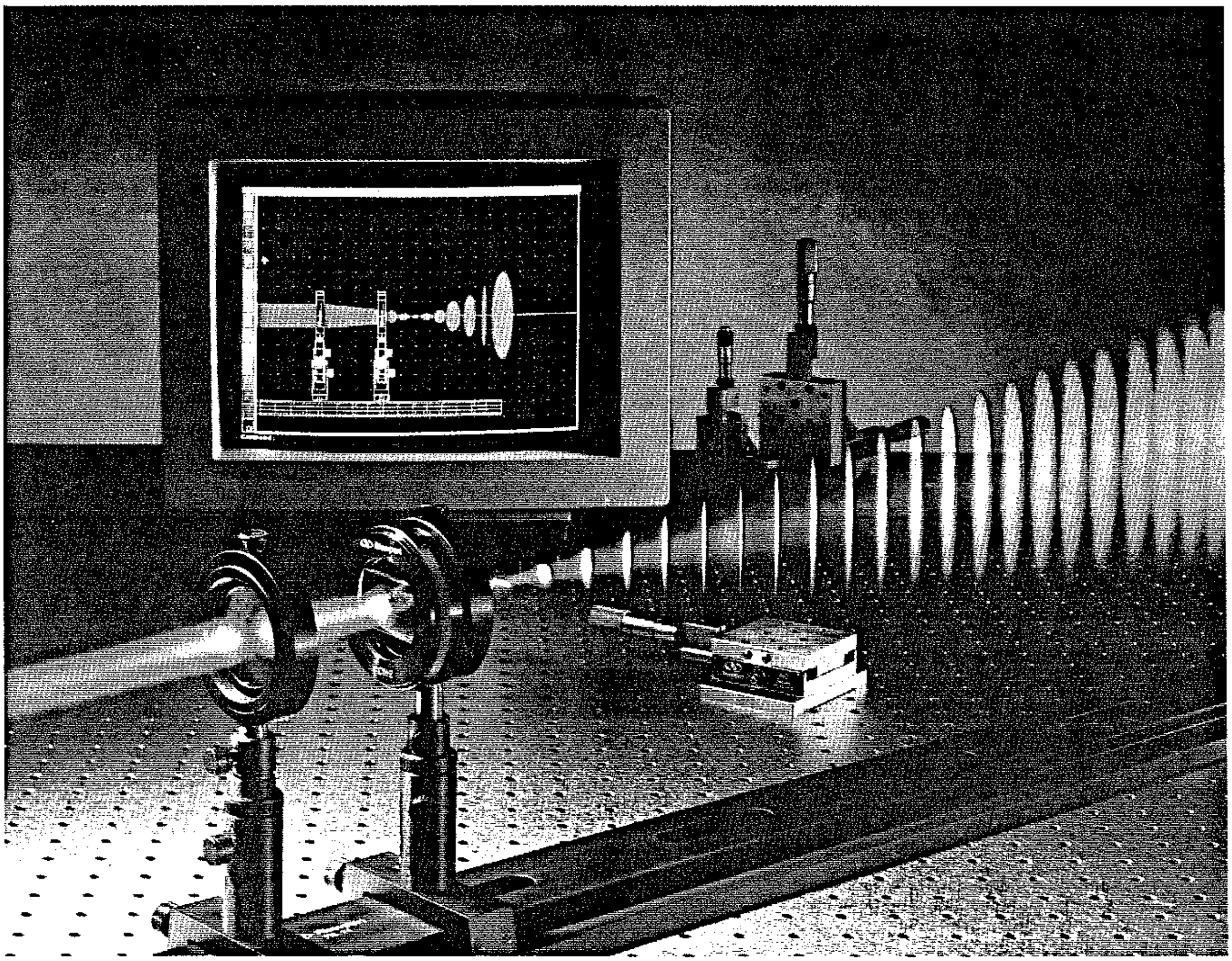
NEWPORT

CATALOG



PRECISION LASER & OPTICS PRODUCTS

RESEARCH • INDUSTRY • EDUCATION

Newport

About this tutorial

This tutorial is intended for the user who, although not formally trained in optics, must make use of optical components to solve a problem. The text provides an elementary review of classical and modern optics emphasizing physical insight over mathematical derivations. The intent is to acquaint the reader with practical applications of commonly available optical components. Due to the constant advance of optical science, no catalog can aspire to be the complete textbook of optics technology. For this reason, Newport refers the interested reader to the following texts:

