited state is proportional to the probability of occupying the state which is proportional to $|c_2(t)|^2$. The relation $U_2(t) = U_2(0)e^{-\gamma t}$ then implies that

$$c_2(t)e^{-i\omega_{21}t} = c_2(0)e^{-\gamma t/2}e^{-i\omega_{21}t}.$$

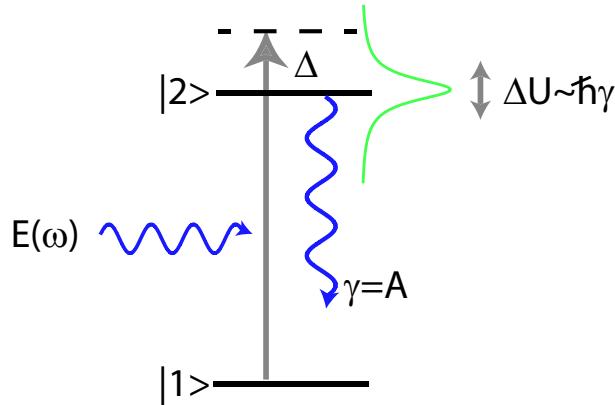


Figure 8.3: Scattering of an electric field $E(\omega)$ from a two-level atom with excited state spontaneous decay rate γ .

We can Fourier decompose $c_2(t)$ as

$$c_2(\omega) \sim \int_0^\infty dt c_2(0)e^{-\gamma t/2} e^{-i\omega_{21}t} e^{i\omega t} = \frac{1}{\frac{\gamma}{2} - i(\omega - \omega_{21})}. \quad (8.7)$$

The integral starts at $t = 0$ since we assume there is no population in the excited level at negative times. The magnitude squared of the coefficient is $|c_2(\omega)|^2 = C \frac{1}{\frac{\gamma^2}{4} + \Delta^2}$ with $\Delta = \omega - \omega_{21}$ and C an as yet unknown normalization coefficient. To fix C we require that

$$1 = \int_{-\infty}^\infty d\omega |c_2(\omega)|^2 = C \int_{-\infty}^\infty d\omega \frac{1}{\frac{\gamma^2}{4} + \Delta^2} = C \frac{2\pi}{\gamma}.$$

Thus we get the normalized lineshape function

$$L(\Delta) = |c_2(\omega)|^2 = \frac{\gamma}{2\pi} \frac{1}{\frac{\gamma^2}{4} + \Delta^2} = \frac{2}{\pi\gamma} \frac{1}{1 + \frac{4\Delta^2}{\gamma^2}}. \quad (8.8)$$

The calculation was formulated in terms of a decay process but we can equally well interpret the probability $P_2 = |c_2(\omega)|^2$ as the likelihood of a photon being absorbed, and the atom thereby excited, as a function of the frequency ω of the radiation. We see that the absorption lineshape $L(\Delta)$ is Lorentzian and falls off as $1/\Delta^2$. The full width at half maximum (FWHM) of the lineshape is $\delta\omega_{\text{FWHM}} = \gamma$ so when the radiation is detuned from resonance by $\pm\gamma/2$ the probability of absorption falls to 1/2 of the peak value.

When the radiation is detuned from the atomic transition frequency, $\Delta = \omega - \omega_{21} \neq 0$. For small detunings $|\Delta| < \sim \gamma$ we may visualize the interaction as a photon of frequency ω being absorbed and the atom excited to state $|2\rangle$. When the atom subsequently decays a photon of energy $\hbar\omega_{\text{sp}} \simeq \hbar\omega_{21}$ is spontaneously emitted. The absorbed and emitted photons need not have exactly the same energy. The difference in energy $\hbar(\omega_{\text{sp}} - \omega)$ is taken up by a change in the center of mass energy of the atom after the absorption-emission event. This effect can be used for laser cooling of atomic motion[73].

When the incident radiation is very far detuned this picture breaks down and we may think of the interaction as a scattering process where the incident and scattered photons have the same frequency. In the two-level approximation the atom ends up in the same internal state after the interaction so this is referred to as Rayleigh scattering, as opposed to Raman scattering which leaves the atom in a different internal state. Note that even in the case of Rayleigh scattering the center of mass atomic energy may change due to the different momenta $\hbar\mathbf{k}$ and $\hbar\mathbf{k}_{\text{sp}}$ of the incident and scattered photons.

The frequency dependence of the scattering cross section is the same as the frequency dependence of the lineshape function. Therefore $\sigma(\Delta)/\sigma(0) = L(\Delta)/L(0)$ so

$$\sigma(\Delta) = \sigma(0) \frac{L(\Delta)}{L(0)} = \sigma(0) \frac{1}{1 + \frac{4\Delta^2}{\gamma^2}}$$

where $\sigma(0)$ is the resonant cross section. We can interpret the effective width of the scattering cross section as the inverse of the resonant lineshape function or $\delta\omega = 1/L(0) = \pi\gamma/2$. This justifies the statement made in the previous section that $\delta\omega \sim A$.

More explicitly we note that the total cross section is

$$\sigma_{\text{total}} = \int_{-\infty}^\infty d\Delta \sigma(\Delta) = \sigma(0) \int_{-\infty}^\infty d\Delta \frac{1}{1 + \frac{4\Delta^2}{\gamma^2}} = \sigma(0) \frac{\pi\gamma}{2}.$$

Comparing with Eq. (8.6), and remembering that $\gamma = A$, we find

$$\sigma(0) = \frac{3}{2\pi} \lambda_{21}^2. \quad (8.9)$$

The derivations used above are heuristic so it is fair to wonder whether or not the resulting lineshape expression (8.8) for the detuning dependent transition rate and cross section is accurate. The justification for this expression comes from solving the problem rigorously using quantum mechanics and density matrix theory.

8.3.1 Saturation effects

When the light intensity is sufficiently strong population is redistributed between ground and excited states. This changes the number of atoms that can absorb an incident photon and therefore the attenuation of a weak beam is larger than that of a strong beam. This effect is known as saturated absorption.

We can readily find the steady-state ground and excited state populations using the rate Eqs. (8.1). In steady state

$$N_2 = \frac{B_{2 \leftarrow 1}^{\text{abs}} \rho_\omega}{A + B_{1 \leftarrow 2}^{\text{ems}} \rho_\omega} N_1.$$

Assuming $g_1 = g_2$ so $B_{1 \leftarrow 2}^{\text{ems}} = B_{2 \leftarrow 1}^{\text{abs}}$ we find

$$\frac{N_2}{N} = \frac{N_2}{N_1 + N_2} = \frac{1}{2} \frac{2B_{2 \leftarrow 1}^{\text{abs}} \rho_\omega / A}{1 + 2B_{2 \leftarrow 1}^{\text{abs}} \rho_\omega / A}.$$

Define a saturation parameter $s = 2B_{2 \leftarrow 1}^{\text{abs}} \rho_\omega / A$ so that

$$\frac{N_2}{N} = \frac{s/2}{1+s}$$

and

$$\frac{N_1}{N} = \frac{1+s/2}{1+s}.$$

We see that for $s = 0$ there is no excited state population ($N_2 = 0$) and as $s \rightarrow \infty$ then $N_2 \rightarrow 1/2$.

A useful quantity is the difference in population of excited and ground states defined by $\Delta N = N_2 - N_1$. This is found to be

$$\Delta N = -N \frac{1}{1+s}.$$

At unit saturation ($s = 1$) we have $N_2/N = 1/4$, $N_1/N = 3/4$ and $\Delta N/N = -1/2$. So unit saturation corresponds to half of the maximum possible excited state population.

The saturation parameter s is also a function of the detuning. It can be written as $s = I/I_{s\Delta}$ where I is the optical intensity and $I_{s\Delta}$ is the saturation intensity at detuning Δ . Since the light-matter interaction gets weaker at finite detuning according to the lineshape function $L(\Delta)$, the saturation intensity must increase inversely proportional to $L(\Delta)$. We can write this as

$$I_{s\Delta} = I_s \frac{L(0)}{L(\Delta)} = I_s \left(1 + \frac{4\Delta^2}{\gamma^2} \right) \quad (8.10)$$

where I_s is the resonant saturation intensity. The intensity dependence of the populations is thus

$$N_1 = N \frac{1 + \frac{I}{2I_{s\Delta}}}{1 + \frac{I}{I_{s\Delta}}} \quad (8.11a)$$

$$N_2 = N \frac{\frac{I}{2I_{s\Delta}}}{1 + \frac{I}{I_{s\Delta}}} \quad (8.11b)$$

$$\Delta N = N_2 - N_1 = -N \frac{1}{1 + \frac{I}{I_{s\Delta}}}. \quad (8.11c)$$

We can express I_s in terms of the linewidth $\gamma = A$ as follows. The intensity per unit frequency I_ω is related to the energy density by $c\rho_\omega = I_{\omega x} + I_{\omega y} + I_{\omega z} = 3I_\omega$ with the intensity along a single direction equal to $I_\omega = c\rho_\omega/3$. We can therefore write $\rho_\omega = 3I_\omega/c$ so that on resonance

$$s_0 = \frac{I}{I_s} = \frac{2B_{2\leftarrow 1}^{\text{abs}}\rho_\omega}{A} = \frac{6\pi^2 c^2}{\hbar\omega_{21}^3} I_\omega.$$

The effective intensity per unit frequency at resonance is $I_\omega = IL(0) = 2I/\pi\gamma$. We therefore arrive at $s_0 = \frac{12\pi c^2}{\hbar\gamma\omega_{21}^3} I$ and using $I_s = I/s_0$ the resonant saturation intensity is

$$I_s = \frac{\hbar\gamma\omega_{21}^3}{12\pi c^2}.$$

It is not hard to show that with degeneracy factors g_1, g_2 this becomes

$$I_s = \frac{g_1}{g_2} \frac{\hbar\gamma\omega_{21}^3}{12\pi c^2} \quad (8.12)$$

and the frequency dependent saturation parameter is

$$s = \frac{g_2}{g_1} \frac{12\pi c^2}{\hbar\gamma\omega_{21}^3} \frac{I}{1 + 4\Delta^2/\gamma^2}.$$

It is convenient to introduce the parameter $\sigma_0 \equiv \sigma(0) = \frac{g_2}{g_1} \frac{3\lambda_{21}^2}{2\pi}$ which we will show below to be the resonant absorption cross section. The saturation parameter can then be expressed as

$$s = \frac{2\sigma_0}{\gamma} \frac{I/(\hbar\omega_{21})}{1 + 4\Delta^2/\gamma^2}.$$

8.3.2 Scattering cross section and absorption

The cross section for absorption be written in terms of a scattering cross section which describes the effective ‘‘scattering’’ area of a single atom. The scattering rate is equal to the rate of spontaneous decay from the excited state. This is given by

$$r_N = \gamma N_2 = \gamma N \frac{\frac{I}{2I_{s\Delta}}}{1 + \frac{I}{I_{s\Delta}}}.$$

The spontaneous scattering rate per atom is thus

$$r = \frac{r_N}{N} = \gamma \frac{\frac{I}{2I_{s\Delta}}}{1 + \frac{I}{I_{s\Delta}}} = \frac{\gamma}{2} \frac{\frac{I}{I_s}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \quad (8.13)$$

Writing this as $r = \frac{I}{\hbar\omega}\sigma$ with σ the scattering cross section we find

$$\sigma = \sigma_0 \frac{\omega/\omega_{21}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}} \quad (8.14)$$

with the low intensity resonant cross section given by

$$\sigma_0 = \frac{\hbar\gamma\omega_{21}}{2I_s}.$$

Using the expression for the saturation intensity from Eq. (8.12), we find

$$\sigma_0 = \frac{g_2}{g_1} \frac{3}{2\pi} \lambda_{21}^2. \quad (8.15)$$

This agrees with the absorption cross section in Eq. (8.9) if we set $g_1 = g_2$.

The rate of scattering is equal to the rate at which energy is lost by a propagating optical beam since the spontaneous scattering occurs in random directions. Note that as $I \rightarrow \infty$ the spontaneous scattering rate tends to a constant $r_{\max} = \gamma/2$. Although the rate of stimulated transitions does not saturate, stimulated transitions do not deplete the propagating beam, and do not contribute to the scattering loss. It is important to note the different intensity dependencies of the scattering rate and the absorption cross section. At high intensity the scattering rate tends to a constant while the absorption cross section tends to zero. The explanation is that at high intensity transitions between the atomic levels are dominated by stimulated processes which do not saturate. Thus the scattering rate per incident photon, which gives the scattering cross section, tends to zero.

If the scattering rate is $\frac{I}{\hbar\omega}\sigma$ per atom then in a unit volume we have a scattering rate of $\frac{I}{\hbar\omega}\sigma N dz$. The differential change in intensity due to propagation over a distance dz is thus

$$dI = -\hbar\omega \frac{I}{\hbar\omega} \sigma N dz$$

or

$$\frac{dI}{dz} = -\sigma N I = -\alpha I$$

where the absorption coefficient is

$$\alpha = \sigma N = \sigma_0 N \frac{\omega/\omega_{21}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \quad (8.16)$$

The intensity decreases exponentially with z as

$$I(z) = I(0)e^{-\alpha z} \quad (8.17)$$

which is often referred to as Beer's law. We see from (8.16) that for all intensities $\alpha > 0$. The light always experiences loss. For this reason two-level atoms are not suitable for achieving optical gain and laser action.

8.4 Refractive index

The refractive index of a gas of two-level atoms can be calculated in analogy to the calculation of the scattering rate in Sec. 8.3. Recall from Eq. (8.7) that the amplitude of the upper level is

$$c_2 \sim \frac{1}{\gamma/2 - i\Delta} = \frac{\gamma/2 + i\Delta}{\gamma^2/4 + \Delta^2}.$$

Note that $\text{Im}(c_2) = \frac{2\Delta}{\gamma} \text{Re}(c_2)$. Keeping only the real part of c_2 we arrived at $\frac{dI}{dz} = -\sigma NI$. Since $I \sim |\mathcal{E}|^2$ with \mathcal{E} the complex field amplitude this is equivalent to

$$\text{Re} \left[\frac{d\mathcal{E}}{dz} \right] = \text{Re} \left[\frac{-\sigma N}{2} \mathcal{E} \right].$$

Since σN is real we can as well write this as

$$\begin{aligned} \frac{d\mathcal{E}}{dz} &= \frac{-\sigma N}{2} \mathcal{E} \\ &= \frac{-\sigma_0 N}{2} \frac{\omega}{\omega_{21}} \frac{1}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}} \mathcal{E}. \end{aligned} \quad (8.18)$$

If we were to repeat the derivation leading to (8.18), but include the imaginary part of c_2 we would arrive at

$$\frac{d\mathcal{E}}{dz} = \frac{-\sigma_0 N}{2} \frac{\omega}{\omega_{21}} \frac{1 + i \frac{2\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}} \mathcal{E}. \quad (8.19)$$

Writing the complex field amplitude in terms of two real variables F, ϕ as $\mathcal{E} = F e^{i\phi}$ we get

$$\begin{aligned} \frac{dF}{dz} &= \frac{-\sigma_0 N}{2} \frac{\omega}{\omega_{21}} \frac{1}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}} F, \\ \frac{d\phi}{dz} &= \frac{-\sigma_0 N}{2} \frac{\omega}{\omega_{21}} \frac{\frac{2\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \end{aligned}$$

Since the phase change per unit length is $d\phi/dz$ which we can set equal to $k\delta n$ the refractive index change due to the atoms is

$$\delta n = \frac{1}{k} \frac{d\phi}{dz} = \frac{-\sigma_0 N}{k} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \quad (8.20)$$

This expression is valid for a not too dense sample of atoms. At high density, as in a solid, the field is given by the incident field plus the field due to the polarization of the medium.

Let's plot the refractive index of Eq. (8.20) as a function of detuning Δ . As we see in Fig. 8.4a) for $\Delta > 0 (< 0)$, $\delta n < 0 (> 0)$. At low intensity, $I/I_s \rightarrow 0$, the maximum change in the index occurs at $|\Delta| = \pm\gamma/2$. There are a few interesting features of the refractive index near a resonance. For $|\Delta| < \gamma/2$ the slope $d\delta n/d\Delta < 0$. This means that the index decreases as the frequency of the light increases. This is opposite to the normal behavior in transparent materials and is referred to as anomalous dispersion. We also see that for sufficiently high density the refractive index can become negative for positive detuning. The condition for a negative index with low intensity light is

$$\frac{-\sigma_0 N}{k} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2}} = -1$$

or

$$N = \frac{k}{\sigma_0} \frac{1 + \frac{4\Delta^2}{\gamma^2}}{\frac{\Delta}{\gamma}}.$$

Putting $\Delta/\gamma = 1/2$ and using parameters for Cs from Fig. 8.4 we find $N = 8.5 \times 10^{19} \text{ m}^{-3}$.

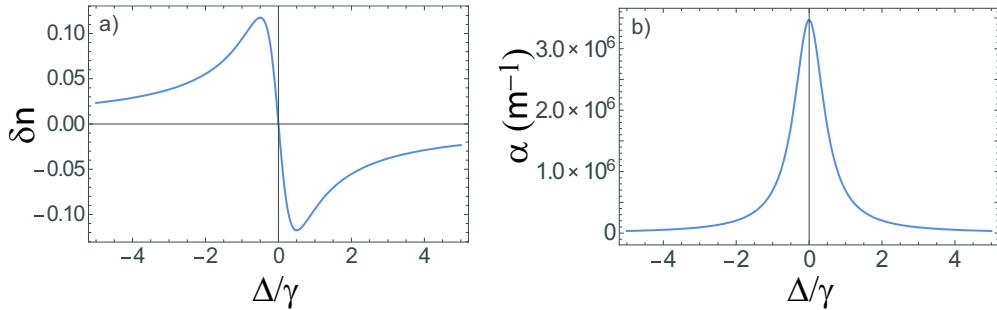


Figure 8.4: a) Change in refractive index from Eq. (8.20) for parameters corresponding to a low density gas of Cs atoms on the D2 transition: $N = 10^{19} \text{ m}^{-3}$, $\lambda = 0.852 \mu\text{m}$, $k = 2\pi/\lambda$, $\sigma_0 = \frac{3}{2\pi}\lambda^2$. The detuning is $\Delta = \omega - \omega_{21}$ and with the assumption that $|\Delta| \ll \omega_{21}$ we put $\omega/\omega_{21} = 1$. b) Absorption coefficient from Eq. (8.16) for the same parameters as in a).

This suggests that it should be possible to create a negative refractive index simply by tuning close to resonance in a moderately dense gas of atoms³. Unfortunately, while Eq. (8.20) is correct we need to keep in mind that there is also very strong absorption near resonance as shown in Fig. 8.4b). At $\Delta/\gamma = 1/2$ the absorption coefficient at the density needed for a negative index implies an energy loss of 99.999% in a distance of a single wavelength. A two-level atom does not provide a useful approach for creating a negative refractive index. More complicated methods based on atoms with multiple levels may be useful since it is possible to take advantage of quantum interference effects to create a large index change with small absorption[74].

8.5 Summary

We have analyzed the interaction of light with a two-level model of matter using the Einstein rate equations. Accounting for the effects of detuning and saturation at finite intensity results in several useful formulae. For convenience we summarize the main results here.

Absorption area law:

$$\int d\omega \sigma(\omega) = \frac{3\lambda_{21}^2}{4} A. \quad (8.6)$$

Normalized, low intensity lineshape:

$$L(\Delta) = \frac{2}{\pi\gamma} \frac{1}{1 + \frac{4\Delta^2}{\gamma^2}}. \quad (8.8)$$

³There is a subtlety in that the condition given here on N is the condition for ϵ to be negative, corresponding to an imaginary index. For true negative refraction it is also necessary that μ be negative which requires modifying the magnetic properties of the medium. This is more difficult since magnetic transitions in real atoms are much weaker than electric transitions.

Population of lower and upper levels:

$$N_1 = N \frac{1 + \frac{I}{2I_{s\Delta}}}{1 + \frac{I}{I_{s\Delta}}}, \quad (8.11a)$$

$$N_2 = N \frac{\frac{I}{2I_{s\Delta}}}{1 + \frac{I}{I_{s\Delta}}}, \quad (8.11b)$$

$$\Delta N = N_2 - N_1 = -N \frac{1}{1 + \frac{I}{I_{s\Delta}}}. \quad (8.11c)$$

Saturation intensity:

$$I_{s\Delta} = I_s \left(1 + \frac{4\Delta^2}{\gamma^2} \right), \quad I_s = \frac{g_1}{g_2} \frac{\hbar\gamma\omega_{21}^3}{12\pi c^2}. \quad (8.10, 8.12)$$

Spontaneous scattering rate per absorber:

$$r = \frac{\gamma}{2} \frac{\frac{I}{I_s}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \quad (8.13)$$

Absorption cross section:

$$\sigma = \sigma_0 \frac{\omega/\omega_{21}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}, \quad \sigma_0 = \frac{g_2}{g_1} \frac{3}{2\pi} \lambda_{21}^2. \quad (8.14, 8.15)$$

Beer's law:

$$I(z) = I(0)e^{-\alpha z}, \quad \alpha = \sigma N. \quad (8.17)$$

Refractive index:

$$\delta n = \frac{-\sigma_0 N}{k} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \quad (8.20)$$

8.6 Microscopic model of the refractive index

The expression for the refractive index found in Sec. 8.4 is based on phenomenological analysis of a two-level absorber interacting with the radiation field. In this section we give an alternative derivation based on the fields produced by accelerated charges as predicted by the Maxwell equations. The derivation has two main parts. First we need to establish the field radiated by a moving charge. We will then use this result to find the refractive index of a material that depends on the density of bound, oscillating charges.

8.6.1 Fields from accelerated charges

To start the calculation recall the usual expression for the electric field at position \mathbf{r} due to a stationary source charge q at position \mathbf{r}_q

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \frac{\mathbf{n}}{|\mathbf{r} - \mathbf{r}_q|^2}.$$

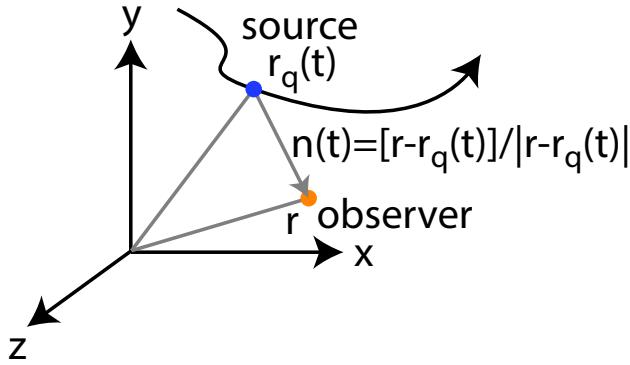


Figure 8.5: Geometry for calculation of field from moving source charge.

Here $\mathbf{n} = \frac{\mathbf{r} - \mathbf{r}_q}{|\mathbf{r} - \mathbf{r}_q|}$ is a unit vector pointing from the source to the observation point as shown in Fig. 8.5. This expression is valid for a stationary charge. For a slowly moving charge with velocity $\mathbf{v}_q = \frac{d\mathbf{r}_q}{dt}$ and $|\mathbf{v}_q| \ll c$ we can use

$$\mathbf{E}(\mathbf{r}) \simeq \frac{q}{4\pi\epsilon_0} \frac{\mathbf{n}}{|\mathbf{r} - \mathbf{r}_q(t)|^2}.$$

If the source motion is relativistic we need to account for the finite propagation time from the source to the observation point since the source position at the time of observation will be different than the position when the observed field was created. To account for relativistic effects we will first calculate the electromagnetic potentials due to the source, and then differentiate to find the fields. In general there will be both an electric and a magnetic field.

The scalar and vector potentials at \mathbf{r}, t due to sources at \mathbf{r}' are

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int d^3r' \frac{\rho(\mathbf{r}', t_{\text{ret}})}{|\mathbf{r} - \mathbf{r}'(t_{\text{ret}})|} \quad (8.21a)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int d^3r' \frac{\mathbf{j}(\mathbf{r}', t_{\text{ret}})}{|\mathbf{r} - \mathbf{r}'(t_{\text{ret}})|}. \quad (8.21b)$$

Here ρ is the charge density and \mathbf{j} is the current density. The sources are to be evaluated at the retarded time defined by

$$t_{\text{ret}} = t - \frac{|\mathbf{r} - \mathbf{r}'(t_{\text{ret}})|}{c}. \quad (8.22)$$

This definition ensures that radiation emitted by the source at position \mathbf{r}' at time t_{ret} reaches the observation point \mathbf{r} at time t . Denoting the position and velocity of a point source with charge q as $\mathbf{r}_q(t), \mathbf{v}_q(t) = \frac{d\mathbf{r}_q(t)}{dt}$, gives charge and current densities

$$\rho(\mathbf{r}, t) = q\delta[\mathbf{r} - \mathbf{r}_q(t)], \quad (8.23a)$$

$$\mathbf{j}(\mathbf{r}, t) = q\mathbf{v}_q(t)\delta[\mathbf{r} - \mathbf{r}_q(t)]. \quad (8.23b)$$

The fields are then calculated from the potentials using

$$\mathbf{E}(\mathbf{r}, t) = -\nabla\phi(\mathbf{r}, t) - \frac{\partial\mathbf{A}(\mathbf{r}, t)}{\partial t}, \quad (8.24a)$$

$$\mathbf{B}(\mathbf{r}, t) = \nabla \times \mathbf{A}(\mathbf{r}, t). \quad (8.24b)$$

With definitions (8.21-8.24) the problem of calculating the fields due to charges in arbitrary motion is reduced to an exercise in mathematics. Although in principle straightforward, evaluating the integrals and derivatives correctly requires some care, as we proceed to show. The first step is to evaluate the potentials using the source definitions (8.23) evaluated at the retarded time t_{ret} . To enforce the correct source time we can write the equation for the scalar potential as

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int d^3 r' \int dt' \frac{\rho(\mathbf{r}', t')}{|\mathbf{r} - \mathbf{r}'(t')|} \delta(t' - t_{\text{ret}}).$$

The additional temporal integration and δ function ensure that the source is evaluated at t_{ret} . Inserting (8.23a) and performing the spatial integral we find

$$\phi(\mathbf{r}, t) = \frac{q}{4\pi\epsilon_0} \int dt' \frac{\delta(t' - t_{\text{ret}})}{|\mathbf{r} - \mathbf{r}_q(t')|}. \quad (8.25)$$

The integral has to be done carefully due to the implicit dependence on t' in the factor $\delta(t' - t_{\text{ret}}) = \delta(t' - t + |\mathbf{r} - \mathbf{r}_q(t')|/c)$. Use the identity $\delta[f(t)] = \sum_n \frac{1}{|df(t)/dt|_{t=t_n}} \delta(t - t_n)$, where t_n are the roots of $f(t)$. The solution of $t' - t + |\mathbf{r} - \mathbf{r}_q(t')|/c = 0$ is $t' = t_{\text{ret}}$ as defined in (8.22) with \mathbf{r}' replaced by \mathbf{r}_q . Then use

$$\begin{aligned} \frac{d}{dt'} \left[t' - t + \frac{|\mathbf{r} - \mathbf{r}_q(t')|}{c} \right] &= 1 + \frac{1}{c} \frac{d}{dt'} |\mathbf{r} - \mathbf{r}_q(t')| \\ &= 1 + \frac{1}{c} \frac{d}{dt'} [r^2 - 2\mathbf{r} \cdot \mathbf{r}_q(t') + \mathbf{r}_q(t') \cdot \mathbf{r}_q(t')]^{1/2} \\ &= 1 + \frac{1}{2c|\mathbf{r} - \mathbf{r}_q(t')|} \frac{d}{dt'} [-2\mathbf{r} \cdot \mathbf{r}_q(t') + \mathbf{r}_q(t') \cdot \mathbf{r}_q(t')] \\ &= 1 + \frac{1}{2c|\mathbf{r} - \mathbf{r}_q(t')|} [-2\mathbf{r} \cdot \mathbf{v}_q(t') + 2\mathbf{r}_q(t') \cdot \mathbf{v}_q(t')] \\ &= 1 - \frac{1}{c|\mathbf{r} - \mathbf{r}_q(t')|} [\mathbf{r} - \mathbf{r}_q(t')] \cdot \mathbf{v}_q(t') \\ &= 1 - \mathbf{n}(t') \cdot \boldsymbol{\beta}_q(t') \end{aligned}$$

where we have introduced $\boldsymbol{\beta}_q(t') = \mathbf{v}_q(t')/c$. Using this in (8.25) we find

$$\phi(\mathbf{r}, t) = \frac{q}{4\pi\epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{r}_q(t_{\text{ret}})|} \frac{1}{1 - \mathbf{n}(t_{\text{ret}}) \cdot \boldsymbol{\beta}_q(t_{\text{ret}})}$$

where $t_{\text{ret}} = t - \frac{|\mathbf{r} - \mathbf{r}_q(t_{\text{ret}})|}{c}$. We can simplify the notation by defining the vector $\mathbf{R}(t) = \mathbf{r} - \mathbf{r}_q(t) = R \mathbf{n}$ and write

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \left[\frac{q}{R - \mathbf{R} \cdot \boldsymbol{\beta}_q} \right]_{\text{ret}} \quad (8.26)$$

where $[...]$ _{ret} means evaluate the quantity inside the brackets at $t = t_{\text{ret}}$. An analogous calculation results in

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \left[\frac{q\mathbf{v}_q}{R - \mathbf{R} \cdot \boldsymbol{\beta}_q} \right]_{\text{ret}}. \quad (8.27)$$

Note that if we set $\mathbf{v}_q = 0$ for a stationary charge we recover $\phi = q/4\pi\epsilon_0 R$ and $\mathbf{A} = 0$. Equations (8.26, 8.27) are known as the Liénard-Wiechert potentials[75, 76].

Differentiating to get the fields still requires a long calculation due to the complicated dependence of the retarded quantities on \mathbf{r} and t . The calculation is done using the chain rule for differentiating together with the expressions

$$\begin{aligned}\frac{\partial t_{\text{ret}}}{\partial t} &= \frac{1}{[1 - \boldsymbol{\beta}_q \cdot \mathbf{n}]_{\text{ret}}}, \\ \nabla t_{\text{ret}} &= \left[\frac{-\mathbf{n}}{cR(1 - \boldsymbol{\beta}_q \cdot \mathbf{n})} \right]_{\text{ret}}.\end{aligned}$$

The result for the fields is

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0} \left[\frac{(\mathbf{n} - \boldsymbol{\beta}_q)(1 - \beta_q^2)}{(1 - \mathbf{n} \cdot \boldsymbol{\beta}_q)^3 R^2} + \frac{\mathbf{n} \times \left[(\mathbf{n} - \boldsymbol{\beta}_q) \times \frac{d\boldsymbol{\beta}_q}{dt} \right]}{c(1 - \mathbf{n} \cdot \boldsymbol{\beta}_q)^3 R} \right]_{\text{ret}}, \quad (8.28a)$$

$$\mathbf{B} = \frac{1}{c} [\mathbf{n} \times \mathbf{E}]_{\text{ret}}. \quad (8.28b)$$

These expressions can be put into a simpler form, originally due to Heaviside[77], and rediscovered by Feynman[78]

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0} \left[\frac{\mathbf{n}}{R^2} + \frac{R}{c} \frac{d}{dt} \frac{\mathbf{n}}{R^2} + \frac{1}{c^2} \frac{d^2}{dt^2} \mathbf{n} \right]_{\text{ret}}, \quad (8.29a)$$

$$\mathbf{B} = \frac{1}{c} [\mathbf{n} \times \mathbf{E}]_{\text{ret}}. \quad (8.29b)$$

We see that the electric field is the sum of three contributions. The normal Coulomb field evaluated at the retarded position of the atom. A correction $\sim \mathbf{v}_q/c$ that increases(decreases) the field when the charge moves towards(away from) the observer, and a term proportional to the acceleration of the charge. The first two terms have $E \sim 1/R^2$ so the Poynting vector $\sim 1/R^4$ and is negligible in the far field where R is large. It is the last term which is independent of R that is responsible for far field radiation. This is the explanation for the statement that radiation is due to accelerating charges.

We conclude that far from the source the radiated fields are

$$\mathbf{E}_{\text{rad}} = \frac{q}{4\pi\epsilon_0 c^2} \left[\frac{d^2 \mathbf{n}}{dt^2} \right]_{\text{ret}}, \quad (8.30a)$$

$$\mathbf{B}_{\text{rad}} = \frac{q}{4\pi\epsilon_0 c^3} \left[\mathbf{n} \times \frac{d^2 \mathbf{n}}{dt^2} \right]_{\text{ret}}. \quad (8.30b)$$

If the motion of the charge is known we can evaluate the fields. An important case is that of a charge that oscillates sinusoidally with position

$$z(t) = z_0 \cos(\omega t).$$

Let the observation point be a distance R from the origin at $\mathbf{r} = R(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))$ as shown in Fig. 8.6. Evaluating Eqs. (8.30) and keeping only the leading terms for $R \gg z_0$ we find

$$\mathbf{E}_{\text{rad}} = \frac{q[-\omega^2 z_0 \cos(\omega t)]_{\text{ret}}}{4\pi\epsilon_0 c^2} \frac{1}{R} \sin(\theta) [\cos(\phi) \cos(\theta) \hat{x} + \sin(\phi) \cos(\theta) \hat{y} - \sin(\theta) \hat{z}], \quad (8.31a)$$

$$\mathbf{B}_{\text{rad}} = \frac{q[-\omega^2 z_0 \cos(\omega t)]_{\text{ret}}}{4\pi\epsilon_0 c^3} \frac{1}{R} \sin(\theta) [-\sin(\phi) \hat{x} + \cos(\phi) \hat{y}]. \quad (8.31b)$$

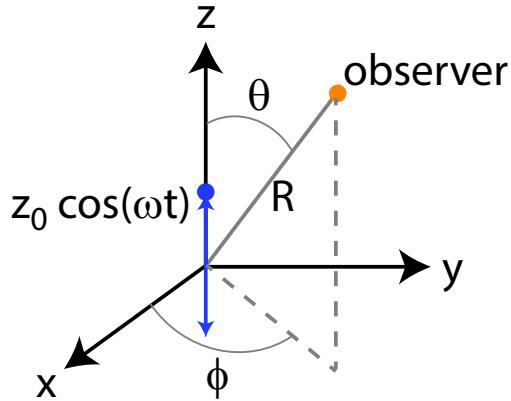


Figure 8.6: Radiation from an accelerated charge.

The factor in the square brackets is the instantaneous acceleration of the source $a = -\omega^2 z(t)$, evaluated at the retarded time. For sinusoidal motion evaluation at the retarded time only amounts to a phase shift of the field, and we will drop the $[...]\text{ret}$ notation for the rest of this section. The fields have an overall $\sin(\theta)$ dependence and therefore vanish for an observer on the z axis. The fields are everywhere linearly polarized. Motion of a charge in a circular orbit produces more general states of polarization.

The corresponding Poynting vector in the far field is

$$\mathbf{S}_{\text{rad}} = \frac{1}{\mu_0} \mathbf{E}_{\text{rad}} \times \mathbf{B}_{\text{rad}}.$$

Evaluating this expression and keeping only the leading terms for $R \gg z_0$ we find

$$\mathbf{S}_{\text{rad}} = \frac{q^2 a^2}{16\pi^2 \epsilon_0 c^3} \frac{1}{R^2} \sin^2(\theta) [\cos(\phi) \sin(\theta) \hat{x} + \sin(\phi) \sin(\theta) \hat{y} + \cos(\theta) \hat{z}]. \quad (8.32)$$

Integrating the outward going Poynting flux times the unit normal vector $\mathbf{e}_r = \sin(\theta) \cos(\phi) \hat{x} + \sin(\theta) \sin(\phi) \hat{y} + \cos(\theta) \hat{z}$ over a bounding sphere of radius R we find the total radiated power

$$P = \frac{q^2 a^2}{6\pi \epsilon_0 c^3}. \quad (8.33)$$

Equation (8.33) for the total radiated power from an accelerated charge was first derived by Larmor[79].

8.6.2 The radiated field from a collection of bound charges

We now introduce a model for the field radiated by many charges, each of which is bound to a fixed position in space. This is a “Lorentz” oscillator model of atoms where we replace each atom by a charge that is harmonically bound and oscillates with natural frequency ω_0 . The natural frequency corresponds to the transition frequency of Einstein’s rate equation model of a two-level absorber.

The classical equation of motion of a bound charge in the presence of a driving field $E \cos(\omega t)$ is

$$m \frac{d^2 x}{dt^2} - m\gamma_b \frac{dx}{dt} = -\kappa x + qE \cos(\omega t) = -\kappa x + q \frac{\mathcal{E}}{2} (e^{-\omega t} + e^{\omega t}). \quad (8.34)$$

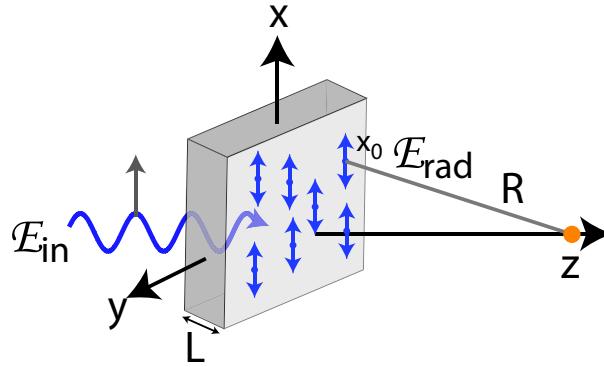


Figure 8.7: Plate with bound charges of density N and surface density $\eta = LN$. The charges are driven by the incident field E_{in} .

We are assuming one dimensional motion along x of a charge q with mass m driven by a field with complex amplitude \mathcal{E} , polarized along \hat{x} . The damping constant γ_b with units of [1/s] accounts for loss of motional energy due to the power radiated by the accelerated charge. The natural frequency is $\omega_0 = \sqrt{\kappa/m}$ where κ is an effective spring constant. Solving for the steady state response $x = \frac{x_0}{2} (e^{-i\omega t} + e^{i\omega t})$ we find

$$x_0 = -\frac{\mathcal{E}q/m}{(\omega^2 - \omega_0^2) - i\gamma_b\omega}.$$

We will be interested in the situation where ω is close to ω_0 and therefore put $\omega - \omega_0 = \Delta$ with $|\Delta| \ll \omega_0$. It follows that

$$\omega^2 - \omega_0^2 = (\omega + \omega_0)(\omega - \omega_0) \simeq 2\omega_0\Delta$$

so

$$x_0 \simeq -\frac{\mathcal{E}q}{m\omega_0} \frac{1}{2\Delta - i\gamma_b} = -\frac{\mathcal{E}q}{m\omega_0} \frac{2\Delta/\gamma_b^2 + i/\gamma_b}{1 + 4\Delta^2/\gamma_b^2}. \quad (8.35)$$

Note that the real part of this last expression is reminiscent of the frequency dependent index of refraction from Eq. (8.20).

To proceed consider a plate of material with thickness L and density of bound charges N as in Fig. 8.7. We will use this model to calculate the phase delay of the field observed at z and compare that with the phase delay due to a plate of refractive index n . Comparing the two expressions will result in a formula for the refractive index in terms of the properties of the bound charges. This approach follows the treatment in the Feynman lectures Vol. I, Ch. 31 [80].

Let a field \mathcal{E}_{in} be incident on the plate from the left as shown in Fig. 8.7. The field seen by each charge is the sum of the incident field and the fields radiated by all the other bound charges. If the radiated fields are weak compared to the incident field then the motion of each charge will be approximately that due to the incident field alone. The transmitted field at a distance z after the plate will be

$$\mathcal{E}_{\text{tr}}(z) = \mathcal{E}_{\text{in}}(z) + \mathcal{E}_{\text{rad}}(z) \quad (8.36)$$

where $\mathcal{E}_{\text{in}}(z) = \mathcal{E}_{\text{in}}(0)e^{i(k_0 z - \omega t)}$, $k_0 = \omega/c$, and \mathcal{E}_{rad} is the radiation field from all the charges in the

plate. Using (8.31a) the electric field radiated by a bound charge at position (x, y) in the plate is

$$\begin{aligned}\mathcal{E}_{\text{rad}}(z) &= \frac{q[-\omega^2 x_0 e^{-i\omega t}]_{\text{ret}}}{4\pi\epsilon_0 c^2} \frac{1}{R} \\ &\times \left[-\frac{1}{1+(x/z)^2} \hat{x} - \frac{xy/z^2}{(1+(x/z)^2)\sqrt{1+(y/z)^2}} \hat{y} - \frac{x/z}{(1+(x/z)^2)\sqrt{1+(y/z)^2}} \hat{z} \right]\end{aligned}$$

with $R = \sqrt{x^2 + y^2 + z^2}$. This follows from (8.31a) with the substitutions $x \rightarrow y, y \rightarrow z, z \rightarrow x, \theta \rightarrow \pi/2 + \tan^{-1}(x/z), \phi \rightarrow \pi/2 - \tan^{-1}(y/z)$. We have also assumed that the plate is thin and $L \ll z$ so we have approximated the plate as a thin layer. We then assume small angles and drop terms proportional to $(x/z)^2, (y/z)^2$ to get

$$\mathcal{E}_{\text{rad}}(z) \simeq \frac{q[\omega^2 x_0 e^{-i\omega t}]_{\text{ret}}}{4\pi\epsilon_0 c^2} \frac{1}{\sqrt{\rho^2 + z^2}} \hat{x}$$

where $\rho^2 = x^2 + y^2$. The field is predominantly \hat{x} polarized since the incident field drives the charges to oscillate along \hat{x} .

The contribution to the field from all the charges is then found by integrating over the plate to get

$$\mathcal{E}_{\text{rad}}(z) = \frac{\omega^2 x_0 q}{4\pi\epsilon_0 c^2} \int_0^\infty d\rho 2\pi\eta \frac{\rho}{\sqrt{\rho^2 + z^2}} [e^{-i\omega t}]_{\text{ret}} \hat{x}.$$

The retarded time is $t_{\text{ret}} = t - R/c$ and using $R^2 = z^2 + \rho^2$ so $\rho d\rho = R dR$ we can write this as

$$\begin{aligned}\mathcal{E}_{\text{rad}}(z) &= \frac{\omega^2 x_0 q}{2\epsilon_0 c^2} e^{-i\omega t} \int_0^\infty d\rho \eta \frac{\rho}{R} e^{i\omega R/c} \hat{x} \\ &= \frac{\omega^2 x_0 q}{2\epsilon_0 c^2} e^{-i\omega t} \int_z^\infty dR \eta e^{i\omega R/c} \hat{x} \\ &= \frac{\omega^2 x_0 q}{2\epsilon_0 c^2} e^{-i\omega t} \left(\frac{-ic}{\omega} \right) \eta e^{i\omega R/c} \Big|_z^\infty \hat{x}.\end{aligned}$$

To be consistent with the assumption of small angles we let the charge density in the plate be zero at infinity which gives the result

$$\mathcal{E}_{\text{rad}}(z) = i \frac{\omega x_0 q \eta}{2\epsilon_0 c} e^{-i\omega t} e^{i\omega z/c} \hat{x}.$$

Inserting the real part of x_0 from Eq. (8.35) we arrive at

$$\begin{aligned}\mathcal{E}_{\text{rad}}(z) &= -i \frac{\omega q N L}{2\epsilon_0 c} \frac{\mathcal{E}_{\text{in}}(0) q}{m\omega_0} \frac{2\Delta/\gamma_b^2}{1+4\Delta^2/\gamma_b^2} e^{-i\omega t} e^{i\omega z/c} \hat{x} \\ &= -i \frac{\omega q N L}{2\epsilon_0 c} \frac{q}{m\omega_0} \frac{2\Delta/\gamma_b^2}{1+4\Delta^2/\gamma_b^2} \mathcal{E}_{\text{in}}(z) \hat{x}.\end{aligned}\tag{8.37}$$

The total field from (8.36) is then

$$\mathcal{E}_{\text{tr}}(z) = \mathcal{E}_{\text{in}}(z) \left(1 - i \frac{\omega q N L}{2\epsilon_0 c} \frac{q}{m\omega_0} \frac{2\Delta/\gamma_b^2}{1+4\Delta^2/\gamma_b^2} \right). \tag{8.38}$$

Note that the radiated field is $\pi/2$ out of phase with the incident field.

