CHAPTER 4

Lenses, Prisms, and Mirrors

In theory, theory and practice are the same. In practice, they're different.

—Anonymous

4.1 INTRODUCTION

Although lasers and nonlinear optics get all the attention, most of the real work in an optical system consists of making the beam you've got into the one you want and routing it where you want it. This prosaic but necessary work is done by lenses, prisms, and mirrors (with an occasional cameo appearance by a grating). In this chapter, we discuss the workings of these simple devices, beginning with what they're made of and how they work. The bulk of the chapter is a heuristic treatment of what they actually *do*, and the surprisingly subtle business of how best to combine them so that they do what you want them to.

Designing a complex lens is a highly specialized art using special tools, and is beyond our scope. Using and combining lenses that others have designed, on the other hand, is a very practical skill that everyone in optics should have.

4.2 OPTICAL MATERIALS

4.2.1 Glass

Glass is remarkable stuff. It can be far more transparent than air and stronger than steel. It is useful for almost every optical purpose from lenses to torsion springs.

By and large, glass is a trouble-free material as well. It comes in a very wide variety of types. The main properties we care about in optical glass are its refractive index, its dispersion (i.e., how much the index changes with wavelength), and its transmittance. For historical reasons, these are specified with respect to certain spectral lines of common elements, which were discovered and labeled by Joseph von Fraunhofer in the solar spectrum. The index usually quoted is n_d, at the d line of helium at 587.6 nm. Older references use the slightly redder D lines of sodium, 589.3 ± 0.3 nm, but it doesn't matter much. The dispersion is quoted as N_{FC},

$$N_{FC} = \frac{n(486.1\text{ nm}) - 1}{n(656.3\text{ nm}) - 1}. \tag{4.1}$$

Building Electro-Optical Systems, Making It All Work, Second Edition, By Philip C. D. Hobbs

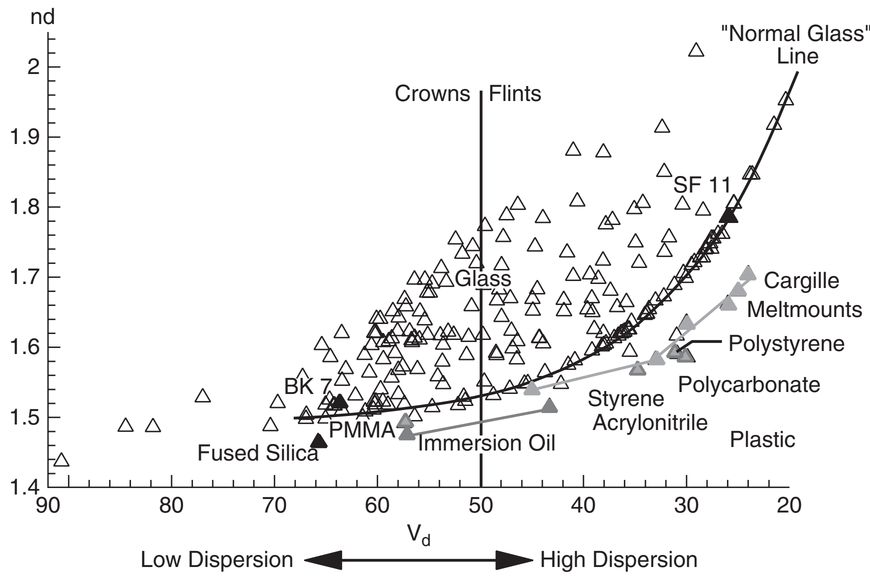

Figure 4.1. Refractive index n_d versus reciprocal dispersive power V_d for Schott optical glasses and common plastics.

The deep red C line at 656.3 nm is the Balmer α line of hydrogen, and the blue-green F line at 486.1 nm is Balmer β. The quantity $n - 1$ governs the power of a glass–air surface, so N_{FC} is the ratio of the powers of a given surface at the F and C lines. The classical way of quoting dispersion is the Abbe number V,

$$V = \frac{n_d - 1}{n_F - n_C}, \tag{4.2}$$

also called the reciprocal dispersive power (a big V means low dispersion). Figure 4.1 is a plot of the index versus dispersion for the optical glasses manufactured by Schott.

By and large, glass has a large coefficient of dispersion in the visible, which is highly variable among different glass types, but it has low temperature coefficients of index and of expansion.

Optical glasses are traditionally divided into two types: crowns, which have low indices (1.5–1.6) and low dispersion, and flints, which have higher indices and dispersion. The classical distinction was that anything whose V was over 50 was a crown, but the two categories have become blurred over the years as new glass formulations have been developed. The most common optical glass is BK7, a borosilicate crown glass with $n = 1.517$. It is popular because it is inexpensive and works well. Glass prices span a huge range—more than 100×.

Glass always has residual optical defects, such as bubbles and striae (long, thin regions of slightly different optical properties). For a critical application such as laser windows, choose a grade of glass whose striae and bubbles are guaranteed to be low enough.

For high quality applications, optical elements are often made of synthetic fused silica, a very pure quartz glass made by a chemical vapor deposition process. Fused quartz, an inferior material, is made by melting natural quartz sand. Fused silica comes in several grades, differing in the density and type of their bubbles and striae, and in their OH content. The O—H bond absorbs at 1.34 μm, 2.2 μm, and especially 2.7 μm. High-OH fused silica is essentially opaque at 2.7 μm. Low-OH fused silica, such as the Infrasil grade from Heraeus Amersil, can be pretty transparent there (depending on its thickness of course).

Fused silica and many types of glass are chemically fairly inert, but different glasses differ significantly. Hostile environments, such as continuous exposure to salt spray, will weather the surface of the glass (and any coatings present), degrading its optical performance. Severely weathered glass may appear to have a whitish crust on top. Fused silica and crown glasses resist weathering quite well, but high index glasses $n \approx 1.8$–2 are often environmentally delicate, as they have less quartz and more lead oxide and other less inert materials. Some of these will darken and weather over time even in ordinary lab conditions. The trade term for chemical inertness is *stain resistance*, and glasses are specified for it. Corrosion is not always a disaster: Fraunhofer discovered that weathering glass slightly increased its transparency—he correctly guessed that weathering the surface produced a layer of lower refractive index, which reduced the surface reflection. Pickling telescope lenses to improve their transmission was popular throughout the 19th century.

4.2.2 Temperature Coefficients of Optical Materials

The subject of temperature effects on optical elements is somewhat subtle. The temperature coefficients of expansion (CTE) and of refractive index (TCN) are both positive, but CTE is specified in normalized (dimensionless) form, whereas TCN is just $\partial n/\partial T$. The time (phase) delay through a piece of dielectric is $(n\ell)/c$, where ℓ is the length of the optical path. The normalized temperature coefficient, $\mathrm{TC_{OPL}}$, is

$$\mathrm{TC_{OPL}} = \frac{1}{n\ell}\frac{\partial(n\ell)}{\partial T} = \frac{\mathrm{TCN}}{n} + \mathrm{CTE}. \tag{4.3}$$

For most glass types, $\mathrm{TC_{OPL}}$ is approximately 10^{-5}/°C or a bit below. Fused silica has a very low CTE—in the 5×10^{-7} range—but a big TCN, about 9×10^{-6}, so with $n = 1.46$, its $\mathrm{TC_{OPL}}$ is 7×10^{-6}. BK7, on the other hand, has a larger CTE, 8×10^{-6}, but a low TCN, only 1.6×10^{-6}; its $\mathrm{TC_{OPL}}$ is 9×10^{-6}. There are a few glasses with negative TCN, such as that used for molded optics by Corning.† The only common solid with a negative TCN is magnesium fluoride (MgF_2). Air's TCN at constant pressure is about -1×10^{-6} (see below). Glass used in solid state laser rods is often specially formulated to achieve a zero $\mathrm{TC_{OPL}}$. Etalons sometimes use two materials with opposite signs of $\mathrm{TC_{OPL}}$, their thicknesses chosen so as to athermalize the path length.

The definition of $\mathrm{TC_{OPL}}$ here is for the phase delay inside the dielectric, which is relevant for discussion of fringes. In discussions of the temperature coefficients of lenses or of a mixture of dielectric and free space, another temperature coefficient is also useful, that of the differential optical path length through the element,

$$G = \mathrm{TCN} + (n-1)\mathrm{CTE}. \tag{4.4}$$

†Mark A. Fitch, Molded optics: mating precision and mass production. *Photonics Spectra*, October 1991.

4.2.3 Air and Other Gases

The expression (4.4) is actually only an approximation, since it leaves out the effects of the refractive index of air. Air's index is nearly 1, but it has surprisingly large temperature and pressure coefficients. An ideal gas behaves strictly as a collection of isolated molecules, whose molecular susceptibilities are constant. Thus the dielectric susceptibility of an ideal gas is strictly proportional to its density, which in turn can be predicted from the ideal gas law. What this means is that χ (and hence $\epsilon_r - 1$) is proportional to pressure and inversely proportional to temperature,

$$(\epsilon_r - 1) \propto \frac{P}{T}. \tag{4.5}$$

Since $\epsilon_r \approx 1$, a binomial expansion shows that

$$\text{TCN} = -\frac{n-1}{T}, \frac{\partial n}{\partial P} = \frac{n-1}{P}. \tag{4.6}$$

For dry air at $T = 288\text{K}$ (15 °C) and $P = 101.325$ kPa (1 atm), $n = 1.00028$, so that TCN $\approx -1.0 \times 10^{-6}$/K and $\partial n/\partial P \approx 2.8 \times 10^{-6}$/kPa. Thus air's TCN is comparable in magnitude to that of BK7. These are small numbers that aren't usually much of a concern, but they become very important in the design of interferometers, especially Fabry–Perot etalons, and in the presence of temperature and pressure gradients. Humidity is a second-order effect, because the water content of moist air is only a couple of percent and the molecular susceptibilities of H_2O, N_2, and O_2 are similar. Helium has a susceptibility about 8 times less than air, with corresponding decreases in the temperature coefficients. Note also that these are *partial* derivatives—TCN is quoted at constant P, and $\partial n/\partial P$ at constant T.

4.2.4 Optical Plastics

Plastic lenses have become very popular lately. They are lightweight, cheap (at least in high volumes), and can be molded with a mounting surface integral with the lens element, which helps a lot with assembly and alignment. Mass-produced aspheric lenses (perhaps with a diffractive element on one side) make it possible to get good performance with fewer elements.

On the other hand, plastic tends to be less uniform than glass, cannot readily be cemented, and is harder to coat. Plastic also has much larger temperature coefficients of expansion (≈150 ppm/°C) and of refractive index (≈100 ppm/°C) than glass. The most popular plastic used is polymethyl methacrylate (PMMA), sold under the trade names Lucite, Perspex, and Plexiglas. Others are polycarbonate (Lexan), cyclic olefin copolymer (COC, sold as Zeonex and Topas), and CR39, used for eyeglasses.

Plastics don't have the variety of optical properties found in glasses. Their indices range from about 1.44 to 1.6, with some of the newest (and very expensive) ones reaching 1.7. They have a narrow but reasonable range of V, 30 to about 58, so that plastic lenses can be achromatized. They have higher internal Rayleigh scatter due to the high molecular weight of the polymer chains. They are less transparent in both the UV and IR than most glasses and are more vulnerable to solvents and other environmental hazards.

UV exposure is especially damaging to some kinds of plastics, causing them to yellow and craze. This is especially true of the polyethylene used in far-IR Fresnel lenses, as used in automatic porch lights. Thermosets such as CR39 ($n_d = 1.50$, $V = 58$) are about the most durable optical plastics.

4.3 LIGHT TRANSMISSION

Alongside refractive index and dispersion, the transmission properties of a material govern its range of applicability. As a rule of thumb, a wide selection of good materials is available between 300 nm and 3 μm; there is some choice between 200 nm and 15 μm; below 200 nm and beyond 15μm, most materials are strongly absorbing, so we take what we can get.