Optics, Eugene Hecht and Alfred Zajac, Addison-Wesley Publishing Co., 1979.

Elements of Modern Optical Design, Donald C. O'Shea, John Wiley & Sons Inc., 1985.

Concepts of Classical Optics, J. Strong, Freeman, 1958.

The following material was written expressly for Newport Corporation by **Dr. William Bridges**, Carl F Braun Professor of Engineering at the **California Institute of Technology**. Dr. Bridges is an OSA fellow and has made numerous contributions to the fields of lasers and electro-optics. His aid in preparing this section is gratefully acknowledged.

Many of the illustrations used in this tutorial were adapted with permission from the book *Optics*, Eugene Hecht and Alfred Zajac, Addison-Wesley Publishing Co., 1979.

The velocity of light, c, in a vacuum is about 3×10^8 meters per second. In other media - for example, glass - the velocity is less. The ratio of **c** to the actual velocity is called the **refractive index, *n***. Thus, in glass with index 1.5 light travels at about 2×10^8 meters per second (fig. 1).

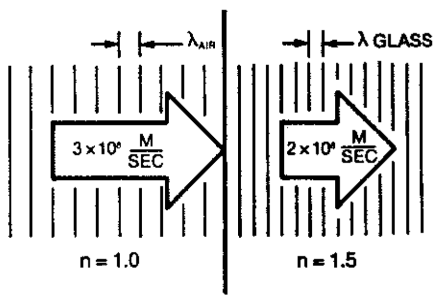


Figure 1.

The color of the light and its frequency in oscillations per second, or **Hertz**, is the same in both media. For example, the frequency of red helium-neon laser light is about 474×10^{12} oscillations per second or 474 Terahertz. Since the frequency is the same in both glass and air, the wavelength must decrease by the same ratio as the velocity, as indicated in the figure. Thus, the helium-neon laser wavelength of 633 nanometers is reduced to 422 nanometers in glass. Typically, glasses and transparent plastics have refractive indices ranging from 1.45 to 2.

We may gain a simple physical picture of this "slowing down" of light in a transparent medium if we picture a medium composed of individual atoms or molecules that can interact with the passing light wave by absorbing and reemitting the light. This absorbed and reemitted light is added to the component passing through at *c* in such a way that the sum is continually phase retarded with respect to the component at *c*. This continuous phase retardation is equivalent to a phase velocity less than *c*.

Snell's Law describes the change in the direction of propagation that occurs when light crosses an interface between two different materials at an angle other than 90° to the interface. The wavefronts must match at the interface, so the direction of the wave in the medium with a higher refractive index must turn toward the normal (fig. 2).

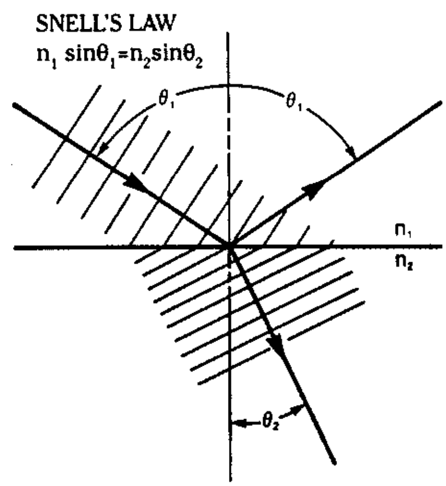


Figure 2.