Let's now compare this expression with the field that would be obtained in the presence of a plate of thickness L and refractive index $n = 1 + \delta n$. The field propagates in vacuum as

$$\mathcal{E}_{\text{in}}(z) = \mathcal{E}_{\text{in}}(0)e^{i(k_0 z - \omega t)}.$$

The field at position z when the plate is present is

$$\mathcal{E}'_{\text{in}}(z) = \mathcal{E}_{\text{in}}(0)e^{i(k_0(z-L) - \omega t)}e^{ik_0(1+\delta n)L} = \mathcal{E}_{\text{in}}(z)e^{ik_0\delta nL}.$$

The field picks up an additional propagation phase $k_0\delta nL$. When the plate is thin the additional phase is small and we can write

$$\mathcal{E}'_{\text{in}}(z) \simeq \mathcal{E}_{\text{in}}(z) (1 + ik_0\delta nL). \quad (8.39)$$

Comparing (8.38) and (8.39) we find

$$\delta n = -\frac{Nq^2}{\epsilon_0 m \omega_0 \gamma_b} \frac{\Delta/\gamma_b}{1 + 4\Delta^2/\gamma_b^2}. \quad (8.40)$$

Let's compare this with the result found from the Einstein rate equations

$$\delta n_{AB} = \frac{-\sigma_0 N}{k} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}} = \frac{-3\lambda^3 N}{4\pi^2} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \quad (8.20)$$

At low intensity the dependence on detuning is the same, if we assume $\gamma_b = \gamma$. We introduced γ_b to account for radiative loss of motional energy which is consistent with γ giving the decay rate of the excited state population in the two-level model. The prefactors still look very different, but are equivalent as we proceed to show.

The Einstein A coefficient can be calculated quantum mechanically. The result for an atom that decays from level 2 to level 1 with angular momentum $J = 0$ is

$$A = \gamma = \frac{e^2 \omega_0^3}{3\pi \epsilon_0 \hbar c^3} \langle 1 | |r| |2 \rangle^2.$$

Here $\langle 1 | |r| |2 \rangle$ is a reduced matrix element that gives the orientation averaged strength of the electric dipole coupling between the levels. The A coefficient varies for different atoms and different choices of levels because the reduced matrix element depends on the spatial structure of the wavefunctions of the participating states. It is convenient to express atomic transitions in terms of oscillator strengths, with unit oscillator strength corresponding to a classical electron oscillator as in Eq. (8.34). The absorption oscillator strength for a $J = 0$ ground state is

$$\bar{f} = \frac{2m_e \omega_0}{3\hbar} \langle 1 | |r| |2 \rangle^2$$

with m_e the electron mass. With this definition the excited state decay rate is

$$\gamma = \frac{e^2 \omega_0^2}{2\pi \epsilon_0 c^3 m_e \bar{f}} = \frac{4\pi^2 e^2}{\epsilon_0 m \lambda^3 \omega_0 \bar{f}}$$

and Eq. (8.40), with $q \rightarrow e$, $m \rightarrow m_e$, and $\omega_0 \rightarrow \omega_{21}$, becomes

$$\delta n = -\frac{\lambda_{21}^3 N}{4\pi^2} \bar{f} \frac{\Delta/\gamma}{1 + 4\Delta^2/\gamma^2}.$$

This expression used a one-dimensional model for the motion of the bound charges. Allowing for motion in three dimensions we get three times larger index change giving

$$\delta n = -\frac{3\lambda_{21}^3 N}{4\pi^2} \bar{f} \frac{\Delta/\gamma}{1 + 4\Delta^2/\gamma^2}. \quad (8.41)$$

We see that near resonance $\omega \approx \omega_{21}$, and at low intensity, Eqs. (8.41) and (8.20) are in agreement for a unit oscillator strength atomic transition. After a long calculation we have exhibited the consistency between the refractive index deduced from the Einstein rate equations for a two-level absorber, and the index that follows from the Maxwell equations as applied to a model of an atom as a bound electron.

Chapter 9

Light Scattering

In this chapter expressions are derived for the scattering of light by spheres.

9.1 Scattering geometry

The geometry is shown in Fig. 9.1. The incident beam propagates along \mathbf{k}_i and the scattered light from the particle to the observation point propagates along \mathbf{k}_s . The plane containing \mathbf{k}_i and \mathbf{k}_s is the scattering plane. Define the normal to the scattering plane by

$$\mathbf{n}_\perp = \frac{\mathbf{k}_i \times \mathbf{k}_s}{|\mathbf{k}_i \times \mathbf{k}_s|}.$$

The unit vector $\mathbf{n}_{i\parallel} = \mathbf{n}_\perp \times \mathbf{k}_i$ is perpendicular to \mathbf{k}_i and lies in the scattering plane. Similarly the unit vector $\mathbf{n}_{s\parallel} = \mathbf{n}_\perp \times \mathbf{k}_s$ is perpendicular to \mathbf{k}_s and lies in the scattering plane.

The incident and scattered electric fields $\mathbf{E}_i, \mathbf{E}_s$ are decomposed into parallel and perpendicular components

$$\begin{aligned}\mathbf{E}_i &= a_{\parallel} \mathbf{n}_{i\parallel} + a_{\perp} \mathbf{n}_\perp \\ \mathbf{E}_s &= b_{\parallel} \mathbf{n}_{s\parallel} + b_{\perp} \mathbf{n}_\perp\end{aligned}$$

where

$$a_{\parallel} = \mathbf{E}_i \cdot \mathbf{n}_{i\parallel}, \quad a_{\perp} = \mathbf{E}_i \cdot \mathbf{n}_\perp.$$

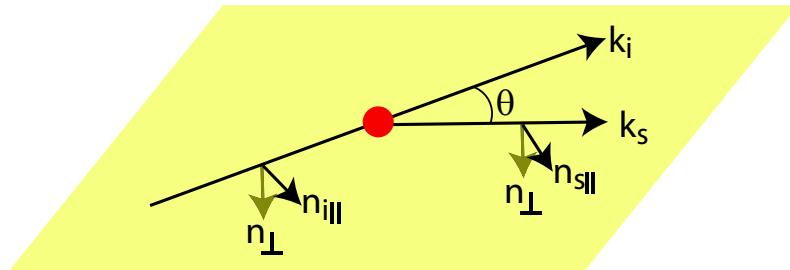


Figure 9.1: Scattering geometry.

When the scattering object is a homogeneous sphere the rigorous solution of this electromagnetic problem results in

$$\begin{aligned} b_{\perp} &= S_1(\theta) a_{\perp} \\ b_{\parallel} &= S_2(\theta) a_{\parallel} \end{aligned}$$

where the scattering angle θ is found from $\mathbf{k}_i \cdot \mathbf{k}_s = k_i k_s \cos(\theta)$, and $0 \leq \theta \leq \pi$. In the far field, or radiation zone, the functions $S_1(\theta), S_2(\theta)$ are convergent infinite sums of Bessel functions and Legendre polynomials. The explicit expressions are

$$S_1(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [a_n \pi_n(\cos(\theta)) + b_n \tau_n(\cos(\theta))], \quad (9.1)$$

$$S_2(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [b_n \pi_n(\cos(\theta)) + a_n \tau_n(\cos(\theta))], \quad (9.2)$$

(9.3)

with

$$\begin{aligned} \pi_n(\cos(\theta)) &= \frac{P_n^1(\cos(\theta))}{\sin(\theta)}, \\ \tau_n(\cos(\theta)) &= \frac{dP_n^1(\cos(\theta))}{d\theta}. \end{aligned}$$

Here P_n^1 is an associated Legendre polynomial. The coefficients a_n, b_n are found from matching boundary conditions on the surface of the sphere. The parameters that define the problem are a sphere of radius a with permittivity and permeability ϵ_1, μ_1 , embedded in a surrounding medium with properties ϵ, μ , illuminated by radiation of wavelength $\lambda = \lambda_0 / \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}}$, with λ_0 the vacuum wavelength. With these parameters we define a size parameter $\alpha = ka$, $k = 2\pi/\lambda$, and a relative refractive index $m = \sqrt{\frac{\epsilon_1\mu_1}{\epsilon\mu}}$. The scattering coefficients are[81, 82]

$$\begin{aligned} a_n &= \frac{m\psi_n(m\alpha) \frac{d\psi_n(m\alpha)}{d\alpha} - \psi_n(m\alpha) \frac{d\psi_n(m\alpha)}{d\alpha}}{m\psi_n(m\alpha) \frac{d\zeta_n^{(1)}(m\alpha)}{d\alpha} - \zeta_n^{(1)}(m\alpha) \frac{d\psi_n(m\alpha)}{d\alpha}}, \\ b_n &= \frac{m\psi_n(m\alpha) \frac{d\psi_n(m\alpha)}{d\alpha} - \psi_n(m\alpha) \frac{d\psi_n(m\alpha)}{d\alpha}}{m\zeta_n(m\alpha) \frac{d\psi_n(m\alpha)}{d\alpha} - \psi_n(m\alpha) \frac{d\zeta_n(m\alpha)}{d\alpha}}, \end{aligned}$$

with $\psi_n(z) = z j_n(z)$, $\zeta_n(z) = z h_n^{(1)}(z)$, $j_n(z) = z^n \left(-\frac{1}{z} \frac{d}{dz}\right)^n \left(\frac{\sin(z)}{z}\right)$ is the spherical Bessel function of the first kind, $y_n(z) = -z^n \left(-\frac{1}{z} \frac{d}{dz}\right)^n \left(\frac{\cos(z)}{z}\right)$ is the spherical Bessel function of the second kind, and $h_n^{(1)}(z) = j_n(z) + iy_n(z)$ is the spherical Hankel function of the first kind.

If either $a_{\perp}(a_{\parallel})$ are zero the incident light is linearly polarized parallel(perpendicular) to the scattering plane and retains the same polarization after scattering. This is a consequence of the spherical symmetry of the scatterer. If both amplitudes are nonzero the incident light may be linearly or elliptically polarized depending on the relative phase of the amplitudes. Linearly polarized light may emerge as elliptically polarized since S_1, S_2 are complex amplitudes with different angle dependent phases. The scattering cross section is

$$\sigma = \frac{\lambda^2}{2\pi} \sum_{n=1}^{\infty} (2n+1) (|a_n|^2 + |b_n|^2). \quad (9.4)$$

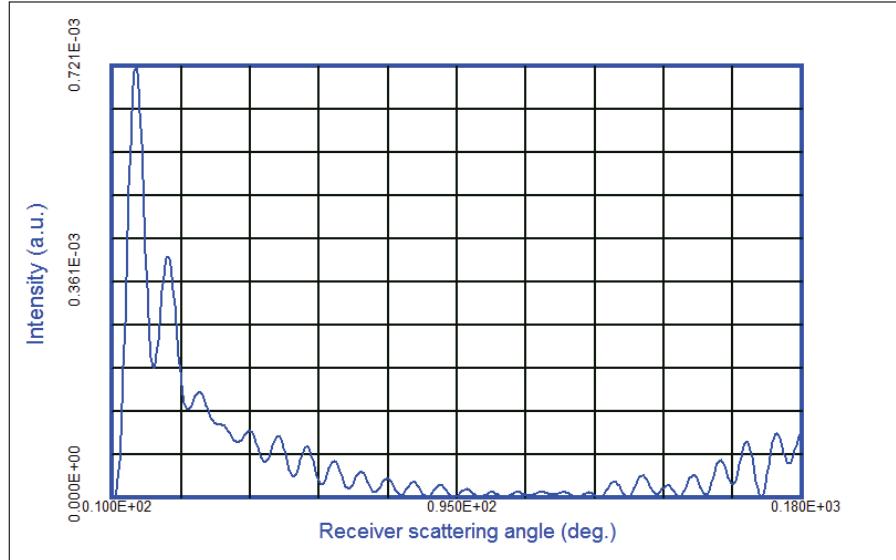


Figure 9.2: Scattered intensity averaged over a circular aperture with half cone angle of 0.1 deg. for a water drop with $n = 1.33$ in air, $d = 5 \mu\text{m}$, and $\lambda = 0.63 \mu\text{m}$.

There are computer codes available for the efficient calculation of the amplitudes. I have used that of W. J. Wiscombe NCAR Technical Note NCAR/TN-140+STR. It turns out that the numerical calculation can be made more efficient by first computing

$$S_{\pm} = S_1 \pm S_2$$

and then using

$$S_1 = \frac{S_+ + S_-}{2}, \quad S_2 = \frac{S_+ - S_-}{2}.$$

This is useful since calculating S_{\pm} first saves some arithmetic operations. An example of scattering by a water drop in air is shown in Fig. 9.2. The pronounced asymmetry with stronger scattering in the forward direction becomes noticeable for particles with $\alpha > 1$. The degree of asymmetry increases with α . There are also a large number of local maxima, referred to as lobes, as a function of the scattering angle. The number of lobes increases proportional to α .

9.2 Rayleigh scattering

When the sphere is small compared to the wavelength of light we can make the approximation that the incident field is uniform across the sphere. In this limit the scattering cross section was derived by Rayleigh[83]. The induced dipole is proportional to the volume of the sphere and the scattered power is proportional to the square of the dipole, so $P \sim a^6$. The scattering cross section for unpolarized incident light is

$$\sigma = \frac{2\pi^5}{3} \frac{d^6}{\lambda^4} \left(\frac{n^2 - 1}{n^2 + 1} \right)^2. \quad (9.5)$$

Here $d = 2a$ is the particle diameter, and $n = \sqrt{\epsilon_1/\epsilon}$ is the relative refractive index. This result can also be found by taking the small a limit of (9.4).

Chapter 10

Dispersive and nonlinear pulse propagation

In this chapter we look at some phenomena due to dispersion and nonlinearity of the refractive index. Dispersion refers to the variation of the refractive index with frequency. This results in different spectral components of a broadband optical field propagating at different velocities. It follows that a pulse that is initially narrow in time will spread after propagation through a dispersive medium. Our study of the two-level atom in Ch. 8 showed that near a resonance the index is dispersive.

Nonlinearity refers to the refractive index depending on the intensity of the light, an effect that is also apparent from the results of Ch. 8. A less obvious effect is that a nonlinear refractive index can lead to the creation of waves of different frequencies. A simple model is based on replacing the harmonically bound electron of Eq. (8.34) with a bound electron in an anharmonic potential

$$m \frac{d^2x}{dt^2} - m\gamma_b \frac{dx}{dt} = -\kappa x - \kappa_2 x^2 + q \frac{\mathcal{E}}{2} (e^{-i\omega t} + e^{i\omega t}). \quad (10.1)$$

With $\kappa_2 = 0$ we found a steady state solution for $z(t)$ oscillating at the driving frequency ω . Allowing for $\kappa_2 \neq 0$ it is not hard to show that the solution can be expanded in a power series in κ_2 of the form

$$z(t) = \sum_{n=0}^{\infty} a_n \kappa_2^n \mathcal{E}^{n+1} e^{-in\omega t} + c.c.$$

with a_n coefficients. The presence of an anharmonic potential leads to oscillation at frequencies $\omega_n = n\omega$. The results of Sec. 8.6.1 imply that the electron will then radiate at frequencies ω_n which leads to interaction of waves with different frequencies. In this chapter we will focus on dispersive effects and nonlinear effects that modify the behavior at a single frequency ω . In Ch. 11 we will consider coupling and interaction of waves of different frequencies.

The actual strength of the nonlinear response, i.e. the coefficients a_n , are determined by the microscopic physics of the light - matter interaction. We can estimate how strong the field must be to induce a nonlinear response by considering an isolated atom. If the applied field is at least comparable to the field inside the atom we may expect the properties to be significantly modified and the response to be nonlinear.

The characteristic field inside an atom is $E_{\text{atom}} = \frac{1}{4\pi\epsilon_0} \frac{e}{a_0^2}$ with a_0 the Bohr radius. This gives an intensity $I = \frac{\epsilon_0 c}{2} E^2 = 4 \times 10^{20} \text{ W/m}^2$, which is an extremely high intensity. Nonetheless nonlinear optical effects can be observed at much lower intensity levels, even with single photons. In a gaseous or solid sample a large number of atoms can interact with a propagating light beam. Consider a

beam of area A propagating through a sample of length L with a density of atoms ρ . Taking $\rho = 10^{20} \text{ 1/m}^3$, $L = 1 \text{ cm}$ and $A = 1 \text{ mm}^2$ gives $N = \rho LA = 10^{12}$ atoms. If the nonlinear effect from each atom is additive this reduces the required intensity to $I \sim 10^8 \text{ W/m}^2$. An example of a strong nonlinear effect due to a laser beam with only mW of power is shown in Fig. 10.2.

The other approach to observing large nonlinear effects is to make the light resonant with an atomic transition. As we saw the absorption cross section for a single resonant absorber is $\sigma \sim \lambda^2$. If the field of a single photon is focused to an area of order λ^2 large effects are possible due to single photons interacting with single atoms. A detailed discussion of the physics involved is beyond the scope of this course. Instead we will proceed with a phenomenological discussion of the macroscopic effects that arise from a nonlinear material response.

10.1 The wave equation

To derive the envelope equation describing dispersive and nonlinear propagation of an optical beam we start with the Maxwell equations

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (10.2a)$$

$$\nabla \times \frac{1}{\mu_0} \mathbf{B} = \frac{\partial \mathbf{D}}{\partial t} \quad (10.2b)$$

where \mathbf{E} is the electric field, \mathbf{B} is the magnetic displacement, and μ is the magnetic permeability (assumed equal to its value in vacuum μ_0). Taking the curl of (10.2a) and using (10.2b) we get

$$\nabla^2 \mathbf{E} - \mu_0 \frac{\partial^2 \mathbf{D}}{\partial t^2} = 0, \quad (10.3)$$

assuming $\nabla \cdot \mathbf{E} = 0$.¹

The electric displacement \mathbf{D} is, in the approximation of an instantaneous response, defined through the constitutive relation

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (10.4)$$

where \mathbf{P} is the electric polarization. In dispersive media where \mathbf{P} is a function of frequency the electric displacement \mathbf{D} is also frequency dependent. The polarization consists of linear and nonlinear contributions and can be written, for media that respond instantaneously, in the form

$$\mathbf{P} = \epsilon_0 \chi^{(1)} \mathbf{E} + \mathbf{P}_{\text{NL}}. \quad (10.5)$$

The electric displacement is then

$$\mathbf{D} = \epsilon_0 \left(1 + \chi^{(1)}\right) \mathbf{E} + \mathbf{P}_{\text{NL}}. \quad (10.6)$$

The linear part of this relation is $\mathbf{D} = \epsilon \mathbf{E}$ which defines the linear refractive index through $\epsilon/\epsilon_0 = n^2$ so

$$n^2 = \frac{\epsilon_0 (1 + \chi^{(1)})}{\epsilon_0} = 1 + \chi^{(1)}$$

and

$$n = \sqrt{1 + \chi^{(1)}} \simeq 1 + \frac{\chi^{(1)}}{2}.$$

This last expression is valid for small index changes.

¹In media with no free charges $\nabla \cdot \mathbf{E} = -(\nabla \ln \epsilon) \cdot \mathbf{E}$, where ϵ is the linear part of the dielectric constant. Thus the assumption of a divergenceless electric field is accurate when the linear properties of the medium change slowly.

10.1.1 Linear propagation in inhomogeneous media

Before describing dispersive and nonlinear propagation it is instructive to consider wave propagation in a medium with spatial modulation of the linear refractive index. In this section we assume that n does not have any frequency dependence. Introducing the refractive index through the definition $n^2 \equiv \epsilon/\epsilon_0$, and using $\mathbf{D} = \epsilon\mathbf{E}$, Eq. (10.3) can be written as

$$\nabla^2\mathbf{E} - \frac{n^2}{c^2} \frac{\partial^2\mathbf{E}}{\partial t^2} = 0, \quad (10.7)$$

where we have made use of the relationship $\mu_0\epsilon_0 = 1/c^2$. We will investigate the effect of a spatially varying refractive index

$$n(\mathbf{r}) = n_0 [1 + n_1(\mathbf{r})], \quad (10.8)$$

where $\mathbf{r} = (x, y, z)$. Although the Maxwell equations are written in terms of real fields it is convenient to introduce a complex notation. We write, therefore,

$$E(\mathbf{r}, t) = \frac{\mathcal{E}(\mathbf{r}, t)}{2} e^{i(k_0 z - \omega_0 t)} + c.c. \quad (10.9)$$

where we have factored out the rapidly varying phase term corresponding to the central frequency ω_0 and the central wavenumber $k_0 = n_0\omega_0/c$ of the beam². In addition we have specialized to the case of a scalar field. Substituting into (10.7) we find that the complex amplitude is governed by the equation

$$\nabla^2\mathcal{E} + 2ik_0 \frac{\partial\mathcal{E}}{\partial z} - \frac{n^2}{c^2} \frac{\partial^2\mathcal{E}}{\partial t^2} + 2i\omega_0 \frac{n^2}{c^2} \frac{\partial\mathcal{E}}{\partial t} = \left(k_0^2 - \frac{\omega_0^2 n^2}{c^2} \right) \mathcal{E}. \quad (10.10)$$

To proceed we make two assumptions. The slowly varying envelope approximation:

$$\frac{\partial^2\mathcal{E}}{\partial t^2} \ll \omega_0 \frac{\partial\mathcal{E}}{\partial t}, \quad \frac{\partial^2\mathcal{E}}{\partial z^2} \ll k_0 \frac{\partial\mathcal{E}}{\partial z},$$

and the assumption of weak variation of the refractive index:

$$n^2(\mathbf{r}) \simeq n_0^2[1 + 2n_1(\mathbf{r})]. \quad (10.11)$$

Equation (10.10) can then be written in the form

$$\frac{\partial\mathcal{E}(\mathbf{r}, t)}{\partial z} + \frac{n_0}{c} \frac{\partial\mathcal{E}(\mathbf{r}, t)}{\partial t} - \frac{i}{2k_0} \nabla_{\perp}^2 \mathcal{E}(\mathbf{r}, t) = ik_0 n_1 \mathcal{E}(\mathbf{r}, t) - \frac{2n_0 n_1(\mathbf{r})}{c} \frac{\partial\mathcal{E}(\mathbf{r}, t)}{\partial t}, \quad (10.12)$$

where $\nabla_{\perp}^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$. The form of the perturbation to the refractive index n_1 depends on the type of nonlinear medium under consideration.

This equation can be simplified further by going to a reference frame moving with the phase velocity of the pulse given by $v_{\phi} = c/n_0$. The moving reference frame is illustrated in Fig. 10.1. We thus change variables to $\zeta(z, t) = z$, $\tau(z, t) = t - \frac{1}{v_{\phi}}z$, and use

$$\frac{\partial}{\partial\zeta} = \frac{\partial}{\partial z} + \frac{1}{v_{\phi}} \frac{\partial}{\partial t},$$

to put (10.12) in the form

$$\frac{\partial\mathcal{E}}{\partial\zeta} - \frac{i}{2k_0} \nabla_{\perp}^2 \mathcal{E} = ik_0 n_1 \mathcal{E} - \frac{2n_0 n_1}{c} \frac{\partial\mathcal{E}}{\partial\tau}. \quad (10.13)$$

²Our notation for the wavenumber in this chapter is $k_0 = k_{\text{vac}}n_0$ or $k = k_{\text{vac}}n$ which is different from the convention in Ch. 5 where we used $k = k_{\text{vac}}$.

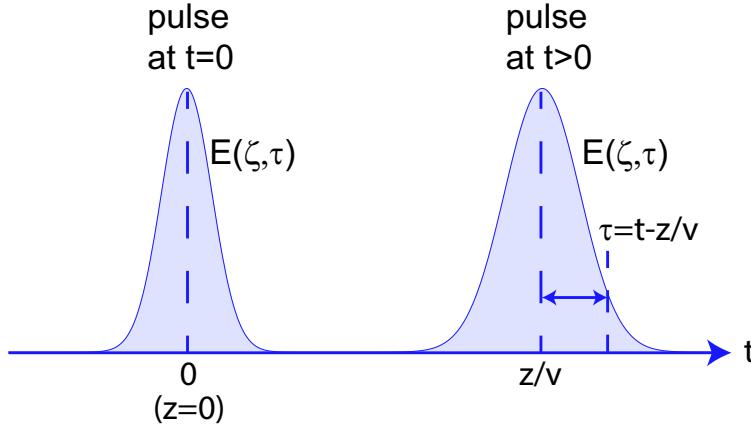


Figure 10.1: Moving reference frame for describing pulse propagation.

The left hand side of this equation is just the paraxial wave equation (5.2) from Chapter 5 which formed the basis of our study of diffraction phenomena. When the refractive index is spatially varying two new terms appear on the right hand side of the wave equation. The first proportional to $n_1 \mathcal{E}$ describes scattering of the field \mathcal{E} off refractive variations n_1 . This acts as a driving term for the field. The second term $n_1 \frac{\partial \mathcal{E}}{\partial t}$ adds an effective dispersion due to the variation in propagation time associated with index variations n_1 .

10.2 Propagation in linear dispersive media

When the refractive index is a function of wavelength or frequency of the light different spectral components propagate at different velocities. This is referred to as dispersion. A traveling wave of the form $E \sim \cos[k(\omega)x - \omega t]$ is characterized by several quantities. The phase velocity is

$$v_\phi \equiv \frac{\omega}{k(\omega)} = \frac{c}{n(\omega)}.$$

The phase velocity can be greater than c if $|n| < 1$.

The group velocity is generally, but not always, the rate at which information is transmitted. It can also be greater than c without violating causality. The group velocity is defined by $v_g \equiv \frac{d\omega}{dk}$. To calculate the group velocity use $k = \frac{\omega}{c} n(\omega)$ and

$$\frac{dk}{d\omega} = \frac{d}{d\omega} \left[\frac{\omega}{c} n(\omega) \right] = \frac{n(\omega)}{c} + \frac{\omega}{c} \frac{dn(\omega)}{d\omega}$$

so that

$$v_g = \frac{d\omega}{dk} = \frac{c}{n(\omega) + \omega \frac{dn(\omega)}{d\omega}} = \frac{c}{n(\omega)} \frac{1}{1 + \frac{\omega}{n(\omega)} \frac{dn(\omega)}{d\omega}} = v_\phi \frac{1}{1 + \frac{\omega}{n(\omega)} \frac{dn(\omega)}{d\omega}}.$$

Since $v_g = v_g(\omega)$ different frequencies propagate with different group velocities which leads to pulse spreading. As we will see below this is typically characterized by the group velocity dispersion coefficient $k'' = \frac{d^2 k}{d\omega^2}$.

In order to evaluate higher order dispersive effects we proceed as follows. We start with (10.3) but assume that the susceptibility ϵ is dispersive. We thus write

$$D(\mathbf{r}, \omega) = \epsilon(\mathbf{r}, \omega) E(\mathbf{r}, \omega), \quad (10.14)$$

where for simplicity we have assumed scalar fields so that the susceptibility ϵ has only a single component ϵ . The fields $D(\mathbf{r}, \omega)$, $E(\mathbf{r}, \omega)$ are the Fourier transforms of the real amplitudes $D(\mathbf{r}, t)$, $E(\mathbf{r}, t)$ that in turn can be written in complex form with the definitions $E(\mathbf{r}, t) = (1/2)\mathcal{E}(\mathbf{r}, t) \exp[i(k_0 z - \omega_0 t)] + c.c.$, etc. Using the Fourier convolution theorem, Eq. (10.14) is equivalent to the following relation between the time domain fields

$$\mathcal{D}(\mathbf{r}, t) e^{i(k_0 z - \omega_0 t)} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\tau \epsilon(t - \tau) \mathcal{E}(\mathbf{r}, \tau) e^{i(k_0 z - \omega_0 \tau)}. \quad (10.15)$$

Equation (10.3) then takes the form

$$\nabla^2 (\mathcal{E}(\mathbf{r}, t) e^{i(k_0 z - \omega_0 t)}) - \mu_0 \frac{\partial^2}{\partial t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\tau \epsilon(t - \tau) \mathcal{E}(\mathbf{r}, \tau) e^{i(k_0 z - \omega_0 \tau)} = 0. \quad (10.16)$$

The time dependent susceptibility can be represented as

$$\epsilon(t - \tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \epsilon(\omega) e^{-i(t-\tau)\omega},$$

where $\epsilon(\omega) = \epsilon_0(1 + \chi^{(1)}(\omega))$. The second term in (10.16) can then be written as

$$\begin{aligned} & -\frac{1}{\epsilon_0 c^2} e^{ik_0 z} \frac{\partial^2}{\partial t^2} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\omega \epsilon(\omega) e^{-i(t-\tau)\omega} \int_{-\infty}^{\infty} d\omega' \mathcal{E}(\mathbf{r}, \omega') e^{-i\omega_0 \tau} e^{-i\tau \omega'} \\ &= -\frac{1}{\epsilon_0 c^2} e^{ik_0 z} \frac{\partial^2}{\partial t^2} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \epsilon(\omega) \mathcal{E}(\mathbf{r}, \omega') e^{-i\omega t} \delta(\omega - \omega_0 - \omega') \\ &= -\frac{1}{\epsilon_0 c^2} e^{ik_0 z} \frac{\partial^2}{\partial t^2} \int_{-\infty}^{\infty} d\omega \epsilon(\omega + \omega_0) \mathcal{E}(\mathbf{r}, \omega) e^{-i(\omega + \omega_0)t} \\ &= e^{i(k_0 z - \omega_0 t)} \int_{-\infty}^{\infty} d\omega k^2(\omega + \omega_0) \mathcal{E}(\mathbf{r}, \omega) e^{-i\omega t}, \end{aligned} \quad (10.17)$$

where the linear dispersion relation

$$k^2(\omega) = \frac{\omega^2}{c^2} \frac{\epsilon(\omega)}{\epsilon_0}$$

has been invoked. Expanding about ω_0 we have

$$k^2(\omega + \omega_0) = k_0^2 + 2k_0 k'_0 \omega + (k_0'^2 + k_0 k''_0) \omega^2 + \dots$$

whence

$$\begin{aligned} & \nabla^2 (\mathcal{E}(\mathbf{r}, t) e^{i(k_0 z - \omega_0 t)}) + e^{i(k_0 z - \omega_0 t)} \left[k_0^2 \mathcal{E}(\mathbf{r}, t) + 2ik_0 k'_0 \frac{\partial \mathcal{E}(\mathbf{r}, t)}{\partial t} - (k_0'^2 + k_0 k''_0) \frac{\partial^2 \mathcal{E}(\mathbf{r}, t)}{\partial t^2} \right] \\ &= 0, \end{aligned} \quad (10.18)$$

where

$$k_0 = k(\omega_0), \quad k'_0 = \left(\frac{dk}{d\omega} \right)_{\omega=\omega_0}, \quad k''_0 = \left(\frac{d^2 k}{d\omega^2} \right)_{\omega=\omega_0}.$$

This equation can be simplified further by going to a reference frame moving with the group velocity of the pulse given by $v_g = 1/k'_0$. We thus put $\zeta = z$, $\tau = t - k'_0 z$, and note that

$$\begin{aligned}\frac{\partial}{\partial \zeta} &= \frac{\partial}{\partial z} + k'_0 \frac{\partial}{\partial t}, \\ \frac{\partial^2 \mathcal{E}}{\partial z^2} - k''_0 \frac{\partial^2 \mathcal{E}}{\partial t^2} &= \left(\frac{\partial}{\partial z} - k'_0 \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial z} + k'_0 \frac{\partial}{\partial t} \right) \mathcal{E} \\ &= \left(\frac{\partial}{\partial z} - k'_0 \frac{\partial}{\partial t} \right) \frac{\partial \mathcal{E}}{\partial \zeta} \\ &\ll k'_0 \frac{\partial \mathcal{E}}{\partial \zeta}.\end{aligned}$$

We thus obtain

$$\frac{\partial \mathcal{E}}{\partial \zeta} - \frac{i}{2k'_0} \nabla_{\perp}^2 \mathcal{E} + \frac{ik''_0}{2} \frac{\partial^2 \mathcal{E}}{\partial \tau^2} = 0. \quad (10.19)$$

This equation has the same mathematical form as the paraxial wave equation (5.2) describing diffraction. We see that temporal dispersion which is proportional to k''_0 is equivalent to spatial dispersion (diffraction) which is proportional to $1/k'_0 = \lambda/2\pi n_0$. This formal similarity means that the solutions we have developed for Fresnel diffraction can be directly applied to the problem of temporal dispersion. Suppose that we are working with plane waves so that $\nabla_{\perp}^2 \mathcal{E} = 0$. Equivalently, if we are considering propagation in a guiding structure such as an optical fiber, the spatial diffraction is controlled by the linear structure so for an eigenmode $\nabla_{\perp}^2 \mathcal{E} \sim \mathcal{E}$ and the linear term can be scaled out of the equation. In either case we arrive at the equation

$$\frac{\partial \mathcal{E}(\tau, \zeta)}{\partial \zeta} + \frac{ik''_0}{2} \frac{\partial^2 \mathcal{E}(\tau, \zeta)}{\partial \tau^2} = 0. \quad (10.20)$$

The Fresnel diffraction solution is given by (5.7) with the replacement $k \rightarrow -1/k''_0$ and the change from a two-dimensional integral over x, y to a one-dimensional integral over τ

$$\mathcal{E}(\tau_2) = \left(\frac{i}{2\pi k''_0 \zeta} \right)^{1/2} \int_{-\infty}^{\infty} d\tau_1 \mathcal{E}(\tau_1) e^{-\frac{i}{2k''_0 \zeta} (\tau_2 - \tau_1)^2}. \quad (10.21)$$

This solution leads to the temporal spreading of pulses in the same way that Gaussian beams diffract and spread in space.

In order to calculate pulse spreading we need a numerical value for k''_0 . This is given in the literature in terms of a parameter

$$D_{\lambda} = \frac{\lambda}{c} \frac{d^2 n}{d \lambda^2} \quad [\text{ps}/(\text{km} \cdot \text{nm})].$$

Using the chain rule we can show that $\frac{d^2 k}{d\omega^2} = \frac{\lambda^3}{2\pi c^2} \frac{d^2 n}{d\lambda^2}$ so

$$k'' = \frac{d^2 k}{d\omega^2} = \frac{\lambda^2}{2\pi c} D_{\lambda}.$$

In this equation $\lambda = \lambda_{\text{vac}}$.

10.3 Wave equation in “Kerr” media

The polarization appearing in Eq. (10.4) for the displacement field consists of linear and nonlinear contributions and can be written, for media that respond instantaneously, in the form

$$\mathbf{P} = \epsilon_0 \chi^{(1)} \mathbf{E} + \mathbf{P}_{\text{NL}}. \quad (10.22)$$

Substituting into (10.3) we find

$$\nabla^2 \mathbf{E} = \mu_0 \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu_0 \frac{\partial^2 \mathbf{P}_{\text{NL}}}{\partial t^2}, \quad (10.23)$$

where $\epsilon \equiv \epsilon_0(1 + \chi^{(1)})$ accounts for the linear refractive index $n_0^2 = \epsilon/\epsilon_0$.

In Kerr media there is a nonlinear contribution to the refractive index proportional to the cube of the electric field. This is written as $P_{\text{NL},i} = \epsilon_0 \sum_{jkl} \chi_{ijkl}^{(3)} E_j E_k E_l$. The tensorial nature of the interaction leads to complicated polarization dependent effects³. Let’s consider a simplified scalar model of the form $P_{\text{NL}} = \epsilon_0 \chi^{(3)} E^3$. Writing $E = \frac{\mathcal{E}}{2} e^{i(k_0 z - \omega_0 t)} + c.c.$ with $k_0 = 2\pi n_0 / \lambda_{\text{vac}}$ and $\omega_0 = k_0 c / n_0$ the field cubed contains terms oscillating at $\pm 3\omega_0, \pm \omega_0$. The term at $-\omega$ is

$$P_{\text{NL}} = \epsilon_0 \chi^{(3)} \frac{|\mathcal{E}|^2 \mathcal{E}}{8} e^{i(k_0 z - \omega_0 t)}$$

and substituting into (10.23) we find after making a slowly varying envelope approximation

$$\frac{\partial \mathcal{E}}{\partial z} + \frac{n_0}{c} \frac{\partial \mathcal{E}}{\partial t} - \frac{i}{2k_0} \nabla_{\perp}^2 \mathcal{E} = i \frac{k_0 \chi^{(3)}}{8n_0^2} |\mathcal{E}|^2 \mathcal{E}.$$

Finally working in a frame moving at the phase velocity of the light we can combine the first order space and time derivatives to give

$$\frac{\partial \mathcal{E}}{\partial \zeta} - \frac{i}{2k_0} \nabla_{\perp}^2 \mathcal{E} = i \frac{k_0 \chi^{(3)}}{8n_0^2} |\mathcal{E}|^2 \mathcal{E}. \quad (10.24)$$

Comparing with (10.8,10.12) we see that the nonlinear term on the right hand side is equivalent to a refractive index perturbation $\delta n = n_0 n_1$ with

$$n_1 = \frac{\chi^{(3)}}{8n_0^2} |\mathcal{E}|^2.$$

When $\chi^{(3)} > 0$ the refractive index increases in regions of high intensity. This leads to focusing of an inhomogeneous field and is called a self-focusing nonlinearity. Conversely when $\chi^{(3)} < 0$ we have self-defocusing. With the variables and coordinates rescaled to dimensionless quantities Eq. (10.24) can be written as

$$\frac{\partial A}{\partial \zeta} - i \nabla_{\perp}^2 A = \pm i |A|^2 A$$

with the \pm sign corresponding to the sign of $\chi^{(3)}$. This is the celebrated nonlinear Schrödinger (NLS) equation which occurs in a wide range of problems in mathematical physics and has played a central role in the development of soliton and nonlinear wave theory. The NLS equation arises naturally in optical media with a Kerr effect giving an intensity dependent refractive index. The Kerr nonlinearity leads to the formation of spatial solitons, temporal solitons, and also combined space-time focusing. Temporal solitons have been extensively investigated in the context of ultrafast data transmission in optical fibers where the nonlinearity can compensate dispersion induced pulse spreading.

³See, for example, the book by Boyd[84] for details.

10.3.1 Kerr effect in atomic media

We have already encountered a situation that leads to a Kerr nonlinearity - light interacting with a two-level atom. The refractive index from Eq. (8.20) is

$$n = 1 - \frac{\sigma_0 N}{k} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}. \quad (8.20)$$

so, in the notation of this chapter,

$$n_1 = -\frac{\sigma_0 N}{k} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2} + \frac{I}{I_s}}.$$

This is not a Kerr nonlinearity, which is cubic in the field, but rather is a more complicated saturable nonlinearity that depends on all powers of the optical field. Note that for $\Delta > 0$, $dn_1/dI > 0$ and for $\Delta < 0$, $dn_1/dI < 0$. As we will see the cases of $\Delta > 0 (< 0)$ correspond to nonlinear self-focusing (self-defocusing).

When $I/I_s \ll 1$ we have a weak nonlinearity and

$$n \simeq \left[1 - \frac{\sigma_0 N}{k} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{1 + \frac{4\Delta^2}{\gamma^2}} \right] + \frac{\sigma_0 N}{k} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{\left(1 + \frac{4\Delta^2}{\gamma^2}\right)^2} \frac{I}{I_s}.$$

The terms in square brackets give the background index n_0 which does not depend on intensity, while the last term is $n_0 n_1$ so

$$n_1|_{I/I_s \ll 1} = \frac{\sigma_0 N}{kn_0} \frac{\omega}{\omega_{21}} \frac{\frac{\Delta}{\gamma}}{\left(1 + \frac{4\Delta^2}{\gamma^2}\right)^2} \frac{I}{I_s}.$$

We recover a Kerr nonlinear response for low intensity, where saturation can be ignored.

10.4 Optical solitons

The nonlinear refractive index leads to some remarkable physics. One example is the optical soliton - a localized beam that propagates without diffraction and stays localized. This is possible because the spreading and wavefront curvature due to diffraction is balanced by nonlinear self-focusing. This problem can be solved exactly for a Kerr nonlinearity in 1+1 dimensions, i.e. one propagation coordinate and one transverse coordinate. The transverse coordinate is position for spatial solitons, and time for temporal solitons.

The governing equation for spatial solitons is

$$\frac{\partial \mathcal{E}}{\partial \zeta} - \frac{i}{2k} \frac{\partial^2 \mathcal{E}}{\partial x^2} = i\kappa |\mathcal{E}|^2 \mathcal{E}. \quad (10.25)$$

The nonlinear coefficient κ is proportional to $\chi^{(3)}$ and for atomic media $\kappa > 0$ for positive detuning.

A solitary wave solution that does not diffract has a constant spatial envelope. Let's seek such a solution in the form

$$\mathcal{E}(x, \zeta) = a(x) e^{ib\zeta}$$

where $a(x)$ is an unknown function and b is a constant. The derivatives are

$$\frac{\partial \mathcal{E}}{\partial \zeta} = ib\mathcal{E}, \quad \frac{\partial^2 \mathcal{E}}{\partial x^2} = e^{ib\zeta} \frac{d^2 a}{dx^2}.$$

Plugging in to (10.25) we get

$$iba(x)e^{ib\zeta} - \frac{i}{2k} \frac{d^2 a(x)}{dx^2} e^{ib\zeta} = i\kappa |a(x)|^2 a(x) e^{ib\zeta}.$$

Assuming $a(x) > 0$ and rearranging we get

$$\frac{d^2 a}{dx^2} = 2kba - 2k\kappa a^3.$$

Try a solution $a = a_0 \operatorname{sech}(a_1 x)$. Plugging in we find that this works if $a_0 = \sqrt{2b/\kappa}$, $a_1 = \sqrt{2kb}$. Thus the solution is

$$\mathcal{E}(x, \zeta) = \sqrt{2b/\kappa} \operatorname{sech}(\sqrt{2kb}x) e^{ib\zeta}.$$

This solution describes a family of profiles with width scaling as $1/\sqrt{2kb}$. The propagation constant b depends on the initial amplitude. Using $\operatorname{sech}(0) = 1$ we have $|\mathcal{E}(0, \zeta)|^2 = \frac{2b}{\kappa}$ so

$$b = \frac{\kappa}{2} |\mathcal{E}(0, \zeta)|^2.$$

We have found an exact solution to the paraxial wave equation that propagates while maintaining a constant shape. This solitary wave is in fact a soliton that has special physical and mathematical properties[85]. The governing equation is integrable and possesses an infinite number of conserved quantities. Solitons collide elastically suffering only phase shifts. The temporal analog of the spatial soliton has been proposed as a way of transmitting pulses through long optical fibers without dispersive spreading[86, 87].

10.5 Amplitude equation for dispersive nonlinear pulse propagation

Nonlinear propagation of pulses with a short duration in time is possible in the presence of material dispersion. The basic amplitude equation is of NLS type⁴. More refined descriptions which include higher order nonlinear and dispersive terms are outside the scope of this discussion.

We can deduce the correct amplitude equation from our results for dispersive but linear pulse propagation, and the NLS equation without dispersion

$$\frac{\partial \mathcal{E}}{\partial \zeta} - \frac{i}{2k_0} \nabla_{\perp}^2 \mathcal{E} + \frac{ik_0''}{2} \frac{\partial^2 \mathcal{E}}{\partial \tau^2} = 0. \quad (10.19)$$

$$\frac{\partial \mathcal{E}}{\partial \zeta} - \frac{i}{2k_0} \nabla_{\perp}^2 \mathcal{E} = i \frac{k_0 \chi^{(3)}}{8n_0^2} |\mathcal{E}|^2 \mathcal{E}. \quad (10.24)$$

These equations suggest that the correct equation for dispersive and nonlinear pulse propagation is

$$\frac{\partial \mathcal{E}}{\partial \zeta} - \frac{i}{2k_0} \nabla_{\perp}^2 \mathcal{E} + \frac{ik_0''}{2} \frac{\partial^2 \mathcal{E}}{\partial \tau^2} = i \frac{k_0 \chi^{(3)}}{8n_0^2} |\mathcal{E}|^2 \mathcal{E}.$$

⁴Our derivation is close to that given in [88].

It can be verified that this is correct by repeating the derivations in Secs. 10.2,10.3 including both dispersive and nonlinear effects from the beginning.