4.3.1 UV Materials

Optical materials don't go very far into the UV. The absolute champion is lithium fluoride, LiF, which dies at about 120 nm, right around the Lyman α line of hydrogen at 121.6 nm. The fluorides of barium (BaF_2), magnesium (MgF_2), and strontium (SrF_2) are nearly as good and are more practical materials—harder, easier to polish, and less vulnerable to water. Water absorption can destroy the UV transmittance of LiF completely.

UV grade fused silica is useful down to 170 nm, but glass craps out at about 300–350 nm. Many types of materials are damaged by exposure to short wave UV (below 320 nm or so); glass will darken, and plastics yellow and craze. Flashlamps and arc lamps have strong UV emission, so this can easily happen even in visible-light applications.

4.3.2 IR Materials

Optical glass and low-OH fused silica are useful out to 3 μm or so. Beyond there, the choices diminish considerably; the best window materials are semiconductors like silicon and germanium, both of which can be made transparent out to 15 μm or further. These materials have high indices, 3.5 for Si and 4 for Ge. This is great for lenses, because with a high index, large aperture lenses can be made with shallowly sloped surfaces, so that aberrations are minimized. It is less helpful for windows, because Fresnel losses are large, and the huge index mismatch makes AR coatings rather narrowband. These materials are opaque in the visible, which is a pain because all your alignment has to be done blind. (Don't underestimate the difficulty of this if you haven't tried it—you'll have a good deal less hair afterwards.)

There exist lower index materials with excellent IR transmission, but most of them are toxic or water soluble. The best ones are diamond (if you can afford it, it goes from 230 nm to 40 μm with a weak interband absorption at 2.6–6.6 μm), zinc selenide (ZnSe), arsenic trisulfide or triselenide glass (As_2S_3 and As_2Se_3), and sodium chloride (NaCl). Good quality synthetic sapphire and calcium fluoride are also good if you can live with their limitations (mild birefringence for sapphire and sensitivity to thermal shock for CaF_2). Others, such as thallium bromoiodide (KRS-5), are sufficiently toxic that only the stout of heart and fastidious of touch should grind and polish them. These materials have the enormous advantage of being transparent in at least part of the visible, which makes using them a great deal easier.

In the far IR, some plastics such as high density polyethylene (HDPE) are reasonable window materials. Their high molecular weight and polycrystalline morphology, which make them highly scattering in the visible, are not a problem in the IR (Rayleigh scattering goes as λ^{-4}). Ordinary polyethylene or PVC (Saran) food wrap makes a good moisture barrier to protect hygroscopic window materials from humidity. These films can be wrapped around the delicate element without affecting its optical properties in the far IR very much (although each type should be tested for absorption before use).

Unlike visible-light optical materials, most IR materials have low dispersion but have huge temperature coefficients of index and expansion compared with glass (silicon's dn/dT is ~170 ppm/K, and while As_2S_3's dn/dT is only about 10 ppm/K, its CTE is 25 ppm/K. Near- and mid-IR absorption depends on molecular vibrational modes.

4.4 SURFACE QUALITY

It isn't just the material that matters, but the surface quality too. Ray bending happens at the surfaces, so they have to be accurate to a fraction of a wavelength to maintain image quality. Total scattered light tends to go as $[4\pi(\text{rms roughness})/\lambda]^2$, so optical surfaces have to be smooth to $\sim \lambda/1000$. The figure error is the deviation of the surface from the specified figure, without regard for small localized errors, which are divided into scratches and digs (small craters or pits). Scratches are usually just that, but digs are often the result of a bubble in the glass having intersected the surface of the element.

A commodity colored glass filter might have a scratch/dig specification of 80/60, which is pretty poor. An indifferent surface polish is 60/40, a good one is 20/10, and a laser quality one (good enough for use inside a laser cavity) is 10/5. The scratch/dig specification largely determines the level of scatter that can occur at the surface and also affects the laser damage threshold and the weathering properties of the glass. It tells something about how numerous and how large the defects can be but is a subjective visual check, not anything that can be easily converted to hard numbers.

Figure error and scattering from scratches and digs are not the only ways that manufacturing variations can mess up your wavefronts. Striae and bubbles in the glass can result in significant wavefront errors, even in a perfectly ground and polished lens, and will produce a certain amount of internal scattering. Rayleigh scattering from the dielectric sets the lower limit.

Lenses are more forgiving of surface errors than mirrors are, because the rays are bent through a smaller angle at the surface. Tipping a mirror by 1° deflects the reflected light through 2°, regardless of incidence angle, whereas tipping a window or a weak lens has almost no effect at all.

An alternative way of saying this is that a figure error of ϵ wavelengths produces a phase error on the order of $\epsilon(n-1)$ waves in a lens and 2ϵ waves in a mirror at normal incidence. If the error is a tilt or a bend, even this reduced error tends to cancel upon exiting the other side of the lens. If the mirror is operated at grazing incidence, to reproduce the ray-bending effect of the lens, the surface roughness sensitivity is reduced equivalently, because k_Z is smaller.

Aside: Surface Error Sensitivity. The effects of roughness or surface error in a high index material such as germanium ($n = 4.0$) can be even worse than in a mirror, when quoted in terms of wavelengths. Since these materials transmit only in the IR, however, the actual sensitivity in waves per micron of surface error is not much different from glass in the visible.

Small-scale roughness produces a phase front with random phase errors. Since $\exp(i\phi) \approx 1 + i\phi$, this appears as additive noise on the wavefront and produces a scattering pattern that is the Fourier transform of the surface error profile. In Section 13.6.9, we'll see an exactly analogous case in which additive noise produces random phase shifts in signals. For the optical case, compute the deviation from perfect specular reflection. Anything that doesn't wind up in the main beam is scatter, so in the thin object approximation (Section 1.3.9) the total integrated scatter (TIS) is

$$\text{TIS} \equiv P_{\text{scat}}/P_{\text{refl}} = 1 - \exp(-(2k_z\langle\Delta z\rangle)^2). \tag{4.7}$$

Quantities like $\langle\Delta z\rangle$ are implicitly filtered to spatial frequencies below k, since evanescent modes don't reduce the specular reflection.

Lenses are often poorly specified for wavefront error, since most of the time simple lenses would be limited by their aberrations even if their figures were perfect. Mirrors are usually specified with a certain degree of flatness ($\lambda/10$ at 633 nm is typical).

4.5 WINDOWS

Although windows are very simple optical elements, they are unusual in that their purpose is not usually optical at all, but environmental: a window separates two volumes that are incompatible in some way. Ordinary windows in a house or an airplane are examples, as are windows in a vacuum chamber, laser tube, or spectrometer cell. Microscope slides and cover slips are also windows. A window is the simplest lens and the simplest prism.

The trick with windows is to make sure that their environmental function does not detract from the working of the optical system. A beam of light is affected by any surface it encounters, so windows should be specified with the same sort of care as lenses and mirrors. There is a large range of choice among windows: material, thickness, coatings, wedge angle, and surface quality. For work in the visible, in a benign environment, choose windows made of an optical crown glass such as BK7 or its relatives. In a corrosive environment, quartz, fused silica, or sapphire are better choices. Filters sold for cameras, for example, UV or Video filters (not skylight or color correcting ones), are pretty flat, have good multilayer coatings for the visible, and are very cheap. Be sure you buy the multilayer coated ones.

4.5.1 Leading Order Optical Effects

The leading order optical effects of windows are shown in Figure 4.2. Rays entering a window are bent toward the surface normal, so images seen through a window appear closer than they are; a window of thickness d and index n shifts the image a distance $\Delta z = d(1 - 1/n)$ but does not change the magnification. The window thus looks from an imaging point of view like a negative free-space propagation. This effect on images is opposite to its effect on the actual optical phase; because light is slower in the window, the phase is delayed by the presence of the window, as though the propagation distance had *increased*. These effects are very useful in imaging laser interferometers, where they allow a nonzero path difference with identical magnification in the two arms—tuning the laser slightly shifts the phase of the interference pattern.

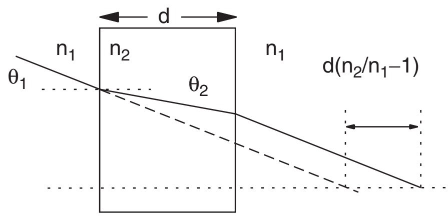

Figure 4.2. The leading order effect of a window is a shift in image position and a phase delay.

4.5.2 Optical Flats

A window that has been polished very flat ($\lambda/20$ or so) is called an *optical flat*. Flats are used as phase references (e.g, in Newton's rings and Fizeau wedges). Most of them have a small built-in wedge angle of 30 arc minutes or so to reduce etalon effects, but you can get them with two flat parallel surfaces.

4.6 PATHOLOGIES OF OPTICAL ELEMENTS

It is normally possible to get lenses, windows, and mirrors of adequate surface quality in a suitable material. The troubles we get into with optical elements don't come so much from random wavefront wiggles due to fabrication errors, but from aberrations, front-surface reflections, and birefringence. Aberrations we'll talk about beginning in Section 9.2.2, but briefly they are higher-order corrections to the paraxial picture of what optical elements do. As for the other two, let's do birefringence first—it's simpler.

4.6.1 Birefringence

(See Section 6.3.2 for more detail.) Birefringence in good quality lenses and windows comes from material properties, as in sapphire or crystalline quartz, and from stress. The stress can be externally applied as in pressure-vessel windows or lenses with poor mounts (e.g., tight metal set screws). It can also be internal to the glass due to nonuniform cooling from the melt or surface modification such as metalliding. The birefringence produced by stress is $n_\perp - n_\parallel$ with respect to the direction of the stress:

$$n_\perp - n_\parallel = K \cdot S \tag{4.8}$$

where S is the stress in N/m^2 (tensile is positive) and K is the piezo-optic coefficient (or stress-optic coefficient). Most optical glasses have piezo-optic coefficients of around $+2 \times 10^{-12}$ m^2/N, so that a compressive stress of 200 N on a 1 mm square area will produce a Δn of around 0.0004. A few glasses have negative or near-zero K values, for example, Schott SF57HHT, whose K is two orders of magnitude smaller ($+2 \times 10^{-14}$ m^2/N). Note that the K value is wavelength dependent, which matters at these low levels.

Stress birefringence comes up especially when we use tempered glass, in which the residual stress is deliberately kept high in order to force surface cracks to remain closed. Colored glass filters are usually tempered, so that it is dangerous to assume that the polarization of your beam will remain constant going through a glass filter.

Good optical glass is carefully annealed to remove most of the residual stress. Fine-annealed glass, the best commercial grade, has less than 12 nm of residual birefringence in a 100 mm thickness. In a normal window of perhaps 3–5 mm thickness, this is usually negligible, but in a large prism it may not be.

Intrinsic birefringence is encountered in windows made of sapphire and some other crystalline materials. It is often vexing because these windows are usually chosen for a good physical reason—sapphire has unusually high mechanical strength and unsurpassed chemical and thermal inertness. Examples are flow cell windows in liquid particle counters, which may encounter liquids such as hydrofluoric acid (HF) solutions, which rapidly destroy quartz and glass, or windows in the plasma chambers used for reactive ion etching. In such cases, we must be grateful that such windows even exist, and make the best of their optical deficiencies.