Dispersion is the term characterizing materials whose refractive index changes with wavelength. All glasses exhibit dispersion to one degree or another, and this property can be both troublesome and helpful in optical systems. Dispersion allows us to separate wavelength components in a prism spectrometer, for example, but it can be harmful if it separates those components around a white-light image in a telescope with simple glass lenses.

We have pictured the physical origin of the refractive index in terms of the interaction of light with an array of atoms or molecules in a material such as glass. Dispersion occurs because those interactions are wavelength dependent. The atoms or molecules are characterized by their resonant frequencies, where they interact most strongly with light. At these resonances, light is strongly absorbed, and the materials are usually quite opaque. Typically, the strong resonances in the glass are in the vacuum ultraviolet region, corresponding to wavelengths near 100 nm. At longer wavelengths, in the visible region, the interactions are typically much weaker, in the "tails" of the strong resonances, and the materials become transparent. However, the interaction with these resonance tails and the phase shifts introduced therefrom increase as the wavelength decreases. Thus, such transparent materials have a higher refractive index at shorter wavelengths (fig. 3).

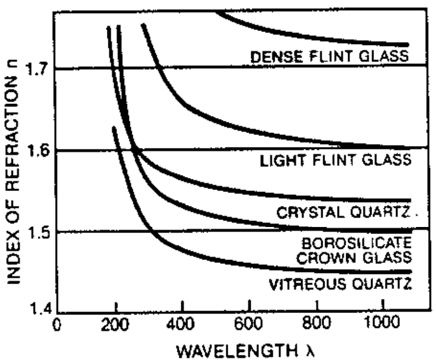


Figure 3.

Normal Dispersion is the term applied to the increase in refractive index with decreasing wavelength. The adjective "normal" is a legacy from the past, when it was used to distinguish this situation from "anomalous dispersion", that is, a decrease in index with a decrease in wavelength. Today, we understand that there is nothing anomalous about "anomalous dispersion"; it is simply another portion of the resonance curve. A more complete discussion of the physical origins of refractive index and dispersion is given in Chapter 3 of Hecht and Zajac.

A Prism is perhaps the most well known optical component exhibiting dispersion in optics. The ability of a prism to disperse the spectrum was already known in the late 1600's. To find the deflection of a light beam by a prism we simply apply Snell's law at each surface, being careful to use the correct refractive index for each wavelength.

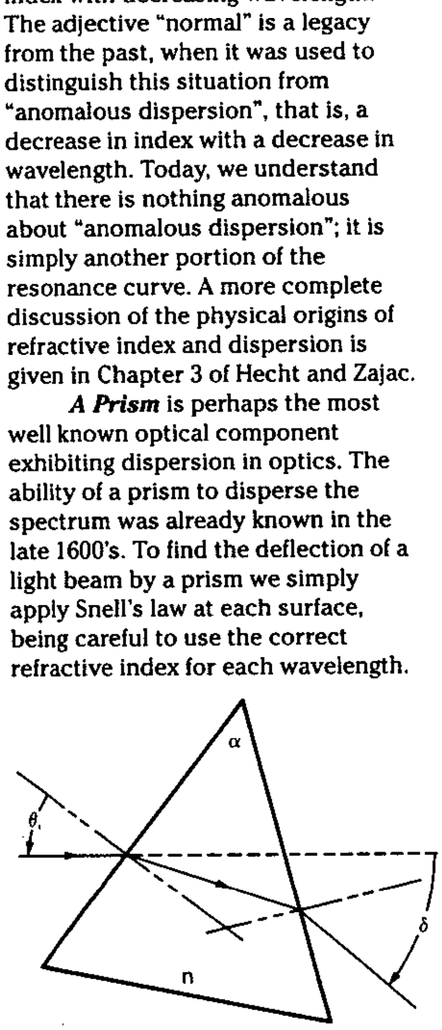


Figure 4

The resulting deviation d is given by (fig. 4):

$$\delta = \theta_i + \sin^{-1}\left[(\sin\alpha)(n^2 - \sin^2\theta_i)^{1/2} - \sin\theta_i\cos\alpha\right] - \alpha$$