In the anomalous dispersion regime $k_0'' < 0$ and with $\chi^{(3)} > 0$ we have self-focusing in space and time. For normal dispersion $k_0'' > 0$, we have focusing in space together with spreading in time. For anomalous dispersion put

$$\tilde{\mathcal{E}} = \sqrt{\frac{k_0 \chi^{(3)}}{8n_0^2}} \mathcal{E}, \quad (\tilde{x}, \tilde{y}) = \sqrt{k_0}(x, y), \quad \tilde{\tau} = \frac{\tau}{\sqrt{|k_0''|}},$$

and drop the tildes to obtain the normalized (3+1)D nonlinear Schrödinger equation

$$\frac{\partial \mathcal{E}}{\partial \zeta} - \frac{i}{2} \nabla^2 \mathcal{E} = i |\mathcal{E}|^2 \mathcal{E}, \quad (10.26)$$

where $\nabla = (\hat{x}\partial/\partial x, \hat{y}\partial/\partial y, \hat{\tau}\partial/\partial \tau)$.

10.6 Filamentation instability

A plane wave propagating in a medium with cubic nonlinearity is unstable with respect to the generation of small scale filaments. Because of its importance as a factor limiting the propagation of high power beams, much work has been devoted to studying the instability. Filamentation may also be identified as the first stage in the development of turbulent fluctuations in the transverse profile of a laser beam.

Time independent propagation of an optical beam $A(r, z)$ in a nonlinear medium is governed by the parabolic equation

$$\left(\frac{\partial}{\partial z} - i \nabla_{\perp}^2 \right) A(r, z) = i s n_2(|A|^2) A(r, z). \quad (10.27)$$

Here $r = (x, y)$ and z are the transverse and axial coordinates, $\nabla_{\perp} = \hat{x}(\partial/\partial x) + \hat{y}(\partial/\partial y)$, $s = \pm 1$, for focusing or defocusing nonlinearities respectively, and n_2 is the nonlinear increment to the refractive index that will be taken to be a function of the intensity $|A^2|$. Equation (10.27) admits plane wave solutions

$$A = A_0 e^{is\beta z}, \quad (10.28)$$

with $\beta = n_2(A_0^2)$. The assumption of A_0 real is sufficiently general.

The nonlinear contribution to the refractive index for Kerr and saturable Kerr media is given by

$$n_2(|A|^2) = |A|^2, \quad (10.29a)$$

$$n_2(|A|)^2 = \frac{|A|^2}{1 + |A|^2}, \quad (10.29b)$$

respectively. The initial linear stage of the filamentation instability may be investigated by putting

$$A(r, z) = A_0 e^{is\beta z} (1 + a e^{\Gamma z + i\mathbf{q} \cdot \mathbf{r}} + b e^{\Gamma z - i\mathbf{q} \cdot \mathbf{r}}). \quad (10.30)$$

Linearizing in the perturbations a, b we get the algebraic equations

$$(\Gamma + iq^2) a = i s \delta n_+, \quad (10.31)$$

$$(\Gamma + iq^2) b = i s \delta n_-. \quad (10.32)$$

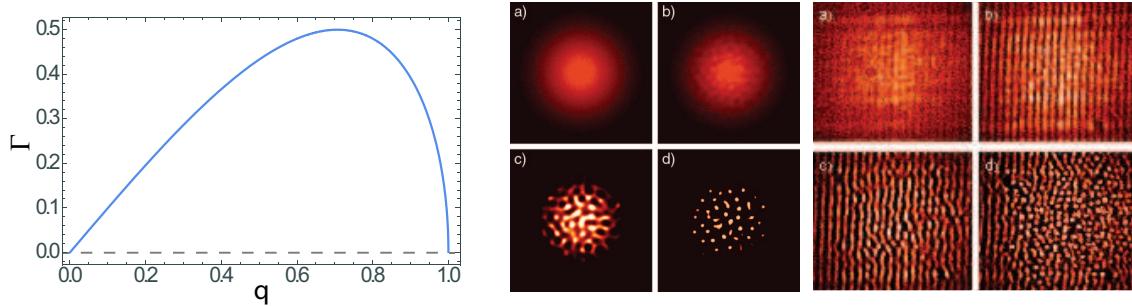


Figure 10.2: Modulation instability gain for Kerr media from Eq. (10.34a) for $2sA_0^2 = 1$. (left) Numerical calculation (center). Observation in a photorefractive crystal from [89] (right).

The linearized perturbations to the refractive index due to the increments a, b are defined by the relation

$$n_2(|A|^2) = n_2(A_0^2) + \delta n_+ e^{\Gamma z + i\mathbf{q} \cdot \mathbf{r}} + \delta n_- e^{\Gamma z - i\mathbf{q} \cdot \mathbf{r}} + O(a^2, b^2). \quad (10.33)$$

For the Kerr nonlinearity $\delta n_+ = A_0^2(a+b^*)$, $\delta n_- = A_0^2(a^*+b)$. For the saturable Kerr nonlinearity the factor A_0^2 should be replaced by $A_0^2/(1+A_0^2)^2$ in the previous expressions. Using Eqs. (10.32) we find

$$\Gamma^2 = q^2 (2sA_0^2 - q^2), \quad (10.34a)$$

$$\Gamma^2 = q^2 \left(2s \frac{A_0^2}{(1+A_0^2)^2} - q^2 \right), \quad (10.34b)$$

for the Kerr and saturable Kerr nonlinearities. Note that for defocusing nonlinearities $s = -1$ and Γ^2 is negative, hence there is no instability. The gain coefficient for Kerr media is shown in Fig. 10.2.

In drift dominated photorefractive media the nonlinearity is proportional to the static electric field created by light induced charge redistribution. The nonlinear increment to the refractive index is given by

$$n_2 = \frac{A_0^2}{1+A_0^2} \left[1 + \frac{\partial \varphi}{\partial x} \right] \quad (10.35)$$

where the equation for the potential φ is[90]

$$\nabla_{\perp}^2 \varphi + \nabla_{\perp} \varphi \cdot \nabla_{\perp} \ln(1+|A|^2) = \frac{\partial}{\partial x} \ln(1+|A|^2). \quad (10.36)$$

Solving a linearized form of Eq. (10.36) we find

$$\delta n_+ = \frac{A_0^2}{(1+A_0^2)^2} \frac{q_x^2}{q^2} (a + b^*), \quad (10.37)$$

$$\delta n_- = \frac{A_0^2}{(1+A_0^2)^2} \frac{q_x^2}{q^2} (a^* + b), \quad (10.38)$$

where $\mathbf{q} = q_x \hat{x} + q_y \hat{y}$. The instability growth rate then takes the form

$$\Gamma^2 = q^2 \left(2s \frac{A_0^2}{(1+A_0^2)^2} \frac{q_x^2}{q^2} - q^2 \right). \quad (10.39)$$

and for the photorefractive nonlinearity

$$h^2 = k_{\perp}^2 \left[\frac{A_0^2}{1 + A_0^2} \frac{k_{\perp x}}{k_{\perp}^2} (k_{\perp x} + i \frac{2A_0}{1 + A_0^2}) - \frac{k_{\perp}^2}{4} \right] \quad (10.40)$$

In the limit of a single transverse dimension (no variation along y) Eq. (10.36) can be integrated and the photorefractive nonlinearity is identical to a saturable Kerr nonlinearity. In this limit Eq.(10.40) also reduces to Eq.(10.34).

Chapter 11

Quadratic nonlinearities

In media without a center of symmetry, i.e. non-centrosymmetric media, the nonlinear $\chi^{(2)}$ coefficient can be nonzero. If there is inversion symmetry then $\chi^{(2)}$ is identically zero. This is illustrated in Fig. 11.1. Let the polarization P_i due to fields E_j, E_k be positive so $P_i = +\epsilon_0 \chi_{ijk}^{(2)} E_j E_k$ in a fixed laboratory frame. Inverting the fields and the crystal the interaction geometry is the same, so the polarization is inverted in the lab frame, giving $P_i = -\epsilon_0 \chi_{ijk}^{(2)} (-E_j) (-E_k)$. If the crystal is centrosymmetric, inverting the crystal again does not change the $\chi^{(2)}$ tensor. Comparing a) and c) in the figure we conclude that

$$\chi_{ijk}^{(2)} E_j E_k = -\chi_{ijk}^{(2)} E_j E_k$$

which implies $\chi_{ijk}^{(2)} = 0$.

Media with nonzero $\chi^{(2)}$ are also referred to as quadratic nonlinear media[91]. They enable a range of wavemixing effects including second harmonic generation (SHG), sum frequency generation (SFG), difference frequency generation (DFG), optical parametric amplification (OPA), optical parametric oscillation (OPO), degenerate optical parametric oscillation (DOPO), and non-degenerate optical parametric oscillation (NDOPO). These effects are widely used for generating tunable sources of coherent light and for generating light beams with quantum optical correlations. We will focus attention on the case of SHG.

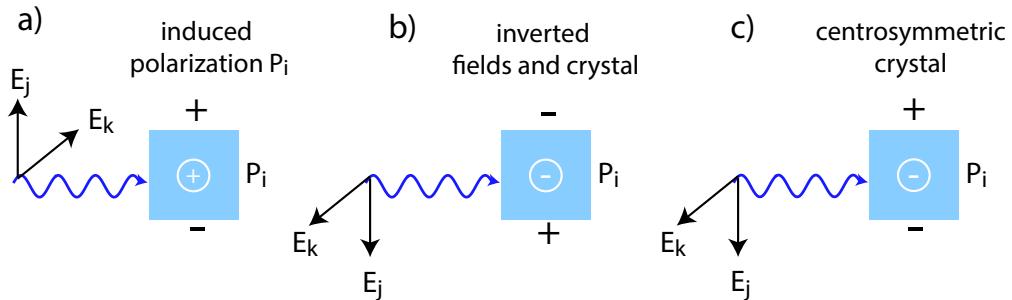


Figure 11.1: Polarization P_i is induced in a medium with quadratic nonlinearity due to fields E_j, E_k . a) the polarization is positive. b) after inverting the fields and the crystal the polarization is negative. c) inverting a centrosymmetric crystal does not change the polarization.

11.1 Parabolic equations describing propagation of parametrically coupled waves

To derive the envelope equations describing propagation of fundamental and second harmonic beams we start with the Maxwell equations

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (11.1)$$

$$\nabla \times \frac{1}{\mu_0} \mathbf{B} = \frac{\partial \mathbf{D}}{\partial t} \quad (11.2)$$

where \mathbf{E} is the electric field, \mathbf{B} is the magnetic displacement, μ is the magnetic permeability (assumed equal to its value in vacuum μ_0) and the electric displacement \mathbf{D} is defined through the constitutive relation

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (11.3)$$

where \mathbf{P} is the electric polarization.

Taking the curl of (11.1a) we have

$$\nabla^2 \mathbf{E} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2}, \quad (11.4)$$

using $\nabla \cdot \mathbf{E} = 0$.¹ The polarization consists of linear and nonlinear contributions

$$\mathbf{P} = \epsilon_0 \chi_L \mathbf{E} + \mathbf{P}_{NL}. \quad (11.5)$$

Substituting into (11.4) we have

$$\nabla^2 \mathbf{E} = \mu_0 \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu_0 \frac{\partial^2 \mathbf{P}_{NL}}{\partial t^2}, \quad (11.6)$$

where $\epsilon \equiv \epsilon_0(1 + \chi_L)$.

11.1.1 Second harmonic generation

The structure of \mathbf{P}_{NL} depends on the physical situation under consideration. We will consider the coupled propagation of fundamental and second harmonic waves of the form

$$E_1(\mathbf{r}, t) = \frac{\mathcal{E}_1(\mathbf{r}, t)}{2} e^{i(k_1 z - \omega_1 t)} + c.c. \quad (11.7)$$

$$E_2(\mathbf{r}, t) = \frac{\mathcal{E}_2(\mathbf{r}, t)}{2} e^{i(k_2 z - \omega_2 t)} + c.c., \quad (11.8)$$

where $\mathbf{r} = (x, y, z)$. Note that when the fields are written in this form the irradiance in units of W/m^2 is given by $I_i = \frac{\epsilon_0}{2} n_i c |\mathcal{E}_i|^2$.

¹In media with no free charges $\nabla \cdot \mathbf{E} = -(\nabla \ln \epsilon) \cdot \mathbf{E}$, where ϵ is the linear part of the dielectric constant. Thus the assumption of a divergenceless electric field is reasonable when the linear properties of the medium change slowly.

Mixing of two fields in a medium with quadratic nonlinearity gives rise to a polarization

$$P_{\text{NL}_i}^{\omega_i=\omega_j \pm \omega_k} = \chi^{(2)\omega_i=\omega_j \pm \omega_k} E_j E_k = \epsilon_0 2 d^{\omega_i=\omega_j \pm \omega_k} E_j E_k \quad (11.9)$$

where $d^{\omega_2=\omega_1+\omega_1}$ is the nonlinear susceptibility and, for simplicity, we have ignored the vectorial nature of the fields. Note the factor of 2 in the definition of d .

In transparent materials the susceptibility is independent of the optical frequency so we may write

$$d^{\omega_2=\omega_1+\omega_1} = d^{\omega_1=\omega_2-\omega_1} \equiv d. \quad (11.10)$$

The total electric field is $E = E_1 + E_2$ so the scalar polarization may be written as

$$\begin{aligned} P_{\text{NL}} &= \epsilon_0 2d(E_1 + E_2)(E_1 + E_2) \\ &= \epsilon_0 2d(E_1^2 + 2E_1 E_2 + E_2^2). \end{aligned} \quad (11.11)$$

Substituting (11.8) in (11.11) we have

$$P_{\text{NL}} = \epsilon_0 2d \left\{ \frac{\mathcal{E}_1^2}{4} e^{i(2k_1 z - 2\omega_1 t)} + \frac{\mathcal{E}_1^* \mathcal{E}_2}{2} e^{i[-k_1 + k_2]z - (-\omega_1 + \omega_2)t} + \text{additional terms} \right\} \quad (11.12)$$

Using Eqs. (11.8,11.12) in (11.6) and collecting terms that oscillate as $e^{-i\omega_1 t}$ and $e^{-i2\omega_1 t}$ gives

$$\nabla^2 (\mathcal{E}_1 e^{ik_1 z}) + \mu_0 \epsilon \omega_1^2 \mathcal{E}_1 e^{ik_1 z} = -\mu_0 \epsilon_0 2 \omega_1^2 d \mathcal{E}_1^* \mathcal{E}_2 e^{i(-k_1 + k_2)z}, \quad (11.13)$$

$$\nabla^2 (\mathcal{E}_2 e^{ik_2 z}) + \mu_0 \epsilon \omega_2^2 \mathcal{E}_2 e^{ik_2 z} = -\mu_0 \epsilon_0 \omega_2^2 d \mathcal{E}_1^2 e^{i2k_1 z}. \quad (11.14)$$

Assuming slowly varying envelopes $\left(\frac{\partial^2 \mathcal{E}_1}{\partial z^2} \ll k_1 \frac{\partial \mathcal{E}_1}{\partial z}, \frac{\partial^2 \mathcal{E}_2}{\partial z^2} \ll k_2 \frac{\partial \mathcal{E}_2}{\partial z}\right)$ Eqs. (11.14) reduce to

$$\frac{\partial \mathcal{E}_1}{\partial z} - \frac{i}{2k_1} \nabla_{\perp}^2 \mathcal{E}_1 = i \frac{\mu_0 \epsilon_0 \omega_1^2 d}{k_1} \mathcal{E}_1^* \mathcal{E}_2 e^{i(-2k_1 + k_2)z}, \quad (11.15)$$

$$\frac{\partial \mathcal{E}_2}{\partial z} - \frac{i}{2k_2} \nabla_{\perp}^2 \mathcal{E}_2 = i \frac{\mu_0 \epsilon_0 \omega_2^2 d}{2k_2} \mathcal{E}_1^2 e^{i(2k_1 - k_2)z}. \quad (11.16)$$

Using $\mu_0 \epsilon_0 = 1/c^2$, where c is the speed of light in vacuum and defining $\Delta k = 2k_1 - k_2$ we have

$$\frac{\partial \mathcal{E}_1}{\partial z} - \frac{i}{2k_1} \nabla_{\perp}^2 \mathcal{E}_1 = i \kappa_1 \mathcal{E}_1^* \mathcal{E}_2 e^{-i\Delta kz}, \quad (11.17)$$

$$\frac{\partial \mathcal{E}_2}{\partial z} - \frac{i}{2k_2} \nabla_{\perp}^2 \mathcal{E}_2 = i \kappa_2 \mathcal{E}_1^2 e^{i\Delta kz}, \quad (11.18)$$

where $\kappa_i = \omega_1 d / c n_i$, with $n_i = c k_i / \omega_i$. For wide beams the terms proportional to ∇_{\perp}^2 can be neglected and the resulting nonlinear ordinary differential equations can be solved exactly in terms of elliptic functions[91].

11.1.2 Cascading

Propagation and coupling of the fundamental and second harmonic waves leads to an effective cubic nonlinearity for the fundamental. This is known as cascading and has been investigated extensively in recent years. To find the size of the effective cubic term neglect diffraction and work in the undepleted pump limit ($|\mathcal{E}_1|^2$ constant). From Eq. (11.18) we then have

$$\frac{\partial \mathcal{E}_2}{\partial z} = i \kappa_2 \mathcal{E}_1^2 e^{i\Delta kz} \quad (11.19)$$

which has solution

$$\mathcal{E}_2(z) = \mathcal{E}_2(0) + \kappa_2 \mathcal{E}_1^2 \left(\frac{e^{i\Delta kz} - 1}{\Delta k} \right). \quad (11.20)$$

Inserting into Eq. (11.17) gives

$$\frac{\partial \mathcal{E}_1}{\partial z} = i\kappa_1 \mathcal{E}_1^* \mathcal{E}_2(0) e^{-i\Delta kz} - \kappa_1 \kappa_2 \frac{\sin(\Delta kz/2)}{(\Delta kz/2)} e^{-i\Delta kz/2} |\mathcal{E}_1|^2 \mathcal{E}_1, \quad (11.21)$$

$$= i\kappa_1 \mathcal{E}_1^* \mathcal{E}_2(0) e^{-i\Delta kz} - \kappa_1 \kappa_2 z \frac{\sin(\Delta kz/2)}{(\Delta kz/2)} e^{-i\Delta kz/2} |\mathcal{E}_1|^2 \mathcal{E}_1. \quad (11.22)$$

The imaginary part of the right hand side of the last equation is equal to $(k_1/n_1)n_2 I \mathcal{E}_1$ (compare (10.12)) whence for an interaction of length L

$$n_2^{SHG} = \frac{2\kappa_1 \kappa_2 L}{k_1 \epsilon_0 c} \frac{\sin(\Delta kL/2)}{(\Delta kL/2)} \sin(\Delta kz/2). \quad (11.23)$$

Choosing $\Delta kL = \pi$ gives

$$n_2^{SHG} = \frac{4\kappa_1 \kappa_2 L}{\pi k_1 \epsilon_0 c}. \quad (11.24)$$

11.2 Conversion efficiency

Solution of Eqs. (11.17) shows that energy can be converted from an optical wave at frequency ω_1 to a wave at the second harmonic frequency. Consider the plane wave limit where Eqs. (11.17) reduce to

$$\frac{\partial \mathcal{E}_1}{\partial z} = i\kappa_1 \mathcal{E}_1^* \mathcal{E}_2 e^{-i\Delta kz}, \quad (11.25)$$

$$\frac{\partial \mathcal{E}_2}{\partial z} = i\kappa_2 \mathcal{E}_1^2 e^{i\Delta kz}, \quad (11.26)$$

where $\kappa_i = \omega_1 d/(cn_i)$. In the weak conversion approximation \mathcal{E}_1 is constant and Eq. (11.26) can be integrated to give

$$I_2(L) = \frac{2\omega_1^2 d^2}{\epsilon_0 c^3 n_1^2 n_2} I_1^2 L^2 \text{sinc}^2\left(\frac{\Delta kL}{2}\right), \quad (11.27)$$

where $\text{sinc}(x) \equiv \sin(x)/x$ and assuming zero input at the second harmonic.

When the input beam is focused tightly it is necessary to use corrections to (11.27) that depend on the beam waist. The correction is known as a Boyd-Kleinman factor[92]. For wide gaussian beams the relation between input power at the fundamental $P_1(0)$ and output power at the second harmonic $P_2(L)$ is found as follows. We have $P_2 = 2\pi \int_0^\infty dr r I_2(r) = b \int_0^\infty dr r I_1^2(r)$, where $b = 2\pi(2\omega_1^2 d^2 / (\epsilon_0 c^3 n_1^2 n_2)) L^2 \text{sinc}^2(\frac{\Delta kL}{2})$. We thus find

$$P_2(L) = P_1^2(0) \frac{1}{\pi w^2} \frac{2\omega_1^2 d^2}{\epsilon_0 c^3 n_1^2 n_2} L^2 \text{sinc}^2\left(\frac{\Delta kL}{2}\right). \quad (11.28)$$

We see that high conversion efficiency requires a small value of the phase mismatch Δk . This can be expressed as

$$\Delta k = 2k_1 - k_2 = 2\pi \left(\frac{2n_1}{\lambda_1} - \frac{n_2}{\lambda_2} \right) = \frac{4\pi}{\lambda_1} (n_1 - n_2).$$

Here λ_1, λ_2 are the vacuum wavelengths at ω_1, ω_2 . We see that phase matching requires $n_1 = n_2$. This is not generally possible for waves with the same polarization due to material dispersion.

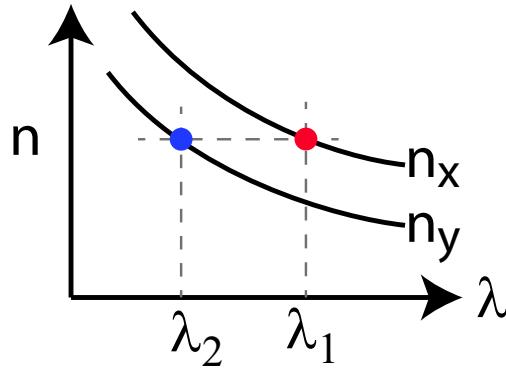


Figure 11.2: Birefringent phase matching. Light at λ_1 is polarized along \hat{x} and light at λ_2 is polarized along \hat{y} .

The most common solution is to use birefringent phase matching in crystals as shown in Fig. ???. Due to the spatial structure of the crystal, waves with different polarization directions have different refractive indices and it is possible to find situations where $n_1 = n_2$ provided the beams are orthogonally polarized. A large number of suitable crystal materials that can be used to satisfy this condition are available[93].

The nonlinear coefficient d can be determined from the measured conversion efficiency through the expression

$$d = \left(\frac{\pi w^2 \epsilon_0 c^3 n_1^2 n_2}{2\omega_1^2} \right)^{\frac{1}{2}} \frac{1}{L \text{sinc}(\frac{\Delta k L}{2})} \frac{\sqrt{P_2(L)}}{P_1(0)}. \quad (11.29)$$

11.3 Boyd-Kleinman theory

When the confocal parameter of the pump beam is comparable to or smaller than the length of the nonlinear medium it is necessary to account for the axial variation of the beam width[92]. We seek gaussian type TEM₀₀ solutions as given by Eq. (6.10) but with axially varying amplitudes of the form

$$\mathcal{E}_1 = A_1(z) \frac{w_1(0)}{w_1(z)} e^{-i\eta_1(z)} e^{i\frac{k_1}{2R_1(z)}r^2} e^{-\frac{r^2}{w_1^2(z)}} = A_1(z) f_1(z, r), \quad (11.30a)$$

$$\mathcal{E}_2 = A_2(z) \frac{w_2(0)}{w_2(z)} e^{-i\eta_2(z)} e^{i\frac{k_2}{2R_2(z)}r^2} e^{-\frac{r^2}{w_2^2(z)}} = A_2(z) f_2(z, r). \quad (11.30b)$$

Inserting (11.30) into (11.18) gives

$$\frac{dA_2(z)}{dz} = i\kappa_2 A_1^2(z) \frac{f_1^2(z, r)}{f_2(z, r)} e^{i\Delta kz}. \quad (11.31)$$

In order to integrate Eq. (11.31) we require that the right hand side is independent of the radial coordinate r . Given a fundamental beam \mathcal{E}_1 Eq. (11.31) is not sufficient to determine the spatial mode of the generated harmonic. We expect that the harmonic mode will be such that the relative phase of the two fields varies at most slowly as the beams propagate axially. If this assumption were not satisfied there would be rapidly varying phase matching resulting in a weak interaction.

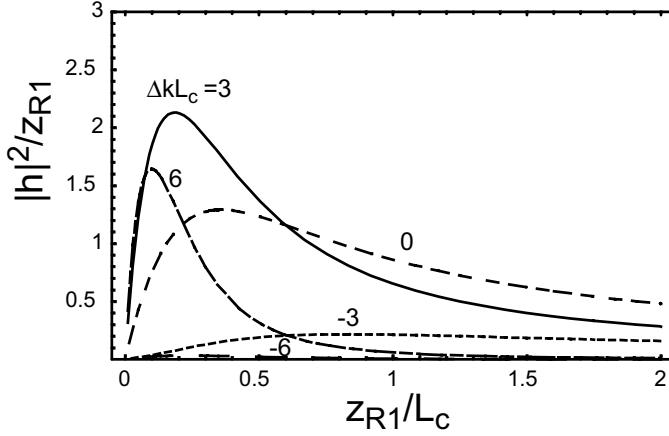


Figure 11.3: Integral appearing in Boyd-Kleinman theory describing harmonic generation as a function of the dimensionless parameters $\Delta k L_c$ and L_{R1}/L_c for $L_c = 1$.

We therefore assume that the generated harmonic beam has the same radius of curvature as the pump beam, or $R_2(z) = R_1(z)$ which implies $L_{R2} = L_{R1}$ and $w_{02}^2 = w_{01}^2/2 + w_{01}^2(\Delta k \lambda_2/(4\pi n_2))$.

We have that

$$\frac{f_1^2}{f_2} = \frac{w_{01}^2 w_2(z)}{w_{02} w_1^2(z)} e^{-i(2\eta_1(z) - \eta_2(z))} e^{i\left(\frac{k_1}{R_1(z)} - \frac{k_2}{2R_2(z)}\right)r^2} e^{-\left(\frac{s_1^2}{w_1^2(z)} - \frac{1}{w_2^2(z)}\right)r^2} \quad (11.32)$$

which reduces, in light of the above ansatz, to

$$\frac{f_1^2}{f_2} = \frac{1}{\left(1 + \frac{z^2}{L_{R1}^2}\right)^{1/2}} e^{-i\eta_1(z)} e^{is_1 r^2} e^{-s_2 r^2} \quad (11.33)$$

with $s_1 = \Delta k/(2R_1(z))$ and $s_2 = 2/w_1^2(z) - 1/(w_1^2(z)/2 + w_1^2(z)(\Delta k \lambda_2/(4\pi n_2)))$. Since the minimum radius of curvature is order the Rayleigh range we have that the maximum values of the r dependent terms are bounded by $s_1 r^2 \sim s_2 r^2 \sim O(\Delta k \lambda_1 r^2 / w_{01}^2)$. The energy of the beams is confined to a transverse radius of order w_1 and optimum values of the phase mismatch satisfy $\Delta k L_c \sim \pi$ so the corrections are $O(\lambda_1/L_c)$ and can be neglected for typical experimental conditions.

The amplitude of the harmonic wave is then found from integrating

$$\frac{dA_2(z)}{dz} = i\kappa_2 A_1^2(z) \frac{e^{i[\Delta kz - \eta_1(z)]}}{\left(1 + \frac{z^2}{L_{R1}^2}\right)^{1/2}} = i\kappa_2 A_1^2(z) \frac{e^{i\Delta kz}}{1 + i\frac{z}{L_{R1}}} \quad (11.34)$$

which gives in the undepleted pump limit

$$A_2(z) = i\kappa_2 A_1^2 h(z) \quad (11.35)$$

with

$$h(z) = \int_{z_0}^z dz' \frac{e^{i\Delta kz'}}{1 + i\frac{z'}{L_{R1}}}. \quad (11.36)$$

The energy conversion efficiency is governed by the integral h across a crystal of length L_c . The integral is maximized when the beam is focused with its waist at the center of the crystal. The

generated power in the harmonic wave is proportional to $|A_2|^2 w_{01}^2$ and the primary figure of merit is the power conversion efficiency which is proportional to $|A_2|^2 w_{01}^2 / (|A_1|^4 w_{01}^4) \sim |h|^2 / w_{01}^2 \sim |h|^2 / L_{R1}$ which is plotted in Fig. 11.3 as a function of the dimensionless ratio L_{R1}/L_c for several values of phase mismatch. Numerical analysis shows that the conversion is maximized for $\Delta k L_c = 3.255$ and $L_{R1} = L_c/5.675$ or $w_{10}^2 = \lambda_1 L_c / (5.675 \pi n_1)$.

The power conversion efficiency is found by manipulations similar to those leading to Eq. (11.28). We find for symmetric focusing

$$E_{NL} \equiv \frac{P_2(L_c/2)}{P_1^2} = \frac{2\omega_1^2 d^2}{\epsilon_0 c^3 n_1^2 n_2} \frac{L_c^2}{\pi w_{10}^2} |\tilde{h}|^2 [\text{W}^{-1}], \quad (11.37)$$

with $\tilde{h} = \int_{-1/2}^{1/2} d\xi \frac{e^{i\Delta k L_c \xi}}{1 + i \frac{L_c}{L_{R1}} \xi}$. In the weak focusing limit Eq. (11.37) should give the same result as Eq. (11.28). Indeed for $L_{R1} \rightarrow \infty$ we find $|\tilde{h}|^2 = \text{sinc}^2(\Delta k L_c / 2)$ so that Eqs. (11.37) and (11.28) agree.

Using the optimum values of phase mismatch and focusing found above, for which $(L_c/L_{R1})|\tilde{h}|^2 = 2 \times 1.068$ we find the maximum possible conversion efficiency obtained for optimum focusing is

$$E_{NL} = 1.068 \left(\frac{2\omega_1^3 d^2 L_c}{\pi \epsilon_0 c^4 n_1 n_2} \right) [\text{W}^{-1}]. \quad (11.38)$$

Equation (11.38) agrees with Eq. (2.22) of Boyd and Kleinman's original paper in the limit where absorption and beam walk-off are neglected. As an example, a 1 cm long a-cut KNbO₃ crystal used for frequency doubling of 0.86 μm light has $d_{\text{eff}} \simeq 11$, and $E_{NL} = 0.023 \text{ W}^{-1}$. A 100 mW pump beam will thus produce 0.2 mW of 0.43 μm radiation in a single pass.

11.3.1 Boyd-Kleinman theory with walk-off

In the absence of walk-off Boyd-Kleinman theory[92] gives a power conversion efficiency for second harmonic generation of (11.37)

$$E_{NL} \equiv \frac{P_2(L_c/2)}{P_1^2} = \frac{2\omega_1^2 d^2}{\pi \epsilon_0 c^3 n_1^2 n_2} \frac{L_c^2}{w_{10}^2} \tilde{h} [\text{W}^{-1}]$$

with

$$\tilde{h} = \left| \int_{-1/2}^{1/2} d\xi \frac{e^{i\Delta k L_c \xi}}{1 + i \frac{L_c}{L_{R1}} \xi} \right|^2. \quad (11.39)$$

Here w_{10} is the waist of the fundamental beam. Numerical calculation of \tilde{h} shows that the optimum conversion is obtained for $\Delta k L_c = 3.255$ and $L_{R1} = L_c/5.675$ or $w_{10}^2 = \lambda_1 L_c / (5.675 \pi n_1)$. We have rewritten the expressions for E_{NL} and \tilde{h} slightly in relation to those in the previous section in order to facilitate comparison with the situation when there is walk-off.

This equation has to be modified in the presence of walk-off of the fundamental beam. Walk-off, or double refraction, is characterized by a dimensionless parameter

$$B = \frac{\sqrt{k_1 L_c}}{2} \delta$$

where $k_1 = 2\pi n_1 / \lambda_1$ and δ is the walk-off angle. In the presence of walk-off the Boyd-Kleinman integral is

$$\tilde{h} = \int_{-1/2}^{1/2} d\xi' \int_{-1/2}^{1/2} d\xi \frac{e^{i\Delta k L_c (\xi - \xi')}}{1 + i \frac{L_c}{L_{R1}} \xi} \frac{e^{-\frac{2L_c}{L_{R1}} B^2 (\xi - \xi')^2}}{1 - i \frac{L_c}{L_{R1}} \xi'}. \quad (11.40)$$

When $B = 0$ Eq. (11.40) reduces to (11.39).

Consider a numerical example of an LBO crystal with $L_c = 2$ cm at $\lambda_1 = 0.821 \mu\text{m}$. We have $n_1 = n_2 = 1.57$, $d_{\text{eff}} = -0.85 \text{ pm/V}$ and ignoring walk-off the optimum focusing condition is $w_{10} = 24.2 \mu\text{m}$, $\Delta k = 163. \text{ m}^{-1}$ giving $E_{\text{NL}} = 6.7 \times 10^{-4}$. An input power of 100 mW should give an output of $6.7 \mu\text{W}$.

For LBO we have $\delta = 11.3 \text{ mrad}$ and the walk-off parameter for the 2 cm crystal is $B = 2.77$. With walk-off, and the same focusing and Δk values we get $E_{\text{NL}} = 1.4 \times 10^{-4}$ so an input power of 100 mW would give an output of $1.4 \mu\text{W}$. However, the optimum focusing with walk-off is achieved with a larger waist. This naively compensates for the loss of beam overlap due to the walk-off. Numerical searching yields optimal parameters of $\Delta k L_c = 2.4$ and $w_{10} = 32. \mu\text{m}$ giving $E_{\text{NL}} = 1.54 \times 10^{-4}$ so an input power of 100 mW would give an output of $1.54 \mu\text{W}$.

For a 1 cm LBO crystal we find the optimum parameters are $\Delta k L_c = 2.45$, $w_{10} = 22.5 \mu\text{m}$ giving $E_{\text{NL}} = 1.05 \times 10^{-4}$. Note that the difference in E_{NL} for 1 and 2 cm crystals is only 50%. This is due to the limitations imposed by walk-off. It is likely that some improvement in E_{NL} could be obtained by moving the pump beam focus away from the center of the crystal towards the input face. This degree of freedom is included in the Boyd and Kleinman analysis but has not been included here.

11.4 Quasi-Phase Matching

Although birefringent phase matching is widely used there are limitations. It is not always possible to find suitable crystals that will phase match at the desired wavelengths. In addition it is usually not possible to use a geometry that gives access to the largest nonlinear coefficients. A more general solution is to use quasi phase matching (QPM) as depicted in Fig. 11.4. In QPM the crystal has its properties modulated with a spatial period Λ which compensates for the momentum mismatch of the interacting waves. Although the idea of QPM dates back to early work on nonlinear optics[91] it only became widely used in the 1990s following advances in crystal processing technology[94].

For quasi-phase matched SHG the phase matching condition is

$$\frac{2n_1}{\lambda_1} - \frac{n_2}{\lambda_2} = \frac{m}{\Lambda}, \quad (11.41)$$

where $n_{1,2}$ is the refractive index at the fundamental and second-harmonic, $\lambda_{1,2}$ are the corresponding wavelengths, Λ is the poling period of the quasi-phase matching, and m is the order of

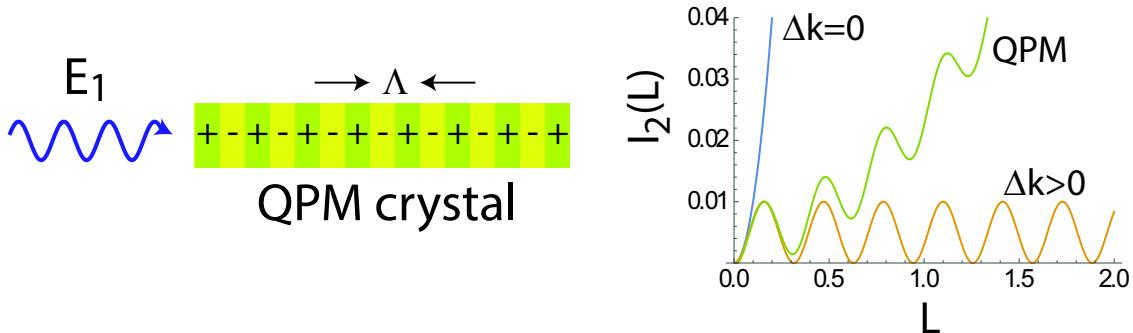


Figure 11.4: A quasi-phase matched crystal is periodically poled with period Λ which results in changing the sign of the $\chi^{(2)}$ coefficient every half period.

the phase matching. Equation (11.41) ensures that the momentum mismatch Δk is compensated by the QPM momentum $K = m2\pi/\Lambda$. Alternatively we recognize that as a function of propagation distance the fundamental and second harmonic waves acquire π relative phase after a distance

$$l_\pi = \frac{\pi}{\Delta k} = \frac{\pi}{2\pi \left(\frac{2n_1}{\lambda_1} - \frac{n_2}{\lambda_2} \right)} = \frac{1}{2} \frac{\Lambda}{m}.$$

The phase mismatch is compensated for in first order phase matching, $m = 1$, by flipping the sign of $\chi^{(2)}$ every $\Lambda/2m$, so the necessary period is $\Lambda = 2ml_\pi$. Although first order phase matching is the most efficient, it is not always possible to make Λ small enough to fully compensate Δk . Instead matching at $m > 1$ allows a larger Λ to be used. This technique can even be used to phase match counterpropagating fundamental and harmonic beams[95], which can lead to interesting spatial instability effects[96, 97].

Let's look at a numerical example of conversion efficiency using periodically poled LiNbO₃. Assuming weak focusing ($\pi w^2/\lambda \gg L$) the output power found from Eq. (11.28) is

$$P_2(L) = P_1^2(0) \frac{1}{\pi w^2} \frac{2\omega_1^2 d_{\text{eff}}^2}{\epsilon_0 c^3 n_1^2 n_2} L^2 \text{sinc}^2\left(\frac{\Delta k L}{2}\right).$$

In congruent LiNbO₃ $d_{33} = -27.2$ pm/V so $d_{\text{eff}} = (2/\pi)d_{33} = -17.3$ pm/V. The effective nonlinearity d_{eff} is less than the actual material coefficient if the poling has a square wave profile, since it is only the amplitude of the fundamental component of the square wave that leads to phase matching.

Take $P_1(0) = 0.1$ W, $L = 2$ cm, $w = 75$ μm, $\lambda_1 = 0.96$ μm, $\Delta k = 0$, and $n_1 \approx n_2 \approx 2.4$, to get $\pi w^2/\lambda = 18.4$ mm and

$$P_2(L) = 0.16 \text{ mW}.$$

Using the optimum Boyd-Kleinman conditions[92] which are $w = \sqrt{\lambda_1 L / (2\pi n)}(1/2.84) = 12.5$ μm and $\Delta k = 3.2/L$ we get

$$P_2(L) = 2.2 \text{ mW}.$$

Say the input source is a 100 mW laser diode and we have 40 mW available at the doubling crystal. Then

$$P_2(L) = 0.35 \text{ mW}$$

using Boyd-Kleinman focusing.

11.5 Intracavity conversion efficiency

We see from the previous examples that the conversion efficiency with low power cw sources is limited to a few percent. Much higher efficiency at low pump power can be obtained by resonant enhancement of the fundamental power in an optical cavity[98, 99]. In order to optimize the conversion efficiency the cavity should be impedance matched by setting the input coupler transmission equal to the linear and nonlinear intracavity losses. Recent advances have demonstrated conversion efficiency as high as 95% [100].

We assume a ring cavity with input coupler transmission T and additional internal linear losses \mathcal{L}_0 . The crystal has conversion efficiency E_{NL} . The input pump power is P_{in} and the intracavity fundamental power is P_1 . When the fundamental is on resonance we can write the intracavity power as

$$P_1 = P_{in} \frac{T}{\left[1 - \sqrt{(1-T)(1-\mathcal{L})}\right]^2}, \quad (11.42)$$

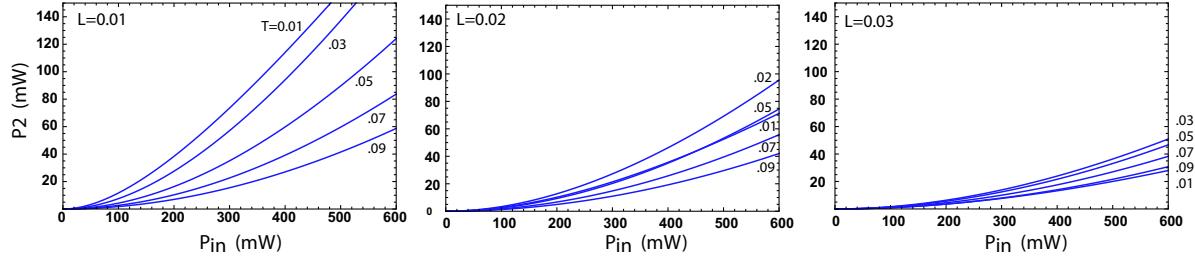


Figure 11.5: Intracavity SHG for different linear loss L and input coupler transmission T . Parameters are for a 2 cm LBO crystal at $0.821 \mu\text{m}$ and optimum focusing which gives $E_{NL} = 0.00015$.

where $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{NL}$ accounts for both linear and nonlinear losses.

The generated harmonic is given by $P_2 = E_{NL}P_1^2$ and we define the conversion efficiency $\epsilon = P_2/P_{in} = E_{NL}P_1^2/P_{in}$. We can therefore write the intracavity power as $P_1 = \sqrt{\epsilon P_{in}/E_{NL}}$. One round trip inside the cavity, excluding the input coupler, results in

$$P_1 \rightarrow (P_1 - E_{NL}P_1^2)(1 - \mathcal{L}_0).$$

Thus

$$\begin{aligned} 1 - \mathcal{L} &= \frac{(P_1 - E_{NL}P_1^2)(1 - \mathcal{L}_0)}{P_1} \\ &= (1 - \sqrt{\epsilon E_{NL}P_{in}})(1 - \mathcal{L}_0) \end{aligned} \quad (11.43)$$

and

$$\begin{aligned} \sqrt{1 - \mathcal{L}} &= \sqrt{1 - \mathcal{L}_0} \sqrt{1 - \sqrt{\epsilon E_{NL}P_{in}}} \\ &\simeq \frac{1}{2} \left(2 - \mathcal{L}_0 - \sqrt{\epsilon E_{NL}P_{in}} \right) \end{aligned} \quad (11.44)$$

where we have used the approximations of $\mathcal{L}_0 \ll 1$, $\sqrt{\epsilon E_{NL}P_{in}} \ll 1$. Combining Eqs. (11.42,11.44) we arrive at

$$\sqrt{\epsilon} = \frac{4T\sqrt{E_{NL}P_{in}}}{[2 - \sqrt{1 - T}(2 - \mathcal{L}_0 - \sqrt{\epsilon E_{NL}P_{in}})]^2}. \quad (11.45)$$

This is a cubic equation for $\sqrt{\epsilon}$ which can be solved in closed form.

As an example we show in Fig. 11.5 the conversion efficiency for a ring cavity with a LBO crystal. When the intracavity linear loss is small ($\mathcal{L}_0 = 0.01$) the conversion is optimized for low input coupler transmission ($T = 0.01$). With higher intracavity loss the optimum conversion occurs for a higher transmission. For moderate and high conversion efficiencies the optimum is when $T \simeq \mathcal{L}$ so that the cavity is impedance matched accounting for both linear and nonlinear (conversion) losses.