Like most common birefringent optical materials, sapphire is *uniaxial* (two of its indices are the same). Its birefringence is fairly small and negative ($n_\perp - n_\parallel = 0.008$) and its index is around 1.8. The phase difference due to randomly oriented sapphire amounts to a few dozen waves for a typical 3 mm thick window in the visible, which is large enough to be very awkward. If we just need to get a beam in or out near normal incidence, we can use so-called *c-axis normal* sapphire windows, where the optic axis is normal to the surfaces, and light entering near normal incidence sees no birefringence. If this is not possible, we must usually choose optical designs that are insensitive to polarization, or compensate for the polarization error with appropriate wave plates or other means. Should this be unreasonably difficult, it is often possible (though usually painful) to use polarization diversity—doing the same measurement at several different polarizations and choosing the ones that work best. Fortunately, it is now possible to obtain quartz windows with a thin coating of transparent amorphous sapphire,† which has the inertness of crystalline sapphire without its pronounced birefringence. There's more on these effects beginning with Section 6.3.1.

4.7 FRINGES

The good news about fringes is that they are very sensitive to many different physical effects—displacement, frequency, temperature, air speed, and so on. Fringes are the basis of a great many highly sensitive measurements, as we saw beginning in Section 1.6. The bad news is that they are very sensitive to many different physical effects. The power of interference fringes to turn small changes into big signals is not limited to the ones we make on purpose, but extends to all the incidental fringes we create by putting things in the way of light as it propagates.

4.7.1 Surface Reflections

All surfaces reflect light. An uncoated glass-to-air surface at normal incidence reflects about 4% of the light reaching it, the proportion generally increasing for higher incidence angles (depending on the polarization). This leads to problems with stray light, etalon fringes, and multiple reflections.

†Research Electro-Optics, Inc.

4.7.2 Etalon Fringes

There are lots of different kinds of fringes, associated with the names of Fizeau, Newton, Haidinger, Fabry and Perot, and so on. All are useful in their place, all will occur accidentally, and all have baleful effects on measurement accuracy, independent of nomenclature. The usual informal term for these unloved stripes is *etalon fringes*. The name should not be allowed to conjure up visions of beautiful uniform beams blinking on and off as the length of a carefully aligned Fabry–Perot cavity changes by λ/2—the fringes we're talking about are not pretty, not uniform, and usually not visually obvious.

Fringes arise from the linear superposition of two fields. In Chapter 1, Eq. (1.68) shows that for two beams having the same shape, and whose phase relationship is constant across them, the combined intensity of two beams is

$$i_{AC}|_{peak} = 2\sqrt{i_{LO} i_{sig}}, \qquad (4.9)$$

where i_{LO} and i_{sig} are the detected photocurrents corresponding to the two beams individually, and i_{AC} isthe amplitude of the interference photocurrent (half the peak-to–valley value). In the case of a single window, etalon fringes arise from the interference of fields reflected from its front and back surfaces. Their interference causes large modulations in the total reflectance of the window, which vary extremely rapidly with wavelength and temperature. If the two reflections are comparable in strength, the net reflectance varies from twice the sum of the two (16% in the case of uncoated glass) to zero. (Since $T + R = 1$ for lossless elements, the transmittance changes from 84% to 100%.) The magnitude of this problem is not commonly realized, which is part of what gives optical measurements their reputation as a black art.

Since the size of the fringes depends on the *amplitude* of the stray reflections, it does not decrease as rapidly as you might imagine with multiple reflections. A two-bounce reflection, whose intensity is only 0.16% of the main beam, can cause a 3.2% p-v change in the reflectance, and even a completely invisible five-bounce beam ($i_5/i_0 \approx 10^{-7}$) can manage a p-v change of 6 parts in 10^4 (with a long path length to boot, which multiplies its sensitivity). In a more complicated system, the possibilities go up very rapidly; the number of possible five-bounce beams goes as the fifth power of the number of elements.

Example 4.1: Polarizing Cube. Consider an ordinary 25 mm polarizing beamsplitter cube, made of BK7 glass ($n_d = 1.517$), with a broadband antireflection (BBAR) coating of 1% reflectance. We'll use it with a HeNe laser at 632.8 nm. As shown in Figure 4.3, if the beam is aligned for maximum interference, the peak-to-valley transmission change is about 4% due to etalon fringes. The cube is 60,000 wavelengths long (120,000 round trip), so it goes from a peak to a valley over a wavelength change of 4 parts in 10^6—0.0026 nm, or 2 GHz in frequency. If the HeNe has a mode spacing of 1 GHz, then a mode jump can produce a transmission change of as much as 2.8% from this effect alone. The temperature effects are large as well. BK7 has a temperature coefficient of expansion of about $+7 \times 10^{-6}$/°C, and its TC of index is $+1.6 \times 10^{-6}$/°C. Thus the TC of optical path length is 8 ppm/°C. At this level, a 1° temperature change will cause a full cycle fringe shift, and for small changes on the slope of the curve, the TC of transmittance is $\pi(8 \times 10^{-6})(120,000)(4\%) \approx 12\%$/°C *for one element alone*. You see the problem—a 1 millidegree temperature shift will make your measurement drift by 120 ppm. Fortunately, we're rarely that well aligned, but being anywhere close is enough to cause big problems.

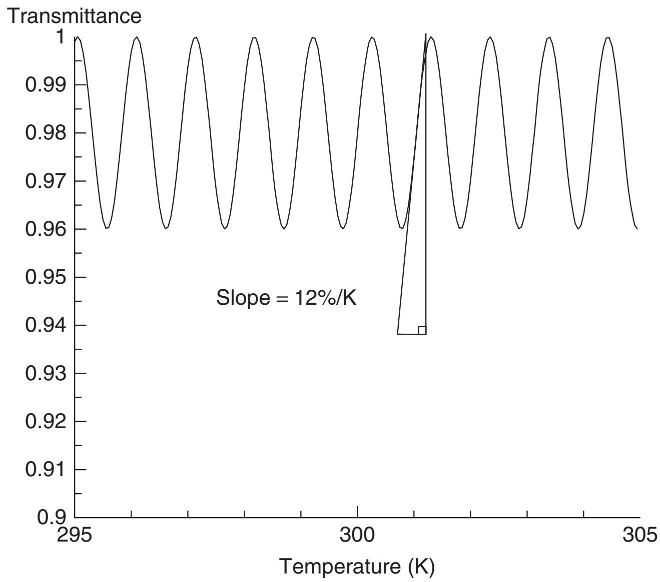

Figure 4.3. Normal-incidence transmission of a polarizing cube for 633 nm HeNe laser light, as a function of temperature.

4.7.3 Getting Rid of Fringes

Since we obviously can't make decent narrowband measurements in the presence of these fringes, we'll just have to get rid of them. Two basic strategies are commonly used: get rid of the fringes altogether, or cause them to smear out and average to some constant value. There is a long list of approaches people use to do these things, because everybody has to do them. None of them works that well, so a highly precise instrument usually relies on a combination of them—wear a belt and suspenders.

Add Strategic Beam Dumps. This sounds like an arms reduction program, but really it's like emergency roof maintenance: put buckets in the living room to catch the leaks. By calculating or finding experimentally where the front-surface reflections go, it is possible to put an efficient beam dump to catch them. See Chapter 5 for more on beam dumps—black paint is *not* usually good enough by itself. The main problem with this is that if you don't catch the stray beams before they hit other optical elements, they multiply in numbers to the point that it is nearly impossible to catch all of them.

Cant the Window. Tipping the element so that the front- and back-surface reflections are offset laterally from one another can be helpful. If the beam is narrow enough that the two will miss each other completely, this is a tolerably complete solution, assuming that there are no higher-order reflections that land on top of each other.

Apply Coatings. Front-surface reflections can be reduced by coating the glass. This is less effective than we would wish, as we saw in the egregious polarizing beamsplitter example above. A really good multilayer coating such as a V-coating can reduce Fresnel reflections to the level of 0.25%, but rarely any further unless the optical materials are specially chosen for the purpose. Such coatings are normally narrowband, but it's narrowband applications that really need them. V-coating the cube of Example 4.1 would reduce the slope to a mere 3%/K.

Come in at Brewster's Angle. Windows and prisms can be used at Brewster's angle, so that the p-polarized reflections go to zero. With highly polarized light and great care in alignment, the reflections can be reduced to very low levels. To reduce the amplitude reflection coefficient of an air–dielectric interface to less than ϵ requires an angular accuracy of

$$\Delta\theta_1 < \epsilon \frac{2n_1 n_2}{n_2 - n_1}, \tag{4.10}$$

so if we require $\epsilon = 10^{-3}$ (reflectivity $= 10^{-6}$) for a glass–air interface, we find that the incidence angle must be within 9 milliradians, or about 0.5 degree. Since the s-polarization reflection coefficient is

$$r_s|_{\theta_B} = \frac{n_2^2 - n_1^2}{n_2^2 + n_1^2}, \tag{4.11}$$

which is larger than the normal incidence value, the polarization must be really pure. With fused silica, $r_s = 13.8\%$, rising to 27.9% for a flint glass with $n = 1.8$.

The other advantage of Brewster incidence is that the residual surface reflection goes off at a large angle, where it can be controlled easily. For a prism, it is more important to control the internal reflection than the external one, because you usually can't put a strategically placed absorber inside an optical element. Thus if you have to choose, AR coat the entrance face and put the exit face at Brewster's angle. You will have to accept the residual stress in the prism causing slight polarization funnies. Brewster angle incidence is never quite what it promises to be.

Cement Elements Together. Elements whose refractive indices are similar can be cemented together, which reduces the surface reflections. If the match is good, for example a plano convex lens and a prism made of the same glass, or a colored glass filter ($n \approx 1.51$–1.55) and a BK7 window, the reflections can be reduced by factors of 100 or more. Another convenient feature is that when the two indices are closely similar, Brewster's angle is 45°, which is often a convenient incidence angle to work at.

Remember that the cement has to be index-matched as well! If the cement has an index midway between the two glass elements, the reflectivities are reduced by an additional factor of 2 on average, but anywhere in between is usually fine. Index oil can be used in lab applications; it comes in standard index increments of 0.002 and can be mixed or temperature controlled to better accuracy than that. Note its high dispersion (see Figure 4.1) and high TCN.

Use Noncollimated Beams. By using a converging or diverging beam, all interfering beams will be in different states of focus. This leads to lots of fringes across the field, so that the average reflectance is stabilized. The average interference term drops only polynomially with defocus, even with Gaussian beams, so more may be needed. With careful design, a reflection can be eliminated by bringing it to a focus at the surface of another element, and placing a dot of India ink or opaque there to catch it—baffle design in its minimalist essence. Remember the astigmatism, spherical aberration, and chromatic aberration caused by putting windows in nonparallel light. (Note: Parallel includes collimated but is more general; it refers to any place where the image point is at infinity.)

4.7.4 Smearing Fringes Out

Use Canted Windows and Low Spatial Coherence. Sometimes the beam is wide, the element thin, or the allowable tilt small, so the reflections won't miss each other completely. Nonetheless, if the spatial coherence of the beam is low enough, canted windows can still be useful. If the reflections are laterally separated by several times λ/NA, where NA is the sine of the minimum divergence angle of the beam at the given diameter, the fringe contrast will be substantially reduced. This works well for narrowband thermal light such as a mercury line.

Use Time Delays and Low Temporal Coherence. The same trick can be played in time. Fringes can be smeared out by taking advantage of the limits of your beam's temporal coherence, with a window several times $c/(n\Delta\nu)$ thick, so that the different optical frequencies present give rise to fringes of different phase, which average to near zero (for narrowband sources, this will not be your thin delicate window). Do remember the coherence fluctuation problem too.

Modulate the Phase. If the source is too highly coherent, etalon fringes can be reduced by wobbling the window or the beam rapidly, as with rotating wedge prisms or rotating diffusers, so that fringe motion is large and rapid compared to the measurement time. Unless this is done really well, it is less effective than it looks. The strength of the fringes depends on the time autocorrelation of the field at a delay corresponding to the round-trip time. If the phases of the two are being varied continuously but slowly compared to optical frequencies, what we get is fringes moving back and forth.