Note that d is a minimum when the light path in the prism is perpendicular to the bisector of the prism angle, that is, when the incident and exit angles are equal, as shown in figure 4. Minimum deviation is commonly used when a prism is employed to disperse light in a spectrometer. This condition also produces a minimum spatial distortion of the exit beam: If the incident beam is circular in cross section, the exit beam will also be circular at minimum deviation. Tilting the prism off minimum deviation spreads out the spectrum even more, but also broadens a circular beam into an elliptical one so that the separation of the wavelengths is actually no better.

Reflection of light from glass surfaces can also be useful or harmful in optical systems. If light strikes an interface between two different materials, a fraction R will be reflected and a fraction T will be transmitted. If the materials on both sides of the interface are lossless, then $R + T = 1$. The values of R and T are functions of the refractive indices of the two materials and the angle and polarization with which the light strikes. These fractions may be calculated by properly matching the electric and magnetic fields of the light waves at the interface. The resulting relationships are called **Fresnel's equations**. For a derivation of these equations, see Chapter 4 of Hecht and Zajac.

The plane of incidence is defined as the plane containing the direction of the incident light and the normal to the surface. We take the direction of polarization as the direction of the electric field. We then distinguish two cases: (1) the incident light is polarized in the plane of incidence, usually denoted by subscripts "parallel", "p" or "||", and (2) the incident light is polarized perpendicular to the plane of incidence, usually denoted by subscripts "perpendicular", "s" (from the German "senkrecht") or "⊥". The reference guide shows R_p and R_s as functions of the angle of incidence for a glass with $n = 1.5$.

As shown on page N-26, when light is incident from the air side of an air-glass interface some notable features are seen:

1 - The reflectivity R is the same at normal incidence for both polarizations, as it should be, since the plane of incidence is undefined if the direction of propagation is the same as the normal to the surface. The value of the reflectivity for normal incidence is

$$R = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2}$$

or about 4% for $n_1 = 1.0$ (air) and $n_2 = 1.5$ (glass).

2 - The reflectivity R is unity for both polarizations at grazing incidence $\theta = 90°$, hence the mirror-like quality of any glass surface, independent of the index of refraction.

3 - The reflectivity of light polarized parallel to the plane of incidence is exactly zero at a particular angle, called **Brewster's angle** or the **polarizing angle**:

$$\theta_B = \arctan\left(\frac{n_2}{n_1}\right)$$

The former name honors the discoverer of this important phenomenon, and the latter refers to an application: If unpolarized light is incident on a glass surface at Brewster's angle, the light reflected is polarized completely perpendicular to the plane of incidence. Of course, according to the curves only about 15% of the perpendicularly-polarized light is reflected at Brewster's angle, so this may not be the best way to produce polarized light. But even this small amount is the reason that polarizing filters can be used to reduce the glare from photographs of automobile windshields and water; The reflected light is partially polarized by reflection near Brewster's angle. Aside from producing polarized light in nature and the laboratory, the Brewster's angle phenomenon is used for optical windows where even the small 4% normal-incidence loss would be prohibitively large, for example the vacuum-sealing windows inside a gas laser resonator. With a carefully oriented Brewster's angle window, the residual losses which are due to scattering, absorption and bulk inclusions, can be made smaller than 0.1% with care. Of course, if your application for a low-loss window requires other than linear polarization, then Brewster's angle windows are not appropriate; Normal windows with good antireflection coatings are the usual solution.

The curves shown in the Reference Guide (pg. N-26) give the fraction of the light **intensity** that is reflected. The fraction of the **electric field** reflected depends upon the indices of the two media and the relative projected area of incidence. Further, electric field reflection coefficients will determine phase differences between incident and reflected waves. Full equations are given in Chapter 4 of Hecht and Zajac. Thus $R + T = 1$ is an expression of the conservations of energy, but $r+t$ is not equal to one, where the lower case letters stand for the field amplitude reflection and transmission coefficients.