11.6 Normalized equations

In addition to SHG other wave mixing situations can be studied. For the subsequent analysis it is convenient to write the propagation equations in normalized form. Put $\mathcal{E}_1 = h_1 A_1$, $\mathcal{E}_2 = h_2 A_2 e^{i\Delta kz}$,

$z = h_3 \tilde{z}$, and $\mathbf{r} = \sqrt{h_4} \tilde{\mathbf{r}}$, where the scalings h_i are to be determined. Substituting into Eqs. (11.17) gives

$$\frac{\partial A_1}{\partial \tilde{z}} - \frac{i}{2k_1} \frac{h_3}{h_4} \tilde{\nabla}_\perp^2 A_1 = i \frac{\omega_1 d}{cn_1} \frac{h_1^* h_2 h_3}{h_1} A_1^* A_2, \quad (11.46)$$

$$\frac{\partial A_2}{\partial \tilde{z}} - \frac{i}{2k_2} \frac{h_3}{h_4} \tilde{\nabla}_\perp^2 A_2 + i\Delta k h_3 A_2 = i \frac{\omega_1 d}{cn_2} \frac{h_1^2 h_3}{h_2} A_1^2. \quad (11.47)$$

We will use the scalings

$$\frac{h_3}{2k_1 h_4} = 1, \quad \frac{\omega_1 d h_1^* h_2 h_3}{cn_1 h_1} = -i, \quad \frac{\omega_1 d h_1^2 h_3}{cn_2 h_2} = i. \quad (11.48)$$

There is no unique choice for the h_i . One possibility is

$$h_1 = \frac{c}{L\omega_1 d} \sqrt{n_1 n_2}, \quad h_2 = -i \frac{cn_1}{L\omega_1 d}, \quad h_3 = L, \quad h_4 = \frac{L}{2k_1}, \quad (11.49)$$

where L is a characteristic length scale. This leads, dropping the tildes, to the scaled equations

$$\frac{\partial A_1}{\partial z} - i\nabla_\perp^2 A_1 = A_1^* A_2, \quad (11.50)$$

$$\frac{\partial A_2}{\partial z} - i \frac{n_1}{2n_2} \nabla_\perp^2 A_2 + i\beta A_2 = -A_1^2, \quad (11.51)$$

where $\beta = \Delta k L$. For future reference we note the conversion factors between intensities and scaled amplitudes: $|A_1| = (L\omega_1 d / (c\sqrt{n_1 n_2})) \sqrt{2I_1 / (\epsilon_0 n_1 c)}$ and $|A_2| = (L\omega_1 d / (cn_1)) \sqrt{2I_2 / (\epsilon_0 n_2 c)}$. The scaled amplitudes are inside a medium with refractive index different from unity. The scaled amplitudes in vacuum are a factor \sqrt{n} larger.

11.6.1 Constant pump solution

Consider the situation where the second harmonic amplitude is zero at $z = 0$. In the limit of weak conversion efficiency we may assume that the pump amplitude at the fundamental frequency changes very slowly with z . We can then solve Eq. (11.50) for the second harmonic field as a function of z , including the effects of diffraction of the second harmonic. Using Fourier transforms it is straightforward to show that the formal solution of Eq. (11.51) is

$$A_2(z) = i\mathcal{F}^{-1} \left[\frac{1 - e^{-i(\beta + \frac{1}{2}k_\perp^2)z}}{\beta + \frac{1}{2}k_\perp^2} \mathcal{F}[A_1^2] \right], \quad (11.52)$$

where \mathcal{F} denotes a Fourier transform in $\mathbf{r} = (x, y)$, and \mathbf{k}_\perp is the transform variable. When $A_1 = A_1^{(0)}$ has no spatial structure, its transform is proportional to $\delta(\mathbf{k}_\perp)$, and we obtain

$$A_2^{(0)}(z) = i \frac{1 - e^{-i\beta z}}{\beta} A_1^{(0)2}. \quad (11.53)$$

For arbitrary A_1 a solution is only possible numerically. When the modulation of A_1 is limited to a finite band of spatial frequencies we can obtain an approximate solution by dropping spatial

derivatives that are higher than 2nd order in \mathbf{r} . This is tantamount to expanding the right hand side of Eq. (11.52) to order k_{\perp}^2 which gives

$$\begin{aligned} A_2(z) &= i\mathcal{F}^{-1} \left[\left(\frac{1 - e^{-i\beta z}}{\beta} + \frac{i\beta z e^{-i\beta z} - 1 + e^{-i\beta z}}{2\beta^2} k_{\perp}^2 \right) \mathcal{F}[A_1^2] \right], \\ &= i \frac{1 - e^{-i\beta z}}{\beta} A_1^2 + i \frac{1 - (1 + i\beta z)e^{-i\beta z}}{2\beta^2} \nabla_{\perp}^2 A_1^2. \end{aligned} \quad (11.54)$$

11.6.2 Effect of a phase shift

In double pass or cavity geometries mirror phase shifts have large effects on the generated second harmonic. Fundamental and harmonic waves propagate from $z = 0 \rightarrow L$. At $z = L$ phase shifts are applied so that $\mathcal{E}_1 \rightarrow \mathcal{E}_1 e^{i\theta_1}$ and $\mathcal{E}_2 \rightarrow \mathcal{E}_2 e^{i\theta_2}$. A second stretch of nonlinear propagation then follows from $z = L \rightarrow 2L$. This describes a double pass geometry where the phases θ_1, θ_2 are due to reflections at a mirror.

Let's calculate the output fields at $z = 2L$ assuming weak nonlinearity. We therefore start by assuming \mathcal{E}_1 is constant (apart from the phase shift θ_1) and calculate \mathcal{E}_2 in this limit. Using the calculated \mathcal{E}_2 we then calculate the first correction to \mathcal{E}_1 .

The propagation equations, without diffraction, are

$$d\mathcal{E}_1/dz = i\kappa e^{-i\Delta kz} \mathcal{E}_1^* \mathcal{E}_2 \quad (11.55)$$

$$d\mathcal{E}_2/dz = i\kappa e^{i\Delta kz} \mathcal{E}_1^2. \quad (11.56)$$

Assuming no input at the harmonic frequency the constant pump solution for \mathcal{E}_2 at $z = L$ is

$$\mathcal{E}_2 = \frac{\kappa \bar{\mathcal{E}}_1^2}{\Delta k} (e^{i\Delta k L} - 1), \quad (11.57)$$

where $\bar{\mathcal{E}}_1$ is the fundamental input field.

We then need to integrate over the interval $L \leq z \leq 2L$. We need to be careful to treat the boundary conditions correctly. Denote the limits approaching L from below (above) as the fields evaluated at L^{\mp} . Then the boundary conditions taking account of the phase jumps are

$$z = L^- \quad \text{fundamental field} = \bar{\mathcal{E}}_1 \quad \mathcal{E}_2 = \frac{\kappa \bar{\mathcal{E}}_1^2}{\Delta k} (e^{i\Delta k L} - 1), \quad (11.58)$$

$$z = L^+ \quad \text{fundamental field} = \bar{\mathcal{E}}_1 e^{i\theta_1} \quad \mathcal{E}_2 = \frac{\kappa \bar{\mathcal{E}}_1^2 e^{i\theta_1}}{\Delta k} (e^{i\Delta k L} - 1). \quad (11.59)$$

The equation for the evolution of the harmonic for $z > L$ is then

$$d\mathcal{E}_2/dz = i\kappa \bar{\mathcal{E}}_1^2 e^{i2\theta_1} e^{i\Delta kz}. \quad (11.60)$$

The solution is

$$\mathcal{E}_2(z) = \frac{\kappa \bar{\mathcal{E}}_1^2 e^{i2\theta_1}}{\Delta k} (e^{i\Delta kz} - e^{i\Delta k L}) + \mathcal{E}_2(L^+) \quad (11.61)$$

$$= \frac{\kappa \bar{\mathcal{E}}_1^2 e^{i2\theta_1}}{\Delta k} (e^{i\Delta kz} - e^{i\Delta k L}) + \frac{\kappa \bar{\mathcal{E}}_1^2 e^{i\theta_2}}{\Delta k} (e^{i\Delta k L} - 1) \quad (11.62)$$

$$= \frac{\kappa \bar{\mathcal{E}}_1^2 e^{i\theta_2}}{\Delta k} [e^{i(2\theta_1 - \theta_2)} (e^{i\Delta kz} - e^{i\Delta k L}) + e^{i\Delta k L} - 1]. \quad (11.63)$$

The solution at $z = 2L$ can be written as $\mathcal{E}_2(2L) = L\kappa g \bar{\mathcal{E}}_1^2$, where

$$g = e^{i\theta_2} \frac{e^{i(2\theta_1-\theta_2)}(e^{i2\xi} - e^{i\xi}) + e^{i\xi} - 1}{\xi}, \quad (11.64)$$

with $\xi = \Delta k L$. For $\theta_1 = \theta_2 = 0$ we get $g = (e^{i2\xi} - 1)/\xi$.

The nonlinear correction to \mathcal{E}_1 at $z = 2L$ is then found by integrating

$$d\mathcal{E}_1/dz = i\kappa e^{-i\Delta kz} \bar{\mathcal{E}}_1^* \mathcal{E}_2(z), \quad (11.65)$$

taking into account the phase jumps. At $z = L^-$ we get

$$\mathcal{E}_1(L^-) = \bar{\mathcal{E}}_1 + i \frac{\kappa^2}{\Delta k} |\bar{\mathcal{E}}_1|^2 \bar{\mathcal{E}}_1 \left[L - \frac{i}{\Delta k} (e^{-i\Delta k L} - 1) \right] \quad (11.66)$$

and at $z = L^+$

$$\mathcal{E}_1(L^+) = \bar{\mathcal{E}}_1 e^{i\theta_1} + i \frac{\kappa^2}{\Delta k} |\bar{\mathcal{E}}_1|^2 \bar{\mathcal{E}}_1 e^{i\theta_1} \left[L - \frac{i}{\Delta k} (e^{-i\Delta k L} - 1) \right] \quad (11.67)$$

The propagation equation for $z > L$ is

$$d\mathcal{E}_1/dz = i\kappa e^{-i\Delta kz} \bar{\mathcal{E}}_1^* e^{-i\theta_1} \mathcal{E}_2(z), \quad (11.68)$$

and the solution can be written as

$$\begin{aligned} \mathcal{E}_1(z) - \bar{\mathcal{E}}_1 e^{i\theta_1} &= i \frac{\kappa^2}{\Delta k} |\bar{\mathcal{E}}_1|^2 \bar{\mathcal{E}}_1 e^{i\theta_1} \left\{ z + \frac{i}{\Delta k} \left[2 - e^{-i\Delta k L} - e^{-i\Delta k(z-L)} \right. \right. \\ &\quad \left. \left. + e^{i(\theta_2-2\theta_1)} \left(e^{-i\Delta k(z-L)} - e^{-i\Delta kz} + e^{-i\Delta k L} - 1 \right) \right] \right\}. \end{aligned} \quad (11.69)$$

The solution at $z = 2L$ is

$$\begin{aligned} \mathcal{E}_1(2L) - \bar{\mathcal{E}}_1 e^{i\theta_1} &= i \frac{\kappa^2}{\Delta k} |\bar{\mathcal{E}}_1|^2 \bar{\mathcal{E}}_1 e^{i\theta_1} \left\{ 2L + \frac{i}{\Delta k} \left[2(1 - e^{-i\Delta k L}) \right. \right. \\ &\quad \left. \left. + e^{i(\theta_2-2\theta_1)} \left(2e^{-i\Delta k L} - e^{-i2\Delta k L} - 1 \right) \right] \right\} \quad (11.70) \end{aligned}$$

$$= L^2 \kappa^2 |\bar{\mathcal{E}}_1|^2 \bar{\mathcal{E}}_1 e^{i\theta_1} \left\{ \frac{2i}{\xi} + \frac{e^{-i\xi} - 1}{\xi^2} \left[2 + e^{i(\theta_2-2\theta_1)} (e^{-i\xi} - 1) \right] \right\}. \quad (11.71)$$

Equation (11.71) gives the correction to the fundamental at $z = 2L$ due to the nonlinearity. When using this result in a meanfield model we do not distinguish between $\bar{\mathcal{E}}_1$ and $\bar{\mathcal{E}}_1 e^{i\theta_1}$. The mean field form of the result is therefore

$$(\Delta \mathcal{E}_1)_{\text{nl}} = L^2 \kappa^2 f |\mathcal{E}_1|^2 \mathcal{E}_1 \quad (11.72)$$

where

$$f = \frac{2i}{\xi} + \frac{e^{-i\xi} - 1}{\xi^2} \left[2 + e^{i(\theta_2-2\theta_1)} (e^{-i\xi} - 1) \right]. \quad (11.73)$$

Neglecting diffraction we have

$$dA_1/dz = i\kappa A_1^* A_2 \quad (11.74)$$

$$dA_2/dz = i\kappa A_1^2 \quad (11.75)$$

Reflection of the fundamental and second harmonic at a mirror, or passage through a dispersive element will introduce a relative phase between them. The phase shift affects the subsequent harmonic conversion.

Put

$$B_1 = A_1 e^{i\phi_1} \quad (11.76)$$

$$B_2 = A_2 e^{i\phi_2} \quad (11.77)$$

Then

$$dB_1/dz = i\kappa B_1^* B_2 e^{i(2\phi_1 - \phi_2)} \quad (11.78)$$

$$dB_2/dz = i\kappa B_1^2 e^{-i(2\phi_1 - \phi_2)} \quad (11.79)$$

In order for the phase shift not to have an effect we need

$$2\phi_1 - \phi_2 = 2m\pi,$$

with m an integer. Thus

$$\phi_1 - \phi_2/2 = m\pi$$

11.6.3 Optical parametric amplification

In non-degenerate parametric amplification a pump at ω_2 interacts with a signal and idler at ω_{\pm} where $\omega_+ + \omega_- = \omega_2$. The governing equations are (need to check the details of the nonlinear coefficients)

$$\frac{\partial \mathcal{E}_+}{\partial z} - \frac{i}{2k_+} \nabla_{\perp}^2 \mathcal{E}_+ = i \frac{\omega_+}{cn_+} d \mathcal{E}_-^* \mathcal{E}_2 e^{-i\Delta kz}, \quad (11.80)$$

$$\frac{\partial \mathcal{E}_-}{\partial z} - \frac{i}{2k_-} \nabla_{\perp}^2 \mathcal{E}_- = i \frac{\omega_-}{cn_-} d \mathcal{E}_+^* \mathcal{E}_2 e^{-i\Delta kz}, \quad (11.81)$$

$$\frac{\partial \mathcal{E}_2}{\partial z} - \frac{i}{2k_2} \nabla_{\perp}^2 \mathcal{E}_2 = 2i \frac{\omega_1}{cn_2} d \mathcal{E}_+ \mathcal{E}_- e^{i\Delta kz}, \quad (11.82)$$

where $\Delta k = k_+ + k_- - k_2$. Note the factor of 2 in the equation for \mathcal{E}_2 . A discussion of the intricacies of correct normalization of nonlinear coefficients can be found in the book by Shen[101].

11.6.4 Combined harmonic generation and parametric amplification

Another type of interaction that is often encountered is second harmonic generation with simultaneous nondegenerate parametric amplification. In this case there are four interacting fields: the fundamental at ω_1 , its second harmonic at $\omega_2 = 2\omega_1$, and the parametric signal and idler pair at ω_{\pm} where $\omega_+ + \omega_- = \omega_2$. In the approximation of equal refractive indices $n_1 = n_2 = n_+ = n_- \equiv n$, and small parametric detuning $|\omega_{\pm} - \omega_1| \ll \omega_1$ all the nonlinear coefficients κ_i are equal and will be denoted as κ . The propagation equations are then

$$\frac{\partial \mathcal{E}_1}{\partial z} - \frac{i}{2k_1} \nabla_{\perp}^2 \mathcal{E}_1 = i\kappa \mathcal{E}_1^* \mathcal{E}_2 e^{-i\Delta kz}, \quad (11.83)$$

$$\frac{\partial \mathcal{E}_+}{\partial z} - \frac{i}{2k_1} \nabla_{\perp}^2 \mathcal{E}_+ = i\kappa \mathcal{E}_-^* \mathcal{E}_2 e^{-i\Delta \tilde{k}z}, \quad (11.84)$$

$$\frac{\partial \mathcal{E}_-}{\partial z} - \frac{i}{2k_1} \nabla_{\perp}^2 \mathcal{E}_- = i\kappa \mathcal{E}_+^* \mathcal{E}_2 e^{-i\Delta \tilde{k}z}, \quad (11.85)$$

$$\frac{\partial \mathcal{E}_2}{\partial z} - \frac{i}{4k_1} \nabla_{\perp}^2 \mathcal{E}_2 = i\kappa \mathcal{E}_1^2 e^{i\Delta kz} + 2i\kappa \mathcal{E}_+ \mathcal{E}_- e^{i\Delta \tilde{k}z}, \quad (11.86)$$

where $\Delta k = 2k_1 - k_2$ and $\Delta \tilde{k} = k_+ + k_- - k_2$.

As in the case of second harmonic generation it is most convenient to work with the equations in normalized form. Put $\mathcal{E}_1 = h_1 A_1$, $\mathcal{E}_2 = h_2 A_2 e^{i\Delta kz}$, $\mathcal{E}_+ = h_+ A_+ e^{i\frac{1}{2}(\Delta k - \Delta \tilde{k})z}$, $\mathcal{E}_- = h_- A_- e^{i\frac{1}{2}(\Delta k - \Delta \tilde{k})z}$, $z \rightarrow Lz$, and $\mathbf{r} \rightarrow \sqrt{L/2k_1} \mathbf{r}$, where the scalings h_i are to be determined. Substituting into Eqs. (11.83) gives

$$\frac{\partial A_1}{\partial z} - i\nabla_{\perp}^2 A_1 = iL\kappa \frac{h_1^* h_2}{h_1} A_1^* A_2, \quad (11.87)$$

$$\frac{\partial A_+}{\partial z} - i\nabla_{\perp}^2 A_+ + i\frac{1}{2}(\beta - \tilde{\beta}) A_+ = iL\kappa \frac{h_-^* h_2}{h_+} A_-^* A_2, \quad (11.88)$$

$$\frac{\partial A_-}{\partial z} - i\nabla_{\perp}^2 A_- + i\frac{1}{2}(\beta - \tilde{\beta}) A_- = iL\kappa \frac{h_+^* h_2}{h_-} A_+^* A_2, \quad (11.89)$$

$$\frac{\partial A_2}{\partial z} - \frac{i}{2} \nabla_{\perp}^2 A_2 + i\beta A_2 = iL\kappa \frac{h_1^2}{h_2} A_1^2 + 2iL\kappa \frac{h_+ h_-}{h_2} A_+ A_-, \quad (11.90)$$

where $\beta = \Delta k L$ and $\tilde{\beta} = \Delta \tilde{k} L$. Again there is no unique choice for the h_i . We will use the scalings

$$h_1 = h_+ = h_- = \frac{1}{L\kappa}, \quad h_2 = -\frac{i}{L\kappa}. \quad (11.91)$$

We then obtain the normalized set

$$\frac{\partial A_1}{\partial z} - i\nabla_{\perp}^2 A_1 = A_1^* A_2, \quad (11.92)$$

$$\frac{\partial A_+}{\partial z} - i\nabla_{\perp}^2 A_+ + i\frac{1}{2}(\beta - \tilde{\beta}) A_+ = A_-^* A_2, \quad (11.93)$$

$$\frac{\partial A_-}{\partial z} - i\nabla_{\perp}^2 A_- + i\frac{1}{2}(\beta - \tilde{\beta}) A_- = A_+^* A_2, \quad (11.94)$$

$$\frac{\partial A_2}{\partial z} - \frac{i}{2} \nabla_{\perp}^2 A_2 + i\beta A_2 = -A_1^2 - 2A_+ A_-. \quad (11.95)$$

In connection with intracavity calculations an alternative form of Eqs. (11.92) will also be useful. Employing the same scalings as in Eq. (11.91), but leaving the exponential phase factors in the definitions of the amplitudes we obtain the set

$$\frac{\partial A_1}{\partial z} - i\nabla_{\perp}^2 A_1 = A_1^* A_2 e^{-i\beta z}, \quad (11.96)$$

$$\frac{\partial A_+}{\partial z} - i\nabla_{\perp}^2 A_+ = A_-^* A_2 e^{-i\tilde{\beta} z}, \quad (11.97)$$

$$\frac{\partial A_-}{\partial z} - i\nabla_{\perp}^2 A_- = A_+^* A_2 e^{-i\tilde{\beta} z}, \quad (11.98)$$

$$\frac{\partial A_2}{\partial z} - \frac{i}{2} \nabla_{\perp}^2 A_2 = -A_1^2 e^{i\beta z} - 2A_+ A_- e^{i\tilde{\beta} z}. \quad (11.99)$$

11.6.5 Constant pump solution with parametric amplification

Consider the situation where the second harmonic amplitude is zero at $z = 0$. In the limit of weak conversion efficiency we may assume that the pump amplitude at the fundamental frequency changes very slowly with z . We can then solve Eq. (11.99) for the second harmonic field as a function of z , including the effects of diffraction of the second harmonic. Using Fourier transforms, and proceeding as in Sec. 11.6.1, it is straightforward to show that the formal solution of Eq. (11.99) is

$$A_2(z) = i\mathcal{F}^{-1} \left[\frac{e^{i\beta z} - e^{-i\frac{1}{2}k_\perp^2 z}}{\beta + \frac{1}{2}k_\perp^2} \mathcal{F}[A_1^2] + 2 \frac{e^{i\tilde{\beta} z} - e^{-i\frac{1}{2}k_\perp^2 z}}{\tilde{\beta} + \frac{1}{2}k_\perp^2} \mathcal{F}[A_+ A_-] \right],$$

where \mathcal{F} denotes a Fourier transform in $\mathbf{r} = (x, y)$, and \mathbf{k}_\perp is the transform variable. When the fields $A_i = A_i^{(0)}$, with $i = 1, +, -$, have no spatial structure, their transforms are proportional to $\delta(\mathbf{k}_\perp)$, and we obtain

$$A_2^{(0)}(z) = i \frac{e^{i\beta z} - 1}{\beta} A_1^{(0)2} + 2i \frac{e^{i\tilde{\beta} z} - 1}{\tilde{\beta}} A_+^{(0)} A_-^{(0)}. \quad (11.100)$$

For pump fields with spatial structure we obtain, (see Eq. (11.54))

$$\begin{aligned} A_2(z) &= i \frac{e^{i\beta z} - 1}{\beta} A_1^2 + 2i \frac{e^{i\tilde{\beta} z} - 1}{\tilde{\beta}} A_+ A_- \\ &\quad - i \frac{1 + i\beta z - e^{i\beta z}}{2\beta^2} \nabla_\perp^2 A_1^2 - 2i \frac{1 + i\tilde{\beta} z - e^{i\tilde{\beta} z}}{2\tilde{\beta}^2} \nabla_\perp^2 (A_+ A_-). \end{aligned} \quad (11.101)$$

Chapter 12

Optical waveguides and fiber optics

Just as wires are used to carry electrical currents around circuits, optical waveguides, defined by spatial variations of the refractive index, can be used to transport light in optical circuits. Optics can also be used for long distance transmission of information. Free space propagation is limited by the need for line of sight, and unavoidable diffractive spreading. The effects of diffraction can be removed by using guided modes in optical fibers which provide the wiring for global telecommunications.

Early work with glass fibers found that absorption and scattering were too large to allow for propagation over more than a few meters. It was realized by Kao and Hockham in 1966[102] that fibers made from fused silica had the potential for very low loss which would enable data transmission over long distances at high rates. The first low loss fibers with attenuation of 7 dB/km were fabricated at Corning in 1970[103]. This was followed by the development of modified chemical vapor deposition at Bell Laboratories in 1974[104] which enabled fabrication of very low loss fibers with attenuation close to 1 dB/km. Modern fibers have reduced this to < 0.2 dB/km in the telecom band near $1.55\text{ }\mu\text{m}$ as seen in Fig. 12.1. The loss increases in the infrared due to molecular absorption bands and in the ultraviolet due to the $1/\lambda^4$ dependence of the Rayleigh scattering cross section of Eq. (9.5). In between there is an absorption minimum near $1.55\text{ }\mu\text{m}$. Despite the low loss, transmission over thousands of km requires reamplification. This has become possible, without detection and regeneration using optical amplifiers. In particular the development of the erbium doped fiber amplifier[105] has made intercontinental fiber optic transmission a reality.

In this chapter we will study the modes of optical waveguides and their dispersion properties which set practical limits on the rate at which data can be transmitted.

12.1 One-dimensional waveguide modes

see notes on class webpage

12.2 Guided modes in a two-dimensional waveguide

see notes on class webpage

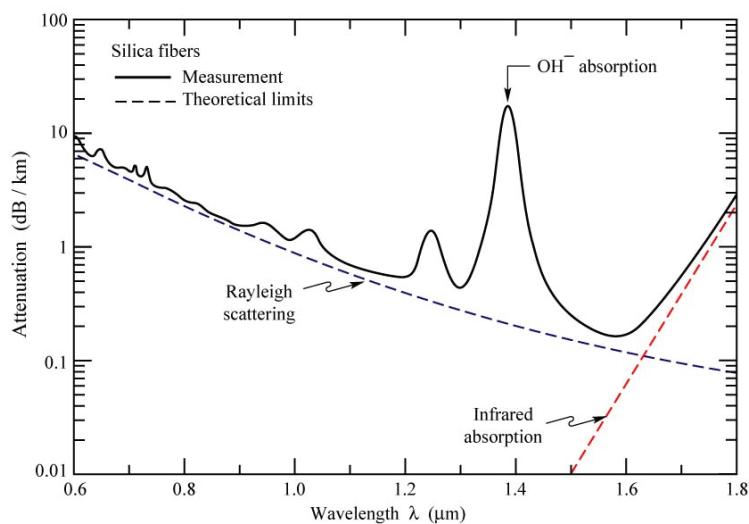


Fig. 22.2. Measured attenuation in silica fibers (solid line) and theoretical limits (dashed lines) given by Rayleigh scattering in the short-wavelength region, and by molecular vibrations (infrared absorption) in the infrared spectral region.

E. F. Schubert
Light-Emitting Diodes (Cambridge Univ. Press)
www.LightEmittingDiodes.org

Figure 12.1: Attenuation of fused silica fibers from E. F. Schubert, Light Emitting Diodes, Cambridge University Press.

Chapter 13

Electro-optics

In this chapter we cover a potpourri of topics in the areas of detectors, modulators, and stabilization schemes. Reference is made to several commercial components for informational purposes only. No recommendation is implied.

Some reference information on electronic components and circuits is provided in Ch. 14.

13.1 Photodetection and signal to noise ratio

Optical detectors convert light into measurable electronic signals. Broadly speaking there are two types of detectors. Thermal detectors that absorb light energy and convert it into measurable heat, and quantum detectors that convert an incident photon into an electrical current. Thermal detectors include thermopiles, and bolometers. Although thermal detectors are most often slow, and not useful for detecting coherence and quantum properties of light, this is not always the case. The superconducting transition edge sensor (TES) uses the heat energy from absorption of a single photon to make a circuit change from a superconducting to a normal state and thereby provides fast and high quantum efficiency photon detection. Originally developed in the 1940s TES detectors have received renewed attention in recent years and are now important components of many quantum optics experiments[106].

Quantum detectors include photovoltaic and photoconductive semiconductor devices, and photomultipliers. There are several possible modes of operation including direct and heterodyne detection. A photovoltaic detector absorbs photons and produces electrical current. The photogenerated current is given by¹

$$I = e\eta \frac{P}{h\nu} G, \quad (13.1)$$

where e is the electronic charge, η is the quantum efficiency of the detector, P is the optical power, h is Planck's constant, ν is the optical frequency, and G is the detector gain. Some detectors such as photodiodes have unit gain. Photomultipliers have secondary emission stages providing large, low noise gain.

Whether or not a signal current I is detectable depends on the noise level and the detection bandwidth. The most important types of electronic noise are shot noise, due to the discreteness of electrical charge, and Johnson noise due to thermal fluctuations. Although these appear to be physically distinct phenomena they can be derived in a unified way[107].

Consider the circuit shown in Fig. 13.1. It consists of a voltage source driving current through a tunnel junction. In time τ there is probability $P_{AB}\tau$ for a charge to pass the junction in the

¹We will use either I or J for photocurrents.

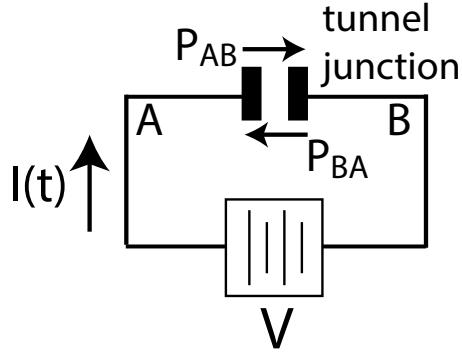


Figure 13.1: Circuit for analyzing shot and thermal noise.

clockwise direction, corresponding to positive current, and probability $P_{BA}\tau$ for a charge to pass the junction in the opposite direction. The integrated current in a time interval τ is

$$\bar{I} = \int_t^{t+\tau} dt I(t).$$

If we measure this current with a sampling frequency $f_s = 1/\tau$ the corresponding bandwidth is $B = f_s/2 = 1/2\tau$. In a short time τ we observe either no current, positive current q/τ , or negative current $-q/\tau$, with q the elementary charge. The relative probabilities are

$$P(\bar{I} = 0) = 1 - (P_{AB} + P_{BA})\tau, \quad P(\bar{I} = q/\tau) = P_{AB}\tau, \quad P(\bar{I} = -q/\tau) = P_{BA}\tau.$$

At each sample time \bar{I} is a random variable and the average current is

$$\begin{aligned} \langle I \rangle &= \sum_k \bar{I}_k P(\bar{I}_k) \\ &= -\frac{q}{\tau} P_{BA}\tau + 0[1 - (P_{AB} + P_{BA})\tau] + \frac{q}{\tau} P_{AB}\tau \\ &= q(P_{AB} - P_{BA}). \end{aligned}$$

The average squared current is

$$\begin{aligned} \langle I^2 \rangle &= \sum_k (\bar{I}_k)^2 P(\bar{I}_k) \\ &= \left(-\frac{q}{\tau}\right)^2 P_{BA}\tau + 0 + \left(\frac{q}{\tau}\right)^2 P_{AB}\tau \\ &= \frac{q^2}{\tau} (P_{AB} + P_{BA}). \end{aligned}$$

Describe the junction as having a resistance R then $\langle I \rangle = V/R$. We can think of a tunneling event as an electron moving between two states: q in A or q in B. The population of charges in these states is n_A and n_B . The number moving from A to B per unit time is

$$N(A \rightarrow B) = n_A P_{AB}$$

and the number moving from B to A per unit time is

$$N(B \rightarrow A) = n_B P_{BA}.$$

In steady state we have $n_A P_{AB} = n_B P_{BA}$. Let state A have energy U_A and state B have energy U_B . Then $U_A - U_B = qV$ and in thermal equilibrium at temperature T

$$\frac{n_A}{n_B} = e^{-qV/k_B T}.$$

Thus

$$\frac{P_{BA}}{P_{AB}} = e^{-qV/k_B T},$$

$$\begin{aligned} \langle I \rangle &= \frac{V}{R} \\ &= q(P_{AB} - P_{BA}) \\ &= qP_{AB} \left(1 - e^{-qV/k_B T}\right), \end{aligned} \quad (13.2)$$

and

$$\begin{aligned} \langle I^2 \rangle &= \frac{q^2}{\tau} (P_{AB} + P_{BA}) \\ &= (2Bq)(qP_{AB})(1 + e^{-qV/k_B T}) \\ &= 2q \frac{V}{R} \frac{1 + e^{-qV/k_B T}}{1 - e^{-qV/k_B T}} B. \end{aligned} \quad (13.3)$$

With Eqs. (13.2,13.3) we can calculate the variance of the current

$$\begin{aligned} \langle (\Delta I)^2 \rangle &\equiv \langle (I - \langle I \rangle)^2 \rangle \\ &= \langle I^2 \rangle - \langle I \rangle^2. \end{aligned}$$

The average current $\langle I \rangle$ will be assumed to be a constant in which case the experimentally relevant variance is simply $\langle (\Delta I)^2 \rangle = \langle I^2 \rangle$. We can evaluate this in two limits. For $T = 0$ and $V \neq 0$ we get

$$\langle I^2 \rangle = 2q \frac{V}{R} B = 2q \langle I \rangle B$$

giving a standard deviation

$$\sigma_{\text{shot}} = \sqrt{\langle (\Delta I)^2 \rangle} = \sqrt{2q \langle I \rangle B}.$$

This is the current shot noise.

For $T > 0$ and $V \rightarrow 0$ we get

$$\begin{aligned} \langle I^2 \rangle &\simeq 2q \frac{V}{R} \frac{2 - qV/k_B T}{qV/k_B T} B \\ &\simeq 4q \frac{V}{R} \frac{k_B T}{qV} B \\ &= \frac{4k_B T}{R} B \end{aligned}$$

giving a standard deviation

$$\sigma_{\text{thermal}} = \sqrt{\frac{4k_B T}{R} B}.$$

This is the thermal or Johnson noise of the current. As we will see in the analysis of photodetectors these fundamental noise sources set lower limits on the size of an optical signal that we can detect.

13.1.1 Thermal noise

Let's recall some basics about useful systems of units and the background thermal noise level of electrical components. It is convenient to introduce a logarithmic scale for measuring electrical quantities. Thus a given power level can be described on a logarithmic scale as a number of decibels or dB, defined by the relation

$$dB \equiv 10 \log_{10} P, \quad (13.4)$$

where P is the power. Equation (13.4) defines a relative scale so that the relation between two power levels P_1 and P_2 measured in dB is simply $10 \log_{10} P_1/P_2$.

An absolute power scale is often needed. Therefore the notation dBm has been introduced, where dBm are simply dB's referenced to a power level of 1 mW. Thus a power P measured in dBm is

$$P[\text{dBm}] \equiv 10 \log_{10} \frac{P[\text{W}]}{0.001}. \quad (13.5)$$

Unless explicitly written otherwise it is always assumed that dBm refer to a *power* and not a voltage or current level.

Let's now specialize to 50Ω systems. The root mean square or rms power load on a resistor R is $P = RI^2 = V^2/R$, where I and V are the rms current and voltage respectively. Thus the power load on a 50Ω resistor can be written in terms of the rms currents and voltages as

$$P_{50\Omega} = 47 + 10 \log_{10} I_{\text{rms}}^2 \quad [\text{dBm}], \quad (13.6)$$

$$= 13 + 10 \log_{10} V_{\text{rms}}^2 \quad [\text{dBm}]. \quad (13.7)$$

We can always express these relations in terms of peak to peak waveforms ($f(t) = (f_{\text{pp}}/2) \cos \omega t$) using $f_{\text{pp}} = 2\sqrt{2}f_{\text{rms}}$.

Electrical components at temperature T exhibit noise due to the random thermal excitation of charges. This is known as Johnson or Nyquist noise, or simply thermal noise. Unless other noise sources are larger the thermal noise sets a lower limit to the detection of weak signals. A detector of resistance R generates a flat noise spectrum (white noise) with a rms power level

$$P_{\text{th}} = 4k_B T B, \quad (13.8)$$

where k_B is Boltzmann's constant and B is the measurement bandwidth. Equivalently we can write

$$I_{\text{th}}^2 = \frac{4k_B T B}{R} \quad (13.9)$$

and

$$V_{\text{th}}^2 = 4k_B T R B. \quad (13.10)$$

Assuming a room temperature of 300 K we note that $k_B T = 0.026 \text{ eV} = 4.16 \times 10^{-21} \text{ J}$ and the thermal noise power is given by

$$P_{\text{th}} = -167.8 + 10 \log_{10} B[\text{Hz}] \text{ dBm}. \quad (13.11)$$

This is generally a small number: the thermal noise power in room temperature 50Ω systems in a 1 MHz bandwidth is about -114. dBm. Equivalently the thermal noise density is $0.91 \text{ nV}/\text{Hz}^{1/2}$.

13.1.2 Relative intensity noise

All optical sources, including lasers, produce light that has nondeterministic temporal variations in the power, i.e. noise. When working with a fixed geometry the intensity of the light I , is simply proportional to the power P . A useful way of characterizing the amount of power noise is the relative intensity noise or RIN value. This is defined by

$$\text{RIN} = \frac{\sqrt{\langle(\delta I)^2\rangle}}{\langle I \rangle} = \frac{\sqrt{\langle(\delta P)^2\rangle}}{\langle P \rangle} \quad (\sqrt{\text{Hz}})^{-1}. \quad (13.12)$$

In this expression $\langle(\delta P)^2\rangle$ is the mean square fluctuation of the power per unit bandwidth so that RIN has units of $(\sqrt{\text{Hz}})^{-1}$. It is common to express the RIN level in dB using the definition

$$\text{RIN}_{\text{dB}} = 10 \log_{10}(\text{RIN}) \quad (\sqrt{\text{Hz}})^{-1}.$$

If we measure the optical power with a photodetector then the detector current J is proportional to the optical power and we can write²

$$\text{RIN}_J = \frac{\sqrt{\langle(\delta J)^2\rangle}}{\langle J \rangle} = \frac{\sigma_J}{\langle J \rangle} \quad (\sqrt{\text{Hz}})^{-1}, \quad (13.13a)$$

$$\text{RIN}_{J,\text{dB}} = 10 \log_{10}(\text{RIN}_J) \quad (\sqrt{\text{Hz}})^{-1}. \quad (13.13b)$$

Observed RIN levels vary widely depending on the type of laser source. Using high performance intensity stabilizing electronics in a feedback loop, a so-called noise eater, it is possible to reduce the RIN to the limiting level set by the shot noise of a coherent state laser beam. In practice most commercial laser sources are substantially noisier than the shot noise limit, and therefore benefit from intensity stabilization.

13.1.3 Shot noise

Given an electrical current J the RMS shot noise current in a bandwidth Δf is

$$\tilde{\sigma}_J = \sqrt{2eJ\Delta f} \quad (\text{A})$$

where $|e|$ is the electronic charge and $\tilde{\sigma}_J$ has units of current. The simplest description of shot noise has a uniform “white” spectrum. The corresponding voltage noise when the current is dropped over a resistance R is

$$\tilde{\sigma}_V = R\sqrt{2eJ\Delta f} \quad (\text{V}).$$

The RMS noise densities per $\sqrt{\text{Hz}}$ are thus

$$\sigma_{J,\text{sn}} = \sqrt{2eJ} \quad (\text{A}/\sqrt{\text{Hz}}), \quad \sigma_{V,\text{sn}} = R\sigma_J = R\sqrt{2eJ} \quad (\text{V}/\sqrt{\text{Hz}}).$$

²The question of whether we should use $10 \times \log_{10}(\text{RIN}_J)$ or $20 \times \log_{10}(\text{RIN}_J)$ in the definition of the logarithmic RIN level may be confusing. The dB value of a current or voltage ratio is $20 \times \log_{10}(J_2/J_1)$ while the dB value of a power ratio is $10 \times \log_{10}(\text{Power}_2/\text{Power}_1)$. In the present context the current is a photodetector current which is proportional to the optical power P . It is therefore convention to write the RIN level as $10 \times \log_{10}(J_2/J_1) = 10 \times \log_{10}(P_2/P_1)$. Be warned though that the other convention, giving twice larger RIN_{dB} values, is also in use.

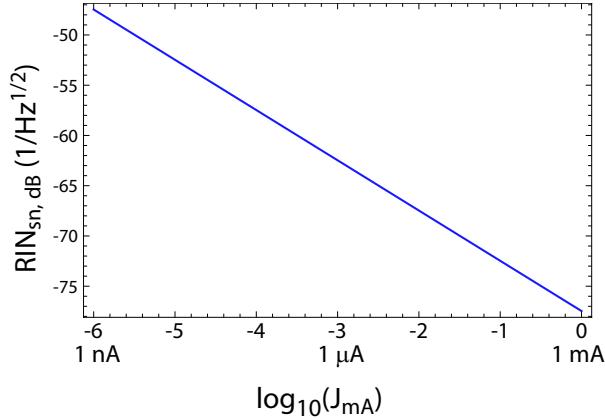


Figure 13.2: Shot noise RIN level.

The RIN level of a current or voltage due to shot noise is

$$\begin{aligned} RIN_{J,sn} &= \frac{\sigma_{J,sn}}{J} = \sqrt{2e/J} \quad (\sqrt{\text{Hz}})^{-1}, \\ RIN_{V,sn} &= \frac{\sigma_{V,sn}}{V} = \frac{\sigma_{J,sn}}{J} = \sqrt{2e/J} \quad (\sqrt{\text{Hz}})^{-1}. \end{aligned}$$

The shot noise level is the same whether we choose to measure a current or a voltage and we will simply label this as RIN_{sn} . Inserting the numerical value of the electronic charge we can write

$$\begin{aligned} RIN_{sn} &= \frac{1.79 \times 10^{-8}}{J_{mA}^{1/2}} \quad (\sqrt{\text{Hz}})^{-1} \\ RIN_{sn,dB} &= -77.5 - 5 \log_{10}(J_{mA}) \quad \text{dB}/\sqrt{\text{Hz}} \end{aligned}$$

where J_{mA} is the numerical value of the current in mA. The shot noise RIN level as a function of photocurrent is shown in Fig. 13.2. A laser power of 2 mW will produce a photocurrent of order 1 mA at the peak sensitivity of a Si photodiode giving $RIN_{sn,dB} = -77.5 \text{ dB}/\sqrt{\text{Hz}}$. To reach $RIN_{sn,dB} = -70 \text{ dB}/\sqrt{\text{Hz}}$ we need a photocurrent of at least 31.6 μA .

13.1.4 Noise equivalent power

Detectors are often specified with a certain noise equivalent power(NEP) per root spectral bandwidth which is the input power for which the output signal to noise ratio (SNR) is unity. An optical power P produces a photocurrent

$$J = \frac{P}{\hbar\omega}\eta e \quad (\text{A}).$$

Here $\frac{P}{\hbar\omega}$ is the number of photons per second incident on the detector, and η is the detection quantum efficiency. The NEP, which has units of $\text{W}/\sqrt{\text{Hz}}$ produces a current with noise spectral density

$$\sigma_{J,NEP} = \frac{NEP}{\hbar\omega}\eta e \quad (\text{A}/\sqrt{\text{Hz}}).$$

A detector is shot noise limited at photocurrent J if the noise spectral density due to the NEP is less than the shot noise RIN level, i.e. if $\sigma_{J,sn} \leq \sigma_{J,NEP}$. We can write this condition as

$$\frac{NEP}{\hbar\omega}\eta e \leq (2eJ)^{1/2}.$$

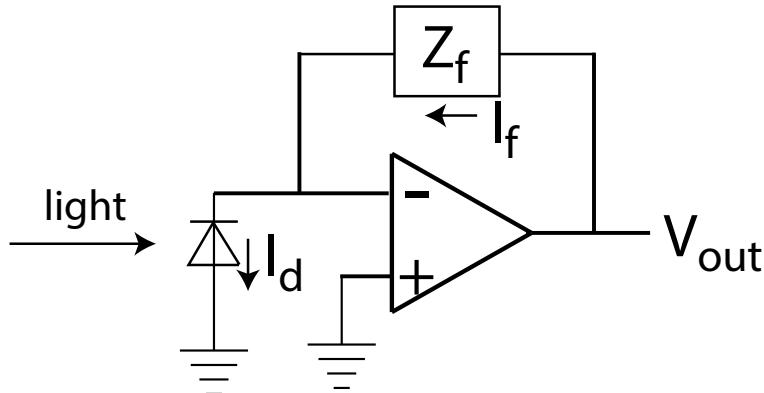


Figure 13.3: Transimpedance photodetector circuit.

Rewriting J in terms of the optical power P we get

$$P \geq \frac{\eta}{2\hbar\omega} NEP^2 \text{ (W).}$$

For example a Thorlabs PDA36A silicon detector has $NEP = 2.3 \times 10^{-12} \text{ W}/\sqrt{\text{Hz}}$ at a gain of 7500 and 1 MHz bw. This NEP value is at the spectral peak of 970 nm where the responsivity is 0.59 A/W. This implies a quantum efficiency of $\eta = 0.75$. At 460 nm the responsivity is 0.23 A/W implying $\eta = 0.62$.

Thus for a 1 MHz bw the power needed to reach the shot noise limit with a PDA36A at 970 nm is $P_{970} \geq 8 \mu\text{W}$ and at 460 nm we have the estimate

$$P_{460} \geq 8. \times (.62/.75) \times (.46/.97) \times (.59/.23)^2 = 20.6 \mu\text{W}.$$

13.2 Detector performance

With these preliminaries out of the way we can analyze the sensitivity of several different detector types and compare direct detection with heterodyne detection.