The kicker is that these fringes won't in general average to zero. For instance, take a triangle wave phase modulation of ± 10.5 cycles, where the unmodulated fields exhibit a bright fringe. Over a modulation cycle, the pattern will traverse 20 bright fringes and 21 dark ones, so that the average fringe amplitude is down by only a factor of 20. If the amplitude changes by 5%, to 10.0 or 11.0 cycles, the average fringe amplitude is 0—assuming that your measurement time is an exact multiple of the modulation period.

Modulate the Frequency. Another way of applying phase modulation is to tune the source wavelength rapidly (e.g., current tuned diode lasers). For modulation slow enough that the entire apparatus sees nearly the same optical frequency, this is nearly the same as the previous case, except that the range of phase shifts attainable is usually lower due to the limited tuning range of inexpensive diode lasers.

If the tuning is fast with respect to the delay between the reflections, the two reflections will be at a different frequency *most of the time*. Since the average frequencies of the two are the same, if the laser is turned on continuously the two instantaneous frequencies have to be the same twice per cycle of modulation (a stopped clock is right twice a day). The autocorrelation thus falls off more slowly than you might hope as the modulation amplitude increases, but nevertheless, this is a very useful trick, especially since by adjusting the modulation frequency and phase, you can put an autocorrelation null at the position of your worst reflection. It is especially good in reducing another etalon effect: mode hopping from diode lasers used in situations where backscatter is unavoidable. (Gating the laser can also improve matters sometimes.)

Put in a Wedge. Fringes are tamer if the two beams are not exactly collinear. Replacing the parallel surfaces with wedged ones makes sure this will be the case, leading to fringes of higher spatial frequency. These fringes will average out spatially if a large detector or a spatial filter is used. If there is enough space available, or the angle can be made large, a simple baffle will separate the two reflections completely. The key to this trick is to make sure that there is no low-spatial frequency component in the fringe pattern. Interference between two unvignetted Gaussian beams is an excellent example; the integral over the detector of the interference term goes to zero faster than exponentially with angular offset.

Allow the beams to be vignetted, or use less smooth pupil functions (e.g., uniform), and all bets are off. Since the pupil function and the vignetting both enter multiplicatively, they work like a frequency mixer in a superhet radio, moving the energy of your harmless high frequency fringe down toward 0, defeating the spatial averaging. Detector nonuniformity is also a problem here. Nonetheless, if this strategy is carefully applied, rejection on the order of 10^4 (electrical) is easily possible.

4.7.5 Advice

You will note that all of these strategies require care, and that all will become less effective very rapidly as the number of optical surfaces becomes larger, or their spacing smaller. Keep your system as simple as possible and, in an instrument where high stability is needed, be prepared to sacrifice aberration correction, which multiplies the number of surfaces. It is usually preferable to make a measurement at half the spatial resolution with 10^3 times better stability.

In a monochromatic system, eliminate all closely spaced, parallel planar surfaces, cement things together a lot, and use mirrors rather than prisms for beam steering.

4.8 MIRRORS

4.8.1 Plate Beamsplitters

These useful devices are just windows with an enhanced reflection coating on one side and often an antireflection coating on the other. They were once made with a very thin (10 nm or less) coating of metal on a glass plate, but such coatings are very lossy and so are no longer widely used. Good beamsplitters use a half-wave stack coating instead. Their problems are the same as those of windows, and the treatment is similar as well. They are available in a wide range of splitting ratios, from 90:10 to 10:90, with a very rough 50:50 being most common.

Beamsplitters are often used to allow illumination and viewing along the same optical path, as in microscope vertical illuminators. In that case, the light encounters the beam-splitter twice, once in reflection and once in transmission. The optimal efficiency is only 25% and occurs with a 50:50 beamsplitter. This maximum is very flat; 60:40 and 40:60 both give you 24%, and even 80:20 gives you 16%. Thus the poor accuracy of the 50:50 beamsplitter is not much of a worry.

Beamsplitters are *always* polarization sensitive. Polarizing plate beamsplitters are available at wavelengths of common high power pulsed lasers: 694 nm ruby and 1064 nm Nd:YAG. The cement used in polarizing cubes is easily damaged by high peak powers. These rely on multilayer $\lambda/2$ stacks oriented near Brewster's angle, so that one polarization is passed nearly completely and the other nearly completely reflected. In these devices, the transmitted beam is much more thoroughly polarized than the reflected one.

A dielectric plate inserted in a converging wave produces significant spherical aberration and, for off-axis points, astigmatism and coma as well (see Section 9.4.1). Either use the first-surface reflection for imaging, and illuminate in transmission, or put the beamsplitter before the objective lens, where the NA is low.

4.8.2 Pellicles

A pellicle is a plate beamsplitter on a diet. It consists of a 2–5 μm thick membrane (typically made of nitrocellulose) stretched across a very flat metal ring, sometimes coated. A pellicle is sufficiently thin that (at least in transmission) the aberrations it introduces are small enough to ignore. It is surprisingly robust mechanically, providing nothing actually hits the membrane.

Pellicles reduce the drift due to etalon fringes by making the etalon very thin, so that the fringe period is large, and its variation with temperature relatively slow. This works well enough for moderately narrowband sources, such as mercury tubes; with lasers, it may or may not, depending on your accuracy requirements. In fact, with broader band sources, pellicles tend to be a nuisance, as their broad fringes make the transmission nonuniform on the scale of thousands of wave numbers (hundreds of nanometers wavelength). Their main benefit in white-light systems is that they essentially eliminate ghost images due to the two reflecting surfaces.

Pellicles are not very flat—1 wave/cm typically, far worse than a decent plate beamsplitter. What's more, pellicles are very sensitive to vibration and air currents, which make them deform. A deformed or vibrating pellicle will not reflect a faithful replica of the incident wavefront; the transmitted beam is almost unaffected by the vibration but still suffers from the nonuniformity. The reflection from a pellicle is strongly angle dependent, varying from about 16% to 0% with angle and wavelength. Cleaning pellicles is difficult—you obviously can't use compressed air or lens paper, but in addition, nitrocellulose softens in ethanol. You can get away with detergent and deionized water or with isopropanol. As with gratings, it's best not to get pellicles dirty in the first place.

4.8.3 Flat Mirrors

Flat mirrors are conceptually the simplest optical elements available and often are the simplest to use, as well. Excellent quality mirrors are available at low cost, for a wide range of wavelengths, and from many suppliers. There are three main dangers in using them: neglect, leading to underspecifying or careless mounting; worst-case design, which although commendable in most fields, is something of a vice in optical systems, since it leads to very expensive overspecification; and blunders such as thumb prints. Mirrors are often more sensitive to contamination than lenses and more difficult to clean. Mirror coatings are discussed in detail in Section 5.2.

Some situations require high surface quality in mirrors: interferometers, flying-spot systems, and the best quality imaging systems are examples. Even there, however, there are lots of places in the system where the best mirrors are not needed. Before the two beams of an interferometer have been split, and especially after they have been recombined, the precision required of the mirrors is less than that of the end mirrors of the interferometer arms. Mirrors made of ordinary float glass (as used in domestic windows) are flat to better than 1 wave per cm of diameter. These very inexpensive mirrors are good enough to pipe light into detectors, for example. If there are more than a few mirrors in use, the total photon efficiency starts to drop dramatically if ordinary protected aluminum is used. Depending on how valuable your photons are, you may be much better off buying more expensive mirrors (or at least more expensively coated ones) if you can't simplify your optical system.

4.9 GLASS PRISMS

Glass prisms are used for dispersing light into its color spectrum, but most often for beam bending and image erection, both of which usually involve internal reflections off one or more faces of the prism. These reflections can be of two sorts: total internal reflection (TIR), in which the light hits the face at an incidence angle greater than the critical angle; or an ordinary reflection from a mirror-coated surface. Which of the two is superior depends on the application. TIR prisms are often used because their efficiency is superior to that of any mirror coating (provided the entrance and exit faces of the prism are AR coated sufficiently well, and the TIR face is really clean). Mirror coatings such as silver or aluminum are used in applications where the TIR condition is violated, where the reflecting surface cannot conveniently be kept clean, or where the phase and polarization shifts on total internal reflection are unacceptable. Some of the more common types of glass prism are shown in Figure 4.4.

4.9.1 Right-Angle and Porro Prisms

Right angle prisms are used for bending a beam through roughly 90° as in Figure 4.4a. Their performance is similar to a simple mirror oriented parallel to the hypotenuse of

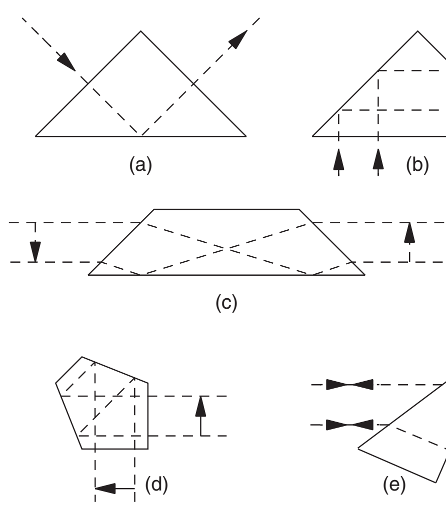

Figure 4.4. Types of glass prisms: (a) and (b) right angle, (c) Dove, (d) penta, and (e) Littrow.

the prism. Light enters normal to one face, bounces off the hypotenuse (either through total internal reflection or by means of a mirror coating), and exits through the other face. This arrangement is commonly used in microscopes, where prisms are more easily aligned and cleaned than mirrors, and where the high efficiency and spectral flatness of TIR reflectors or metallic silver coatings is important. Another advantage of prisms for microscopes is that the bulky optic is mostly inside the optical path, whereas the thick substrate of a mirror is entirely outside it. This makes a prism system mechanically more compact, an important attribute of parts that must slide in and out of the optical path.

The other way of using a right angle prism is shown in Figure 4.4b, which is the typical operating mode of the Porro prism (which is just a gussied-up right angle prism). Here the beam is reflected through 180° in one axis and undeviated in the other. The 180° angle is constant irrespective of rotations of the prism about an axis coming out of the page. This property is purely geometrical—the 180° is made up of two 90° reflections that add; rotating the prism through an angle ϕ will decrease the effect of the first reflection by ϕ while increasing the second one by exactly the same amount. Even the surface refraction cancels out, since the light emerges through the hypotenuse at exactly the same angle it entered at.

Porro prisms have a big enough incidence angle for TIR, so they are usually uncoated. There is a polarization shift on TIR, since the *s* and *p* polarizations have different phase shifts. Porro prisms are usually used in L-shaped pairs, one for up–down reversal and one for left–right, so as to produce an erect image. Provided that the prisms are at right angles to one another, *s* polarization for one bounce is *p* for the other, so the polarization shift cancels out.

Right angle prisms have one major pathology, which is the one that runs through this chapter (and much of the first part of this book in fact): severe etalon fringes due to the coincidence of the incident light and the front-surface reflections. We all have a natural desire for the beam path to be as simple as possible, without complicated alignment at odd angles. Unfortunately, this desire leads to lots of collimated beams and perpendicular surfaces, which makes for lots of etalon fringes. It is thus in direct conflict with our other natural desire, namely, to have our gizmos work when they're done.

4.9.2 Dove Prisms

The Dove prism is an image rotating prism, which also inverts the image left–right. It is a cylinder of square cross section and has faces cut at Brewster's angle to its central axis. Light propagating along the axis is refracted at the faces, bounces off one side, and exits through the other face, being refracted back to axial propagation in the process. If the prism is rotated about its axis, the image rotates at twice the speed of the prism. Interestingly, the polarization of the light does *not* rotate—it stays more or less the same (see Section 6.2.4 for why). Because of the two symmetrical refractions, there is no *angular* chromatic aberration, as in a simple prism, but there is *lateral*—different colors emerge moving in parallel directions but offset laterally from each other.