Also on page M-22 are reflectance curves for the same glass-air interface, but with the light incident from the glass side. Brewster's angle is again evident in the figure, but at about 34° instead of 56°, in order to satisfy Snell's law across the boundary. Note that 34°+56°=90°. This provides a handy way to remember Brewster's angle: The angle between transmitted P-pol and reflected S-pol beams is 90°. The figure also shows the same reflectivity for both polarizations at normal incidence, ≈4%, the same as the air-glass interface.

The new feature is that the reflectivity goes to unity at an angle less than 90°, about 42° for glass with $n = 1.5$. This is known as the critical angle. For angles greater than the critical angle, the light is 100% reflected interior to the glass, hence the phenomenon is called **total internal reflection** or **TIR**. We can understand this phenomenon readily from Snell's law, if we note that light incident at the critical angle from the glass side corresponds to a transmitted angle of 90°:

$$n_1 \sin\theta_C = n_2 \sin(90°) = n_2$$

$$\theta_C = \arcsin\left(\frac{n_2}{n_1}\right)$$

$$\theta_C \approx 42° \quad n_2 = 1(\text{air}), n_1 = 1.5(\text{glass})$$

Angles of incidence greater than the critical angle yield no solutions to Snell's law in real numbers. The light has no place to go, so to speak. Total internal reflection is often used when reflectivity values near 100% are required. TIR is by itself lossless; The only losses are caused by scattering or absorption by defects or contamination of the dielectric surface. One disadvantage in using TIR mirrors are that the reflection takes place inside the glass medium. There may be some losses or reflections associated with getting the light into the glass in the first place. Since the critical angle for most common glasses is of the order of 42°, it is common to use 45° as the incident angle, thus turning the light through 90°. Roof prisms and cornercube reflector are thus made, using two TIR reflections. Roof prisms are also used in pairs, with four TIR reflections to displace a light ray without deviation, as in binoculars .

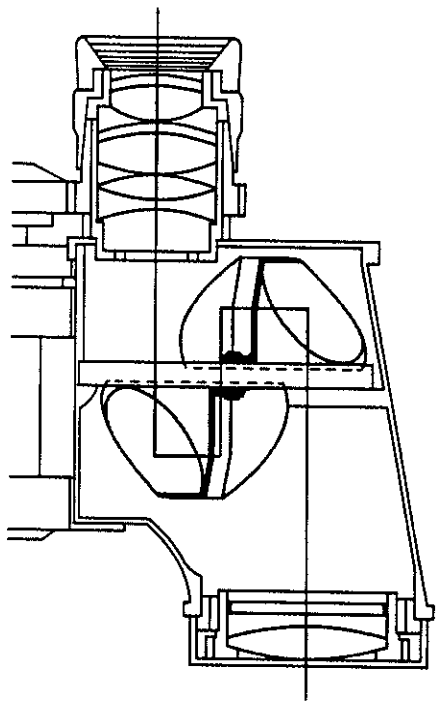


Figure 5.

The small amount of light reflected from an air-glass interface may not seem important, but it can make the difference in critical applications. The Reference Guide curves show that at near normal incidence the amount of light reflected is only about 4%. However, it is quite common for an optical system to contain five or more lenses, for a total of ten or more glass-air interfaces. The light transmitted through such a sequence would be $(1-.04)^{10}$, or about 66%. This amount of lost light might be critical to your optical system in two ways: the loss itself, or the damage the missing 34% might do bouncing around from surface to surface inside! A modern multi-element camera lens, for example, would suffer greatly on both counts if nothing were done. Fortunately, it is possible to coat optical surfaces with layers of dielectric materials a fraction of wavelength thick to greatly reduce the amount of light reflected. In the simplest case of a single layer coating, the reflection from the air-coating interface is exactly canceled by the reflection from the coating-glass interface. To do this, both amplitude and phase must be correct, requiring two degrees of freedom: the thickness and the refractive index of the coating. For exact cancellation, the refractive index of the coating must be the geometric mean of the refractive indices of air and glass, or $\sqrt{n_{GLASS}}$, and the thickness must be one quarter wavelength, measured in the coating material. Such a coating can exhibit exactly zero reflection at only one wavelength, that for which the thickness is $\lambda/4$. Since we may be unable to find a material with a refractive index of exactly $\sqrt{n_{GLASS}}$ we may also be unable to obtain exactly zero reflection at any wavelength. To gain the additional degrees of freedom necessary to use practical coating materials and to tailor the wavelength response, modern antireflection coatings are usually made up of a number of layers of differing index and differing thicknesses. Three such coatings are common in optics:

A single $\lambda/4$ coating of MgF_2, with R<1.3% on glass of index 1.5.

A multi-layer broad-band coating, with R<0.5% average over a 300 nm band.

A narrow-band "V-coating" for specific laser wavelengths, with R<0.2% at the design wavelength.

To return to our example of the five element lens with ten glass-air surfaces, we would have overall transmissions of 88%, 95% and 98% respectively if all of the surfaces were coated with one of the three coatings listed above. If the above lens were designed to be a laser collimator, a 50% increase in transmitted energy would be realized by coating the lenses with a V-coat. Since laser (or any) optical energy is considerably more expensive than coating, antireflection coatings are a very cost effective method of improving performance.

Lenses

Lenses are some of the most common elements in optical systems. Refraction at a curved surface can be used to converge or diverge a beam of light, form an image or take a Fourier transform, among other things. The most obvious attributes that a lens possesses are its diameter and its focal length (the distance from the lens at which parallel rays converge to a common point); and, of course, whether the lens is positive (converges parallel rays) or negative (diverges parallel rays). It would greatly simplify optics if this were all we needed to specify when buying a lens. Alas, there is no single lens with a given diameter and focal length that fits everyone's application. The art of optical design is to specify the right lens parameters for the specific job at hand.

For example, we've already seen that the index of refraction of glass is a function of wavelength. Because of this **dispersion**, the focal length of a simple glass lens is different for different colors. This property, termed **chromatic aberration**, would be of little concern if you are using a monochromatic source like a laser in your system, but it might be of critical concern if you are trying to design a telescope that will work with white light.

Another situation is illustrated below. We usually think of an "ideal lens" as imaging faithfully the information in one plane onto another plane. But what if your application requires imaging the surface of a sphere onto a plane? The "ideal" lens wouldn't be ideal for your needs at all.

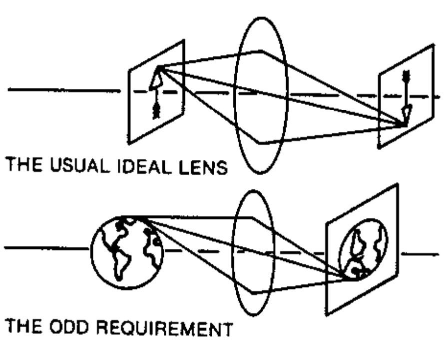


Figure 6.

To pick the right lens, you must first know exactly what you want it to do, and how well you need to do it.

Why "how well?" Can't we do the job exactly? Unfortunately, even if your application called for imaging a plane onto a plane, you will still not be able to find the "ideal" lens that will do the job perfectly. You can come as close as you wish, depending on how much money you wish to spend on the job. *But the art of optical design is in doing the job as well as needed for the minimum cost.*

The Cost of a lens is influenced by many factors. First, it is much less expensive to make a lens with spherical or plane surfaces than any other shape. The curved surfaces on a lens are usually made by grinding the initially flat glass blank against a tool with the desired radius of curvature. Many lenses are usually ground at the same time by a single tool for the sake of economy. It is easy to see that only spherical surfaces can slide freely over one another in such a situation. This simple physical fact makes accurate spherical surfaces the natural product of the grinding process.