13.2.1 Photovoltaic detector

A standard circuit for a photovoltaic detector uses an opamp in a transimpedance configuration to convert the detector photocurrent into a voltage which can then be conveniently filtered and amplified with additional circuitry. Using low noise op-amp circuits the minimum detectable optical power and frequency response are essentially determined by the characteristics of the detector and the transimpedance opamp.

An equivalent circuit model is shown in Fig. 13.3. If we assume the opamp is ideal then the -input is effectively at ground potential so $I_f = V_{\text{out}}/Z_f = I_d$, $V_{\text{out}} = Z_f I_d$ and the transimpedance gain is

$$G = \left| \frac{V_{\text{out}}}{I_d} \right| = |Z_f|.$$

If Z_f is just a resistor then the value of the resistor determines the gain.

In order to quantitatively calculate the noise characteristics of this circuit we need a more detailed description of the photodiode as shown in Fig. 13.4. For a first calculation we will ignore

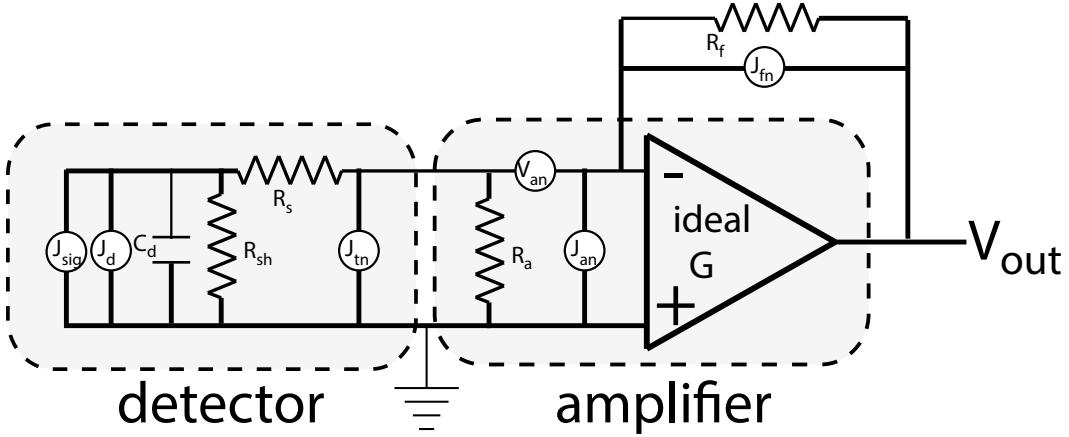


Figure 13.4: Noise model of transimpedance photodetector circuit. The components are: J_{sig} = detector signal current, J_d = detector dark current, C_d = detector capacitance, R_{sh} = detector shunt resistance, R_s = detector series resistance, J_{tn} = detector Johnson noise current, J_{fn} = feedback resistor Johnson noise current, J_{an} = op amp input current noise, V_{an} = op amp input voltage noise, R_a = amplifier input resistance, G = gain of ideal amplifier, R_f = feedback resistor.

type	$R_{800\text{nm}}$ (A/W)	J_d (nA)	R_s (kΩ)	R_{sh} (GΩ)	C_d (pF)
Hamamatsu S3071, $\phi 5$ mm	0.55	0.5	?	18.	120
Thorlabs FDS100 3.6×3.6 mm	0.52	1.0	?	?	24
OSI Optoelectronic PIN-6DP $\phi 4.57$ mm	0.57	?	?	2.	2000

Table 13.1: Si photodetector parameters.

the frequency dependence and just calculate the low frequency noise. The output noise power is

$$\sigma_V^2 = \left(\frac{2e(J_{\text{sig}} + J_d)R_{sh}^2}{R_s^2 + R_{sh}^2} + J_{tn}^2 + J_{fn}^2 + J_{an}^2 \right) R_f^2 + V_{an}^2 \left(1 + \frac{R_f}{R_{sh} + R_s} \right)^2 \quad (\text{V}^2/\text{Hz}).$$

Using $J_{tn}^2 = \frac{4k_B T}{R_{sh} + R_s}$, $J_{fn}^2 = \frac{4k_B T}{R_f}$ we get

$$\sigma_V^2 = \left(\frac{2e(J_{\text{sig}} + J_d)R_{sh}^2}{R_s^2 + R_{sh}^2} + \frac{4k_B T}{R_{sh} + R_s} + \frac{4k_B T}{R_f} + J_{an}^2 \right) R_f^2 + V_{an}^2 \left(1 + \frac{R_f}{R_{sh} + R_s} \right)^2 \quad (\text{V}^2/\text{Hz}).$$

This expression assumes that $R_a \gg R_s, R_{sh}$ which is true for JFET op amps with $R_a \sim 10^{13} \Omega$. Using $R_{sh} \gg R_s$ and $R_{sh} \gg R_f$ this simplifies to

$$\sigma_V^2 = \left(2e(J_{\text{sig}} + J_d) + \frac{4k_B T}{R_f} + J_{an}^2 \right) R_f^2 + V_{an}^2 \quad (\text{V}^2/\text{Hz}).$$

Let's now evaluate the output noise power for a few different detector and op-amp combinations. Some detectors and opamps are listed in Tables 13.1, 13.2. Note that there is a tradeoff between voltage and current noise in the opamps.

Calculations of the low frequency output noise are shown in Fig. 13.5 for several different opamps and detector photocurrents. We see that the shot noise is detectable with any of the opamps when $R_f > \sim 15 \text{ k}\Omega$. These noise calculations show only the low frequency noise spectra since we have neglected finite opamp bandwidth and noise peaking effects. It is also important to look at the frequency dependence of the noise ♣?? to be added...

type	V_{an} (nV/ $\sqrt{\text{Hz}}$) at 1 kHz	J_{an} (pA/ $\sqrt{\text{Hz}}$)	C (pF)	$R_{f,\max}$
LF356	12	0.01	3	1.2 M Ω
AD8512	8	? (guess 0.02)	11.5	~ 1 M Ω
AD8675	2.8	0.3	9.6	9.3 k Ω
OPA209	2.2	0.5	?	4.4 k Ω
OPA1611	1.1	1.7	?	0.65 k Ω

Table 13.2: Op amp parameters. $R_{f,\max} = V_{an}/J_{an}$ is the feedback resistance for which the current noise is equal to the voltage noise.

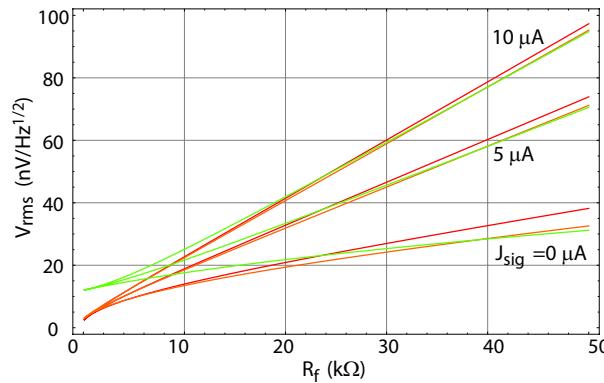


Figure 13.5: Output noise voltage for a Thorlabs FDS100 detector with LF356(green), AD8675 (orange), OPA209(red) opamps.

13.2.2 Photoconducting detector

Photoconductors are devices that exhibit a variable conductivity, or resistance, depending on the incident light intensity. A typical circuit configuration is shown in Fig. 13.6. They are often used for detection of infrared radiation. Using Eq. (13.1) the detector current responsivity is

$$\mathcal{R}_I = \frac{I}{P} = \frac{e\eta}{h\nu} G. \quad (13.14)$$

The voltage responsivity is

$$\mathcal{R}_V = R_{eq}\mathcal{R}_I = \frac{e\eta}{h\nu} G R_{eq}, \quad (13.15)$$

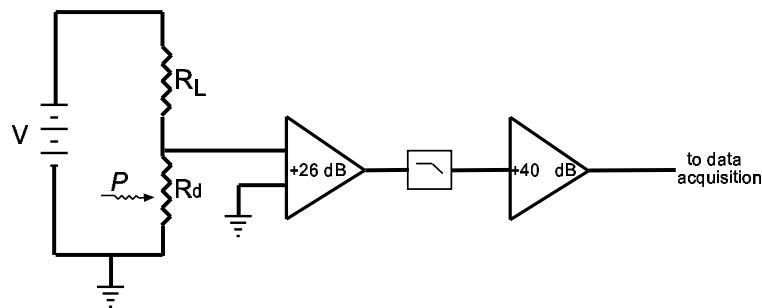


Figure 13.6: Biasing circuit for a photoconductive detector followed by two gain stages.

where $R_{\text{eq}} = R_d R_L / (R_d + R_L)$ is the equivalent resistance of the detector and load resistors in parallel. The photoconductive gain is a dimensionless quantity given by $G = \tau / \tau_{\text{tr}}$ where τ is the lifetime of a photogenerated charge carrier, and τ_{tr} is the transit time across the device. Since the transit time depends on the applied electric field the gain increases with bias current. See, for example, Dereniak and Boreman[108]. It is advantageous to cool photoconductive detectors to increase the carrier lifetime and thereby the gain to values much larger than unity.

In a direct detection experiment the signal to noise ratio of the electronic signal power is

$$z = \frac{I^2}{I_{\text{th}}^2 + I_{\text{gr}}^2 + I_{\text{pa}}^2 + I_{1/\text{f}}^2}, \quad (13.16)$$

where

$$I_{\text{th}}^2 = \frac{4k_B T B}{R_{\text{eq}}}, \quad (13.17)$$

$$I_{\text{gr}}^2 = 4e^2 G^2 B \left(\eta \frac{P + P_{\text{bg}}}{h\nu} + g_{\text{th}} \right), \quad (13.18)$$

$$I_{\text{pa}}^2 = \frac{4k_B T B}{R_{\text{eff,pa}}}, \quad (13.19)$$

are the thermal, generation recombination, and preamplifier noise powers respectively, and $I_{1/\text{f}}^2$ is the $1/f$ noise. In the above P_{bg} is the optical power due to background radiation, g_{th} is the thermal (phonon assisted) generation rate of charge carriers, and $R_{\text{eff,pa}}$ is the noise effective input impedance of the preamplifier³. If the detected signal is demodulated at an intermediate frequency far from dc the $1/f$ noise can be assumed negligible. Equation (13.16) can then be rewritten in the form

$$z = \frac{1}{4B} \frac{e^2 G^2 g}{\frac{k_B T}{R_{\text{eff}}} + e^2 G^2 (g + g_{\text{bg}} + g_{\text{th}})}, \quad (13.20)$$

where $R_{\text{eff}} = R_{\text{eq}} R_{\text{eff,pa}} / (R_{\text{eq}} + R_{\text{eff,pa}})$ is the combined noise effective impedance of the detector, the load resistor, and the preamplifier, $g = \eta P / h\nu$ is the carrier generation rate due to the signal, and $g_{\text{bg}} = \eta P_{\text{bg}} / h\nu$ is the carrier generation rate due to background radiation.

Depending on which term in the denominator of Eq.(13.20) dominates detection will be thermal or background limited. Consider first the case where thermal noise in the detection circuit dominates. Neglecting the 2nd term in the denominator and setting the signal to noise ratio to unity we find for the minimum detectable power,

$$P_{\text{min,th}} = \sqrt{\frac{4k_B T B}{R_{\text{eff}}}} \frac{h\nu}{e\eta G}. \quad (13.21)$$

Estimates for a room temperature HgCdTe detector (typical parameters: $G = 5 \times 10^{-3}$, $\eta = 0.6$, area= 1 mm², $R_d = 50 \Omega$, $R_L = 950 \Omega$) and a 10.6 μm signal wavelength corresponding to a CO₂ laser give $P_{\text{min,th}} \sim 1$ nW in a 1 Hz bandwidth.

It is instructive to calculate how large an optical power is required for detection to be background limited. The thermal carrier generation rate is unknown, but as a first approximation we may set it equal to the background rate g_{bg} . The background rate is found by integrating the Planck distribution for thermal radiation times the spectral dependence of the photodetector sensitivity. As a rough estimate we may take the effective spectral bandwidth of the detector as 1 μm and

³The noise effective input impedance is $R_{\text{eff,pa}} = R_{\text{pa}} 10^{-NF/10}$, where NF is the noise figure in dB.

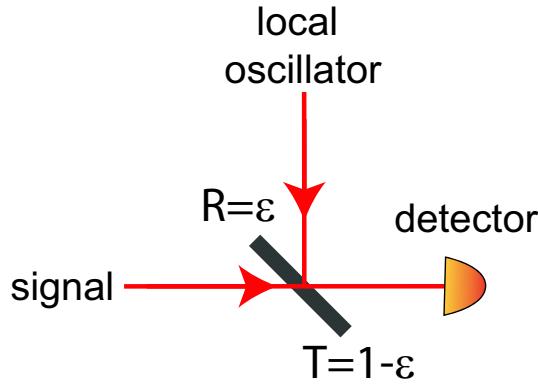


Figure 13.7: A weak signal and strong local oscillator are combined for heterodyne detection.

using the curve given in Dereniak and Boreman p.69, estimate $\mathcal{P}_{\text{bg}} \sim 3 \times 10^{-5} \text{ W}$ for a 1 mm^2 detector. The background generation rate is then $g_{\text{bg}} = 9.6 \times 10^{14} \text{ sec}^{-1}$ using $\eta = 0.6$. Equating the terms in the denominator of Eq. (13.20) gives

$$g^* = \frac{k_B T}{R_{\text{eff}} e^2 G^2} - 2g_{\text{bg}} \quad (13.22)$$

for the carrier generation rate g^* at which the signal dependent noise is equal to the thermal noise. We find $g^* = 3.2 \times 10^{20} \text{ sec}^{-1}$ or, equivalently, for the corresponding optical power $\mathcal{P}^* = 10 \text{ W}$. Since \mathcal{P}^* is considerably greater than the damage level of a 1 mm^2 detector (which is $\sim 1 \text{ W}$) it is not possible to reach the signal noise limit in direct detection with the room temperature infrared detectors.

13.2.3 Heterodyne detection

In order to improve the sensitivity and reduce the effect of background noise we adopt a heterodyne detection scheme. A weak signal to be measured is combined with a much stronger local oscillator at the surface of the detector as shown in Fig. 13.7. Using a beamsplitter with transmission $T = 1 - \epsilon$ close to unity and small reflection $R = \epsilon$ most of the weak signal reaches the detector.

Equation (13.16) for the signal to noise ratio then takes the form

$$z = \frac{I_{\text{hd}}^2}{I_{\text{lo}}^2 + I_{\text{th}}^2 + I_{\text{gr}}^2 + I_{\text{pa}}^2 + I_{1/\text{f}}^2}, \quad (13.23)$$

with

$$I_{\text{hd}} = \frac{\eta e}{h\nu} \eta_{\text{hd}} \sqrt{PP_{\text{lo}}} G, \quad (13.24)$$

$$I_{\text{lo}}^2 = 4e^2 G^2 B \frac{\eta P_{\text{lo}}}{h\nu}, \quad (13.25)$$

where P_{lo} is the optical power of the local oscillator and η_{hd} is a factor less than one that accounts for the efficiency of the heterodyne mixing at the detector surface. The heterodyne efficiency lies between 0 and 1 and can be defined as

$$\eta_{\text{hd}} = \frac{1}{2} \frac{\int_A d\mathbf{r} \mathcal{E}_s \mathcal{E}_{\text{lo}}^* + \mathcal{E}_s^* \mathcal{E}_{\text{lo}}}{\left[\int_A d\mathbf{r} |\mathcal{E}_s|^2 \int_A d\mathbf{r} |\mathcal{E}_{\text{lo}}|^2 \right]^{1/2}}$$

Here the integral is over the detector surface A , \mathcal{E}_s is the signal field, and \mathcal{E}_{lo} is the local oscillator field. We can estimate the requirement on alignment of the beams assuming that the signal propagates along z , normal to the detector surface, and the local oscillator propagates at a small angle θ in the $x - z$ plane. We can then put

$$\begin{aligned}\mathcal{E}_s &= A_s e^{i(kz-\omega t)} \\ \mathcal{E}_{lo} &= A_{lo} e^{i(k \cos(\theta)z + k \sin(\theta)x - \omega t)}\end{aligned}$$

so that

$$\mathcal{E}_s \mathcal{E}_{lo}^* = A_s A_{lo}^* e^{ik[(1-\cos(\theta))z - \sin(\theta)x]}.$$

Let the detector be positioned at $z = 0$ and take $A_s = A_{lo}$ giving

$$\eta_{hd} = \frac{|A_s|^2}{2|A_s|^2 A} \left[\int_A d\mathbf{r} e^{-ik \sin(\theta)x} + c.c. \right] = \frac{1}{A} \int_A d\mathbf{r} \cos[k \sin(\theta)x].$$

For a circular detector of radius a we find

$$\eta_{hd} = 2 \frac{J_1(ak \sin(\theta))}{ak \sin(\theta)}.$$

Since $J_1(x)/x \rightarrow 1/2$ as $x \rightarrow 0$ we recover unit heterodyne efficiency as $\theta \rightarrow 0$ or $a \rightarrow 0$. For a finite sized detector $\eta_{hd} > .9$ for $ak \sin(\theta) < 0.91$. This corresponds to

$$\sin(\theta) < \frac{0.91}{ak}.$$

Taking $a = 1$ mm, $\lambda = 1$ μm we require $\theta < 1.4 \times 10^{-4}$ rad = 0.008 deg. We see that extremely good alignment of the interfering beams is necessary to take advantage of heterodyne detection with a finite sized detector.

Since η_{hd} requires some care to measure we will for simplicity assume in the following that it has a numerical value of 1. Defining in the same fashion as above a carrier generation rate due to the local oscillator g_{lo} , and again neglecting $1/f$ noise, we can rewrite Eq. (13.23) as

$$z = \frac{1}{4B} \frac{\eta_{hd}^2 e^2 G^2 g g_{lo}}{\frac{k_B T}{R_{\text{eff}}} + e^2 G^2 (g + g_{lo} + g_{bg} + g_{th})}. \quad (13.26)$$

As we found for the case of direct detection thermal noise is always dominant for room temperature photoconductive detectors. We can therefore simply write

$$\begin{aligned}z &= \frac{R_{\text{eff}}}{4k_B T B} \eta_{hd}^2 e^2 G^2 g g_{lo}, \\ &= \frac{R_{\text{eff}}}{4k_B T B} \left(\frac{\eta \eta_{hd} e G}{h \nu} \right)^2 P P_{lo}\end{aligned} \quad (13.27)$$

for the signal to noise ratio. We see that the power of the optical signal for which the signal to noise ratio is unity is inversely proportional to the local oscillator power and is given by

$$P_{\min,hd} = \frac{4k_B T B}{R_{\text{eff}}} \left(\frac{h \nu}{\eta \eta_{hd} e G} \right)^2 \frac{1}{P_{lo}}. \quad (13.28)$$

Setting P_{lo} to 0.1 times the power at which detector damage occurs, or 0.1 W, we find a minimum detectable optical power of 1.3×10^{-14} W, assuming a 1 Hz detection bandwidth. This is about

detection scheme		0.1	bandwidth [MHz]	
			1.	10.
direct	$\mathcal{P}_{\min,\text{th}}$ [W]	3.8×10^{-7}	1.2×10^{-6}	3.8×10^{-6}
power limited heterodyne	$\mathcal{P}_{\min,\text{hd}}$ [W]	1.3×10^{-12}	1.3×10^{-11}	1.3×10^{-10}
quantum limited heterodyne	$\mathcal{P}_{\min,\text{ql}}$ [W]	1.2×10^{-14}	1.2×10^{-13}	1.2×10^{-12}

Table 13.3: Minimum detectable power in Watts for different detection limits and bandwidths with an infrared photoconductive detector.

five orders of magnitude less than the minimum detectable power found above for direct detection which demonstrates the superiority of the heterodyne configuration.

Nonetheless with a cooled infrared detector it is possible to achieve signal, or quantum, noise limited operation. In this case the dominant noise source is that generated by the local oscillator. Hence, from Eq. (13.26), the quantum limited signal to noise ratio is

$$z_{\text{ql}} = \frac{1}{4B} \eta_{\text{hd}}^2 g. \quad (13.29)$$

Setting z_{ql} to unity we find the minimum detectable power in a 1 Hz bandwidth to be

$$\mathcal{P}_{\min,\text{ql}} = \frac{4Bh\nu}{\eta\eta_{\text{hd}}^2} = 1.2 \times 10^{-16} \text{ W}. \quad (13.30)$$

In Table 13.3 we compare the minimum detectable powers for several different bandwidths for the cases of thermally limited direct detection, local oscillator power limited heterodyne detection, and quantum limited detection. We conclude that a cooled infrared detector operated in heterodyne mode offers the best sensitivity.

13.2.4 Photomultiplier

Photomultipliers consist of a sensitive photocathode followed by multiple dynode stages. The basic idea is shown in Fig. 13.8. Photoelectrons emitted from the cathode are accelerated to the dynode where they release multiple electrons, which are then accelerated to the next dynode, and so on. There can be up to 15 dynodes in common photomultipliers. This provides very low noise secondary amplification together with high gain. Although the quantum efficiency of photocathodes is typically small compared to the peak of photovoltaic detectors the low noise and high gain of a photomultiplier makes them useful for detection of low light levels.

The large gain amplifies both the signal and the shot noise of the photocathode current which dominates over thermal noise sources. It is then not hard to show that the minimum detectable signal is

$$P_{\min} = \frac{\hbar\omega}{\eta} \sqrt{\frac{I_d B}{e}} \quad (13.31)$$

where I_d is the photocathode dark current.

Using a heterodyne detection scheme to evade the dark current limit the minimum detectable signal is

$$P_{\min,\text{hd}} = \frac{\hbar\omega B}{\eta}. \quad (13.32)$$

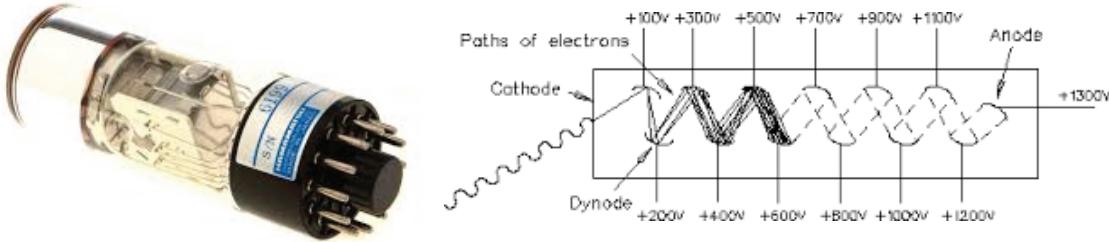


Figure 13.8: Signal amplification in a photomultiplier.

13.2.5 Noise in electron multiplying detectors

In the preceding section we have implicitly assumed that the gain process due to secondary electron emission is noiseless. This assumption needs to be looked at more closely. The following analysis shows that photomultipliers with a moderate number of dynodes can have an excess noise factor as small as 0.25 dB. On the other hand devices that have a large number of secondary gain elements each with a small gain tend to have an excess noise factor of 3 dB.

Consider a detector which accepts m input electrons and through a gain process produces n output electrons. Call the probability distribution of the input electrons $P(m)$ and $G(n)$ the probability of getting n output electrons from one input electron due to the gain process. With these definitions the conditional probability of m input electrons giving n output electrons is

$$P(n|m) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_{m-1}=0}^{\infty} G(n_1)G(n_2)\dots G(n_{m-1})G(n - n_1 - n_2 - \dots - n_{m-1}).$$

The expected value of n given m is the conditional mean

$$\begin{aligned} \langle n \rangle_m &= \sum_{n=0}^{\infty} n P(n|m) \\ &= \sum_{n=0}^{\infty} n \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_{m-1}=0}^{\infty} G(n_1)G(n_2)\dots G(n_{m-1})G(n - n_1 - n_2 - \dots - n_{m-1}) \\ &= \sum_{n'=0}^{\infty} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_{m-1}=0}^{\infty} (n' + n_1 + n_2 + \dots + n_{m-1}) G(n_1)G(n_2)\dots G(n_{m-1})G(n') \\ &= m \langle G \rangle, \end{aligned} \tag{13.33}$$

where we have used $n' = n - n_1 - n_2 - \dots - n_{m-1}$. The expected mean number of output electrons is therefore

$$\begin{aligned} \langle n \rangle &= \sum_{n=0}^{\infty} \langle n \rangle_m P(m) \\ &= \sum_{n=0}^{\infty} m \langle G \rangle P(m) \\ &= \langle G \rangle \langle m \rangle. \end{aligned} \tag{13.34}$$

The result can also be derived naively by assuming that the gain and the number of input electrons are independent so that $\langle n \rangle = \langle Gm \rangle = \langle G \rangle \langle m \rangle$.

We can also calculate the conditional mean square of the number of output electrons as

$$\begin{aligned}
\langle n^2 \rangle_m &= \sum_{n=0}^{\infty} n^2 P(n|m) \\
&= \sum_{n=0}^{\infty} n^2 \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_{m-1}=0}^{\infty} G(n_1)G(n_2)\dots G(n_{m-1})G(n - n_1 - n_2 - \dots - n_{m-1}) \\
&= \sum_{n'=0}^{\infty} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_{m-1}=0}^{\infty} (n' + n_1 + n_2 + \dots + n_{m-1})^2 G(n_1)G(n_2)\dots G(n_{m-1})G(n') \\
&= m\langle G^2 \rangle + m(m-1)\langle G \rangle^2.
\end{aligned} \tag{13.35}$$

The expected mean square number of output electrons is therefore

$$\begin{aligned}
\langle n^2 \rangle &= \sum_{n=0}^{\infty} \langle n^2 \rangle_m P(m) \\
&= \sum_{n=0}^{\infty} [m\langle G^2 \rangle + m(m-1)\langle G \rangle^2] P(m) \\
&= \langle G^2 \rangle \langle m \rangle + \langle G \rangle^2 \langle m^2 \rangle - \langle G \rangle^2 \langle m \rangle.
\end{aligned} \tag{13.36}$$

The variance of the output count is therefore

$$\begin{aligned}
\langle (\Delta n)^2 \rangle &= \langle n^2 \rangle - \langle n \rangle^2 \\
&= \langle G \rangle^2 \langle (\Delta m)^2 \rangle + \langle (\Delta G)^2 \rangle \langle m \rangle.
\end{aligned} \tag{13.37}$$

We see that the output variance is the input variance times the square of the mean gain, plus an additional term due to fluctuations in the gain.⁴

The gain noise factor is defined as

$$F \equiv \frac{\langle (\Delta n)^2 \rangle}{\langle G \rangle^2 \langle (\Delta m)^2 \rangle} = 1 + \frac{\langle (\Delta G)^2 \rangle}{\langle G \rangle^2} \frac{\langle m \rangle}{\langle (\Delta m)^2 \rangle}.$$

In the ideal case of constant deterministic gain $\langle (\Delta G)^2 \rangle = 0$ and $F = 1$.

Consider now a detection device with several stages of gain, as in a photomultiplier. Each stage has gain g_j and associated gain variance $\sigma_j^2 \equiv \langle g_j^2 \rangle - \langle g_j \rangle^2$. We will assume as in the analysis leading to Eq. (13.37) that each stage is statistically independent and calculate the variance of the output after r stages. Consider first the output after the second stage assuming a single electron incident on the first stage. The mean composite gain after the second stage is just $\langle G_2 \rangle = \langle g_1 \rangle \langle g_2 \rangle$ and using (13.37) we see that the output variance is

$$\sigma_{G_2}^2 = \langle g_2 \rangle^2 \sigma_1^2 + \langle g_1 \rangle \sigma_2^2.$$

If we then add a third stage we find

$$\begin{aligned}
\langle G_3 \rangle &= \langle g_1 \rangle \langle g_2 \rangle \langle g_3 \rangle \\
\sigma_{G_3}^2 &= \langle g_3 \rangle^2 (\langle g_2 \rangle^2 \sigma_1^2 + \langle g_1 \rangle \sigma_2^2) + \langle g_1 \rangle \langle g_2 \rangle \sigma_3^2 \\
&= (\langle g_1 \rangle \langle g_2 \rangle \langle g_3 \rangle)^2 \left(\frac{\sigma_1^2}{\langle g_1 \rangle^2} + \frac{\sigma_2^2}{\langle g_1 \rangle \langle g_2 \rangle^2} + \frac{\sigma_3^2}{\langle g_1 \rangle \langle g_2 \rangle \langle g_3 \rangle^2} \right).
\end{aligned} \tag{13.38}$$

⁴This result is sometimes referred to as the Burgess variance[109]. It was discussed in the context of electron amplifying devices in [110]. The result was also published as early as 1938 in [111].

Thus for r stages we have

$$\begin{aligned}\langle G_r \rangle &= \prod_{j=1}^r \langle g_j \rangle \\ \sigma_{G_r}^2 &= \langle G_r \rangle^2 \left(\frac{\sigma_1^2}{\langle g_1 \rangle^2} + \frac{\sigma_2^2}{\langle g_1 \rangle \langle g_2 \rangle^2} + \dots + \frac{\sigma_r^2}{\langle g_1 \rangle \langle g_2 \rangle \dots \langle g_{r-1} \rangle \langle g_r \rangle^2} \right).\end{aligned}\quad (13.39)$$

This expression can be simplified if we assume that all stages are statistically equivalent and put $g = \langle g_j \rangle$ and $\sigma_g^2 = \sigma_j^2$. Then

$$\begin{aligned}\sigma_{G_r}^2 &= \langle G_r \rangle^2 \frac{\sigma_g^2}{g^2} \sum_{j=1}^r \frac{1}{g^{j-1}} \\ &= \langle G_r \rangle^2 \frac{\sigma_g^2}{g(g-1)} \left(1 - \frac{1}{g^r} \right).\end{aligned}\quad (13.40)$$

The noise figure of the amplifier with r equivalent gain stages is therefore

$$F = 1 + \frac{\sigma_g^2}{g(g-1)} \left(1 - \frac{1}{g^r} \right) \frac{\langle m \rangle}{\langle (\Delta m)^2 \rangle}.$$

As an example a photomultiplier may have $r = 10$ dynodes and a gain of $\langle g \rangle = 4$ per dynode giving a total current gain of about 10^6 . In this high gain limit and assuming the input photoelectrons have Poissonian statistics the noise figure is

$$F = 1 + \frac{1}{g-1} \left(1 - \frac{1}{g^r} \right) \approx 1 + \frac{1}{3} = \frac{4}{3}$$

or 1.25 dB. We see that photomultipliers can have very low noise gain.

A different limit applies to devices such as “staircase” avalanche photodiodes or electron multiplying CCD chips (EMCCD). In EMCCD devices there are many gain stages, $r > 100$, each with a very low gain, $g \sim 1.01$. In this limit we write $g = 1 + \delta$ so that $\langle g \rangle = 1 + \langle \delta \rangle$, $\langle g^2 \rangle = \langle \delta^2 \rangle + 2\langle \delta \rangle + 1$, and $\langle (\Delta g)^2 \rangle = \langle (\Delta \delta)^2 \rangle$.

Eq. (13.40) for the gain variance then takes the form

$$\begin{aligned}\sigma_{G_r}^2 &= \langle G_r \rangle^2 \frac{\langle (\Delta \delta)^2 \rangle}{\langle 1 + \delta \rangle \langle \delta \rangle} \left(1 - \frac{1}{\langle 1 + \delta \rangle^r} \right) \\ &\simeq \langle G_r \rangle^2 \frac{\langle (\Delta \delta)^2 \rangle}{\langle \delta \rangle} \frac{r \langle \delta \rangle}{1 + r \langle \delta \rangle}.\end{aligned}\quad (13.41)$$

Using the approximation that $r \langle \delta \rangle \gg 1$ and assuming the statistics of δ are Poissonian gives

$$\sigma_{G_r}^2 \simeq \langle G_r \rangle^2. \quad (13.42)$$

The excess noise factor is therefore $F = 1 + 1 = 2$, or 3 dB. Commercial EMCCD detectors appear to meet or exceed this estimate. For example a device with $G = 1000$ and $r = 500$ stages has $\delta = 0.0139$ and $r\delta = 507$ so the above approximations are all accurate.



Figure 13.9: Examples of piezoelectric modulators. Three point control of a mirror holder (left), angular tilt of a mirror (center), a long bimorph actuator for mm scale motion (right).

13.3 Optical Modulators

Several different devices can be used to control and modulate the amplitude, polarization, frequency, or phase of an optical beam. Modulators are also used to imprint images onto light fields and to scan beams. Devices include mirrors mounted on piezoelectric transducers, liquid crystal devices, acousto-optic modulators, electro-optic modulators, and spatial light modulators. In this section we survey some of the basic types of modulators.

13.3.1 Piezoelectric device

The piezoelectric effect relates material stress to voltage. An applied mechanical stress generates a voltage. Conversely an applied voltage generates a mechanical stress which results in a change in physical size. This can be used to modulate the phase of an optical beam by attaching a reflecting surface to a piece of piezoelectric material, or by using the piezoelectric to control an optical holder as shown in Fig. 13.9. There is a limit to the allowable applied voltage before damage occurs which limits the available motion. Common elements work with voltages up to 1 kV and have motion ranging from about $1 \mu\text{m} - 1 \text{ mm}$. The speed of modulation is relatively slow since it involves mechanical motion. Small devices have been used to phase modulate optical beams at frequencies up to $\sim 100 \text{ kHz}$.

Frequency modulation requires a continuous time derivative of the optical phase. An optical field with frequency ω_0 can be shifted to $\omega = \omega_0 + a$ by applying a phase shift $\phi(t) = at$. Although it is not possible to have a continuous linear ramp of the phase due to the limit on the piezoelectric motion, approximate frequency shifting is possible using a serrodyne technique. This involves applying a triangular voltage ramp giving a linear phase shift, with a rapid hop of the position back to zero before the next tooth of the waveform. Provided the phase shift of each linear ramp is a multiple of 2π and the time needed for the hop back to zero is small compared to the ramp time this method can be highly efficient.

Since the motion per applied volt can be very small piezoelectric devices can achieve position resolution down to the nm range. This sensitivity makes piezo mounted mirrors very useful for frequency tuning of optical resonators and varying path lengths in interferometers. Piezoelectric devices can also be used for beam pointing. Mirrors with two piezoelectric actuators mounted behind them can tilt along two orthogonal axes and thereby provide 2D beam steering. Piezoelectrics are also used for x,y,z and x,y,z plus tilt control of mechanical devices.

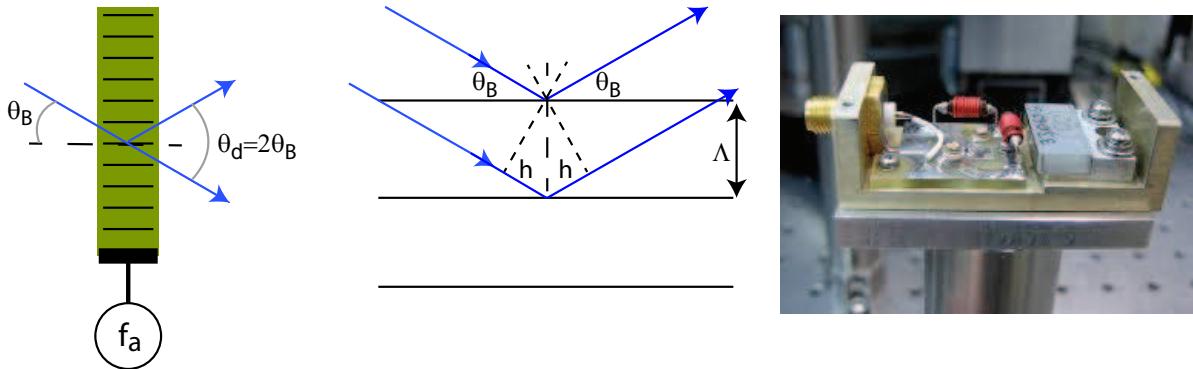


Figure 13.10: Bragg diffraction in an acousto-optic deflector with acoustic wavelength Λ . A picture of a real device is shown on the right.

13.3.2 Acousto-optic modulator

The acousto-optic modulator (AOM) is widely used for frequency, amplitude, and pointing control of optical beams. The basic principle involves diffraction of a beam from a piece of transparent material which supports propagation of an acoustic wave. An electrical transducer is attached to the material and application of an electrical signal at frequency f_a generates an acoustic wave with angular frequency $\Omega = 2\pi f_a$ propagating at the acoustic velocity v_a . The wavelength of the acoustic wave is $\Lambda = v_a/f_a$ and the wavenumber is $K = 2\pi/\Lambda = \Omega/v_a$.

The acoustic wave results in periodic variation of the strain inside the optical material. The dielectric constant changes proportional to the applied strain in materials that exhibit a photoelastic effect. The tensorial nature of photoelasticity is complicated and details can be found in books on crystal structure⁵. We will simply assume a local scalar response such that the refractive index change is $\delta n = pA_a$ where A_a is the amplitude of the acoustic wave and p is proportional to the relevant component of the photoelastic tensor.

Consider a device with acoustic wave velocity v_a at an applied frequency f_a as shown in Fig. 13.10. Light of frequency ω and wavelength λ is diffracted by the wave through an angle $\theta_d = 2\theta_B$. At the Bragg angle θ_B the light reflected from multiple planes interferes constructively giving a strong diffracted intensity. Referring to the figure the condition for this to occur is $2hkn = m2\pi$ where n is the refractive index inside the medium and the integer m specifies the diffraction order. Substituting $h = \Lambda \sin(\theta_B)$ we get the condition

$$\sin(\theta_B) = m \frac{\pi}{nk\Lambda} = m \frac{\lambda/n}{2\Lambda} = m \frac{\lambda/n}{2} \frac{f_a}{v_a}.$$

The deflection angle is

$$\theta_d = 2\theta_B = 2 \sin^{-1} \left(m \frac{\lambda/n}{2} \frac{f_a}{v_a} \right) \simeq m(\lambda/n) \frac{f_a}{v_a}.$$

The deflector can be used as a beam scanner giving parallel displacements by placing the device in the front focal plane of a lens.

The deflected light is frequency shifted by an amount which can be calculated from energy conservation. The acoustic phonons have energy $E = \hbar\Omega$ and momentum $\hbar K = \hbar\Omega/v_a$. Energy

⁵See for example the book by Nye[112].

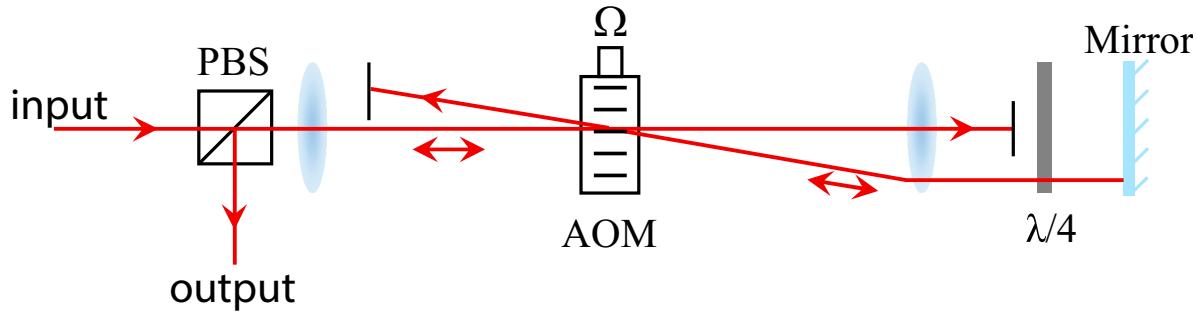


Figure 13.11: Double pass geometry for frequency shifting without angular deflection.

and momentum conservation dictate that the light diffracted into order m satisfies

$$\omega_d = \omega + m\Omega, \quad \mathbf{k}_d = \mathbf{k} + m\mathbf{K}.$$

We see that the angular deflection is accompanied by a change in frequency. This effect can be used to change the direction of the beam for spatial scanning or for changing the frequency of the light. In the latter case it is sometimes inconvenient that the frequency change is accompanied by an unwanted angular deflection. The angular deflection can be corrected for by retroreflecting the deflected light back through the modulator. After two passes in the geometry shown in Fig. 13.11 there is no angular deflection and a frequency shift of 2Ω .

When the incident angle deviates from the Bragg angle the diffracted intensity falls off. In order to understand the angular sensitivity we can use our earlier result Eq. (5.41) for diffraction from a periodic array. The general result for the diffracted intensity in the back focal plane of a lens with focal length f is

$$I(x_2) = |\mathcal{F}[U(x)]|^2 \frac{\sin^2\left(\frac{Nk\Lambda x_2}{2f}\right)}{\sin^2\left(\frac{k\Lambda x_2}{2f}\right)}$$

where $U(x)$ is the transmission function of a unit cell, $\Lambda = v_a/f_a$ is the acousto-optic grating period, and N is the number of unit cells. For the acousto-optic modulator $U(x) = e^{i\epsilon \sin(x/\Lambda)}$ with $\epsilon = k(\delta n)L$, δn is the peak refractive index change caused by the acoustic wave and L is the width of the device along the propagation direction of the light. We can estimate the number of unit cells as $N = w/\Lambda$ with w the optical beam width.

The deflection angle is approximately $\theta = x_2/f$ for small angles so

$$I(x_2) = |\mathcal{F}[U(x)]|^2 \frac{\sin^2\left(\frac{Nk\Lambda\theta}{2}\right)}{\sin^2\left(\frac{k\Lambda\theta}{2}\right)}$$

and the width of the diffraction peak is roughly

$$\delta\theta = 2\frac{2\pi}{Nk\Lambda} = \frac{2\lambda}{N\Lambda} = \frac{2\lambda}{w}.$$

This is precisely what we expect for a Gaussian beam of width w which has a far field diffraction angle $\theta_{\text{diff}} \sim \lambda/w$. In order for the diffracted light to be concentrated in a single Bragg order we require that $\theta_B \gg \theta_{\text{diff}}$. This translates to the condition

$$\frac{\lambda f_a}{n v_a} \gg \frac{\lambda}{w}$$

or

$$w \gg \frac{nv_a}{f_a} = n\Lambda. \quad (13.43)$$

The index of refraction of acousto-optic devices is generally in the range of 1 – 3 and Eq. (13.43) says that the beam width must be large compared to the acoustic wave period. This is a necessary condition for Bragg diffraction that efficiently couples the incident light into a single diffracted order. In the opposite limit multiple orders can be excited⁶.

The dependence of the diffracted power on the device parameters can be found from a coupled wave analysis which is conceptually similar to wave mixing in $\chi^{(2)}$ media which was treated in Ch. 11. For sufficiently long interaction lengths it is possible to transfer close to 100% of the light into the Bragg diffracted beam.

13.3.3 Time bandwidth product

If the device has length L along the direction of propagation of the acoustic wave the time to fill the acoustic medium with a new frequency is $T = L/v_a$. The time bandwidth product is defined as $TBW = T\Delta f_a$ where Δf_a is the allowable range of acoustic wave frequencies. This range is set by some combination of material parameters, the piezo transducer, and associated electronic circuit elements.

Suppose the incident optical beam has a Gaussian intensity profile with waist $w_0 = L/q_1$ with q_1 some constant. The output is Fourier transformed by a lens of focal length f to give a waist in the back focal plane of $w = (\lambda f/\pi w_0)$. The change in position of the center of the output beam as the frequency is changed by Δf_a is $\Delta x = f\Delta\theta_B = (f\lambda/v_a)\Delta f_a$. The number of output spots separated by $q_2 w$ is therefore

$$N = \frac{\Delta x}{q_2 w} = \frac{(f\lambda/v)\Delta f_a}{q_2(\lambda f/\pi w_0)} = \frac{\pi w_0 \Delta f_a}{q_2 v} = \frac{\pi}{q_1 q_2} (T\Delta f_a) = \frac{\pi}{q_1 q_2} TBW.$$

Let us assume $q_1=4$. A smaller q_1 would give more resolvable spots but would lead to clipping of the intensity at the aperture of the Bragg cell since $w_0 = L/q_1$. To get good separation of the output beams we will take $q_2 = 2$ which corresponds to an intensity crosstalk between output beam centroids of $e^{-2(q_2 w)^2/w^2} = e^{-2q_2^2} = 3.3 \times 10^{-4}$. With these choices the number of resolvable spots is $N = \frac{\pi}{8} TBW$. Acousto-optic devices are available with TBW up to several thousand so a large number of resolvable spots can be accessed. This capability renders these devices very useful for optical signal processing tasks.