4.9.3 Equilateral, Brewster, and Littrow Prisms

An equilateral prism is commonly used for spectral dispersion. For use with polarized light, employment of a Brewster prism, in which the light enters and leaves near Brewster's angle, is generally superior. A Littrow prism Figure 4.4(e) is one in which the light enters at Brewster's angle and is reflected at normal incidence from the second face of the prism. Light of a certain wavelength thus returns along the path of the incident light, with other wavelengths dispersed on one side or the other. Such a prism is nice because it avoids having the beam direction pass through inconveniently arbitrary angles, and leads to a compact light path with few bends. Such prisms are commonly used as cavity mirrors in argon ion lasers. The laser can oscillate only at the wavelength at which the light path retraces itself. Littrow prisms are not particularly vulnerable to etalon fringes, because the front-surface reflections go off at large angles. The residual external reflection can be got rid of easily with a beam dump, and the internal one controlled with a patch of black wax or other index-matched absorber placed where it hits the prism surface.

There are several types of compound dispersing prisms, of which the Amici prism is representative. It has alternating triangles of high and low dispersion glass cemented together, oriented like the teeth of a bear trap with the low dispersion prisms forming one row of teeth and the high dispersion ones the other. This allows multiplication of the dispersing power without the beam having to go in circles. The cemented construction allows the internal surfaces to work near grazing, for high dispersion, without the large surface reflections and sensitivity to figure errors. Such high dispersion prisms have been superseded almost entirely by diffraction gratings, except in oddball applications where polarization sensitivity or overlap of grating orders is a problem and the linearity of the dispersion is not critical.

4.9.4 Pentaprisms

A pentaprism (Figure 4.4(d)) is an image erecting prism that maintains a constant 90° deviation between incoming and outgoing rays, independent of their incidence angle. The physical basis of this is two reflections in one plane, as in the porro prism. The beam undergoes two reflections from mirrors that are at an accurate 45° to one another (as shown in Figure 4.4d). Unless the prism is made of a very high index material ($n > 2.5$ or so), the steep incidence angle makes TIR operation impossible, so the reflecting faces are normally silvered. The entrance and exit facets are both normal to the beam direction, so pentaprisms produce etalon fringes but don't exhibit much chromatic distortion in parallel light. (See Section 4.10.)

4.9.5 Other Constant-Angle Prisms

There's nothing special about 90° or 180° as far as constant deviation prisms are concerned—the only requirement is to have two reflections from surfaces rigidly held together. To make a 60° constant deviation, for example, you can use a 30-60-90 degree prism. Send the beam in near normal to the short face. It will bounce off the hypotenuse (by TIR) and the long side, then exit through the hypotenuse at normal incidence, deviated by 60° exactly. The incidence angle on the long side is only 30°, so it must be silvered unless $n > 2.0$.

4.9.6 Wedges

A wedge prism is used for performing small deviations (up to perhaps 20°) in the pointing of a beam. Two such identical wedges mounted coaxially and independently rotatable can be used to point a beam anywhere in a cone of 40° half-angle (except for a small zone around 0°, caused by the inevitable slight mismatch between the angles of the two prisms). This adjustment is considerably more compact and robust than a reflection from two mirrors, but is somewhat less convenient to adjust and (like all refracting devices) more prone to etalon fringes.

The tendency to produce fringes is better controlled in wedge prisms than in most other refracting devices, since the surfaces are not parallel and the surface reflections can often be isolated by applying black wax or other index-matched absorbing material in strategic locations, or merely by making sure that none even of the high-order surface reflections can reenter the main optical path.

4.9.7 Roof Prisms

Most of the standard prism types are occasionally modified by replacing one face with a *roof*—a pair of surfaces at 90° to one another. The effect is identical to cementing a right angle prism, hypotenuse-first, on top of the standard prism: an additional left–right inversion takes place. The most common type is the Amici roof prism, which is a right angle prism with a roof. It has the useful property of reflecting a beam through 90° without inverting it left-to-right. In imaging applications, the roof prism must be made very carefully, because the ridge of the roof appears right in the middle of the field of view; any imperfections will be quite obvious.

4.9.8 Corner Reflectors and Cats' Eyes

A corner reflector (aka corner cube or retroreflector) is a constant 180° deviation prism. These useful devices come in two varieties: hollow ones, which are built up from three flat mirrors accurately perpendicular to one another, and solid ones, which are ground from a block of glass. They have the threefold symmetry of a cube about its body diagonal, but since the beam comes in and out at opposite sides, the optical symmetry is sixfold—looking into the corner cube, you see your own eye cut by six radial segments like an equatorial slice of an orange.

Solid retroreflectors may use TIR or may be coated with metal. The hollow ones tend to work better (stock items are available with 2 arc seconds tolerance, vs. 20 to 50 for solid). Solid ones have poorer transmitted beam quality and suffer from etalon fringes and multiple reflections; those that rely on TIR also cause large polarization shifts. On the other hand, solid retroreflectors are considerably easier to clean, very rugged, and can be used as a vernier adjustment of polarization by rotating them slightly. The polarization changes where the beam crosses a segment boundary, and the shift is big enough to cause weird apodizations in polarizing applications. The phase is also not continuous across the boundary, which will mess up focused or interferometric measurements. The net of all this is that corner cubes work great, but if you want to do anything fancy with the returned beam, it has to fit completely within one of the six 60° orange segments. It follows that the displacement of the returning beam axis has to be at least two beam diameters or so.

A retroreflector successively inverts k_x, k_y, and k_z of the incoming beam on each reflection. The amplitude profile is reflected through the axis of the cube, and $\mathbf{k}_{\text{out}} = -\mathbf{k}_{\text{in}}$. For collimated beams, this is identical to the action of a *cat's eye*—a lens with a mirror surface at its focus. It is different for converging or diverging beams, of course, since the back focus of the lens arrangement is reimaged at the back focal plane, whereas the retroreflector looks like a free-space propagation, so that the light would converge or diverge considerably before returning. This approximate equivalence is useful in building focused beam interferometers such as the ISICL sensor of Example 1.12. With a lens in one arm and a corner reflector in the other, no fine alignment is necessary, apart from collimation.

4.9.9 Beamsplitter Cubes

A beamsplitter cube works the same way as a plate beamsplitter, except that the reflective coating is deposited on the hypotenuse of a right angle prism, and another one is cemented on top of it, forming a cube with a reflective surface at its diagonal. They are most commonly polarizing, so that one linear polarization is reflected and the other transmitted, similarly to the polarizing type of plate beamsplitter.

Cube beamsplitters are very widely used, much more widely than their merits deserve. The advantages of the cube, namely, no beam deviation and easy mounting, are nowhere near sufficient to make up for the deficiency we saw in Example 4.1: severe etalon fringes in both the transmitted and reflected beams. If you do laser-based measurements, these infernal devices will make your life miserable.

On the other hand, with broadband sources and low spectral resolution, etalon fringes are not normally a problem, so cubes are a good choice. Even with lasers a bit of tweaking can help; if your beams are narrow, canting the cube slightly will help by making the reflections miss each other laterally, but you can't go too far or the polarization quality degrades badly. For experiments where you want good polarization and don't mind tweakiness, a cube mounted on a three-axis tilt table (such as a small prism table) can often be adjusted to give polarization purity of 1 part in 10^5 or even better in the transmitted beam.

Relaxing the requirement for no beam deviation can improve matters quite a lot more. If the faces are polished at an angle of a few degrees to one another, much better control of the reflections can be achieved. Such near-cubes are not catalog products, unfortunately, but can be synthesized for lab purposes by removing the coating from one or two faces and attaching wedge prisms with UV epoxy or index oil. There is only a very small index discontinuity at the surface, so the repolished surface doesn't have to be perfect, making hand work feasible if you don't have an optical shop.

Like all reflection type polarizers, polarizing cubes have problems with the polarization purity of the reflected light; more on that appears in Chapter 6.

4.9.10 Fresnel Rhombs

In Section 1.2.6 we saw that beyond the critical angle, the reflection coefficients of p- and s-polarized light both have magnitude 1 but have different phases, and that the phases depend only on n and θ_i, as given by (1.15),

$$\delta = \delta_s - \delta_p = -2\arctan\left[\frac{\cos\theta_i\sqrt{\sin^2\theta_i - n_2^2/n_1^2}}{\sin^2\theta_i}\right]. \tag{4.12}$$

To get a 90° retardation requires that $n_2/n_1 > 1/(\sqrt{2}-1) = 2.41$, which while not impossible is inconveniently high. A Fresnel rhomb, shown in Figure 4.5, does the trick by using two TIR reflections from a glass–air interface to produce an achromatic quarter-wave retarder.

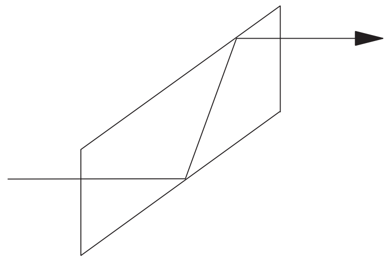

Figure 4.5. The Fresnel rhomb functions as an achromatic quarter-wave retarder.

Retarders in general are discussed in Section 6.9, but briefly, a quarter-wave retarder can be used to change linear into circular polarization and back again. Most retarders produce phase shifts by time delaying one polarization with respect to the other, so that the retardation depends strongly on the wavelength. Fresnel rhombs do not, so that apart from material dispersion, their phase retardation is constant with wavelength. This property makes them unique. The two reflections make this another constant-deviation prism—0° this time. Two rhombs cemented together like a V make an achromatic half-wave retarder with no beam deviation.

The retardation of a rhomb depends mildly on field angle, though less than that of a birefringent retarder (since when one angle goes down the other goes up). The main problem is the the long path length in glass. A 10 mm aperture rhomb made of BK7 has a path length in glass of $(10\text{ mm})(2\ \sec(54^\circ)) = 34$ mm, so that material nonuniformities produce retardation variations and phase wiggles across the aperture. Big rhombs are thus less accurate than other retarders for narrowband applications, besides being heavy.

4.10 PRISM PATHOLOGIES

Glass prisms are pretty trouble-free devices. They share the normal problems of any thick piece of dielectric, namely, residual birefringence, material nonuniformity, and etalon fringes. In polychromatic measurements, chromatic effects are also important.

A thick piece of glass is not as uniform as an air space of the same size, so that the waveform quality is poorer. Apart from polarization funnies, a prism has the same effect on an image as a window whose thickness is the length of the optical path inside the prism, unfolded (see Section 3.11.15). If the entrance and exit angles from the prism are not equal, the equivalent window has a wedge angle as well. Reflecting prisms such as pentaprisms and Fresnel rhombs can unfold to a very long path in glass. This means, for example, that a pentaprism placed in a converging or diverging beam will introduce large amounts of spherical aberration if care is not taken.

A big chunk of dispersive dielectric will cause lots of chromatic aberration if either the entrance and exit angles are different or the incident light is not parallel (focused at infinity).

4.11 LENSES

The sign conventions used in lens design are simple, but tend to be hard to remember, because they are completely arbitrary. Here are the four rules.

Sign Conventions in Lens Design

1. The object is at the extreme left of the drawing, and the image at the right (not so for mirrors of course).
2. The radius of a curved surface is positive if it is convex toward the left.
3. Distances along the direction of ray propagation are positive. If the ray would have to back up to get from the object to the lens or from the lens to the image, the distance is negative. Both d_o and d_i are positive when the image is real (true also for mirrors).
4. Phase increases with extra propagation distance; a ray that has to travel further than it should to get to a reference surface has a more positive phase, and so a positive aberration coefficient.

Glass lenses have been used for over a thousand years, since transparent glass became available. The fact that useful lenses could be made in the early days of glassmaking is an excellent indication of their forgiving qualities; for such precise artifacts, lenses are remarkably hard to get wrong. The main danger to the beginner is getting the signs backwards.