13.3.4 Electro-optic modulator

Another widely used device is the electro-optic modulator (EOM). A laser beam propagates through a crystal which has a nonzero Pockels coefficient. This implies that the refractive index changes proportional to the local electric field. Applying an external voltage to the modulator changes the refractive index and if this is done periodically in time a time dependent phase modulation is achieved and frequency sidebands are imprinted on the laser beam. The result is conceptually equivalent to what can be achieved with a piezoelectric device. The practical difference is that

⁶The condition in Eq. (13.43) is necessary but not sufficient for achieving Bragg diffraction. It is also necessary that the interaction length along the direction of the optical beam be sufficiently large. If this is not the case then multiple diffraction orders will be observed even though (13.43) is satisfied. The appearance of multiple orders is referred to as the Raman-Nath limit and a separate analysis is required to determine the sufficient condition for Bragg diffraction.

while piezoelectrics are limited to ~ 100 kHz, electro-optic modulation is possible from DC up to tens of GHz. In addition the tensorial nature of the Pockels effect can be used to modulate the polarization state of light. In combination with a polarizing beam splitter this can be used for amplitude control or for fast on/off switching of a light beam.

The dependence of the refractive index on the applied electric field can be written as

$$n(E) = n(0) + a_1 E + \frac{1}{2} a_2 E^2 + \dots$$

The a_1 coefficient is due to the linear electro-optic or Pockels effect. The a_2 coefficient is due to the quadratic or Kerr effect. This is the same nonlinear effect that we discussed in Sec. 10.3 for the case where the modulating field was at optical frequencies. It is conventional to write this as

$$n(E) = n - \frac{1}{2} r n^3 E - \frac{1}{2} s n^3 E^2 + \dots$$

Here $n = n(0)$ and r, s are the electro-optics coefficients and r is only non-zero in non centrosymmetric media.

The reason for this particular way of writing the index response is that in anisotropic media it is convenient to work with the impermeability defined as

$$\eta = \frac{\epsilon_0}{\epsilon} = \frac{1}{n^2}.$$

Changes in η are then written as

$$\begin{aligned} \Delta\eta &= \frac{d\eta}{dn} \Delta n \\ &= \left(\frac{-2}{n^3}\right) \Delta n \\ &= \left(\frac{-2}{n^3}\right) \left(-\frac{1}{2} r n^3 E - \frac{1}{2} s n^3 E^2\right) \\ &= rE + sE^2. \end{aligned}$$

Thus the field dependent impermeability is

$$\eta(E) = \eta(0) + rE + sE^2.$$

In anisotropic media η, r, s are all tensors so

$$\eta_{ij}(E) = \eta_{ij}(0) + \sum_k r_{ijk} E_k + \sum_{k,l} s_{ijkl} E_k E_l.$$

The symmetry properties of the r, s tensors depend on the crystal symmetry, but we will not discuss this further here.

Let's consider the effect of phase modulation of a beam $\mathcal{E}_{\text{in}}(z, t) = \mathcal{E}_0 e^{i(k_0 n_0 z - \omega_0 t)}$. We apply a voltage $V_{\text{mod}} = V_0 \sin(\Omega t)$ to the electro-optic crystal of length L . This results in an index change $n_1(t) = dE_{\text{mod}}(t) = dV_{\text{mod}}(t)/w$ where $d = -\frac{1}{2} r n^3$ quantifies the electro-optic response and w is the width of the crystal. The field exiting the modulator is

$$\begin{aligned} \mathcal{E}_{\text{out}}(z = L, t) &= \mathcal{E}_0 e^{i(k_0 [n_0 + n_1(t)] L - \omega_0 t)} \\ &= \mathcal{E}_0 e^{-i\omega_0 t} e^{i k_0 n_0 L} e^{i k_0 n_1(t) L} \\ &= \mathcal{E}_{\text{in}}(L, t) e^{i \eta \sin(\Omega t)} \end{aligned}$$

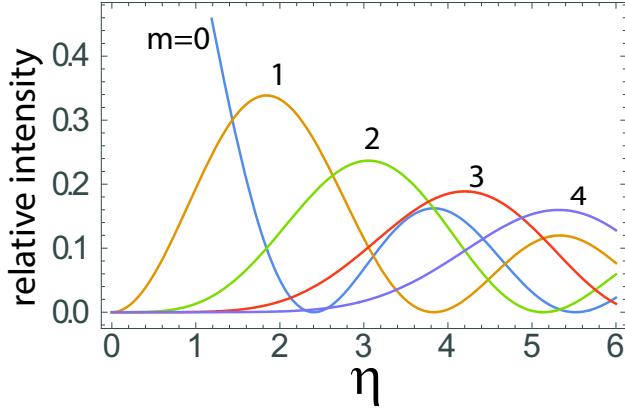


Figure 13.12: Relative intensity in EOM order m as a function of the maximum phase shift η .

with $\eta = k_0 L d V_0 / w$. The dimensionless parameter η is proportional to the applied voltage, and determines the strength of the phase modulation. The above expression for the modulated field implicitly assumes that the transit time is short compared to the time over which the modulating field changes, i.e. $T = L n_0 / c \ll 1/\Omega$. If this is not the case the optical field sees a time varying field during the transit. This puts a limit on the frequency bandwidth of electro-optic modulators. This limit can be partially overcome using a traveling wave modulator where both the optical field and the modulating field are propagating.

The exponential can be written as a sum of harmonics using the identity

$$e^{ia \sin(b)} = \sum_{m=-\infty}^{\infty} J_m(a) e^{imb}$$

to arrive at

$$\mathcal{E}_{\text{out}}(L, t) = \mathcal{E}_{\text{in}}(L, t) \sum_{m=-\infty}^{\infty} J_m(\eta) e^{im\Omega t}. \quad (13.44)$$

We see that the output field is a phase modulated version of the input with frequency components at $\omega_0 \pm m\Omega$. Note that the intensity is unchanged since

$$\begin{aligned} |\mathcal{E}_{\text{out}}(L, t)|^2 &= |\mathcal{E}_{\text{in}}(L, t)|^2 \left| \sum_{m=-\infty}^{\infty} J_m(\eta) e^{im\Omega t} \right|^2 \\ &= |\mathcal{E}_{\text{in}}(L, t)|^2 \sum_{m=-\infty}^{\infty} J_m(\eta) e^{im\Omega t} \sum_{n=-\infty}^{\infty} J_n(\eta) e^{-in\Omega t} \\ &= |\mathcal{E}_{\text{in}}(L, t)|^2 e^{i\eta \sin(\Omega t)} e^{i\eta \sin(-\Omega t)} \\ &= |\mathcal{E}_{\text{in}}(L, t)|^2. \end{aligned}$$

Therefore placing a detector that is sensitive to the optical intensity at the output of the EOM will not reveal any modulation. On the other hand if we separate the different frequency components, for example using an optical resonator, the modulation will emerge.

The Fourier transform of the output field is

$$\begin{aligned}\mathcal{E}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \mathcal{E}_0 e^{ik_0 n_0 L} e^{-i\omega_0 t} \sum_{m=-\infty}^{\infty} J_m(\eta) e^{im\Omega t} e^{i\omega t} \\ &= \mathcal{E}_0 e^{ik_0 n_0 L} \sum_{m=-\infty}^{\infty} J_m(\eta) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt e^{-i\omega_0 t} e^{im\Omega t} e^{i\omega t} \\ &= \sqrt{2\pi} \mathcal{E}_0 e^{ik_0 n_0 L} \sum_{m=-\infty}^{\infty} J_m(\eta) \delta(\omega - (\omega_0 - m\Omega)).\end{aligned}$$

which is a sum of spectral components separated by Ω .

The relative intensity in the m^{th} sideband which is shifted by $m\Omega$ from the carrier is $|J_m(\eta)|^2$. Figure 13.12 shows the relative intensity of the first few sidebands as a function of the modulation strength η which is also the maximum phase shift. The maximum intensity that can be put into the first sideband is 0.34 at $\eta = 1.84$. At this modulation strength 68% of the energy will be in the $m = \pm 1$ sidebands, 10% will be left in the carrier, 20% will be in the $m = \pm 2$ sidebands and the remaining 2% will be in higher order sidebands, mostly $m = \pm 3$. The value of the applied voltage to achieve the maximum first order modulation is

$$V_0 = 1.84 \frac{w}{k_0 L d} = 0.293 \frac{w\lambda}{L d}.$$

Taking $L = 1$ cm, $w = 0.5$ mm, $\lambda = 1$ μm , and $d = 6.7$ pm/V for LiNbO₃ we find $V_0 = 2190$ V. We see that for conveniently sized crystals the voltage required for strong modulation is large. The voltage requirement can be reduced substantially by reducing w . This is achieved by propagating the beam in a guided fiber mode so that w and V_0 can be reduced by up to two orders of magnitude.

One drawback of the EOM is that it produces double sideband modulation when driven with a sine wave. If a pure frequency shift is desired the serrodyne technique described in Sec. 13.3.1 can be used. In this way frequency shifts of several hundred MHz have been achieved with EOMs[113, 114, 115].

13.4 Laser stabilization

Experiments such as high resolution spectroscopy and optical clocks require stabilization of optical sources. This is often performed by comparing the frequency of the laser to be stabilized to a resonance of an optical cavity or an atomic transition frequency. In the following sections we describe several approaches: the Pound-Drever-Hall method in Sec. 13.4.1 and spatial tilt-locking in Sec. 13.4.3.

13.4.1 Pound-Drever-Hall locking

Laser frequency stabilization is important for applications in spectroscopy and metrology. One widely used method is to “lock” the frequency of a laser to a reference provided by an atomic transition or a resonance of a stable optical cavity. The latter approach originates from work of Pound in the microwave domain[116], and was later extended to locking of lasers[12]. The method is now widely used and referred to as Pound-Drever-Hall or PDH locking. As an example PDH locks play a major role in the operation and success of the LIGO gravitational wave observatory.

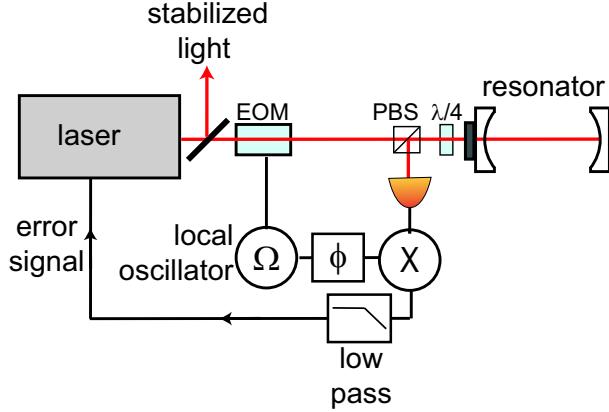


Figure 13.13: Arrangement for Pound-Drever-Hall stabilization of a laser to a reference resonator.

The layout of a PDH locking setup is shown in Fig. 13.13. The laser to be stabilized is passed through an EOM that imparts sidebands to the carrier at $\pm\Omega$. The field after the EOM can be written as

$$\mathcal{E}_{\text{EOM}} = \left[\mathcal{E}_0 + \mathcal{E}_+ e^{i\phi_+} e^{-i\Omega t} + \mathcal{E}_- e^{i\phi_-} e^{i\Omega t} \right] e^{-i\omega t} \quad (13.45)$$

We assume the modulation is weak so that only the carrier at ω and the first order sidebands at $\omega \pm \Omega$ have significant amplitude, and take $\mathcal{E}_0, \mathcal{E}_+, \mathcal{E}_-$ to be real with relative phases of the sidebands defined by ϕ_+, ϕ_- . The light is then coupled to a Fabry-Perot resonator and the reflected light is separated using a half-wave plate and polarizing beam splitter (PBS). The combination of the wave plate, reflection from the cavity, and the PBS form an optical circulator. The optical field amplitude impinging on the detector can then be written as

$$\mathcal{E}_{\text{det}} = \left[r_0 e^{i\chi_0} \mathcal{E}_0 + r_+ \mathcal{E}_+ e^{i(\phi_+ + \chi_+)} e^{-i\Omega t} + r_- \mathcal{E}_- e^{i(\phi_- + \chi_-)} e^{i\Omega t} \right] e^{-i\omega t}.$$

Here r_0, r_+, r_- and χ_0, χ_+, χ_- are the cavity reflection amplitudes and phase shifts for the carrier and sidebands.

The reflected light is detected resulting in a voltage

$$\begin{aligned} V_{\text{det}}(t) &= \eta_d |\mathcal{E}_{\text{det}}|^2 \\ &\simeq \eta_d [r_0^2 |\mathcal{E}_0|^2 + \mathcal{E}_0 \mathcal{E}_+ r_0 r_+ \cos(\Omega t + \chi_0 - \chi_+ - \phi_+) + \mathcal{E}_0 \mathcal{E}_- r_0 r_- \cos(\Omega t - \chi_0 + \chi_- + \phi_-)] \end{aligned}$$

where $\eta_d = \eta_q R$ is the product of the detector quantum efficiency η_q and the responsivity R . We have neglected additional terms $\sim \mathcal{E}_+^2, \mathcal{E}_-^2, \mathcal{E}_+ \mathcal{E}_-$ which are assumed small. This is mixed with a phase shifted copy of the reference oscillator $V_{\text{lo}} = \cos(\Omega t + \phi)$ which results in terms near zero frequency, terms that oscillate at Ω and terms that oscillate at 2Ω . The mixer output is low pass filtered leaving only the dc term

$$V_{\text{error}} = \frac{1}{2} \mathcal{E}_0 \mathcal{E}_+ r_0 r_+ \cos(\chi_0 - \chi_+ - \phi_+ - \phi) + \frac{1}{2} \mathcal{E}_0 \mathcal{E}_- r_0 r_- \cos(\chi_0 - \chi_- - \phi_- + \phi).$$

The error signal is effectively due to the interference of the sidebands with the carrier accounting for the detuning dependent phase of the cavity reflectivity.

So far we have kept the analysis general. For an ideal EOM $\phi_+ = 0$ and $\phi_- = \pi$ and using (13.44) for the field amplitudes we find

$$V_{\text{error}} = \frac{1}{2} \mathcal{E}_{\text{in}}^2 J_0(\epsilon) J_1(\epsilon) r_0 [r_+ \cos(\chi_0 - \chi_+ - \phi) - r_- \cos(\chi_0 - \chi_- + \phi)] \quad (13.46)$$

with \mathcal{E}_{in} the incident field on the EOM and ϵ the modulation index. Let's now specify the cavity reflectivities and phases in terms of the detuning of the laser from the cavity resonance $\nu = (\omega - \omega_c)/(2\pi)$. The amplitude and phase of the cavity reflection are

$$\begin{aligned} r &= \left[\frac{\frac{R_{\text{in}}(1-\mathcal{R}^{1/2}/R_{\text{in}})^2}{(1-\mathcal{R}^{1/2})^2} + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(\pi\nu/\nu_{\text{FSR}})}{1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(\pi\nu/\nu_{\text{FSR}})} \right]^{1/2}, \\ \chi &= \tan^{-1} \left[\frac{(R_{\text{in}} - 1) \sin(2\pi\nu/\nu_{\text{FSR}})}{(R_{\text{in}} + 1)[1 - \cos(2\pi\nu/\nu_{\text{FSR}})]} \right]. \end{aligned}$$

These expressions follow from those given in Secs. 3.2.3, 3.2.5 with $\mathcal{R} = R_{\text{in}}R_{\text{out}}R_{\text{cav}}$, and $\mathcal{F} = \frac{\pi\mathcal{R}^{1/4}}{1-\mathcal{R}^{1/2}}$. In the following we will simplify to the case of a critically coupled symmetric cavity with $R_{\text{cav}} = 1$ and $R_{\text{out}} = R_{\text{in}}$. In this case $\mathcal{F} = \pi R_{\text{in}}^{1/2}/(1 - R_{\text{in}})$ and

$$\begin{aligned} r &= \frac{\frac{2\mathcal{F}}{\pi} |\sin(\pi\nu/\nu_{\text{FSR}})|}{\left[1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(\pi\nu/\nu_{\text{FSR}})\right]^{1/2}}, \\ \chi &\simeq \tan^{-1} \left[\frac{\frac{-\pi}{2\mathcal{F}} \sin(2\pi\nu/\nu_{\text{FSR}})}{1 - \cos(2\pi\nu/\nu_{\text{FSR}})} \right]. \end{aligned}$$

For small detuning, $\nu \frac{\mathcal{F}}{\nu_{\text{FSR}}} \ll 1$ the approximate expressions are

$$\begin{aligned} r &\simeq 2\mathcal{F} \frac{\nu}{\nu_{\text{FSR}}} = 2 \frac{\nu}{\nu_c}, \\ \chi &\simeq 2\mathcal{F} \frac{\nu}{\nu_{\text{FSR}}} = 2 \frac{\nu}{\nu_c} \end{aligned}$$

where we have introduced the cavity FWHM linewidth $\nu_c = \nu_{\text{FSR}}/\mathcal{F}$. Expanding (13.46) to lowest order in ν we find an analytical approximation to the error signal

$$V_{\text{error}} \simeq \frac{4\mathcal{E}_{\text{in}}^2 J_0(\epsilon) J_1(\epsilon)}{\pi} \frac{\nu}{\nu_c} \mathcal{F} \frac{\sin(\pi\Omega/\omega_{\text{FSR}})}{\left[1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2(\pi\Omega/\omega_{\text{FSR}})\right]^{1/2}} \frac{\frac{\pi}{2\mathcal{F}} \cos(\phi) \cos(\pi\Omega/\omega_{\text{FSR}}) - \sin(\phi)}{\left[1 + \left(\frac{\pi}{2\mathcal{F}}\right)^2 \cot^2(\pi\Omega/\omega_{\text{FSR}})\right]^{1/2}}. \quad (13.47)$$

We see that in a high finesse cavity the error signal amplitude is optimized by choosing $\phi = \pi/2$ for the local oscillator phase. Furthermore the slope of the error signal is proportional to the frequency offset normalized to the cavity linewidth $dV_{\text{error}}/d(\nu/\nu_c) \sim \text{constant}$ in a high finesse cavity. With $\phi = \pi/2$ and $\mathcal{F} \gg 1$ we find

$$V_{\text{error}} \simeq -2\mathcal{E}_{\text{in}}^2 J_0(\epsilon) J_1(\epsilon) \frac{\nu}{\nu_c}. \quad (13.48)$$

Finally choosing $\epsilon = 1.08$ we get a maximal slope of $V_{\text{error}} \simeq -0.68 \mathcal{E}_{\text{in}}^2 \frac{\nu}{\nu_c}$.

This signal can be used in a feedback loop to stabilize the frequency of the laser light relative to the resonator. The error signal using the full dependence on ν from Eq. (13.46) is shown in Fig. 13.14. We see in agreement with the linearized analysis that the largest error signal and steepest slope near resonance are obtained for $\phi = \pi/2$ which corresponds to a local oscillator $V_{\text{lo}} = \cos(\Omega t + \phi) = -\sin(\Omega t)$. This could be anticipated since the EOM modulated light has sidebands with positive and negative amplitudes corresponding to quadrature phase modulation of

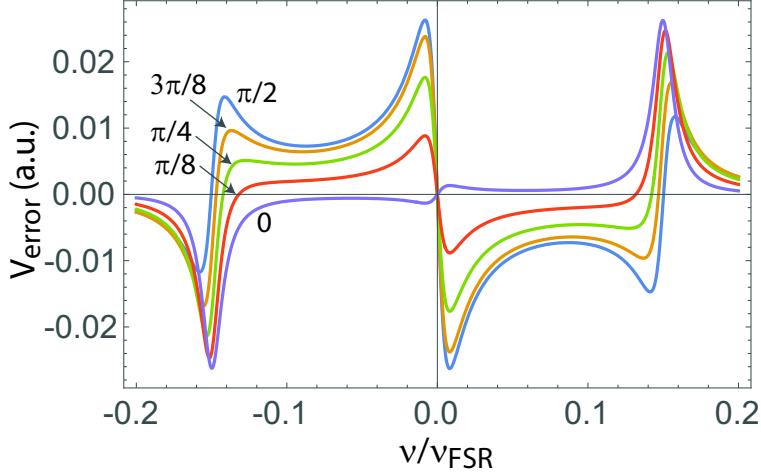


Figure 13.14: PDH error signal for $R_{\text{in}} = R_{\text{out}} = 0.95$, $R_{\text{cav}} = 1$, giving a cavity finesse of $F = 61$, $(\Omega/2\pi)/v_{\text{FSR}} = 0.1$. The curves are labeled with the local oscillator phase ϕ .

the carrier. The optimal reference oscillator should be in phase with the modulated carrier, which implies $\phi = \pi/2$. In practice the phase shift needs to be adjusted empirically due to time delays in the optical and electronic signal paths.

Alternatively, if the laser is stable, but the resonator is not, the error signal can be fed back to a piezo actuator holding one of the resonator mirrors to stabilize the resonator length. This can be used to transfer stability from a stable laser to a less stable one. The resonator is locked to the stable laser and then the unstable laser is locked to the resonator. In this way the relative frequency of the two lasers can be stabilized to a high degree.

13.4.2 Frequency response of PDH error signal

The actual performance of a PDH lock depends on the noise level of the error signal and the characteristics of the frequency dependent gain in the electronic feedback loop. We will primarily be interested in the error signal response due to a laser fluctuation at noise frequency f_N . We start the analysis by writing the EOM output field (13.45) as

$$\mathcal{E}_{\text{EOM}} = \mathcal{E}_0 [1 + a(t)] e^{-i[\omega t - b(t)]} [J_0(\epsilon) + J_1(\epsilon)e^{-i\Omega t} - J_1(\epsilon)e^{i\Omega t}]. \quad (13.49)$$

Here $a(t), b(t)$ are the time dependent amplitude and phase of the laser field and we have assumed the maximum noise frequency is small compared to the inverse transit time of the EOM. We will primarily be interested in the PDH response to frequency fluctuations of the laser of the form

$$\omega \rightarrow \omega + \delta\omega(t) = \omega + \int d\omega_N [s_N \cos(\omega_N t) - c_N \sin(\omega_N t)]$$

with c_N, s_N coefficients that determine the amplitude of the fluctuations at dither frequency ω_N . Neglecting amplitude fluctuations the corresponding phase term is

$$\begin{aligned} b(t) &= \int_0^t dt' \int d\omega_N [s_N \cos(\omega_N t') - c_N \sin(\omega_N t')] \\ &= \int d\omega_N \int_0^t dt' [s_N \cos(\omega_N t') - c_N \sin(\omega_N t')] \\ &= \int d\omega_N \left[c_N \frac{\cos(\omega_N t)}{\omega_N} + s_N \frac{\sin(\omega_N t)}{\omega_N} \right] \\ &= \int d\omega_N |s| \sin(\omega_N t + \theta_s). \end{aligned}$$

where we have introduced the phase noise amplitude $s = \frac{s_N + i c_N}{\omega_N} = |s| e^{i\theta_s}$ with $\theta_s = \tan^{-1}(c_N/s_N)$.

To evaluate the error signal response to a single spectral component of the fluctuations we use (13.49) with $a(t) = 0$ and $b(t) = c_N \frac{\cos(\omega_N t)}{\omega_N} + s_N \frac{\sin(\omega_N t)}{\omega_N}$ which describes phase noise at offset frequency ω_N with in phase and quadrature amplitudes c_N, s_N . We then expand the fluctuation into harmonics as

$$e^{ib(t)} = e^{i[c_N \cos(\omega_N t) + s_N \sin(\omega_N t)]/\omega_N} = \sum_{n=-\infty}^{\infty} i^n J_n(c_N/\omega_N) e^{in\omega_N t} \sum_{m=-\infty}^{\infty} J_m(s_N/\omega_N) e^{im\omega_N t}.$$

Assuming the noise is weak we can write this as

$$\begin{aligned} e^{ib(t)} &\simeq [J_0(c_N/\omega_N) + i J_1(c_N/\omega_N) e^{i\omega_N t} + i J_1(c_N/\omega_N) e^{-i\omega_N t}] \\ &\quad \times [J_0(s_N/\omega_N) + J_1(s_N/\omega_N) e^{i\omega_N t} - J_1(s_N/\omega_N) e^{-i\omega_N t}] \\ &\simeq 1 + \frac{1}{2\omega_N} (s_N + i c_N) e^{i\omega_N t} + \frac{1}{2\omega_N} (-s_N + i c_N) e^{-i\omega_N t} \\ &= 1 + \frac{s}{2} e^{i\omega_N t} - \frac{s^*}{2} e^{-i\omega_N t}. \end{aligned}$$

The EOM output field is then

$$\mathcal{E}_{\text{EOM}} = \mathcal{E}_{\text{in}} [J_0(\epsilon) + J_1(\epsilon) e^{-i\Omega t} - J_1(\epsilon) e^{i\Omega t}] e^{-i\omega t} \left[1 + \frac{s}{2} e^{i\omega_N t} - \frac{s^*}{2} e^{-i\omega_N t} \right], \quad (13.50)$$

while the field reflected from the cavity and detected is given by

$$\begin{aligned} \mathcal{E}_{\text{det}} &= \mathcal{E}_{\text{in}} e^{-i\omega t} \left[J_0(\epsilon) \left(r_0 e^{i\chi_0} + \frac{sr_{0N}}{2} e^{i\chi_{0N}} e^{i\omega_N t} - \frac{s^* r_{0\bar{N}}}{2} e^{i\chi_{0\bar{N}}} e^{-i\omega_N t} \right) \right. \\ &\quad + J_1(\epsilon) \left(r_+ e^{i\chi_+} + \frac{sr_{+N}}{2} e^{i\chi_{+N}} e^{i\omega_N t} - \frac{s^* r_{+\bar{N}}}{2} e^{i\chi_{+\bar{N}}} e^{-i\omega_N t} \right) e^{-i\Omega t} \\ &\quad \left. - J_1(\epsilon) \left(r_- e^{i\chi_-} + \frac{sr_{-N}}{2} e^{i\chi_{-N}} e^{i\omega_N t} - \frac{s^* r_{-\bar{N}}}{2} e^{i\chi_{-\bar{N}}} e^{-i\omega_N t} \right) e^{i\Omega t} \right]. \quad (13.51) \end{aligned}$$

Here r_0, r_+, r_- and χ_0, χ_+, χ_- are the cavity reflection amplitudes and phase shifts for the carrier and sidebands, r_{0N}, r_{+N}, r_{-N} and $\chi_{0N}, \chi_{+N}, \chi_{-N}$ describe the cavity response at frequencies $\omega + \omega_N, \omega + \Omega + \omega_N, \omega - \Omega + \omega_N$ and the terms with N replaced by \bar{N} are the same except $\omega_N \rightarrow -\omega_N$.

The photodetector voltage signal will be mixed with a local oscillator of frequency Ω and low pass filtered. The photodetector spectral components near Ω to leading order in s are

$$\begin{aligned} V_{\text{det}}(t) = & \eta_d |\mathcal{E}_{\text{in}}|^2 J_0(\epsilon) J_1(\epsilon) \\ & \times \{2r_0 r_+ \cos(\Omega t + \chi_0 - \chi_+) - 2r_0 r_- \cos(\Omega t - \chi_0 + \chi_-) \\ & + r_0 r_{+N} |s| \cos((\Omega - \omega_N)t + \chi_0 - \chi_{+N} + \theta_s) + r_0 r_{-\bar{N}} |s| \cos((\Omega - \omega_N)t - \chi_0 + \chi_{-\bar{N}} + \theta_s) \\ & - r_+ r_{0\bar{N}} |s| \cos((\Omega - \omega_N)t - \chi_+ + \chi_{0\bar{N}} - \theta_s) - r_- r_{0N} |s| \cos((\Omega - \omega_N)t + \chi_- - \chi_{0N} - \theta_s) \\ & + r_+ r_{0N} |s| \cos((\Omega + \omega_N)t - \chi_+ + \chi_{0N} + \theta_s) + r_- r_{0\bar{N}} |s| \cos((\Omega + \omega_N)t + \chi_- - \chi_{0\bar{N}} + \theta_s) \\ & - r_0 r_{+\bar{N}} |s| \cos((\Omega + \omega_N)t + \chi_0 - \chi_{+\bar{N}} - \theta_s) - r_0 r_{-N} |s| \cos((\Omega + \omega_N)t - \chi_0 + \chi_{-N} - \theta_s)\}. \end{aligned}$$

The detector output is then mixed with a phase shifted copy of the reference oscillator $V_{\text{lo}} = \cos(\Omega t + \phi)$ and low pass filtered to obtain

$$\begin{aligned} V_{\text{error}} = & \eta_d |\mathcal{E}_{\text{in}}|^2 J_0(\epsilon) J_1(\epsilon) \\ & \times \{2r_0 r_+ \cos(\chi_0 - \chi_+ - \phi) - 2r_0 r_- \cos(-\chi_0 + \chi_- - \phi) \\ & + r_0 r_{+N} |s| \cos(\omega_N t - \chi_0 + \chi_{+N} - \theta_s + \phi) + r_0 r_{-\bar{N}} |s| \cos(\omega_N t + \chi_0 - \chi_{-\bar{N}} - \theta_s + \phi) \\ & - r_+ r_{0\bar{N}} |s| \cos(\omega_N t + \chi_+ - \chi_{0\bar{N}} + \theta_s + \phi) - r_- r_{0N} |s| \cos(\omega_N t - \chi_- + \chi_{0N} + \theta_s + \phi) \\ & + r_+ r_{0N} |s| \cos(\omega_N t - \chi_+ + \chi_{0N} + \theta_s - \phi) + r_- r_{0\bar{N}} |s| \cos(\omega_N t + \chi_- - \chi_{0\bar{N}} + \theta_s - \phi) \\ & - r_0 r_{+\bar{N}} |s| \cos(\omega_N t + \chi_0 - \chi_{+\bar{N}} - \theta_s - \phi) - r_0 r_{-N} |s| \cos(\omega_N t - \chi_0 + \chi_{-N} - \theta_s - \phi)\}. \end{aligned}$$

Let us now evaluate the error signal assuming the average laser frequency is locked to the cavity resonance so $\nu = 0$ which implies $\chi_0 = 0$, $r_0 = 0$, $r_+ = r_-$, $\chi_+ = -\chi_-$, $r_{0N} = r_{0\bar{N}}$, $\chi_{0N} = -\chi_{0\bar{N}}$, $r_{+N} = r_{-\bar{N}}$, $r_{+\bar{N}} = r_{-N}$, $\chi_{+N} = -\chi_{-\bar{N}}$, $\chi_{+\bar{N}} = -\chi_{-N}$. Setting $\phi = \pi/2$ for the maximum error signal we find

$$\begin{aligned} V_{\text{error}} = & 2\eta_d |\mathcal{E}_{\text{in}}|^2 J_0(\epsilon) J_1(\epsilon) |s| r_{+N} \\ & \times [\sin(\omega_N t - \chi_+ + \chi_{0N} + \theta_s) - \sin(\omega_N t + \chi_+ + \chi_{0N} + \theta_s)]. \end{aligned}$$

This expression can be evaluated numerically to find the error signal in response to phase noise. We can derive an analytical approximation with the assumption $\omega_N \ll \omega_{\text{FSR}}$ by expanding the error voltage to order ν_N^2 which results in

$$V_{\text{error}} = V_0 |s| \frac{\frac{\nu_N}{\nu_c} \sin(\omega_N t + \theta_s) + 2 \left(\frac{\nu_N}{\nu_c} \right)^2 \cos(\omega_N t + \theta_s)}{1 + 2\nu_N^2/\nu_c^2}$$

with

$$V_0 = \frac{4\pi\eta_d |\mathcal{E}_{\text{in}}|^2 J_0(\epsilon) J_1(\epsilon) \sin(2\pi\Omega/\omega_{\text{FSR}})}{\mathcal{F} \{ [1 + 2(\pi/2\mathcal{F})^2 - \cos(2\pi\Omega/\omega_{\text{FSR}})] [1 + (\pi/2\mathcal{F})^2 - [1 - (\pi/2\mathcal{F})^2] \cos(2\pi\Omega/\omega_{\text{FSR}})] \}^{1/2}}.$$

It is useful to rewrite the error signal as

$$V_{\text{error}} = V_0 |s| \frac{\nu_N}{\nu_c} \frac{\left(1 + \frac{4\nu_N^2}{\nu_c^2} \right)^{1/2}}{1 + \frac{2\nu_N^2}{\nu_c^2}} \sin(\omega_N t + \theta_s + \phi_s) \quad (13.52)$$

with

$$\phi_s = \tan^{-1} \left(\frac{2\nu_N}{\nu_c} \right).$$

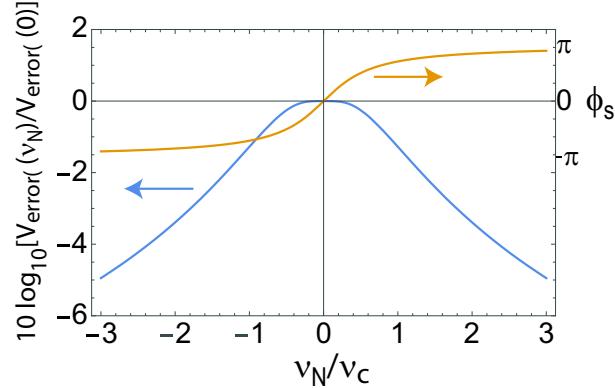


Figure 13.15: Response of PDH detector to frequency fluctuations.

The frequency dependence of the response is contained in the factor⁷

$$|s| \frac{\nu_N}{\nu_c} \frac{\left(1 + \frac{4\nu_N^2}{\nu_c^2}\right)^{1/2}}{1 + \frac{2\nu_N^2}{\nu_c^2}} = \frac{|s_N + ic_N|}{\nu_c} \frac{\left(1 + \frac{4\nu_N^2}{\nu_c^2}\right)^{1/2}}{1 + \frac{2\nu_N^2}{\nu_c^2}} \sim \frac{|s_N + ic_N|}{\nu_N}.$$

Recall that $|s| = |s_N + ic_N|/\omega_N$ with c_N, s_N the frequency fluctuation amplitudes at dither frequency Ω_N . As shown in Fig. 13.15 the error signal in response to frequency fluctuations rolls off as a function of ν_N/ν_c decaying at large frequencies $\sim 1/\nu_N$ which is a 10 dB/octave voltage roll off. On the other hand the response to phase fluctuations of magnitude $|s|$ asymptotes to a constant at large ν_N . Thus the PDH error signal acts as a frequency detector at low frequencies and a phase detector at frequencies that are large compared to the cavity linewidth. This behavior was emphasized in the original Drever-Hall paper from 1983.

It is instructive to also look at the dependence of V_0 on the choice of modulation frequency Ω . Figure 13.16 shows the normalized V_0 for several different values of the finesse. While Ω should be large enough to allow for efficient filtering and separation of signal bands near DC and around the modulation frequency choosing Ω larger than needed serves to decrease the strength of the error signal. This is in contrast to the response to the unmodulated carrier given by Eq. (13.47) which is essentially independent of Ω for Ω at least a few times larger than ω_c .

⁷We note that the algebraic form of this expression differs from other published results we have found including W. Nagourney, *Quantum electronics for atomic physics*, Oxford University Press, Oxford (2010), Eq. (4.65); T. Day, E. K. Gustafson, and R. L. Byer, IEEE J. Qu. Electr. **28**, 1106 (1992); D. Shoemaker, et al. Opt. Lett. **14**, 609 (1989).

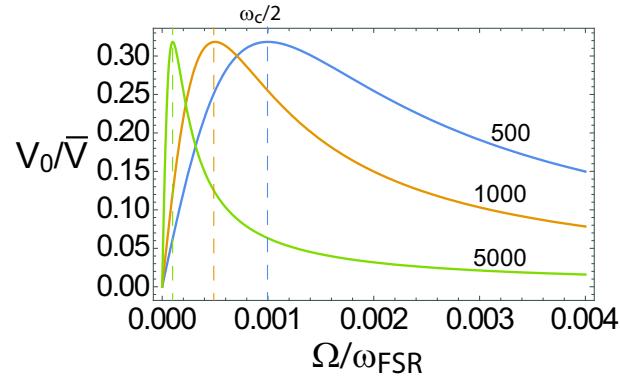


Figure 13.16: Response of PDH detector normalized to $\bar{V} = 4\pi\eta_d|\mathcal{E}_{\text{in}}|^2 J_0(\epsilon)J_1(\epsilon)$ for $\mathcal{F} = 500, 1000, 5000$ as a function of modulation frequency. The vertical dashed lines show the corresponding values of $\omega_c/2$.

13.4.3 Tilt-Locking

An interesting method of locking an optical field to a reference cavity uses tilt-locking[117]. Let the one-dimensional optical field consist of two modes as

$$A(x, t) = A_{00}(x, t) + A_{01}(x, t)$$

where $A_{00}(x, t) = e^{-x^2} e^{-\omega t} e^{i\phi_0}$ and $A_{01}(x) = xe^{-x^2} e^{-\omega t}$. A large detector will give the signal

$$\begin{aligned} V &\sim \int dx |A|^2 \\ &= \int dx [|A_{00}|^2 + |A_{01}|^2 + A_{00}A_{01}^* + A_{00}^*A_{01}] \\ &= V_{00} + V_{01} + V_{\text{beat}}. \end{aligned}$$

Since the 00 and 01 modes are orthogonal, as can be easily checked, $V_{\text{beat}} = 0$, so $V = V_{00} + V_{01}$.

Suppose now we replace the large detector with a split detector that produces a signal V_+ that is the integrated intensity for $x > 0$ and V_- that is the integrated intensity for $x < 0$. Then

$$\begin{aligned} V_+ &= \frac{V_{00}}{2} + \frac{V_{01}}{2} + \int_0^\infty dx (A_{00}A_{01}^* + A_{00}^*A_{01}) \\ V_- &= \frac{V_{00}}{2} + \frac{V_{01}}{2} + \int_{-\infty}^0 dx (A_{00}A_{01}^* + A_{00}^*A_{01}). \end{aligned}$$

The difference signal obtained by subtracting the photodiode outputs is thus

$$\begin{aligned} V_d &= V_+ - V_- \\ &= \int_0^\infty dx [A_{00}A_{01}^* + A_{00}^*A_{01}] - \int_{-\infty}^0 dx [A_{00}A_{01}^* + A_{00}^*A_{01}] \\ &= \int_0^\infty dx [A_{00}(x)A_{01}^*(x) + A_{00}^*(x)A_{01}(x) - A_{00}(-x)A_{01}^*(-x) - A_{00}^*(-x)A_{01}(-x)]. \end{aligned}$$

Then use $A_{00}(x) = A_{00}(-x)$ and $A_{01}(x) = -A_{01}(-x)$ to get

$$\begin{aligned} V_d &= 2 \int_0^\infty dx [A_{00}(x)A_{01}^*(x) + A_{00}^*(x)A_{01}(x)] \\ &= 2 \int_0^\infty dx [A_{00}(x)A_{01}^*(x) + A_{00}^*(x)A_{01}(x)] \\ &= 4 \cos(\phi_0) \int_0^\infty dx x e^{-2x^2} \\ &= \cos(\phi_0). \end{aligned}$$

Note that for $\phi_0 = \pi/2$ the difference signal is zero (this is the case sketched in Fig. 1 of the paper by Shaddock, et al.), whereas for other relative phase shifts the difference will be positive or negative. The value of ϕ_0 will, for fixed cavity length, depend on the tilt angle. Increasing the tilt angle will increase the magnitude of ϕ_0 . If we now tune the cavity (or change the laser frequency) then ϕ_0 will change to $\phi = \phi_0 + \theta_{00}$ where θ_{00} is the phase shift of the reflected 00 mode. The phase of the 01 mode does not change since it has odd parity and does not couple to the resonator mode. Thus the difference signal will be

$$V_d = \cos(\phi_0 + \theta_{00}).$$

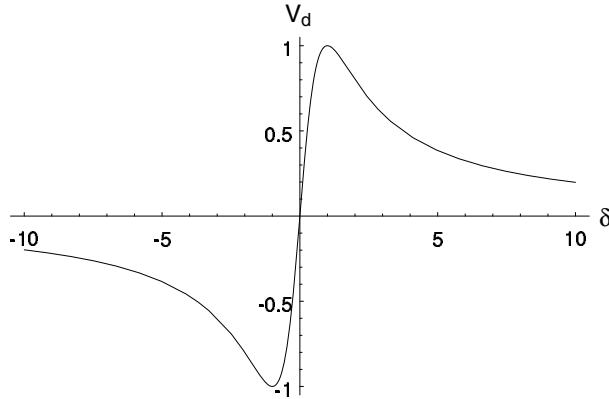


Figure 13.17: Error signal as a function of cavity tuning.

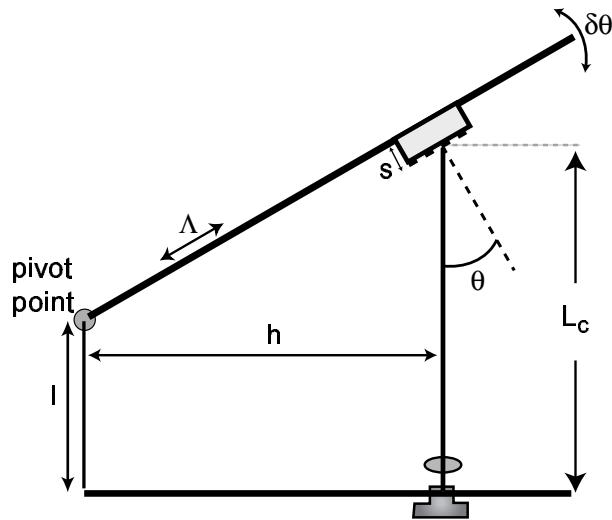


Figure 13.18: Geometrical layout of a Littrow laser cavity.

If we choose the working point at $\phi_0 = \pi/2$ and the cavity has a high finesse so that $\theta_{00} \simeq 2 \tan^{-1}(\delta)$ where $\delta = (\omega - \omega_c)/\gamma$ is the normalized cavity detuning. Then

$$V_d \sim \sin [2 \tan^{-1}(\delta)] = 2 \frac{\delta}{1 + \delta^2}.$$

Figure 13.17 shows V_d which has the desired dispersive shape. For a quantitative comparison with experiment the full formulas for the Fabry-Perot phase shift as a function of cavity tuning should be used. This type of angle dependent phase shift can also be used for stabilization of the angle of a laser beam relative to a reference resonator[118].

13.4.4 Littrow laser scanning

A convenient approach to building a tunable laser diode is to use feedback from a diffraction grating in the so-called Littrow configuration where the first order diffracted order propagates back towards the laser as shown in Fig. 13.18. Tilting the grating changes the angle θ as well as the cavity length L_c . When the geometry is chosen correctly the optical frequency determined by the diffraction

grating angle stays in resonance with the changing cavity length so that the laser can be tuned continuously without mode hops.

The conditions for synchronous scanning are found as follows. The grating equation for the Littrow configuration of diffraction into the -1 order is $\sin(\theta) = \lambda_d/(2\Lambda)$ where λ_d is the optical wavelength, Λ is the diffraction grating period, and θ is the angle the beam makes with the grating normal. The roundtrip cavity phase for a beam of wavelength λ that starts at the laser, propagates to the grating, and returns to the laser, is

$$\phi = \phi_0 + 2\frac{2\pi}{\lambda}L_c - \frac{2\pi}{\Lambda}x$$

where $L_c = l + h \tan(\theta) - s/\cos(\theta)$, $x = h/\cos(\theta) - s \tan(\theta)$ is the coordinate in the grating plane where the beam meets the grating, and ϕ_0 is a fixed phase offset. We assume that at wavelength λ_d the horizontal offset h has been chosen such that the total round trip phase is $\phi = 2\pi n$, with n an arbitrary integer. The cavity is then on resonance.

When the grating is rotated the wavelength $\lambda_d = 2\Lambda \sin(\theta)$ changes. In order to keep the cavity on resonance we require that $d\phi/d\theta = 0$. This condition can be written as

$$\begin{aligned} \frac{d\phi}{d\theta} &= \frac{4\pi}{\lambda} \frac{dL_c}{d\theta} - \frac{4\pi L_c}{\lambda^2} \frac{d\lambda}{d\theta} - \frac{2\pi}{\Lambda} \frac{dx}{d\theta} \\ &= \frac{2\pi}{\Lambda} \left[\frac{h}{\sin(\theta)} - L_c \frac{\cos(\theta)}{\sin^2(\theta)} \right]. \end{aligned} \quad (13.53)$$

The cavity phase variation is zero for $L_c = h \tan(\theta)$ or $l = s/\cos(\theta)$.

Lack of precision in constructing the laser will nonetheless result in a lack of synchronism during a wavelength scan. To evaluate how much error in the offset distance l can be tolerated we need to know the width of the resonances due to the diffraction grating and cavity length. Typical diffraction efficiencies for gratings used in the Littrow configuration are 10-50 % resulting in a relatively low finesse Fabry Perot longitudinal mode structure. The width of the diffraction grating reflectivity as a function of laser frequency can be estimated as $\Delta\theta = \theta/p$ where the number of lines of the diffraction grating that are illuminated by a beam of width d is given by $p = d/(\Lambda \cos(\theta))$. Thus $\Delta\theta = \theta \cos(\theta)\Lambda/d$ and the number of longitudinal cavity modes inside the diffraction peak is roughly

$$\begin{aligned} m &= \Delta\theta \left| \frac{\delta\nu_c}{\delta\theta} \right| \frac{1}{\nu_{\text{FSR}}} \\ &= \theta \frac{N}{p}. \end{aligned} \quad (13.54)$$

The longitudinal mode index is roughly $N = 2L_c/\lambda$ which gives

$$m = \theta \frac{2L_c \Lambda \cos(\theta)}{\lambda d}.$$

Using $\lambda = .78 \mu\text{m}$, $\Lambda = 1/1200 \text{ mm}$, and $d = 4 \text{ mm}$ we get $\theta = 28^\circ$ and $m \simeq 6.9$. Thus there are always a handful of longitudinal modes under the peak of the diffraction response which implies that the cavity will oscillate at any value of L_c .

Returning to the question of synchronicity a naive estimate is that the laser will mode hop when $|\delta\nu_d - \delta\nu_c| \sim \nu_{\text{FSR}}$. Assuming $l = l_0 + y$ we can show that the scan distance $\delta\nu$ before the synchronicity error is ν_{FSR} is given by

$$\delta\nu = \nu_{\text{FSR}} \frac{L_c}{y}.$$

Using the above parameters and $y = 1$ mm gives $\delta\nu = 150$ GHz. Although this estimate is quite crude we expect that building the cavity with the error in l not more than ~ 1 mm should allow scans of 10's of GHz.

13.4.5 Frequency comb stability

The optical frequency of one tooth of an optical frequency comb is given by

$$f = Nf_{\text{rep}} + f_{\text{off}}$$

where N is the index of the comb tooth, f_{rep} is the cavity repetition rate, and f_{off} is the offset frequency from the nearest tooth which comes from the $f - 2f$ lock. The uncertainty in f due to the uncertainty of the repetition rate and offset lock circuits is

$$\delta f = N\delta f_{\text{rep}} + \delta f_{\text{off}}.$$

Consider a Menlo Systems Yb fiber comb with $f_{\text{rep}} = 250$ MHz. Typical numbers for the comb lock uncertainties are $\delta f_{\text{rep}} = 3$ mHz and $\delta f_{\text{off}} = 5$ Hz. We use this to look at a 920 nm light beam with frequency 326 THz, so $N = 1.3 \times 10^6$. The lock uncertainties therefore imply

$$\delta f = 3.9 \text{ kHz}.$$

In addition each tooth of the comb is broadened to about $\delta f_{\text{tooth}} = 100$ kHz due to phase noise. The δf set by the lock uncertainties can only be reached after averaging for a time t such that $\delta f_{\text{tooth}}/\sqrt{\delta f_{\text{tooth}}t} \sim \delta f$ or

$$t \sim \frac{\delta f_{\text{tooth}}}{(\delta f)^2} = 6.6 \text{ ms}.$$

This time is short enough that it is not of concern for our application of stabilizing a cw laser which is narrowed using a stable cavity.

In addition to this uncertainty there is the uncertainty due to the instability of the low frequency reference oscillator. The repetition rate is referenced to a to a Novus OXCO with $f_{\text{osc}} = 10$ MHz. The oscillator is slaved to a GPS timing signal for long term stability. The repetition rate multiplier is $N_{\text{osc}} = f_{\text{rep}}/f_{\text{osc}} = 25$. The stability floor due to the external reference is

$$\delta f' = (NN_{\text{osc}})\delta f_{\text{osc}}.$$

The Novus has $\delta f_{\text{osc}}/f_{\text{osc}} = 4 \times 10^{-10}$ at 1 sec. giving

$$\delta f' = 130 \text{ kHz}.$$

At 10 sec. the Novus improves to $\delta f_{\text{osc}}/f_{\text{osc}} = 2.5 \times 10^{-11}$ giving

$$\delta f' = 8.1 \text{ kHz}.$$

Influence of the reference oscillator phase noise has not been considered ♠??

A more stable reference oscillator would make a big difference at 1 sec. averaging time, but at 10 sec. we could only win a factor of about 2, given our current comb lock uncertainties.

Chapter 14

Electronics

This chapter is a collection of basic reference information about electronic components and circuits.

14.1 Linear components

Circuits are described in terms of their voltage V and current I . Basic linear components are resistors, symbol R , capacitors, symbol C , and inductors symbol L .

The impedance relation for resistors (Ohms law) is $I = V/R$ or

$$Z_R = \frac{V}{I} = R.$$

For capacitors we use $V = Q/C$ with Q the charge. Since $I = dQ/dt$ we have $V = \int dt I(t)/C$ or

$$\frac{dV}{dt} = \frac{I}{C}.$$

Now introduce Fourier transform pairs

$$\begin{aligned} V(\omega) &= \int dt V(t) e^{i\omega t} \\ V(t) &= \int d\omega V(\omega) e^{-i\omega t}. \end{aligned}$$

Fourier transforming gives $-i\omega V(\omega) = I(\omega)/C$ and the impedance is

$$Z(\omega) = \frac{V(\omega)}{I(\omega)} = \frac{i}{\omega C}.$$

For inductors we use $V = L \frac{dI}{dt}$ so $V(\omega) = -i\omega L I(\omega)$ and the impedance is

$$Z(\omega) = \frac{V(\omega)}{I(\omega)} = -i\omega L.$$

It is common to use electrical engineering notation which assumes a time dependence $V(t) \sim e^{j\omega t}$ and impedances $Z_R = R$, $Z_C = -j/\omega C$, $Z_L = j\omega L$. With this convention a positive time delay corresponds to a negative $\arg[Z]$.

Components in series have their impedances summed,

$$Z = Z_1 + Z_2.$$

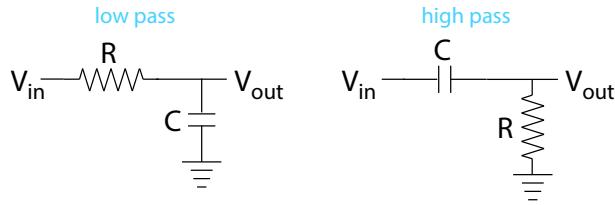


Figure 14.1: Low and high pass filters.

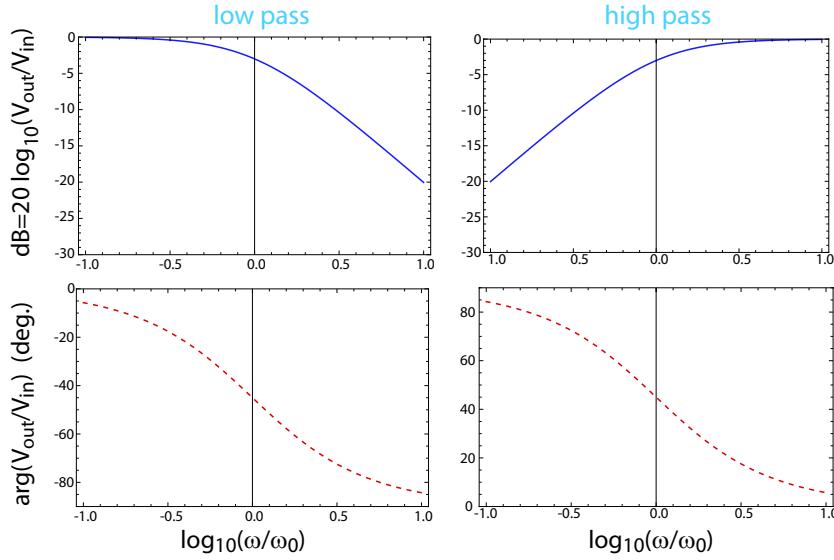


Figure 14.2: Amplitude and phase response of low and high pass filters.

Components in parallel combine as $1/Z = 1/Z_1 + 1/Z_2$ so

$$Z = \frac{1}{1/Z_1 + 1/Z_2} = \frac{Z_1 Z_2}{Z_1 + Z_2}.$$

14.1.1 Low and high pass filters

First order low and high pass filters with R and C are shown in Fig. 14.1. For the low pass filter $Z_{LP} = Z_C/(Z_R + Z_C) = \frac{1}{1+j\omega RC}$. Define the RC time constant $\tau = RC$ so

$$Z_{LP} = \frac{1}{1 + j\omega\tau}.$$

For the high pass filter

$$Z_{HP} = Z_R/(Z_R + Z_C) = \frac{j\omega\tau}{1 + j\omega\tau}.$$

The amplitude and phase are shown in Fig. 14.2. We see that the power is down by 3 dB at $\omega_0 = 1/\tau$ and falls off at 20 dB/decade. The phase lag is 45 deg. at $\omega = 1/\tau$ and asymptotes to 90 deg. at high frequency.

If we want a faster falloff, say 40 dB /decade we can cascade two filters. The response is

$$Z_2 = Z^2 = \frac{1}{(1 + j\omega/\tau)^2}.$$

At $\omega = 1/\tau$ we have 6 dB attenuation and 90 deg. phase shift. At higher frequency the power falls off as 40 dB/decade and the asymptotic phase lag is 180 deg.

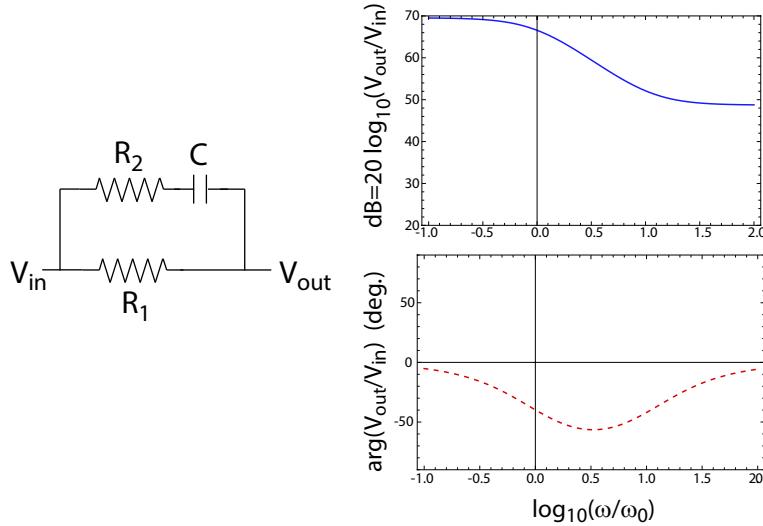


Figure 14.3: Phase advance circuit. The response is shown for $R_1 = 3000$ Ohms, $R_2 = 300$ Ohms, $C = 5$ nF giving $\omega_0/2\pi = 9650$ Hz.

14.1.2 Phase advance

Although we can add more filter stages to achieve faster decay the phase lag makes such filters unstable when they are included in a feedback loop. We can partially compensate with a phase advance circuit as in Fig. 14.3. This has the transfer function

$$Z = R_1 \frac{1 + jR_2C\omega}{1 + j(R_1 + R_2)C\omega}.$$

Putting $\tau = (R_1 + R_2)C$ this becomes

$$Z = R_1 \frac{1 + j \frac{R_2}{R_1 + R_2} \omega \tau}{1 + j \omega \tau}.$$

When $R_2 < R_1$ this looks like a low frequency low pass filter with a higher frequency high pass filter. The overall effect is a low pass filter with a constant amplitude and zero phase lag at high frequencies. The response is shown in Fig. 14.3.

14.2 Operational amplifiers

Operational amplifiers add gain and can be used to achieve filtering functions that are not readily available using only passive components.

14.2.1 Inverting configuration

When the input voltage is connected to the negative input the transfer function of the op-amp circuit shown in Fig. 14.4 is

$$H(\omega) = -Z_f(\omega)/Z_{\text{in}}(\omega).$$

Let's assume the feedback block is configured as parallel R_f, C_f so

$$Z_f = \frac{R_f}{1 + j\omega\tau}$$

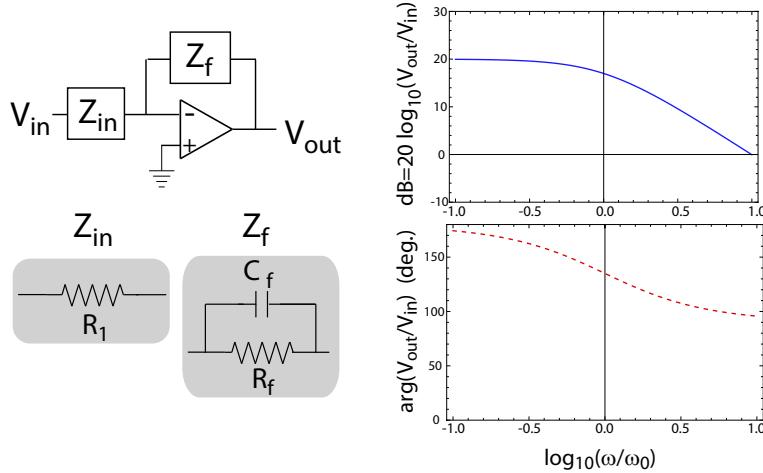


Figure 14.4: Inverting opamp. The calculated response is for $R_f = 10 \text{ k}\Omega$, $C_f = 1 \text{ nF}$, $R_1 = 1 \text{ k}\Omega$.

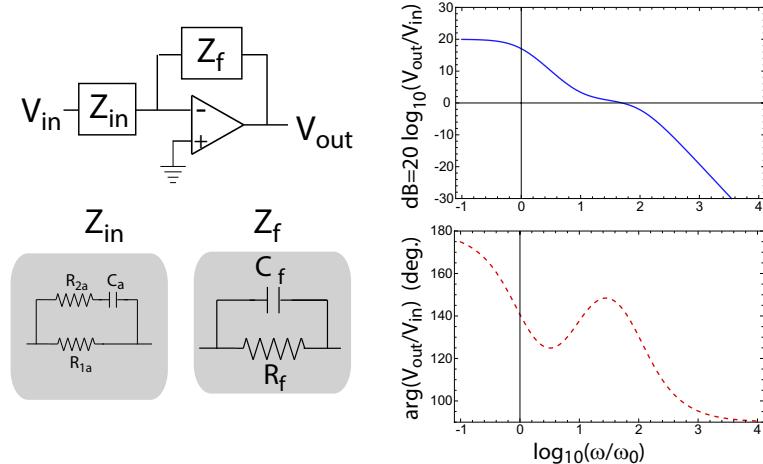


Figure 14.5: Inverting opamp with phase advance input. The response is shown for $R_f = 10. \text{ k}\Omega$, $C_f = 1 \text{ nF}$, $R_{1a} = 1. \text{k}\Omega$, $R_{2a} = 100. \Omega$, $C_a = 1 \text{ nF}$, giving $\omega_f/2\pi = 15.9 \text{ kHz}$, $\omega_a/2\pi = 144.7 \text{ kHz}$.

with $\tau = R_f C_f$. The input block is $Z_1 = R_1$ so the amplifier transfer function is

$$Z = -\frac{R_f/R_1}{1 + j\omega\tau}.$$

The response is shown in Fig. 14.4. As in the passive RC filters there is a roll off at -20 dB/decade and a 90 deg. phase shift across the resonance. The advantage compared to a passive filter is the presence of gain.

To reduce the phase shift we can switch the Z_{in} block to the phase advance circuit of Fig. 14.3. The result is shown in Fig. 14.5.

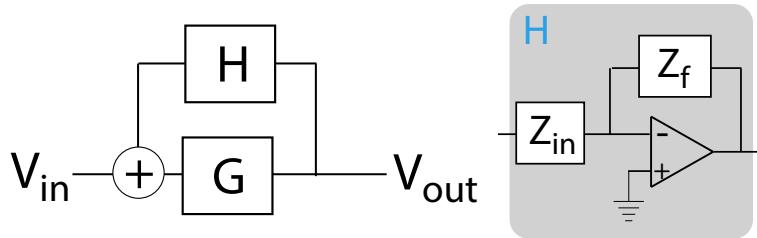


Figure 14.7: a Arrangement for locking a laser beam to an optical cavity using a dither lock. b shows the H and G circuit components.

14.2.2 Electro-optical servo loops

Suppose we use the opamp integrator as part of a closed loop feedback servo. We now refer to the integrator transfer function as $H(\omega)$ and consider the circuit shown in Fig. 14.6. The time domain voltages satisfy

$$G(V_{in} + HV_{out}) = V_{out}.$$

Solving for the output we find the closed loop response

$$Z_{CL} = \frac{V_{out}}{V_{in}} = \frac{G}{1 - GH}.$$

The mapping of an actual optical stabilization circuit onto a closed loop topology is not always obvious. An example for a dither lock to a cavity is given in Fig. 14.7.

Another example is a laser noise eater where H is defined by a photodetector and inverting integrator (as in the previous example) and the actuator G is an acousto-optic modulator. A simple model for G is a lowpass filter with corner frequency ω_G so

$$G = \frac{G_0}{1 + j\omega/\omega_G}$$

where G_0 accounts for the dc response of the actuator. The closed loop response is then

$$\begin{aligned} Z_{CL} &= \frac{G_0}{1 + j\omega/\omega_G} \frac{1}{1 + \frac{1}{1+j\omega/\omega_G} \frac{R_f/R_1}{1+j\omega\tau}} \\ &= \frac{G_0}{1 + j\omega/\omega_G + \frac{R_f/R_1}{1+j\omega\tau}}. \end{aligned}$$

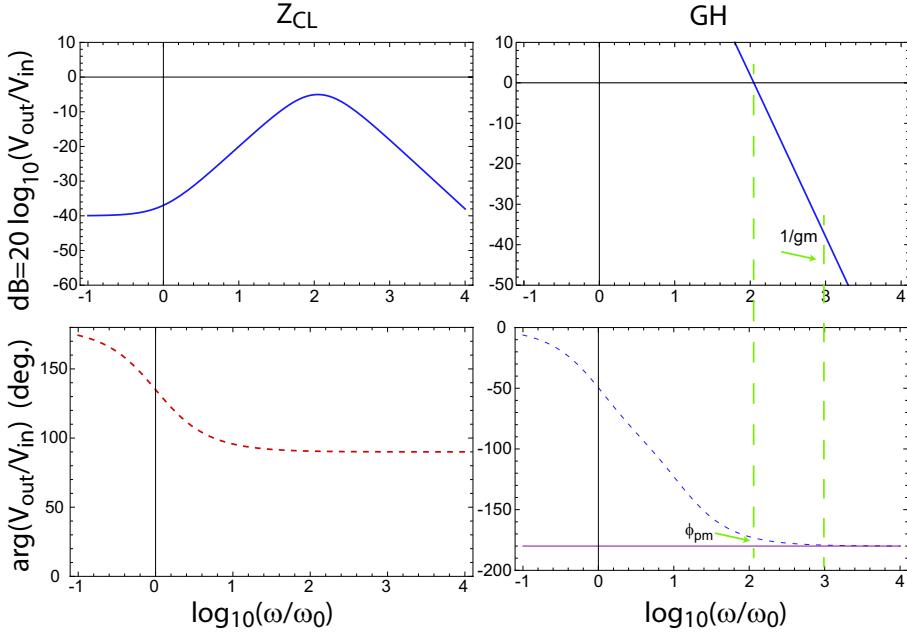


Figure 14.8: Noise eater response with closed loop transfer function (left) and GH (right). Parameters: $G_0 = -10.$, $R_1 = 1. \text{ k } \Omega$, $R_f = 100. \text{ k } \Omega$, $C_f = 0.2 \text{ nF}$, giving $\omega_0/2\pi = 8. \text{ kHz}$ and $\omega_G/2\pi = 100 \text{ kHz}$.

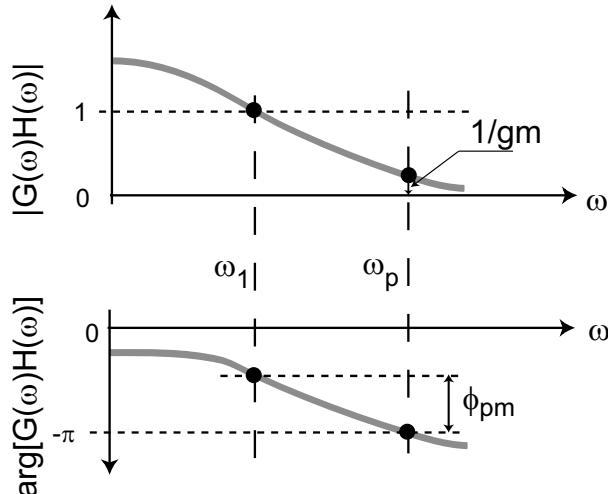


Figure 14.9: Gain and phase margins.

For the laser noise eater $G_0 < 0$ since the beam is attenuated.

An example of a calculated response is shown in Fig. 14.8. The stability of this circuit depends on GH never being equal to -1 since that would give a pole in the response and oscillation. The relative stability can be quantified in terms of gain and phase margins which are shown in Fig. 14.9. The phase margin is the amount by which the phase differs from $-\pi$ at the frequency of unity gain. The inverse gain margin is the value of the gain at the frequency where the phase reaches $-\pi$. Ideally the phase and gain margins should be as large as possible for stable operation.

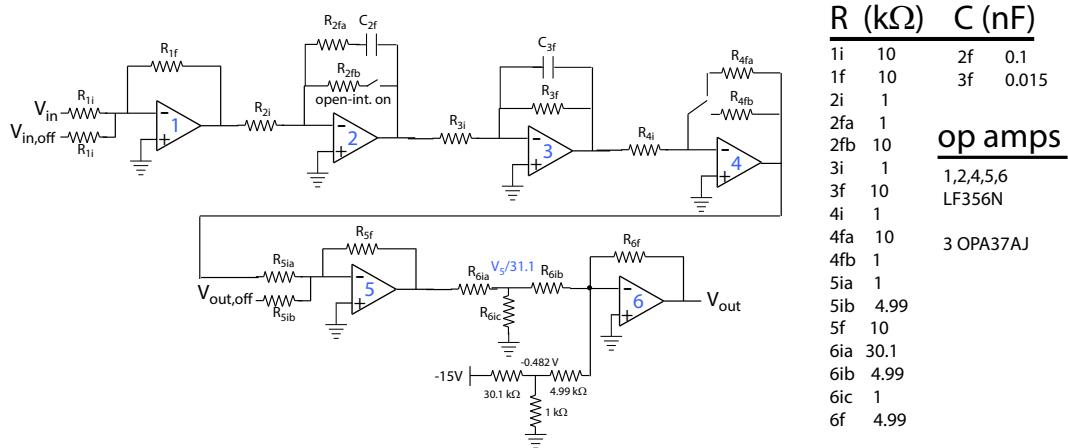
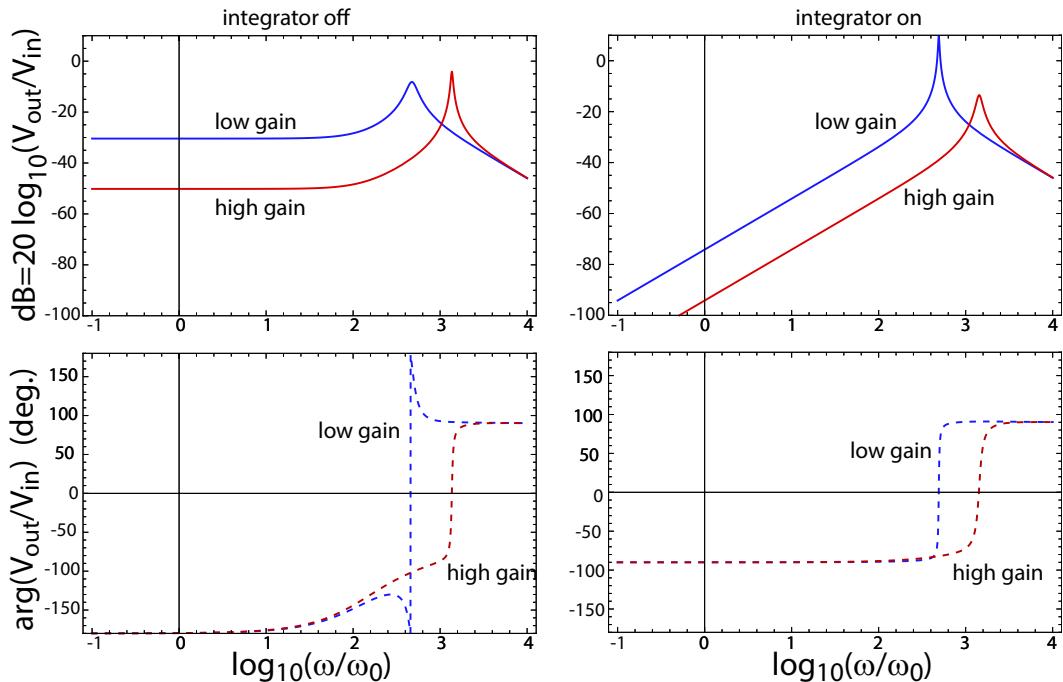


Figure 14.10: Noise eater circuit.

Figure 14.11: Closed loop response of fast noise eater circuit. The actuator is modeled as a low pass filter with $\omega_G/2\pi = 50$ kHz and dc gain $G_0 = -1$. The logarithmic frequency scale is referenced to $\omega_0 = 2\pi \times 1$ kHz.

14.3 Analysis of a noise eater

An example of a fast noise eater circuit is shown in Fig. 14.10. The gain is divided into six opamp stages. The component values are given in the figure. Stages 1,4,5,6 are flat gain with 1,5 providing input and output offsets. Stage 2 can be switched to integrate mode, stage 3 is low pass gain, and stage 4 has switchable gain. The divider network and fixed offset voltage before stage 6 serve to limit the maximum possible output to avoid damage to the AOM driver.

Figure 14.11 shows the closed loop response. Whether or not the circuit is unstable depends on the actual value of the actuator gain.

14.3.1 Measuring RIN and Power spectrum units

RIN can be measured with an electronic spectrum analyzer, for example the Stanford Research SR785. This unit lists the following as equivalent levels:

$$34 \text{ } dBV_{\text{pk}} \quad 31 \text{ } dBV_{\text{rms}} \quad 50.1 \text{ } V_{\text{pk}} \quad 35.4 \text{ } V_{\text{rms}}$$

This makes sense with the following definitions. For a sine wave $V = a \sin(\omega t)$ we have

$$V_{\text{pk}} = a$$

$$V_{\text{rms}} = \left(\frac{1}{T} \int_0^T dt' V^2(t') \right)^{1/2} = \frac{a}{\sqrt{2}}.$$

so

$$V_{\text{rms}} = \frac{V_{\text{pk}}}{\sqrt{2}}$$

which agrees with the last two SR785 equivalents. If we define

$$dBV_{\text{pk}} = 20 \log_{10}(V_{\text{pk}}), \quad dBV_{\text{rms}} = 20 \log_{10}(V_{\text{rms}})$$

then $50.1 \text{ } V_{\text{pk}}$ is $34.0 \text{ } dBV_{\text{pk}}$ or $31.0 \text{ } dBV_{\text{rms}}$.

In order to extract a RIN spectrum from SR785 measurements we do the following. Take spectra on the SR785 using PSD (power spectral density) units of $dB(V_{\text{rms}}/\sqrt{\text{Hz}})$. When using PSD units the displayed voltages have been normalized by the bandwidth of each bin. Note that this value is $20 \log_{10}(V_{\text{rms}}/\sqrt{\text{Hz}})$ so we will divide it by 2 to get $10 \log_{10}(V_{\text{rms}}/\sqrt{\text{Hz}})$.

The RIN_{dB} value is therefore

$$\text{RIN}_{\text{dB}} = 0.5 \times dB(V_{\text{rms}}/\sqrt{\text{Hz}}) - 10 \log_{10}(V_{\text{dc}}).$$

Here $dB(V_{\text{rms}}/\sqrt{\text{Hz}})$ is the numerical value displayed by the SR785 and V_{dc} is the time averaged dc value of the signal. This can be separately measured or extracted from the dc component of the SR785 spectrum, i.e. $V_{\text{dc}} = 10^{dB(V_{\text{rms}}/\sqrt{\text{Hz}})/20}$ where the argument in the exponent is the zero frequency entry of the spectrum. Alternatively one can choose not to use PSD units and normalize out the bandwidth dependence by hand.

Appendix A

Fourier analysis

Fourier analysis is an indispensable tool for describing optical wave propagation. This appendix provides a review.

A.1 Fourier transforms

Any function of time $f(t)$ can be thought of as a superposition of frequency components $g(\omega)$ that represent pure oscillations at frequency ω . If $f(t)$ is perfectly periodic, say $f(t) = \cos(\omega_0 t)$, then there is a single frequency component, $g(\omega) \sim \delta(\omega - \omega_0)$ and we can write

$$f(t) \sim \int_0^\infty d\omega g(\omega) \cos(\omega t) = \int_0^\infty d\omega \delta(\omega - \omega_0) \cos(\omega t) = \cos(\omega_0 t).$$

A more complicated function of time is $f(t) = \sum_j c_j \cos(\omega_j t)$ where c_j are amplitude coefficients. In this case $g(\omega) \sim \sum_j c_j \delta(\omega - \omega_j)$ and

$$f(t) \sim \int d\omega g(\omega) \cos(\omega t) = \int d\omega \sum_j c_j \delta(\omega - \omega_j) \cos(\omega t) = \sum_j c_j \cos(\omega_j t).$$

As a specific example consider a superposition of odd multiples of ω_0 with $c_j = (-1)^j / (2j + 1)$ for $j = 0, 1, 2, 3, \dots$ so

$$g(\omega) \sim \sum_{j=0}^N \frac{(-1)^j}{2j+1} \delta[\omega - (2j+1)\omega_0]$$

where $N + 1$ is the number of frequency components. The time dependence is

$$f(t) \sim \int d\omega g(\omega) \cos(\omega t) = \sum_{j=0}^N \frac{(-1)^j}{2j+1} \cos[(2j+1)\omega_0 t].$$

Figure A.1 shows $f(t)$ for increasing values of N . We see that the coefficients c_j lead to an accurate representation of a square wave, with the accuracy increasing as N increases.

The spectral components span a frequency range of $\omega_0 \leq \omega \leq (2N + 1)\omega_0$. An increase in ω_0 would lead to a decrease in the period of the square wave, and vice versa. This type of inverse relationship between $f(t)$ and $g(\omega)$ is generic. A narrow $f(t)$ implies a broad $g(\omega)$ while a broad $f(t)$ implies a narrow $g(\omega)$. We can quantify the range of $f(t)$ and $g(\omega)$ by their variances $\sigma_t^2 = \langle (t - \langle t \rangle)^2 \rangle$, $\sigma_\omega^2 = \langle (\omega - \langle \omega \rangle)^2 \rangle$, where the angle brackets denote integrals over the independent

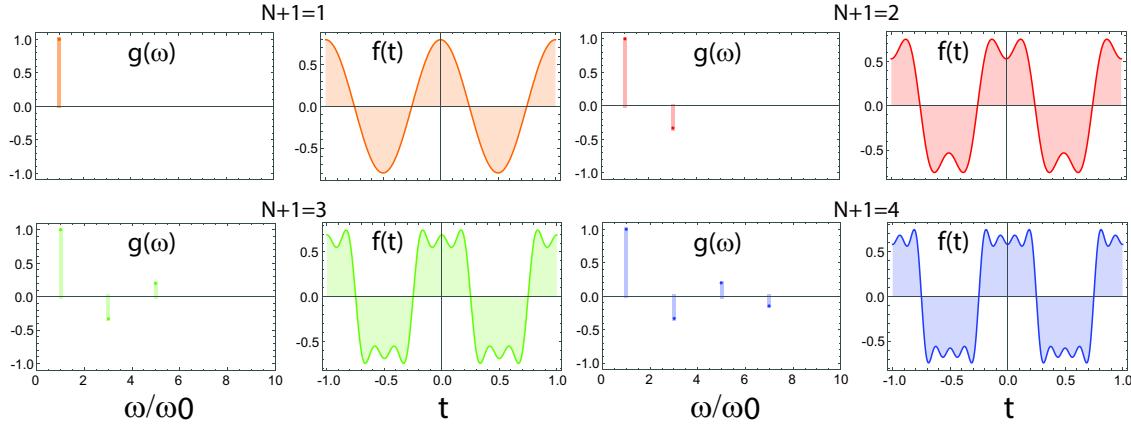


Figure A.1: Fourier series representation of a square wave for 1, 2, 3, 4 spectral components.

variable with the weighting functions $|f|^2$ and $|g|^2$. It is a fundamental result of Fourier analysis that the product $\sigma_t \sigma_\omega \geq 1/2$ for any f, g that are transform pairs. In the context of optics this implies a physical limit to the resolution of a lens. In the context of quantum mechanics this is a mathematical statement of Heisenberg's uncertainty relation that says a particle cannot simultaneously be localized in position and momentum. As we will see the uncertainty limit is saturated for Gaussian functions.

This representation in terms of components $\cos(\omega t)$ can only describe even functions of t . To include odd functions we add components proportional to $\sin(\omega t)$. More generally we wish to represent complex functions of t , such as the optical field amplitude. To do this we introduce the complex Fourier transforms

$$g(\omega) = \mathcal{F}[f(t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt f(t) e^{i\omega t} \quad (\text{A.1a})$$

$$f(t) = \mathcal{F}^{-1}[g(\omega)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega g(\omega) e^{-i\omega t} \quad (\text{A.1b})$$

We can demonstrate that these definitions are compatible since

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega g(\omega) e^{-i\omega t} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt' f(t') e^{i\omega t'} \right] e^{-i\omega t} \\ &= \int_{-\infty}^{\infty} dt' f(t') \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{i\omega(t'-t)} \right]. \end{aligned}$$

The contents of the square brackets in the last line are a representation of the delta function

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{i\omega t}. \quad (\text{A.2})$$

Using this in the previous equations gives

$$f(t) = \int_{-\infty}^{\infty} dt' f(t') \delta(t' - t) = f(t)$$

as expected.

The delta function is a generalized function that can be represented in an infinity of different ways. To prove the validity of (A.2) introduce

$$h_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/(2\sigma^2)}$$

and for any function $f(t)$ define the integral

$$I = \int_{-\infty}^{\infty} dt h_\sigma(t) f(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dt f(t) e^{-t^2/(2\sigma^2)}.$$

Change variables to $t' = t/\sigma$ so that

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt' f(\sigma t') e^{-t'^2/2}.$$

Now take the limit as $\sigma \rightarrow 0$ giving

$$I = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt' f(\sigma t') e^{-t'^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt' f(0) e^{-t'^2/2} = f(0).$$

Since this is true for any $f(t)$ for which the integral exists we must have that

$$\lim_{\sigma \rightarrow 0} h_\sigma(t) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/(2\sigma^2)} = \delta(t).$$

Returning to the demonstration of (A.2) we can use the Fourier transform relations to calculate

$$g_\sigma(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt h_\sigma(t) e^{i\omega t} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/(2\sigma^2)} e^{i\omega t} = \frac{1}{\sqrt{2\pi}} e^{-\sigma^2\omega^2/2}.$$

Using the inverse relation we have

$$h_\sigma(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega g_\sigma(\omega) e^{-i\omega t} = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-\sigma^2\omega^2/2} e^{-i\omega t}.$$

Then take the limit as $\sigma \rightarrow 0$ giving

$$\lim_{\sigma \rightarrow 0} h_\sigma(t) = \lim_{\sigma \rightarrow 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-\sigma^2\omega^2/2} e^{-i\omega t} = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega t} = \delta(t).$$

The delta function is an even function of its argument so we can equally well write

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{i\omega t}$$

which proves (A.2). It should be noted that the word “prove” has been used in a casual sense as we have switched the orders of integrals and assumed infinite integrals always exist. A mathematician would require more rigor, but for the purpose of physics calculations these relations are correct. It should be emphasized that the use of t, ω as conjugate variables is just one possible choice. When describing optical diffraction using the methods of Fourier optics we will work with the pair x, k_x .

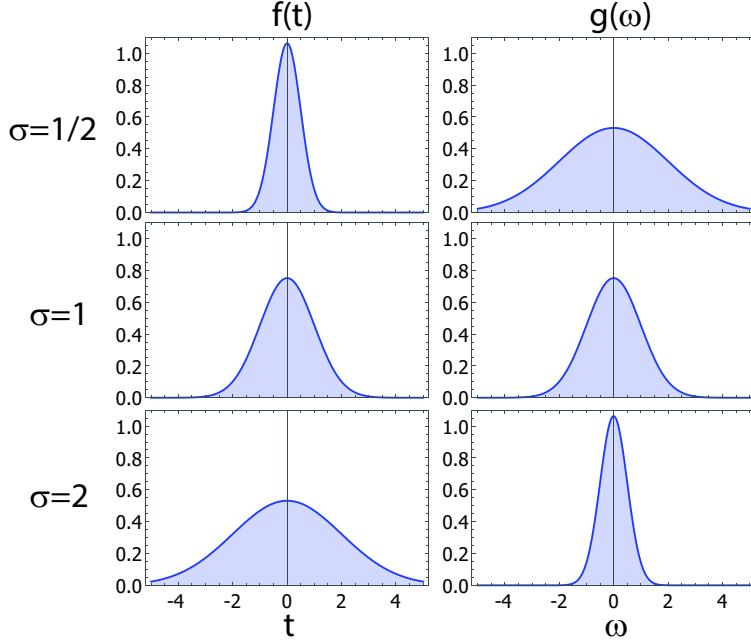


Figure A.2: Gaussian function $f(t)$ and the Fourier transform $g(\omega)$ for $\sigma = 1/2, 1, 2$.

A.2 Self-reciprocal functions

A function $f(t)$ and its transform $g(\omega)$ typically have very different functional forms. It is, however, possible to find functions that are the same as their Fourier transform, apart from a scale factor. Such functions are called Fourier self-reciprocal[55], and they play an important role in optical wave propagation since the fundamental solution of the paraxial wave equation is self-reciprocal.

The most important self-similar function is the Gaussian. Consider

$$f(t) = \frac{1}{\pi^{1/4}\sigma^{1/2}} e^{-t^2/(2\sigma^2)}$$

which has been normalized so that $\int_{-\infty}^{\infty} dt |f(t)|^2 = 1$. The Fourier transform is

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/(2\sigma^2)} e^{i\omega t} = \frac{\sigma^{1/2}}{\pi^{1/4}} e^{-\sigma^2\omega^2/2}$$

which has the same normalization $\int_{-\infty}^{\infty} d\omega |g(\omega)|^2 = 1$. We see that $f(t)$ and $g(\omega)$ have the same functional form but different scale factors as shown in Fig. A.2. The widths of the functions are inversely related since

$$\begin{aligned} \langle t \rangle &= \int_{-\infty}^{\infty} dt t |f|^2 = 0, \\ \langle t^2 \rangle &= \int_{-\infty}^{\infty} dt t^2 |f|^2 = \frac{\sigma^2}{2}, \\ \langle \omega \rangle &= \int_{-\infty}^{\infty} d\omega \omega |g|^2 = 0, \\ \langle \omega^2 \rangle &= \int_{-\infty}^{\infty} d\omega \omega^2 |g|^2 = \frac{1}{2\sigma^2}. \end{aligned}$$

The product of the standard deviations is then

$$\sigma_t \sigma_\omega = \sqrt{\sigma_t^2} \sqrt{\sigma_\omega^2} = \sqrt{\langle t^2 \rangle - \langle t \rangle^2} \sqrt{\langle \omega^2 \rangle - \langle \omega \rangle^2} = \frac{\sigma}{\sqrt{2}} \frac{1}{\sqrt{2}\sigma} = \frac{1}{2}.$$

It can be shown that for any Fourier transform pair $\sigma_t \sigma_\omega \geq 1/2$. The Gaussian function saturates this limit.

Another example of a Fourier self-similar function is $f(x) = \text{sech}(ax)$. It is interesting to note that $f(x)$ is also the fundamental solution of the focusing nonlinear Schrödinger equation.

A.3 Fourier identities

Some useful properties of Fourier transforms, the delta function, and some integrals are given on the following pages.