A lens images one space (the *object space*) into another, the *image space*. Since light propagation is time-reversal symmetric, lenses work fine backwards too; thus the choice of which is the object and which the image is somewhat arbitrary, so the two are often lumped together as *conjugate points*, or just *conjugates*. We covered the paraxial thin lens case in Section 1.3; here we go into the more general case.

4.11.1 Thin Lenses

The simple lens, usually made of glass and having spherical surfaces, is the most useful basic optical component. Although they are not in themselves adequate as imaging devices, except well into the infrared or at very low numerical aperture, they can be built up into lens systems that perform remarkably well. The simplest approximation to what a lens does is the *thin-lens approximation*, where the total thickness of a lens is vanishingly small. How thin is thin in real life? The thickness of the lens has to be small compared to the depth of focus of the beam you're using,

$$d < \frac{\lambda}{\mathrm{NA}^2}, \tag{4.13}$$

so that not knowing just where the ray bending is really going on doesn't affect the results.

A thin lens is characterized by its focal length f or equivalently by its *power* P, which is $1/f$. If its radii of curvature are r_1 and r_2, it has a focal length f (in air) given by the so-called *lensmaker's equation*,

$$\frac{1}{f} = (n-1)\left[\frac{1}{R_1} - \frac{1}{R_2}\right]. \tag{4.14}$$

From the rules in the text box, both radii are measured from the right side of the lens, so that for a biconvex lens r_1 is positive and r_2 negative; thus they both contribute positive power in (4.14). The powers of thin lenses placed in contact add.

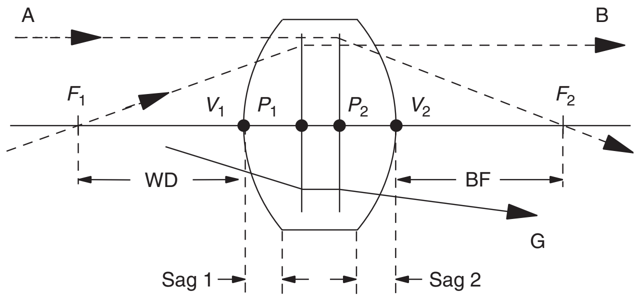

Figure 4.6. A thick lens acts like a thin lens, but operating between the *principal planes*. In the thin-lens limit, the principal planes collapse at the center of the lens. (Adapted from Kingslake.)

This approximation is convenient for initial layout of an optical system, but the effects of thickness must be put in before the real optomechanical design is done; the condition (4.13) is extremely stringent, requiring that a 1 mm thick lens of 10 mm diameter have a focal length $f \gg 800$ mm when used with a collimated red HeNe beam.

4.11.2 Thick Lenses

Fortunately this ferocious restriction can be got round easily; Gauss himself recognized that a lens of finite thickness has imaging properties very similar to those of a thin lens, except for the location error, and that this error can be eliminated by splicing in a small section of hyperspace, as shown in Figure 4.6†.

The planes P_1 and P_2 are the *principal planes* of the lens and intersect the lens axis at the *principal points*. Notionally, a paraxial ray coming from the left is undeviated until it hits the first principal plane P_1. It is then magically translated to P_2 at exactly the same lateral position (*height*) and then bent as if a thin lens of the same focal length were at P_2. The focal points F_1 and F_2 are the front and back foci, respectively. The axial distance from the left vertex of the lens to F_1 is the *front focal distance* or *working distance*, and that from the right vertex to F_2 is the *back focal distance* (also confusingly called the *back focus*). These are what you'd measure with a caliper and are tabulated by the lens manufacturer. The back focus is nearly always less than the focal length (BF $<$ FL).

If the refractive indices of the media on both sides of the lens are the same, then $f_1 = f_2$. If not, the Lagrange invariant (see Section 9.2.9) can be used to show that $n_1/f_1 = n_2/f_2$.

The lensmaker's equation can be generalized to the case of a single thick lens;

$$P = \frac{1}{f} = (n-1)\left[\frac{1}{R_1} - \frac{1}{R_2} + \frac{t}{n}\frac{n-1}{R_1 R_2}\right], \tag{4.15}$$

where t is the thickness of the lens, measured from vertex to vertex. The front and back focal distances are

$$\mathrm{WD} = f_1\left[1 - \frac{t(n-1)}{nR_1}\right] \tag{4.16}$$

†Rudolf Kingslake, *Lens Design Fundamentals*. Academic Press, Orlando, FL, 1978, p. 49.

and

$$\mathrm{BF} = f_2\left[1 - \frac{t(n-1)}{nR_2}\right], \tag{4.17}$$

and the separation 2δ between the principal planes is

$$2\delta = t + \mathrm{BF} - \mathrm{WD} = \frac{t(n-1)}{n}. \tag{4.18}$$

It is reasonable to take the center of the lens to be halfway between the front and back foci.

The axial distance from the vertex of a surface to its rim is called the *sagitta* or *sag* for short. The same term is used for an off-axis point, but there it is not the sag of the surface but the sag of that point, so confusion seldom arises. For a spherical surface of radius R and element diameter d,

$$\mathrm{sag} = R\left[1 - \sqrt{1 - \frac{d^2}{2R^2}}\right] \approx \frac{d^2}{4R}. \tag{4.19}$$

Example 4.2: Biconvex Lens. Consider a 100 mm $f/4$ symmetrical biconvex lens (25 mm diameter), made of BK7. From the lensmaker's equation, $R_1 = 200(1.517 - 1) = 103.4$ mm. Over a 25 mm diameter, each surface will have a sag of about $25^2/(412)$ or 1.5 mm. If we leave 1 mm edge width, then $t \approx 4$ mm. If we take that thickness, then the thick-lens equation gives us (in units of meters and diopters)

$$10 \text{ diopters} = 0.517\left[\frac{2}{R} + \frac{0.004(0.517)}{1.517R^2}\right], \tag{4.20}$$

where the second term is expected to be a small perturbation. We can either use the quadratic formula or just use the approximate value of 0.1034 m we got before to plug into the correction term; either way, we get $R = 104.7$ mm. Since the refractive index is uncertain at the level of $\pm$ 0.002, and most of the time the tolerance on focal length is a couple of percent, the iterative method works fine. Note also that we have discovered a useful rule of thumb: for glass of $n = 1.5$, the surface radii are equal to f for an equiconvex lens ($f/2$ and ∞ for a plano-convex).

Note that the temperature coefficients of index and of expansion (both positive) fight each other in (4.15); as T increases, the radii, thickness, and index all normally go up. This is in contrast to what we saw earlier for the temperature coefficient of optical path length. Since the net effect can be made positive or negative, it is possible to *athermalize* even a single element lens, so its focal length is nearly constant with temperature. Of course, the mounting and other mechanical parts must be considered in an athermalized design.

A thick lens can easily be put into the *ABCD* matrix formulation. Consider a thick lens of focal length f whose principle planes are at $\pm\delta$. The *ABCD* matrix for this is composed of a thin lens $L(f)$ with a (negative) free-space propagation operator $Z(-\delta)$ on each side:

$$LT(f;\delta) = Z(-\delta)L(f)Z(-\delta) = \begin{bmatrix} 1+\dfrac{\delta}{f} & -\left(2\delta - \dfrac{\delta^2}{f}\right) \\ -\dfrac{1}{f} & +\dfrac{\delta}{f} \end{bmatrix} \tag{4.21}$$

One subtle but important point: you might think that the symmetry of the operator form in (4.21) would mean that the lens can be put in backwards without any problem, but that isn't so. For an asymmetric lens, the front and back focal distances are different, so putting the lens in a mount backwards will move the center of the lens, and so cause a focus shift. It also changes the aberrations. This somewhat subtle effect leads to a huge gotcha if the lens is nearly, but not quite, symmetric; the asymmetry may not be immediately obvious, leading to blunders in assembly.

4.11.3 Fast Lenses

A lens with a short focal ratio (diameter/focal length) produces a bright image, due to concentrating light from a large angular range. In photographic applications, this allows a short exposure time, so that the lens is said to be *fast*. Fast lenses bend rays sharply, for which they need highly curved (*steep*) surfaces. Unfortunately, aberrations increase rapidly with increasing incidence angles of the rays; making a fast lens with good image quality is challenging. A rule of thumb to minimize spherical aberration is to minimize the maximum incidence angle of any ray on any surface. Thus when using a single element lens to focus a beam, use a plano-convex one with its flat side toward the focus. Use a double-convex lens for 1:1 imaging, and a meniscus lens in a converging beam.

Aside: Short and Fast. If you're wondering why we bother with *fast* and *short* instead of, say, *large*, it's another one of those historical quirks. A rather slow lens whose focal length is 8 times its diameter is said to be an $f/8$ lens, pronounced "eff eight." The aperture rings of camera lenses just say "8." People who think this is 1/8 would say that the ratio is small, while those who think it's 8 would say it was large. Everybody knows what *fast* and *short* mean—fast exposures and a short focal length for the lens diameter. Increasing and reducing the aperture with a diaphragm are always known as *opening up* and *stopping down*, also from camera lore; the detentes on an aperture ring are known as *stops* and go in integral powers of $\sqrt{2}$: 1.4, 2, 2.8, 4, 5.6.... Doubling the aperture would be "opening up 2 stops." Don't confuse this with the other use of *stop*, as in *field stop* and *aperture stop*—this could get muddled, since when stopping down the aperture, you're adjusting the aperture stop.

4.11.4 Lens Bending

The lensmaker's equation shows that the power of a lens can be distributed between the two surfaces by increasing one radius while decreasing the other. This procedure is called *lens bending* and leads to lenses of the sorts shown in Figure 4.7. Lenses of different types have identical paraxial properties, but their aberrations differ when the focal ratio is shorter. Lens bending is the most important degree of freedom in lens design.

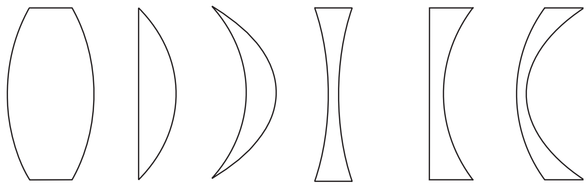

Figure 4.7. Lens types: (a) double convex, (b) plano-convex, (c) positive meniscus, (d) double-concave, (e) plano-concave, and (f) negative meniscus.

4.11.5 Dependence of Aberrations on Wavelength and Refractive Index

A single element 1-inch $f/2$ glass lens bent for minimum spherical aberration has about 10 waves RMS error at 588 nm. If it had an index of 4, that would be 1 wave; at 10.6 μm, it's 0.05 waves—diffraction limited.† The time-delay differences between different components do not change with wavelength, apart from dispersion, but as the wavelength gets longer, these differences become small compared to a wavelength, which is what *diffraction limited* means. Another way of looking at it is that as λ increases, the diffraction spot grows until it eventually dwarfs the geometric errors.

4.11.6 Aspheric Lenses

Because of the limitations of simple spherical lenses, it is natural to consider two possible ways of improving their performance: using them in combination, and relaxing the seemingly astrological constraint of spherical surfaces. Aspheric lenses can indeed perform better than simple spheres, and in low precision applications such as condensers, or large volume applications such as disposable cameras (where the lenses are made by a sophisticated plastic-over-glass molding process), they can be an excellent solution. The plastic-over-glass approach minimizes problems with the temperature coefficient and poor transparency of the plastic. Aspheric lenses are also commonly made by hot-pressing a few waves of asphericity into a glass preform (Corning) and by a sol-gel process based on tetramethyl orthosilicate (TMOS), which can be turned into a gel consisting of pure silica and then cured and baked to make it fully dense (Geltech). One-off custom aspheres are difficult to make and so are too expensive for most purposes. Molded glass aspheres can have good optical performance (e.g., a single element 4 mm 0.55 NA laser collimating lens (Corning 350160) with λ/20 RMS wavefront error—a Strehl ratio of around 0.95).

It is not really that a particular asphere is so very awkward to fabricate, at least not more so than another one; rather, what is at work is the strong tendency of any surface being polished to become spherical. This tendency is largely responsible for the fact that a small optical polishing shop producing components with surface accuracies measured in nanometers usually looks rather lower tech than an auto garage. The precision comes from the lens grinder's skill, the ease of testing the particular property sought (i.e., focusing), and from the surface's own seeming desire for sphericity.‡

Making aspheric lenses or mirrors requires resisting this desire, either by generating the surface with computer numerically controlled (CNC) diamond machining, or by

†*McGraw-Hill Encyclopedia of Lasers and Optical Technology*, p. 530.

‡Large optical shops nowadays have big surface generating machines that can polish many lenses at once.

nonuniform grinding and polishing, combined with iteration after iteration of polishing and careful interferometric measurement using a precisely made null corrector. Both procedures are expensive, and diamond machining has the additional disadvantage that the tool marks left behind tend to scatter large amounts of light when the element is used with visible light (it is much less of a problem in the IR).

4.11.7 Cylinder Lenses

The most common type of asphere is the cylindrical lens. These are widely available and relatively inexpensive, but their optical quality is notoriously poor. Grinding a lens with one accurately circular cross section and one accurately rectangular one is nontrivial.

Few applications of cylindrical lenses really require high accuracy, fortunately. Cylinders are often used as light buckets, directing light to a slit, as in a spectrograph, or to a linear photodiode array. Their one common, moderate accuracy application is in anamorphic pairs for correcting gross astigmatism or distortion, as in diode laser collimators; a better choice for this application is the combination of anamorphic prisms and controlled misalignment of the collimator. Despite the nice catalog pictures, cylinder lenses are lousy, so don't design systems requiring accurate ones.

4.12 COMPLEX LENSES

4.12.1 Achromats and Apochromats

For nearly all dielectric materials at nearly all wavelengths, the dispersion coefficients are positive; that is, n increases as λ decreases. The only exceptions are deep within absorption bands, where you won't want to use the stuff anyway. Thus it is not possible to color correct merely by sandwiching two plates of opposite dispersion.

On the other hand, you can color correct lenses by putting a positive lens of a high dispersion material next to a negative lens of low dispersion material, or vice versa. Let's take the positive lens case. A powerful positive lens made from crown glass next to a weaker negative lens made from flint glass produces a weakened but still positive two-element lens. If the powers of the two elements are adjusted correctly, then as we go to shorter λ, the increasing positive power is balanced by the increasing negative power, so that the net power of the combination is nearly constant. With two elements, the lens can be color corrected at two wavelengths and is called an *achromat*. Exploiting the different shapes of the dispersion curves of different glasses, color correction can be extended to three or more wavelengths, giving much lower average error over the wavelength band of interest; such a lens is called an *apochromat*.

A side benefit of needing two lenses to get color correction is that the extra degrees of freedom can be used to improve the monochromatic aberrations as well; a commercial achromat has so much less spherical aberration that its wavefront error will generally be 10 or more times better than a single-element lens of the same diameter and focal length.

Example 4.3: Achromatic Doublet. Suppose we want to make an achromatic 200 mm $f/8$ lens, corrected so that the F and C wavelengths come to a common focus. We'll use BK7 ($n_d = 1.51673$, $n_F = 1.52224$, $n_C = 1.51432$, so $N_{FC} = 1.01539$ and $V = 65.24$) for the front element, and SF11 ($n_d = 1.78446$, $n_F = 1.80645$, $n_C = 1.77599$, so $N_{FC} =$

1.03925 and $V = 25.75$) for the rear. The crown glass is usually more durable than the flint, so it is put on the exposed side unless there is a compelling reason not to. Real lens designers do this by exact ray tracing. We'll do it paraxially with the lensmaker's equation:

$$\begin{aligned} P_{\text{tot}} &= P_1 + P_2 = (n_1 - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + (n_2 - 1)\left(\frac{1}{R_3} - \frac{1}{R_4}\right) \\ &= P_{1d}\frac{(n_1 - 1)}{(n_{1d} - 1)} + P_{2d}\frac{(n_2 - 1)}{(n_{2d} - 1)}, \end{aligned} \tag{4.22}$$

where R_1 and R_2 are the surface radii of the first element, and R_3 and R_4 are those of the second element.

The achromatic condition requires $P_F = P_C$, which leads to

$$\frac{P_{1d}}{P_{2d}} = -\frac{V_1}{V_2} = -2.534 \tag{4.23}$$

and hence $P_{1d} = 1.652P$, $P_{2d} = -0.652P$. It is slightly simpler to express this in terms of the ratio of P_{1C}/P_{2C} and N_{FC}. For this lens, we require a positive element 1.65 times stronger than the combination, and a negative element 0.65 times as strong. We haven't specified anything about the bending of this lens, so we can put most of the power in the buried surfaces; nevertheless, it is difficult to make a cemented achromat faster than $f/1.4$ this way. It is possible to distribute the chromatic correction across the groups of a more complicated design to achieve good color correction at short focal ratios.

4.12.2 Camera Lenses

Camera lenses are a wonderful and inexpensive resource for optical instrument builders. For a hundred bucks or so, you get a very well corrected lens, preadjusted, mounted, and tested, on its own focusing ring and with an aperture diaphragm. Mating bayonet mounts are available to attach them to your system.

Bargains like that are not common in optics, so take advantage of it while it lasts. Camera lenses tend to have lots of elements, so their stray light and etalon fringe performance is not as good as simpler designs. Use the slower, fixed focal length designs rather than the superduper fast ones, the extreme wide angles, or the zooms; they have better image quality and a lot fewer elements to scatter and make fringes. Ordinary camera lenses are best for distant objects; the macrolenses are better for magnifications of 1:10 to 1:1. For higher magnifications, turn the lens around—for 1:1 to 10:1, use a macrolens backwards, and for magnifications greater than 10×, use an ordinary lens backwards (watch out for the soft coatings on the rear elements—they're not as durable as the hard front-surface coatings). For large magnifications, you can also use enlarger lenses, which are somewhat cheaper.

Camera lenses are often described as having, say, 13 elements in 6 groups (about typical for a zoom lens in 35 mm format). That means that there are 13 bits of glass, but that some of them are cemented together, leaving only $6 \times 2 = 12$ air–glass surfaces. Since the air–glass surfaces scatter more light, this is worth knowing; a 13-element lens with 13 groups would have 26 surfaces, so its surface reflections would likely be much worse. The etalon fringes in a lens like this would daunt the bravest, but even in a white-light application with good coatings, around a quarter of the total light would be bouncing around inside the lens barrel, much of it eventually arriving at the image plane to reduce the contrast and introduce artifacts. The 13/6 lens would be about half that bad.

4.12.3 Microscope Objectives

A lot of optical instruments can be cobbled together from microscope parts. Microscopes are high image quality systems built in modular fashion. Their optical trains must be designed to support this use, and those are just the qualities we want in instrument prototypes. Usually we just use the objectives, but sometimes an eyepiece or a trinocular head is useful too—you can't align what you can't see (see Section 11.8).

Microscope objectives are specified by magnification and numerical aperture. To find out what the actual focal length is, divide the magnification into the *tube length*, which is nearly always 200 mm, or 160 mm for old designs. Thus a modern 20×, 0.5 NA objective normally has a focal length of 10 mm. The working distance will be significantly less than this, which is sometimes very inconvenient in instruments; we often have to get other things in between the sample and the lens. Since this gets worse with shorter focal lengths, and since the NA and not f controls the resolution, 20×, 0.5 NA is the most all-round useful microscope objective for instruments.

Long working distance objectives are available; the longest ones come from Mitutoyo and are big enough to use as antiaircraft shells. They get rapidly more expensive as the aperture and working distance increase.

Some high NA microscope lenses come with an adjustable correction for cover glass thickness, which dials in a controllable amount of spherical aberration to correct for that introduced by the cover slip. This can be useful in other situations as well.

Microscope lenses exhibit severe longitudinal chromatic aberration; different colors come to focus at different depths. This is a trade-off based on the characteristics of the human visual system, which has poorer spatial resolution in the blue, but is obnoxious in some instrument applications, such as white light scanning microscopy. For such applications, and for use in the UV or IR, where microscope lenses are tough to get, you can use an all-mirror microscope objective, the *Schwarzschild objective*.

In choosing a microscope objective, know what it is you need. For applications not requiring the highest quality, such as the condenser side of a spatial filter, use a commodity objective such as the cheap ones from American Optical, Newport, Swift, or several others. For imaging, or on the collection side of a spatial filter, a good objective such as a Nikon, Olympus, Reichert, Leitz, or Zeiss will work much better. Japanese objectives tend to be nicely corrected in the objective itself, which makes them useful for other purposes where you don't want the microscope attached.

4.12.4 Infinity Correction

A lens designed to operate with its image at infinity is said to be *infinity corrected*. Most modern microscope objectives are infinity corrected, because the resulting parallel light exhibits no chromatic errors when being piped through bending prisms, and the system aberrations do not depend strongly on where subsequent optical components are located. These properties make infinity corrected lenses extremely useful in building instruments.

Camera lenses run backwards are another example of infinity correction—most of them are designed to have the object at infinity, but by time reversal symmetry, this is exactly equivalent to having the object at the film plane and the image at infinity. A pair of camera lenses operated nose to nose makes a good transfer lens, for example, to image the center of one scan mirror onto the center of another one, or to image an acousto-optic cell at the pupil of a microscope objective to make a scanning microscope. Note that the antireflection coating on the back of the lens is often much softer than the one on the front, and so much more easily damaged in cleaning. Similarly, the glass itself is often more delicate, and of course all the mechanical works are exposed to damage or corrosion.

4.12.5 Focusing Mirrors

Curved mirrors can do nearly anything lenses can but have a completely different set of design trade-offs. A focusing mirror causes much larger ray bending than a lens of the same focal length; this makes it much more sensitive to surface inaccuracies and misalignment than a lens, but also can lead to a more compact optical system, due to the opportunity for folding the light path. Folding leads to two problems, however; obscuration, as mirrors partially shadow one another, and light leakage, as far off-axis light can often get into the detector without having traversed the entire optical system. Baffles can eliminate leakage, but eliminating obscuration requires the use of off-axis aspheric mirrors, which are very difficult to align among other faults.

Mirrors exhibit no dispersion, so chromatic aberration is eliminated in all-mirror systems; the expected improvement in optical quality is not always realized, since it is much easier to make multielement lens systems than multielement mirror systems. On the other hand, for UV and IR use, it is delightfully easy to be able to focus the system with a HeNe laser and be sure that it will be perfectly focused at 335 nm or 1.06 μm. Most lenses have poorly corrected longitudinal chromatic aberration, so that different wavelengths come to focus at slightly different depths. Where spatial filtering is used with broadband illumination (e.g., real-time confocal scanning optical microscopy, where a disc full of pinholes is spun to make an array of moving spots), longitudinal chromatic is extremely objectionable, so mirror systems make a lot of sense.

Because there are no transparent surfaces in an all-mirror system, there is no opportunity for etalon fringes to form, which can be a very important advantage.

Where there are fewer surfaces, there are fewer degrees of freedom to optimize, and mirrors are much more seriously affected by surface errors than lenses are. For these reasons, aspheric mirrors are much more common than aspheric lenses. Overall, mirrors are wonderful in special situations such as working with invisible light, but there is nothing like an optical system based on focusing mirrors to make you appreciate the forgiving qualities of lenses.