A.3.1 Transforms on the full line

Transform pair:

$$f(x) = \mathcal{F}^{-1}[g(k)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk g(k) e^{ikx} \quad (\text{A.3a})$$

$$g(k) = \mathcal{F}[f(x)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx f(x) e^{-ikx} \quad (\text{A.3b})$$

Plancherel:

$$\int_{-\infty}^{\infty} dx f_1^*(x) f_2(x) = \int_{-\infty}^{\infty} dk g_1^*(k) g_2(k) \quad (\text{A.4a})$$

$$\int_{-\infty}^{\infty} dx |f(x)|^2 = \int_{-\infty}^{\infty} dk |g(k)|^2 \quad (\text{A.4b})$$

$$(\text{A.4c})$$

Convolution:

$$f_1(x) * f_2(x) \equiv \int_{-\infty}^{\infty} dx' f_1(x') f_2(x - x') \quad (\text{A.5a})$$

$$\mathcal{F}[f_1(x) * f_2(x)] = \sqrt{2\pi} g_1(k) g_2(k) \quad (\text{A.5b})$$

Correlation:

$$f_1(x) \star f_2(x) \equiv [f_1(-x)]^* * f_2(x) = \int_{-\infty}^{\infty} dx' [f_1(-x)]^* f_2(x - x') \quad (\text{A.6a})$$

$$\mathcal{F}[f_1(x) \star f_2(x)] = \sqrt{2\pi} [g_1(k)]^* g_2(k) \quad (\text{A.6b})$$

Multiple dimensions:

$$f(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} d\mathbf{k} g(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \quad (\text{A.7a})$$

$$g(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} d\mathbf{r} f(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \quad (\text{A.7b})$$

Quantum mechanics version:

$$f(\mathbf{r}) = \frac{1}{(2\pi\hbar)^{3/2}} \int_{-\infty}^{\infty} d\mathbf{p} g(\mathbf{p}) e^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}} \quad (\text{A.8a})$$

$$g(\mathbf{p}) = \frac{1}{(2\pi\hbar)^{3/2}} \int_{-\infty}^{\infty} d\mathbf{r} f(\mathbf{r}) e^{-(i/\hbar)\mathbf{p}\cdot\mathbf{r}} \quad (\text{A.8b})$$

Transform pairs:

When $f(x) \leftrightarrow g(k)$ then:

$$f(x - x_0) \leftrightarrow e^{-ikx_0}g(k) \quad (\text{A.9a})$$

$$e^{ik_0x}f(x) \leftrightarrow g(k - k_0) \quad (\text{A.9b})$$

$$\frac{df(x)}{dx} \leftrightarrow ikg(k) \quad (\text{A.9c})$$

$$xf(x) \leftrightarrow i\frac{dg(k)}{dk} \quad (\text{A.9d})$$

$$x\frac{df(x)}{dx} \leftrightarrow -g(k) - k\frac{dg(k)}{dk} \quad (\text{A.9e})$$

$$\delta(x - x_0) \leftrightarrow \frac{1}{\sqrt{2\pi}}e^{-ikx_0} \quad (\text{A.10a})$$

$$e^{ik_0x} \leftrightarrow \sqrt{2\pi}\delta(k - k_0) \quad (\text{A.10b})$$

$$e^{-ax^2+bx} \leftrightarrow \frac{1}{\sqrt{2a}}e^{(b-ik)^2/(4a)} \quad (\text{A.10c})$$

$$xe^{-ax^2+bx} \leftrightarrow \frac{b-ik}{(2a)^{3/2}}e^{(b-ik)^2/(4a)} \quad (\text{A.10d})$$

$$x^2e^{-ax^2+bx} \leftrightarrow \frac{2a + (b-ik)^2}{(2a)^{5/2}}e^{(b-ik)^2/(4a)} \quad (\text{A.10e})$$

$$\frac{1}{1+x^2} \leftrightarrow \sqrt{\frac{\pi}{2}}e^{-|k|} \quad (\text{A.11a})$$

$$\frac{1}{(1+x^2)^2} \leftrightarrow \sqrt{\frac{\pi}{8}}(1+|k|)e^{-|k|} \quad (\text{A.11b})$$

A.3.2 Transforms on the half line

For functions defined on the half line there are sin and cos transform pairs:

$$f(x) = \mathcal{F}_c^{-1}[g(k)] = \sqrt{\frac{2}{\pi}} \int_0^\infty dk g(k) \cos(kx) \quad (\text{A.12a})$$

$$g(k) = \mathcal{F}_c[f(x)] = \sqrt{\frac{2}{\pi}} \int_0^\infty dx f(x) \cos(kx) \quad (\text{A.12b})$$

$$f(x) = \mathcal{F}_s^{-1}[g(k)] = \sqrt{\frac{2}{\pi}} \int_0^\infty dk g(k) \sin(kx) \quad (\text{A.13a})$$

$$g(k) = \mathcal{F}_s[f(x)] = \sqrt{\frac{2}{\pi}} \int_0^\infty dx f(x) \sin(kx) \quad (\text{A.13b})$$

A.3.3 Delta function

Definition:

$$\delta(x) = 0 \quad x \neq 0, \quad (\text{A.14a})$$

$$= \infty \quad x = 0, \quad (\text{A.14b})$$

$$\int_{-\infty}^{\infty} dx f(x)\delta(x) = f(0) \quad (\text{A.14c})$$

Properties:

$$\int_{-\infty}^{\infty} dx \delta(x) = 1 \quad (\text{A.15a})$$

$$\int_{-\infty}^{\infty} dx f(x) \delta(x - x_0) = f(x_0) \quad (\text{A.15b})$$

$$\delta(x - x_0) = \delta(x_0 - x) \quad (\text{A.16a})$$

$$\delta(ax) = \frac{1}{|a|} \delta(x) \quad (\text{A.16b})$$

$$\delta(g(x)) = \sum_j \frac{1}{|g'(x_j)|} \delta(x - x_j), \quad \text{where } x_j \text{ are the roots of } g \quad (\text{A.16c})$$

$$\frac{d^n}{dx^n} \delta(x) = \frac{(-1)^n n!}{x^n} \delta(x) \quad (\text{A.16d})$$

Representations:

$$\delta(x) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad (\text{A.17a})$$

$$= \lim_{\sigma \rightarrow 0} \frac{1}{\pi} \frac{\sigma}{x^2 + \sigma^2} \quad (\text{A.17b})$$

$$= \lim_{\sigma \rightarrow 0} \frac{1}{\pi x} \sin(x/\sigma) \quad (\text{A.17c})$$

$$= \lim_{\sigma \rightarrow 0} \frac{\sigma}{2\pi} \frac{\sin^2 \frac{x}{2\sigma}}{(x/2)^2} \quad (\text{A.17d})$$

$$\delta(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\sqrt{i\pi\varepsilon}} e^{i\frac{x^2}{\varepsilon}} \quad (\text{A.18})$$

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ikx} \quad (\text{A.19})$$

If $u_n(x)$ form a complete set then

$$\delta(x - x') = \sum_n u_n(x) u_n^*(x') \quad (\text{A.20})$$

Multiple dimensions:

Cartesian coordinates:

$$\delta(\mathbf{r}) = \delta(x)\delta(y)\delta(z) \quad (\text{A.21a})$$

Spherical coordinates:

$$\delta(\mathbf{r} - \mathbf{r}') = \frac{1}{r^2} \delta(r - r') \delta(\cos \theta - \cos \theta') \delta(\phi - \phi') \quad (\text{A.22a})$$

$$= \frac{1}{r^2} \delta(r - r') \frac{\delta(\theta - \theta')}{\sin \theta} \delta(\phi - \phi') \quad (\text{A.22b})$$

$$\nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) = -4\pi \delta(\mathbf{r} - \mathbf{r}') \quad (\text{A.22c})$$

A.3.4 Definite integrals:

$$\int_{-\infty}^{\infty} dx e^{-x^2} = \sqrt{\pi} \quad (\text{A.23})$$

$$\int_{-\infty}^{\infty} dx e^{-ax^2} = \frac{\sqrt{\pi}}{a^{1/2}}, \quad Re[a] > 0 \quad (\text{A.24a})$$

$$\int_{-\infty}^{\infty} dx x^2 e^{-ax^2} = \frac{\sqrt{\pi}}{2a^{3/2}}, \quad Re[a] > 0 \quad (\text{A.24b})$$

$$\int_{-\infty}^{\infty} dx x^4 e^{-ax^2} = \frac{3\sqrt{\pi}}{4a^{5/2}}, \quad Re[a] > 0 \quad (\text{A.24c})$$

$$\int_{-\infty}^{\infty} dx x^6 e^{-ax^2} = \frac{15\sqrt{\pi}}{8a^{7/2}}, \quad Re[a] > 0 \quad (\text{A.24d})$$

$$\int_{-\infty}^{\infty} dx x^p e^{-ax^2} = \frac{1 + (-1)^p}{2a^{\frac{p+1}{2}}} \Gamma\left(\frac{p+1}{2}\right), \quad Re[a] > 0, Re[p] > -1, \quad (\text{A.24e})$$

The following are formally divergent but useful

$$\int_{-\infty}^{\infty} dx e^{iax^2} = \sqrt{\frac{i\pi}{a}}, \quad a \text{ real} \quad (\text{A.25a})$$

$$\int_{-\infty}^{\infty} dx e^{i(ax^2+bx)} = \sqrt{\frac{i\pi}{a}} e^{-ib^2/(4a)}, \quad a \text{ real} \quad (\text{A.25b})$$

$$\int_{-\infty}^{\infty} dx \frac{1}{1+bx^2} = \frac{\pi}{\sqrt{b}} \quad (\text{A.26a})$$

$$\int_{-\infty}^{\infty} dx \frac{e^{-ax^2}}{1+bx^2} = \frac{\pi}{\sqrt{b}} e^{a/b} \text{Erfc}(\sqrt{a/b}), \text{ where } \text{Erfc}(x) = 1 - \text{Erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2} \quad (\text{A.26b})$$

$$\int_{-\infty}^{\infty} dx \frac{\sin(ax)}{ax} = \frac{\pi}{a} \quad (\text{A.27a})$$

$$\int_{-\infty}^{\infty} dx \left[\frac{\sin(ax)}{ax} \right]^2 = \frac{\pi}{a} \quad (\text{A.27b})$$

$$\int_{-\infty}^{\infty} dx \left[\frac{\sin(ax)}{ax} \right]^3 = \frac{3\pi}{4a} \quad (\text{A.27c})$$

$$\int_0^\infty dx e^{-ax} = \frac{1}{a}, \quad Re[a] > 0 \quad (\text{A.28a})$$

$$\int_0^\infty dx xe^{-ax} = \frac{1}{a^2}, \quad Re[a] > 0 \quad (\text{A.28b})$$

$$\int_0^\infty dx x^2 e^{-ax} = \frac{2}{a^3}, \quad Re[a] > 0 \quad (\text{A.28c})$$

$$\int_0^\infty dx x^3 e^{-ax} = \frac{6}{a^4}, \quad Re[a] > 0 \quad (\text{A.28d})$$

$$\int_0^\infty dx x^p e^{-ax} = \frac{\Gamma(p+1)}{a^{p+1}}, \quad Re[a] > 0, Re[p] > -1, \quad (\text{A.28e})$$

Appendix B

Window deformation

Experiments that involve imaging objects inside vacuum chambers rely on optically transparent windows. Typical glasses deform under the differential pressure inside and outside the vacuum chamber. Deflection of the surface perpendicular to the plane of the window causes optical aberrations and it is therefore useful to quantify it using elasticity theory.

Consider a circular plate of radius a and thickness h . The plate is uniformly loaded perpendicular to its surface with intensity q . The deflection perpendicular to the surface at radius r is $w(r)$. The magnitude of the deflection depends on the boundary conditions at the edge of the plate. For a plate clamped at the edge, so the deflection vanishes at $r = a$, the solution is (Timoshenko [119], Eq. (62))

$$w(r) = \frac{q}{64D} (a^2 - r^2)^2.$$

Here $D = \frac{Eh^3}{12(1-\nu^2)}$ is the flexural rigidity, E is the modulus of elasticity, and ν is Poisson's ratio so

$$w(r) = \frac{3(1-\nu^2)q}{16E} \frac{(a^2 - r^2)^2}{h^3}. \quad (\text{B.1})$$

The units of q are force per unit area or pressure and the units of E are pressure so w has units of length as it should. The deflection is maximal at the center of the plate where it takes the value

$$w_0 = \frac{3(1-\nu^2)q}{16E} \frac{a^4}{h^3}. \quad (\text{B.2})$$

We see that the deflection is proportional to the fourth power of the radius and the inverse cube of the thickness.

As an example values for pyrex are $\nu = 0.20$, and $E = 62.75 \times 10^9$ (Pa). A vacuum window is subject to a load intensity which is simply the atmospheric pressure so $q = P_{\text{at}} = 1.01 \times 10^5$ (Pa) and

$$w_0 = 2.9 \times 10^{-7} \frac{a^4}{h^3} (\text{m}).$$

The dependence of w_0 on radius for several standard thicknesses is shown in Fig. B.1.

Deformation of the window leads to image aberrations that are severe as regards reflected light and less so for transmitted light. Window deformation can be expressed as an effective change in the optical path length δL which can be expanded in powers of the radius as[120]

$$\delta L = L_2 \left(\frac{r}{a}\right)^2 + L_4 \left(\frac{r}{a}\right)^4 + L_6 \left(\frac{r}{a}\right)^6. \quad (\text{B.3})$$

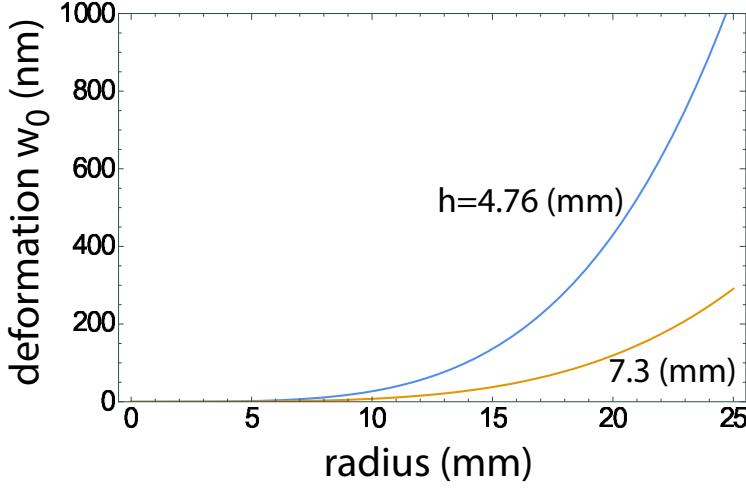


Figure B.1: Maximal deformation of a pyrex vacuum window.

On axis the window is deflected, but there is no change in path length since we are neglecting pressure induced compression of the window material and photoelastic changes to the refractive index. The coefficients for a clamped circular window are

$$\begin{aligned} L_2 &= \frac{n - n_0}{2} \frac{9P_{\text{at}}^2 (1 - \nu^2)^2}{16E^2} \frac{a^6}{h^5}, \\ L_4 &= -2L_2, \\ L_6 &= L_2. \end{aligned}$$

The deformation coefficients are related to optical parameters as follows. The focal length of the window is

$$f = -\frac{a^2}{2L_2} = -\frac{16E^2}{9(n - n_0)(1 - \nu^2)^2 P_{\text{at}}^2} \frac{h^5}{a^4}.$$

For reasonable parameters the focal length is so long as to be of negligible impact. Even taking extreme parameters such as $a = 10$ cm, $h = 1$ mm and $n = 1.47$ for pyrex, $n_0 = 1$ for vacuum, we find $f = -15.8$ m.

Of greater importance are window induced aberrations. One way of characterizing this is in terms of the reduction of peak intensity of a Gaussian beam focused through the window. Considering the number of phase reversals that occur up to a radius a due to the r^4 term in (B.3) the minimum window thickness needed such that the on axis intensity does not decrease by more than a factor of two is[120]

$$h_{\min} \simeq 1.7 \left[\frac{2(n - 1)P_{\text{at}}^2 a}{E^2 \lambda} \right]^{1/5} a. \quad (\text{B.4})$$

(This result uses $\nu = 0.3$ and $n_0 = 1$.) We see that the required thickness increases only slightly faster than a linear proportionality with the radius. Figure B.2 shows the dependence of h_{\min} on window radius.

This result only accounts for the r^4 term in (B.3) and is therefore approximate. It is argued in [120] that the r^6 term is unlikely to significantly change the result. In any case a conservative design approach ought to include a safety factor in the thickness relative to h_{\min} .

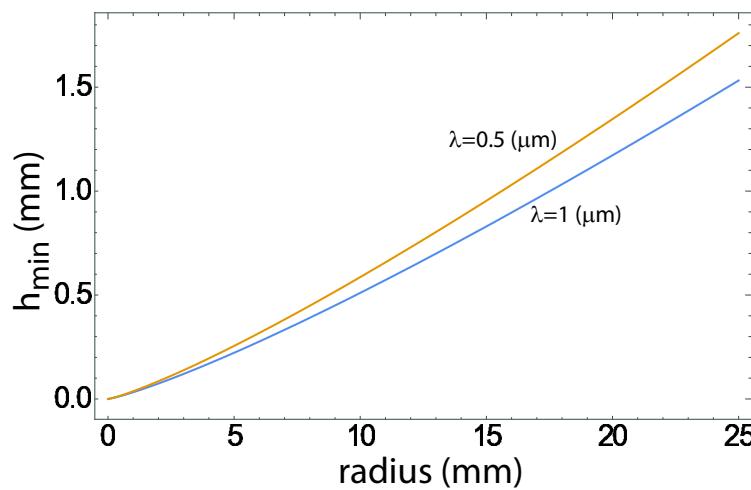


Figure B.2: Minimum window thickness such that the peak intensity of a focused Gaussian beam is at least half of the unaberrated value.

Appendix C

Books on optics

Many books have been written on the subject of optics. Here is a sampling of what is available sorted by topical emphasis.

Classic and older texts

M. Born and E. Wolf, *Principles of optics*, 7th expanded ed., (Cambridge University Press, Cambridge, 1999).

R. W. Ditchburn, *Light*, (Dover, 2011).

P. Drude, *The theory of optics*, (Dover, New York, 1959).

G. R. Fowles, *Introduction to modern optics*, 2nd ed., (Dover, New York, 1975).

F. A. Jenkins and H. E White, *Fundamentals of optics*, 4th ed., (McGraw-Hill, New York, 2001).

V. Ronchi, *Optics the science of vision*, (Dover, New York, 1991).

A. Sommerfeld, *Optics*, (Academic Press, New York, 1964).

J. Strong, *Concepts of classical optics*, (W. H. Freeman and company, San Francisco, 1958).

J. Valasek, *Elements of optics*, (McGraw-Hill, New York, 1928).

Modern texts with broad coverage

S. A. Akhmanov and S. Yu. Nikitin, *Physical optics*, (Clarendon Press, Oxford, 1997).

G. Brooker, *Modern classical optics*, (Oxford University Press, Oxford, 2003).

G. Chartier, *Introduction to optics*, (Springer, New York, 2005).

B. D. Guenther, *Modern Optics*, 2nd ed., (Oxford University Press, Oxford, 2015).

- E. Hecht, *Optics*, 5th ed., (Pearson, 2017).
- I. R. Kenyon, *The Light Fantastic: A modern introduction to classical and quantum optics*, 2nd ed., (Oxford University Press, Oxford, 2008).
- M. V. Klein and T. E. Furtak, *Optics*, 2nd ed., (Wiley, New York, 1986).
- W. Lauterborn, T. Kurz, and M. Wiesenfeldt, *Coherent optics fundamentals and applications*, 2nd ed., (Springer, Berlin, 2003).
- A. Lipson, S. G. Lipson, and H. Lipson, *Optical physics*, 4th ed., (Cambridge University Press, Cambridge, 2011).
- M. Mansuripur, *Classical optics and its applications*, 2nd ed., (Cambridge University Press, Cambridge, 2009).
- D. Meschede, *Optics light and lasers*, 2nd ed., (Wiley-VCH, Weinheim, 2007).
- J. R. Meyer-Arendt, *Introduction to classical and modern optics*, 4th ed., (Addison-Wesley, 1994).
- E. E. Miller and F. L. Roesler, *Applied optics*, (UW Madison, 1999).
- J. Peatross and M. Ware, *Physics of light and optics*, (available online at <http://optics.byu.edu/textbook.aspx>, 2015 edition).
- F. L. Pedrotti, L. M. Pedrotti, and L. S. Pedrotti, *Introduction to optics*, 3rd ed., (Benjamin-Cummings, 2006).
- B. E. A. Saleh and M. C. Teich, *Fundamentals of photonics*, 2nd ed., (Wiley-Interscience, Hoboken, 2007).
- K. K. Sharma, *Optics: Principles and Applications*, (Academic Press, Burlington, 2006).
- D. A. Steck, *Classical and Modern Optics*, available online at <http://steck.us/teaching> (revision 1.4.32, 25 June 2012).
- F. Träger, Ed., *Springer handbook of lasers and optics*, 2nd ed., (Springer, Berlin, 2012).
- A. Yariv and P. Yeh, *Photonics Optical electronics for modern communications*, 6th ed., (Springer, Berlin, 2000).
- M. Young, *Optics and Lasers: Including fibers and optical waveguides*, 5th ed., (Springer, Berlin, 2000).

Geometrical optics and optical design

- A. E. Conrady, *Applied optics and optical design, part 1*, (Dover, New York, 1985).

A. E. Conrady and H. Kingslake, *Applied optics and optical design, part 2*, (Dover, New York, 1988).

E. L. Dereniak and T. D. Dereniak, *Geometrical and trigonometric optics*, (Cambridge University Press, Cambridge, 2008).

A. Gerrard and J. M. Burch, *Introduction to matrix methods in optics*, (Dover, Mineola, 1994).

J. F. Nye, *Natural focusing and fine structure of light*, (Institute of Physics Publishing, Bristol, 1999).

D. C. O'Shea, *Elements of modern optical design*, (John Wiley, New York, 1985).

R. R. Shannon, *The art and science of optical design*, (Cambridge University Press, Cambridge, 1997).

Fiber, guided wave, integrated, and nano optics

J. A. Arnaud, *Beam and fiber optics*, (Academic Press, 1976).

R. G. Hunsperger, *Integrated optics: theory and technology*, 3rd ed., (Springer, Berlin, 1991).

J. Jahns and S. Helfert, *Introduction to Micro- and Nanooptics*, (Wiley-VCH, Weinheim, 2012).

J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic crystals Molding the flow of light*, (Princeton University Press, Princeton, 2008).

G. Lifante, *Integrated photonics: fundamentals*, (Wiley, Chichester, 2003).

D. Marcuse, *Theory of dielectric optical waveguides*, 2nd ed., (Academic Press, Boston, 1991).

E. G. Neumann, *Single-mode fibers fundamentals*, (Springer, Berlin, 1988).

L. Novotny and B. Hecht, *Principles of Nano-Optics*, 2nd ed., (Cambridge University Press, Cambridge, 2012).

M. Skorobogatiy and J. Yang, *Fundamentals of Photonic Crystal Guiding*, (Cambridge University Press, Cambridge, 2009).

A. W. Snyder and J. D. Love, *Optical Waveguide Theory*, (Kluwer Academic Publishers, Dordrecht, 2000).

Fourier optics and imaging

B. B. Baker and E. T. Copson, *The mathematical theory of Huygens' principle*, 3rd ed., (Chelsea Publishing, New York, 1987).

J. M. Cowley, *Diffraction physics*, 3rd revised ed., (Elsevier, Amsterdam, 1995).

- J. D. Gaskill, *Linear systems, Fourier transforms, and optics*, (John Wiley, New York, 1978).
- J. W. Goodman, *Introduction to Fourier optics*, 3rd ed., (Roberts & Company, Englewood, 2005).
- M. Gu, *Advanced optical imaging theory*, (Springer-Verlag, Berlin, 2000).
- A. Papoulis, *Systems and transforms with applications in optics*, (McGraw-Hill, 1968).
- J. J. Stamnes, *Waves in Focal Regions: Propagation, Diffraction, and Focusing of Light, Sound, and Water Waves*, (Taylor & Francis, New York, 1986).

Laser resonator optics and interferometers

- Y. A. Anan'ev, *Laser resonators and the beam divergence problem*, (Adam Hilger, Bristol, 1992).
- G. Hernandez, *Fabry-Perot interferometers*, (Cambridge University Press, Cambridge, 1986).
- N. Hodgson and H. Weber, *Optical Resonators: Fundamentals, Advanced Concepts, Applications*, 2nd ed., (Springer, Berlin, 2005).
- W. Koechner, *Solid-state laser engineering*, 6th ed., (Springer, 2006).

A. E. Siegman, *Lasers*, (University Science Books, Mill Valley, 1986).

J. M. Vaughan, *The Fabry-Perot Interferometer: History, Theory, Practice and Applications*, (Taylor & Francis, 1989).

Light scattering

- C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles*, (Wiley, New York, 1983).
- G. Gouesbet and G. Gréhan , *Generalized Lorenz-Mie Theories*, (Springer, Berlin, 2011).
- M. Kerker, *The Scattering of Light, and Other Electromagnetic Radiation*, (Academic Press, New York, 1969).
- M. I. Mishchenko, L. D. Travis, and A. A. Lacis, *Scattering, absorption, and emission of light by small particles*, (Cambridge University Press, Cambridge, 2002).
- M. I. Mishchenko, L. D. Travis, and A. A. Lacis, *Multiple scattering of light by particles, Radiative transfer and coherent backscattering*, (Cambridge University Press, Cambridge, 2006).

- A. V. Osipov and S. A. Tretyakov, *Modern electromagnetic scattering theory with applications*, (Wiley, Chichester, 2017).
- H. C. van de Hulst, *Light scattering by small particles*, (Dover, New York, 1981).

Polarization

R. M. A. Azzam and N. M. Bashara, *Ellipsometry and polarized light*, (North Holland, Amsterdam, 1989).

C. Brosseau, *Fundamentals of polarized light, A statistical optics approach*, (Wiley, New York, 1998).

D. H. Goldstein, *Polarized light*, 3rd ed., (CRC Press, Boca Raton, 2011).

G. P. Können, *Polarized light in nature*, (Cambridge University Press, Cambridge, 1985).

Statistical optics and speckle phenomena

J. W. Goodman, *Statistical optics*, (John Wiley, New York, 1985).

J. W. Goodman, *Speckle phenomena in optics, theory and applications*, (Roberts & Company, Greenwood Village, 2007).

E. L. O'Neill, *Introduction to statistical optics*, (Dover, New York, 1991).

B. Ya' Zel'dovich, A. V. Mamaev, and V. V. Shkunov, *Speckle-wave interactions in application to holography and nonlinear optics*, (CRC Press, Boca Raton, 1995).

Specialized topics

L. Allen, S. M. Barnett, M. J. Padgett, *Optical Angular Momentum*, (IOP Press, Bristol, 2003).

J. Chaves, *Introduction to nonimaging optics*, (CRC Press, Boca Raton, 2008).

A. J. Devaney, *Mathematical foundations of imaging, tomography and wavefield inversion*, (Cambridge University Press, Cambridge, 2012).

J. Glückstad and D. Palima, *Generalized phase contrast: Applications in optics and photonics*, (Springer, Berlin, 2010).

P. C. D. Hobbs, *Building electro-optical systems: making it all work*, 2nd ed., (Wiley, Hoboken, 2009).

U. Leonhardt and T. Philbin, *Geometry and light the science of invisibility*, (Dover, New York, 2010).

J. F. Nye, *Natural Focusing and Fine Structure of Light: Caustics and Wave Dislocations*, (IOP Press, Bristol, 1999).

W. T. Welford, *Useful optics*, (University of Chicago Press, Chicago, 1991).

Bibliography

- [1] V. G. Veselago, *The electrodynamics of substances with simultaneously negative values of ϵ and μ* , Sov. Phys. Uspekhi **10**, 509 (1968).
- [2] J. B. Pendry, *Negative refraction makes a perfect lens*, Phys. Rev. Lett. **85**, 3966–3969 (2000).
- [3] R. C. Jones, *A new calculus for the treatment of optical systems I. description and discussion of the calculus*, J. Opt. Soc. Am. **31**, 488 (1941).
- [4] G. G. Stokes, *On the composition and resolution of streams of polarized light from different sources*, Trans. Cambridge Philos. Soc. **9**, 399 (1852).
- [5] M. Born and E. Wolf, *Principles of Optics, sixth edition* (Pergamon Press, Oxford) (1980).
- [6] C. K. Hong, Z. Y. Ou, and L. Mandel, *Measurement of subpicosecond time intervals between two photons by interference*, Phys. Rev. Lett. **59**, 2044–2046 (1987).
- [7] F. Goos and H. Hänchen, *Ein neuer und fundamentaler versuch zur totalreflexion*, Ann. Phys. **436**, 333 (1947).
- [8] A. Fresnel, *Mémoire sur les modifications que la réflexion imprime à la lumière polarisée*, Mémoires de l'Académie des sciences de l'Institut de France **11**, 373 (1832).
- [9] A. A. Michelson and F. G. Pease, *Measurement of the diameter of a Orions with the interferometer*, Astrophys. J. **53**, 249 (1921).
- [10] C. Fabry and A. Pérot, *Theorie et applications d'une nouvelle methode de spectroscopie interferentielle*, Ann. Chim. Phys. **16**, 115 (1899).
- [11] C. J. Hood, H. J. Kimble, and J. Ye, *Characterization of high-finesse mirrors: Loss, phase shifts, and mode structure in an optical cavity*, Phys. Rev. A **64**, 033804 (2001).
- [12] R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, A. J. Munley, and H. Ward, *Laser phase and frequency stabilization using an optical resonator*, Appl. Phys. B **31**, 97 (1983).
- [13] R. M. Waxler and G. W. Cleek, *Effect of temperature and pressure on refractive-index of some oxide glasses*, J. Res. Natl. Bur. Stand. **77A**, 755 (1973).
- [14] A. A. Michelson, *The relative motion of the earth and the luminiferous ether*, Am. J. Science **22**, 120 (1881).
- [15] L. Zehnder, *Ein neuer interferenzrefraktor*, Z. für Instrumentenkunde **11**, 275 (1891).
- [16] L. Mach, *Ueber einen interferenzrefraktor*, Z. für Instrumentenkunde **12**, 89 (1892).

- [17] H. Fizeau, *Sur les hypothèses relatives à l'éther lumineux*, Compt. Rend. **33**, 349 (1851).
- [18] G. Sagnac, *L'éther lumineux démontré par l'effet du vent relatif d'éther dans un interféromètre en rotation uniforme*, Compt. Rend. **157**, 708 (1913).
- [19] H. A. Macleod, *Thin-film optical filters*, 4th ed. (CRC Press, Boca Raton) (2010).
- [20] H. W. Kogelnik and T. Li, *Laser beams and resonators*, Appl. Opt. **5**, 1550 (1966).
- [21] G. A. Massey and A. E. Siegman, *Reflection and refraction of Gaussian light beams at tilted ellipsoidal surfaces*, Appl. Opt. **8**, 975 (1969).
- [22] D. M. Kane, *Ti-Sapphire laser cavity mode and pump-laser mode calculations*, Appl. Opt. **33**, 3849 (1994).
- [23] T. Freegarde and C. Zimmermann, *On the design of enhancement cavities for second harmonic generation*, Opt. Commun. **199**, 435 (2001).
- [24] L. Seidel, *Zur dioptrik. ueber die entwicklung der glieder 3ter ordnung, welche den weg eines ausserhalb der ebene der axe gelegenen lichtstrahles durch ein system brechender medien bestimmen*, Astr. Nachr. **43**, 289 (1856).
- [25] L. Seidel, *Zur dioptrik. ueber die entwicklung der glieder 3ter ordnung, welche den weg eines ausserhalb der ebene der axe gelegenen lichtstrahles durch ein system brechender medien bestimmen*, Astr. Nachr. **43**, 305 (1856).
- [26] L. Seidel, *Zur dioptrik. ueber die entwicklung der glieder 3ter ordnung, welche den weg eines ausserhalb der ebene der axe gelegenen lichtstrahles durch ein system brechender medien bestimmen*, Astr. Nachr. **43**, 321 (1856).
- [27] W. Sellmeier, *Zur Erklärung der abnormen Farbenfolge im spectrum einiger Substanzen*, Ann. d. Phys. **219**, 272 (1871).
- [28] B. Schmidt, Central-Zeitung f. Optik u. Mechanik **52** (1931).
- [29] C. Huygens, *Traite de la lumiere*, An English translation is available at www.gutenberg.org. (1690).
- [30] A. Fresnel, *Deuxième mémoire sur la diffraction de la lumière*, Ann. de chim. et de phys. **2**, 239 (1816).
- [31] A. Tafove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed. (Artech House Publishers, Boston) (2005).
- [32] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco) (1968).
- [33] G. Kirchhoff, *Zur theorie der lichtstrahlen*, Ann. d. Phys. **254**, 663 (1883).
- [34] L. Rayleigh, *On the passage of waves through apertures in plane screens, and allied problems*, Phil. Mag. **43**, 259 (1897).
- [35] A. Sommerfeld, *Zur mathematischen theorie der beugungserscheinungen*, Nachr. Kgl. Wiss. Gottingen **4**, 338 (1894).
- [36] A. Sommerfeld, *Mathematische theorie der diffraction*, Math. Ann. **47**, 317 (1896).

- [37] A. Sommerfeld, *Optics*, Academic Press, London (1964).
- [38] E. Wolf and E. W. Marchand, *Comparison of the Kirchhoff and the Rayleigh-Sommerfeld theories of diffraction at an aperture*, J. Opt. Soc. Am. **54**, 587 (1964).
- [39] E. Wolf, *Electromagnetic diffraction in optical systems i. an integral representation of the image field*, Proc. Roy. Soc. A **253**, 349 (1959).
- [40] B. Richards and E. Wolf, *Electromagnetic diffraction in optical systems ii. structure of the image field in an aplanatic system*, Proc. Roy. Soc. A **253**, 358 (1959).
- [41] P. Debye, *Das verhalten von lichtwellen in der nähe eines brennpunktes oder einer brennlinie*, Ann. Physik **30**, 755 (1909).
- [42] P. Baues, *Huygens' principle in inhomogeneous isotropic media and a general integral equation applicable to optical resonators*, Opto-Electr. **1**, 37 (1969).
- [43] S. A. Collins, Jr., *Lens-system diffraction integral written in terms of matrix optics*, J. Opt. Soc. Am. **60**, 1168 (1970).
- [44] H. T. Yura and S. G. Hanson, *Optical beam wave propagation through complex optical systems*, J. Opt. Soc. Am. A **4**, 1931 (1987).
- [45] G. B. Airy, *On the diffraction of an object-glass with circular aperture*, Trans. Cambr. Phil. Soc. **5**, 283 (1835).
- [46] G. T. D. Francia, *Resolving power and information*, J. Opt. Soc. Am. **45**, 497 (1955).
- [47] M. Tsang, R. Nair, and X.-M. Lu, *Quantum theory of superresolution for two incoherent optical point sources*, Phys. Rev. X **6**, 031033 (2016).
- [48] W. H. F. Talbot, *Facts relating to optical science, No. iv*, Philos. Mag. **9**, 401 (1836).
- [49] K. Patorski, *The self-imaging phenomenon and its applications*, Prog. Opt. **27**, 1 (1989).
- [50] J. Wen, Y. Zhang, and M. Xiao, *The Talbot effect: recent advances in classical optics, nonlinear optics, and quantum optics*, Adv. Opt. Phot. **5**, 83 (2013).
- [51] E. v. Lommel, Abh. der K. Bayer Akad. der Wissen. **15**, 233 (1885).
- [52] V. N. Pochernyaev, *Integrals of products of Bessel functions for applied electrodynamic problems*, Ukrain. Math. J. **47**, 658 (1995).
- [53] M. J. Piotrowicz, M. Lichtman, K. Maller, G. Li, S. Zhang, L. Eisenhower, and M. Saffman, *Two-dimensional lattice of blue-detuned atom traps using a projected Gaussian beam array*, Phys. Rev. A **88**, 013420 (2013).
- [54] M. Saffman and M. Lichtman, *System and method for optical confinement of atomic particles*, US patent 9,355,750, granted May 31, 2016 .
- [55] E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, 2nd ed. (Oxford University Press, Oxford) (1948).
- [56] L. G. Gouy, *Sur une propriété nouvelle des ondes lumineuses*, CR Hebdomadiaries Séances Acad. Sci. **110**, 1251 (1890).

- [57] L. G. Gouy, *Sur la propagation anomale des ondes*, Annales des Chimie et de Physique **24**, 145 (1891).
- [58] T. D. Visser and E. Wolf, *The origin of the Gouy phase anomaly and its generalization to astigmatic wavefields*, Opt. Commun. **283**, 3371 (2010).
- [59] M. W. Beijersbergen, L. Allen, H. E. L. O. van der Veen, and J. P. Woerdman, *Astigmatic laser mode converters and transfer of orbital angular momentum*, Opt. Commun. **96**, 123 (1993).
- [60] M. A. Bandres and J. C. Gutiérrez-Vega, *Ince-Gaussian beams*, Opt. Lett. **29**, 144 (2004).
- [61] J. Durnin, J. J. Miceli, and J. H. Eberly, *Diffraction-free beams*, Phys. Rev. Lett. **58**, 1499–1501 (1987).
- [62] M. A. Bandres, J. C. Gutiérrez-Vega, and S. Chávez-Cerda, *Parabolic nondiffracting optical wave fields*, Opt. Lett. **29**, 44 (2004).
- [63] J. C. Gutiérrez-Vega, M. D. Iturbe-Castillo, and S. Chávez-Cerda, *Alternative formulation for invariant optical fields: Mathieu beams*, Opt. Lett. **25**, 1493 (2000).
- [64] H. Tanner, *Zoom lenses for Gaussian beams*, Opt. Technol. **1**, 24 (1968).
- [65] D. Slepian and H. Pollak, *Prolate spheroidal wavefunctions, Fourier analysis and uncertainty - I*, Bell Syst. Tech. J. **40**, 43 (1961).
- [66] G. D. Boyd and J. P. Gordon, *Confocal multimode resonator for millimeter through optical wavelength masers*, Bell Syst. Tech. J. **40**, 489 (1961).
- [67] G. D. Boyd and H. Kogelnik, *Generalized confocal resonator theory*, Bell Syst. Tech. J. **41**, 1347 (1962).
- [68] H. W. Kogelnik, E. Ippen, A. Dienes, and C. V. Shank, *Astigmatically compensated cavities for cw dye lasers*, IEEE J. Qu. Electr. **8**, 373 (1972).
- [69] A. Einstein, *On the quantum theory of radiation*, Phys. Zeits. **18**, 121 (1917).
- [70] J. P. Gordon, H. J. Zeiger, and C. H. Townes, *Molecular microwave oscillator and new hyperfine structure in the microwave spectrum of Nh₃*, Phys. Rev. **95**, 282–284 (1954).
- [71] J. P. Gordon, H. J. Zeiger, and C. H. Townes, *The maser – new type of microwave amplifier, frequency standard, and spectrometer*, Phys. Rev. **99**, 1264–1274 (1955).
- [72] T. H. Maiman, *Stimulated optical radiation in Ruby*, Nature **187**, 493 (1960).
- [73] H. J. Metcalf and P. van der Straten, *Laser cooling and trapping* (Springer-Verlag, New York) (1999).
- [74] J. Kästel, M. Fleischhauer, S. F. Yelin, and R. L. Walsworth, *Tunable negative refraction without absorption via electromagnetically induced chirality*, Phys. Rev. Lett. **99**, 073602 (2007).
- [75] A. Liénard, *Electric and magnetic field produced by a moving charged point*, L'Eclairage Electrique **16**, 5, 53, 106 (1898).

- [76] E. Wiechert, *Electrodynamical laws*, Archives Neerlandaises des Sciences Exactes et Naturelles **5**, 549 (1900).
- [77] O. Heaviside, *The radiation from an electron moving in an elliptic, or any other orbit*, Nature **69**, 342 (1904).
- [78] R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman lectures on physics **II** (1963).
- [79] J. Larmor, *On the theory of the magnetic influence on spectra; and on the radiation from moving ions*, Phil. Mag. **44**, 503 (1897).
- [80] R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman lectures on physics **I** (1963).
- [81] L. Lorenz, *Lysbevægelsen i og uden for en af plane lysbølger belyst kugle*, Kong. Dansk. Videnskab. Selskab. Skrifter **6**, 1 (1890).
- [82] G. Mie, *Beiträge zur optik trüber medien, speziell kolloidaler metallösungen*, Ann. d. Phys. **25**, 377 (1908).
- [83] L. Rayleigh, *On the light from the sky, its polarization and colour*, Phil. Mag. **41**, 107 (1871).
- [84] R. W. Boyd, *Nonlinear optics* (Academic, San Diego) (1992).
- [85] V. E. Zakharov and A. B. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, ZhETF **61**, 118 (1971), [J. Exp. Theor. Phys. **34**, 62 (1972)].
- [86] A. Hasegawa and F. Tappert, *Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. i. anomalous dispersion*, Appl. Phys. Lett. **23**, 142 (1973).
- [87] L. F. Mollenauer, R. H. Stolen, and J. P. Gordon, *Experimental observation of picosecond pulse narrowing and solitons in optical fibers*, Phys. Rev. Lett. **45**, 1095–1098 (1980).
- [88] N. N. Akhmediev and A. Ankiewicz, *Solitons, nonlinear pulses and beams* (Chapman and Hall, London) (1997).
- [89] A. V. Mamaev, M. Saffman, D. Z. Anderson, and A. A. Zozulya, *Propagation of light beams in anisotropic nonlinear media: From symmetry breaking to spatial turbulence*, Phys. Rev. A **54**, 870 (1996).
- [90] A. A. Zozulya and D. Z. Anderson, *Propagation of an optical beam in a photorefractive medium in the presence of a photovoltaic nonlinearity or an externally applied electric field*, Phys. Rev. A **51**, 1520 (1995).
- [91] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, *Interactions between light waves in a nonlinear dielectric*, Phys. Rev. **127**, 1918–1939 (1962).
- [92] G. D. Boyd and D. A. Kleinman, *Parametric interaction of focused Gaussian light beams*, J. Appl. Phys. **39**, 3597–3639 (1968).
- [93] V. G. Dmitriev, G. G. Gurzadyan, and D. N. Nikogosyan, *Handbook of Nonlinear Optical Crystals, third revised edition* (Springer, Berlin) (1999).
- [94] M. M. Fejer, G. A. Magel, D. H. Jundt, and R. L. Byer, *Quasi-phase-matched second harmonic generation: tuning and tolerances*, IEEE J. Qu. Electr. **28**, 2631–2654 (1992).

- [95] Y. J. Ding and J. B. Khurgin, *Backward optical parametric oscillators and amplifiers*, IEEE J. Qu. Electr. **32**, 1574–1582 (1996).
- [96] G. D’Alessandro, P. S. J. Russell, and A. A. Wheeler, *Nonlinear dynamics of a backward quasi-phase-matched second-harmonic generator*, Phys. Rev. A **55**, 3211–3218 (1997).
- [97] P. M. Lushnikov, P. Lodahl, and M. Saffman, *Transverse modulational instability of counter-propagating quasi-phase-matched beams in a quadratically nonlinear medium*, Opt. Lett. **23**, 1650–1652 (1998).
- [98] A. Ashkin, G. Boyd, and J. Dziedzic, *Resonant optical second harmonic generation and mixing*, IEEE J. Qu. Electr. **2**, 109 (1966).
- [99] E. S. Polzik and H. J. Kimble, *Frequency doubling with $KNbO_3$ in an external cavity*, Opt. Lett. **16**, 1400 (1991).
- [100] S. Ast, R. M. Nia, A. Schönbeck, N. Lastzka, J. Steinlechner, T. Eberle, M. Mehmet, S. Steinlechner, and R. Schnabel, *High-efficiency frequency doubling of continuous-wave laser light*, Opt. Lett. **36**, 3467 (2011).
- [101] Y. R. Shen, *The principles of nonlinear optics* (Wiley, Hoboken) (2003).
- [102] K. C. Kao and G. A. Hockham, *Dielectric-fibre surface waveguides for optical frequencies*, Proc. IEEE **113**, 1151 (1966).
- [103] F. P. Kapron, D. B. Keck, and R. D. Maurer, *Radiation losses in glass optical waveguides*, Appl. Phys. Lett. **17**, 423 (1970).
- [104] J. B. MacChesney, P. B. O’Connor, and H. M. Presby, *A new technique for the preparation of low-loss and graded-index optical fibers*, Proc. IEEE **62**, 1280–1281 (1974).
- [105] R. J. Mears, L. Reekie, S. B. Poole, and D. N. Payne, *Low-threshold tunable-CW and Q-switched fiber laser operating at 1.55 μm*, Electr. Lett. **22**, 159 (1986).
- [106] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, *Detecting single infrared photons with 93% system efficiency*, Nat. Photon. **7**, 210 (2013).
- [107] L. Callegaro, *Unified derivation of Johnson and shot noise expressions*, Am. J. Phys. **74**, 438 (2006).
- [108] E. L. Dereniak and G. D. Boreman, *Infrared detectors and systems* (Wiley, New York) (1996).
- [109] R. E. Burgess, *Homophase and heterophase fluctuations in semiconducting crystals*, Discuss. Faraday Soc. **28**, 151 (1959).
- [110] L. Mandel, *Image fluctuations in cascade intensifiers*, Brit. J. Appl. Phys. **10**, 233 (1959).
- [111] W. Shockley and J. R. Pierce, *A theory of noise for electron multipliers*, Proc. IRE **26**, 321 (1938).
- [112] J. F. Nye, *Physical properties of crystals* (Oxford University Press, Oxford) (1985).
- [113] R. Houtz, C. Chan, and H. Müller, *Wideband, efficient optical serrodyne frequency shifting with a phase modulator and a nonlinear transmission line*, Opt. Expr. **17**, 19235 (2009).

- [114] D. M. S. Johnson, J. M. Hogan, S.-W. Chiow, and M. Kasevich, *Broadband optical serrodyne frequency shifting*, Opt. Lett. **35**, 745 (2010).
- [115] R. Kohlhaas, T. Vanderbruggen, S. Bernon, A. Bertoldi, A. Landragin, and P. Bouyer, *Robust laser frequency stabilization by serrodyne modulation*, Opt. Lett. **37**, 1005 (2012).
- [116] R. V. Pound, *Electronic frequency stabilization of microwave oscillators*, Rev. Sci. Instr. **17**, 490 (1946).
- [117] D. A. Shaddock, M. B. Gray, and D. E. McClelland, *Frequency locking a laser to an optical cavity by use of spatial mode interference*, Opt. Lett. **24**, 1499 (1999).
- [118] N. M. Sampas and D. Z. Anderson, *Stabilization of laser beam alignment to an optical resonator by heterodyne detection of off-axis modes*, Appl. Opt. **29**, 394 (1990).
- [119] S. Timoshenko and S. Woinowsky-Krieger, *Theory of plates and shells*, 2nd Ed. (McGraw-Hill, New York) (1959).
- [120] M. Sparks and M. Cottis, *Pressure-induced optical distortion in laser windows*, J. Appl. Phys. **44**, 787 (1973).

Index