Aside: Off-Axis Performance of Focusing Mirrors. Fast mirrors have amazingly bad oblique aberrations, and off-axis ones are the worst. For example, a 25 mm diameter, $f/1$, 90° off-axis paraboloid exhibits a spot size of approximately 30% of the off-axis distance—if you go 100 microns from the axis, the spot grows from the diffraction limit to 30 μm diameter. You really can't run any sort of field angle at all with those things.

4.12.6 Anamorphic Systems

An anamorphic system is one whose magnifications in x and y differ. The main uses of these are to correct for perspective distortion caused by oblique imaging, as in a picture of the ground taken from the side of an aeroplane, and to circularize elliptical beams from diode lasers. There are two main types: prism or grating systems, and telescopes made from cylindrical lenses. When a beam encounters a surface, its edges define an illuminated patch. If the beam comes in near grazing incidence, its illuminated patch will be greatly elongated. The law of reflection guarantees that a reflected beam will have the same shape as the incident one, but if the beam is refracted or diffracted at the surface, this is no longer the case. On refraction, a beam entering near grazing incidence will leave near the critical angle, elongated. Two prisms are often used together, with the second one bending the beam back to its original propagation direction. On diffraction, a beam entering near grazing can be made to leave near normal, so that a properly chosen grating can substitute for a 90° folding mirror. This idea is used in commercial beam circularizers based on anamorphic prisms.

Cylindrical telescopes can easily be made in any desired magnification, can be slightly defocused in order to correct astigmatism, and do not offset the beam axis, but this is as far as their advantages go. It is a great deal easier to make good prisms than good cylindrical lenses, and the astigmatism correction can be done by mildly misaligning the collimating lens.

4.12.7 Fringe Diagrams

Fringe diagrams are very nearly useless for good quality ($< \lambda/4$) optics. A good quality optical component will have a surface accuracy of a small fraction of a wavelength. Such a small deviation produces only very small irregularities in the spacing and direction of the fuzzy fringes in the diagram. Localized errors are reasonably easily spotted, but global ones, such as astigmatism, are very hard to see by eye, especially since the strong visual asymmetry caused by the fringes running in one direction masks asymmetric errors—you can see a $\lambda/4$ kink by sighting along the fringes, but good luck seeing a $\lambda/4$ spacing error spread out over 20 periods. For evaluating the quality of an optical component, use a phase shifting measuring interferometer if at all possible. Failing that, a Foucault knife-edge test using a razor blade is good for qualitative work, for example, examining a batch of lenses for bad units. If you really need to get quantitative data from a hard copy fringe diagram, either scan it into a computer and digest it in software, or use a parallel rule and a pencil to construct the axes of the fringes, and measure their straightness and separation. It really would be much more useful if optics vendors would ship their fringe diagrams on CD or make them available for downloading, but the author is not hanging about waiting for the day.

4.13 OTHER LENS-LIKE DEVICES

There are other devices besides lenses and curved mirrors that can make beams converge or diverge. These are based on refractive index gradients or on diffraction.

4.13.1 GRIN Lenses

Section 9.2.6 shows how a refractive index gradient causes light to bend. It turns out that a circular cylinder whose index decreases parabolically moving out from the core works like a lens. The ray bending happens continuously throughout the bulk, rather

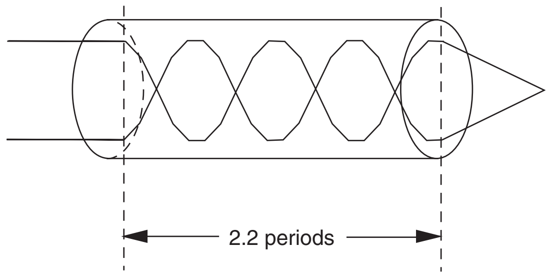

Figure 4.8. A GRIN lens exhibits periodic foci along its length.

than happening abruptly at the surface, as shown in Figure 4.8. Such a device is called a graded-index (GRIN) lens, or *GRIN rod*. Because the bending happens continuously, a length of GRIN rod exhibits periodically spaced images down its axis, alternating between erect and inverted. If it is cut to an even number of half-periods, a point on one surface is imaged onto the other surface; if it is a quarter period shorter, the image is at infinity. Thus making a GRIN lens of any desired focal length (up to a quarter of a period) is just a matter of cutting the rod to the right length.

At one time, GRIN lenses were quite poor—good enough for coupling light in and out of fibers but not for imaging. Recently, they have been developed to the point where their imaging properties are quite respectable. Fiber coupling is probably still the most common application, but borescopes and endoscopes are now often made from long GRIN rods instead of many sets of relay lenses. Besides simplicity and robustness, GRIN rods avoid the accumulation of field curvature that plagues designs with cascaded relays (e.g., periscopes).

Another approach to using index gradients is to fuse together slabs of optical glass of slightly different index.† When a spherical surface is ground into the high index side, the power of the margins of the lens is automatically weakened by the gradual decrease of n. This weakening can be chosen so as to cancel the spherical aberrations of the surface, and so aspheric-quality images can be obtained with a single spherical element.

Aside: Birefringence of GRIN Rods. GRIN rods are usually made by diffusing dopants in from the outside of a plain glass rod. This results in an axially symmetric pattern of residual mechanical stress, so that GRIN rods are actually birefringent, with a pattern not found in nature. This is sometimes important.

4.13.2 Fresnel Zone Plates, Diffractive Lenses, and Holographic Optical Elements

Recently, as a result of improvements in optical modeling software and in the molding of plastic and sol-gel glass, it has become possible to fabricate not only aspheric lenses but lenses with diffractive properties: for example, an aspheric lens with a phase grating on the other surface (zone plates and general holographic elements are discussed in Chapter 7). These diffractive lenses can have unique properties. Although the power

†Gradium glass, made by LightPath Technologies, Albuquerque, NM.

of the refractive surface decreases as the wavelength increases, that of the diffractive element increases; thus it can be used to cancel the chromatic aberration of the refractive surface, resulting in a novel element, an achromatic singlet. The possibilities inherent in such a capability have only begun to be assimilated, and such elements should be considered any time more than 10,000 units are needed. Before becoming too breathless with excitement, however, remember the drawbacks: plastics have huge temperature coefficients of index and of thermal expansion; getting good aberration correction over a large field is very difficult with only two surfaces of a low index material; and you absolutely must have very high diffraction efficiency, which is very difficult to maintain over a wide bandwidth (e.g., the visible spectrum). Also see Section 7.9.9 for reasons to keep the diffractive power small. The effective V number of a DOE can be found from the grating equation: $V = \lambda_d/(\lambda_f - \lambda_c) = -3.452$.

4.13.3 Fresnel Lenses

A mirror does all its work at the surface—all that stuff underneath is only there to keep the surface accurately in place. You can make a lightweight mirror by removing unnecessary material from the blank. This isn't quite as easy to do with a lens, since the material in the middle is there to preserve phase coherence by making all the rays arrive with the same delay. For crude light bucket applications, however, an idea of the same sort leads to the *Fresnel lens*. A Fresnel lens is a whole bunch of concentric annular lenses, with the same focal length, as shown in Figure 4.9. Because of all the sharp edges, there's lots of scatter, and because of the loss of phase coherence, the image quality is very poor.[†] Fresnel lenses can't normally be coated, either. Thus their efficiency is poor—as low as 50% in some cases.

On the other hand, they are light and compact, and nonimaging applications such as condensers aren't sensitive to their poor image quality. Projectors are the largest use of Fresnel lenses, but you can make a solar furnace this way, too—up to 2500 times solar concentration has been demonstrated. Off-axis Fresnel lenses are available, or can be made easily from a big one with snips or a razor knife.

When using a fast Fresnel lens, make sure to put the side with the ridges toward the more distant conjugate. Otherwise, the outer rings will exhibit TIR, and no light will get through them. Really steep conventional lenses exhibit this behavior too. The image quality of a Fresnel lens, nothing much to start with, gets dramatically worse with increasing field angle. (This can often be helped by shifting the aperture stop well out ahead of the Fresnel lens.)

4.13.4 Microlens Arrays

Lenses are sometimes used in arrays, to produce an array of very small, poor images. Microlenses as small as 10 μm in diameter are often deposited on top of interline transfer CCDs, to corral light into the small active area of each pixel. Somewhat larger lenses can be used to measure wavefront tilt as a function of position (the Shack–Hartmann technique), from which the wavefront can be reconstructed, more or less. (Even in thermal light, where the etalon fringes aren't so bad, there are never enough pixels per microlens

[†] Since there's no point knocking oneself out in a lost cause, Fresnel lenses are also manufactured to very loose tolerances.

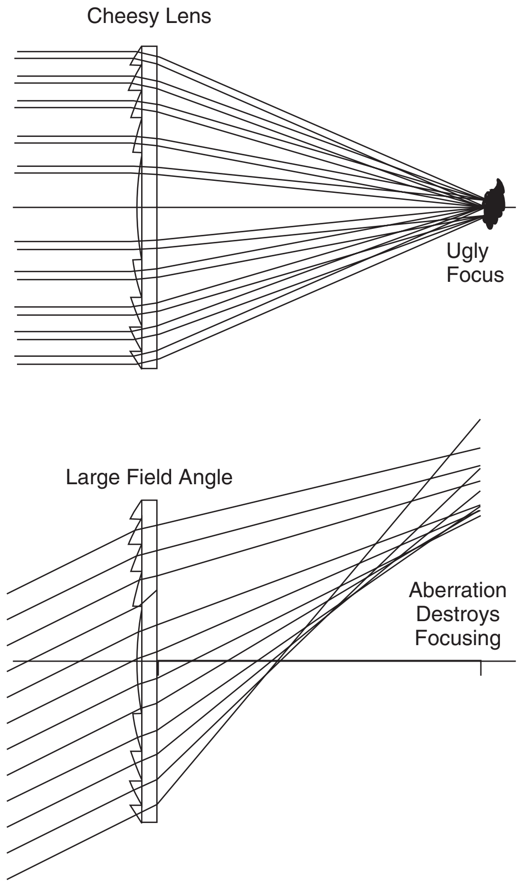

Figure 4.9. Fresnel lens.

to do a decent job of the reconstruction, unfortunately; the Shack–Hartmann measures the local slope of the wavefront, so to reconstruct it you have to extrapolate, which is always fraught with problems.)

Another interesting class of microlens array applications relies on the moiré pattern between two microlens arrays of slightly different pitch, which can synthesize the equivalent of a single short-focus lens. Standard microlens products are available (e.g., from WaveFront Sciences). They are typically 0.2 mm to a few millimeters in diameter, with focal lengths of 1–100 mm. They are generally plano-convex singlets, of course, and so their numerical apertures are limited even given their small Fresnel numbers. The existence of standard products makes microlens arrays good candidates for building into real systems.

4.13.5 Axicons

In addition to flat surfaces and spheres, there exists a class of conical prisms called *axicons*, as shown in Figure 4.10. They are generally made by single-point diamond turning, because otherwise it's difficult to get an accurate surface. Typical uses of an axicon are

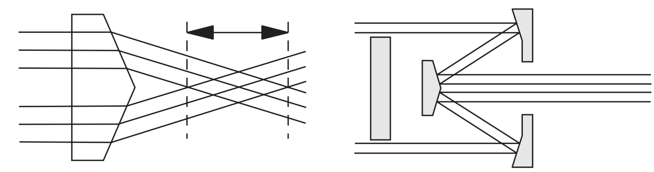

Figure 4.10. An axicon converts between filled and annular beams, or between a collimated beam and a J_0 Bessel beam.

sending a laser beam through a Schwarzschild (Cassegrain) microscope objective without hitting the central obstruction, turning annular beams from unstable laser resonators into uniform beams, and, less respectably, making J_0 Bessel beams (misnamed "nondiffracting") from uniform ones. Aligning axicons is very fiddly, especially in angle. The cone beam dump of Section 5.6.10 is also an axicon.