It is possible to grind other shapes, called **aspheric surfaces**, by a variety of special techniques, but invariably these must be ground one at a time, greatly increasing the cost of the lens. And since aspheric surfaces may possess a unique symmetry axis, there is an increased cost associated with aligning the axes of the two sides of an aspheric lens if both the sides are aspheric. With two spherical surfaces, there is no such critical alignment.

Material selection is also an important factor in lens cost. Glass comes in a wide variety of grades and optical parameters. Among the important parameters is the wavelength range of low-loss transmission. Glasses which transmit well in the ultraviolet or infrared portion of the electromagnetic spectrum generally are more expensive than those which transmit only in the visible portion. Pure fused silica is often used for UV and IR applications, since it has the widest wavelength transmission range of all silica-based glasses. Even here, however, there are different grades of fused silica in which certain impurities are specifically controlled to reduce absorption at specific wavelengths. If your optical system must operate near one of these absorption regions, you may have to select lenses made from ultra-pure material, and this requirement will increase the cost of the lenses.

Optical Homogeneity of the glass required for your lens will also affect the cost. Glasses are specified by their **striae** specification, a measure of how uniform the refractive index is throughout the glass. Striae manifest themselves as streaks of slightly differing refractive index, frozen in the glass as it cooled. They may or may not be important in your optical system. Optical glasses such as BK-7 can be made striae-free by annealing. It is more difficult to make striae-free pure fused silica because of its much higher melting temperature, so premium fused silica lenses are typically more expensive than their BK-7 counterparts.

Cosmetic quality specifications also affect the cost of a lens. Surface quality is usually specified as the **scratch and dig** number - such as 20-10 or 10-5 - which represents the visibility of scratches and digs (short pit-like defects) over the aperture of the optic. Highly polished surfaces are created by successively decreasing the coarseness of the grinding compound grains as the surface forms. More changes of abrasive and longer polishing times are required to achieve finer surfaces. Close tolerances on surface finish result in higher lens cost not only from the longer polishing time, but also from the inspection process necessary to certify the specification.

The usual reason for specifying good surface finish is to *reduce scattered light* in the optical system. You should evaluate how your system is affected by scattered light before specifying a scratch and dig number. For example, scratches and digs are large in comparison to an unexpanded laser beam. Optics for such applications usually require the best surfaces obtainable to reduce scattering losses. At the other extreme, lenses for a recollimating telescope may well be quite tolerant of scattered light, and the least expensive standard specification may suffice.

The size of a lens also greatly affects its cost. There is a "cheapest size," on the order of one inch in diameter for ordinary glass lenses, where fabrication costs roughly equal material costs. Lenses larger than this are more costly since they require more material, and fewer can be manufactured at a time. Lenses smaller than the most economical size also become expensive, because they are harder to handle, and dimensional tolerances are harder to maintain.

Tolerances on dimensions directly affect the cost of a lens. Diameters can usually be ground quite precisely, within a few thousandths of an inch with commonly available equipment. However, tight tolerances on focal length require custom tooling to generate the correct radius of curvature and constant testing of the finished lenses against master test plates. A large quantity of lenses must be manufactured before the custom tooling and test plate cost is offset. *It is less expensive to use stock values of focal length and build adjustment into your optical system than it is to order a custom lens.* As we will see later, it may be advantageous to use two inexpensive stock lenses to do the job of one expensive custom lens, and gain some additional degrees of freedom to deal with aberrations in the bargain.

Shorter focal length lenses are also generally more expensive than longer focal length lenses, although a better measure is the ratio of focal length to diameter, termed the **f-number** and often written **f/**. For example, a 50 mm diameter lens with a focal length of 200 mm would be described as an f/4 lens. Generally, the smaller the f/, the more costly the lens. Note that the f/ of a lens may be different depending on its use. If you use only the central 10 mm of the above lens, for example, then you are using it at f/20, not the f/4 that it is capable of at its full diameter.

To restate it: *The art of optical design is to find the lowest-cost lens or combination of lenses that will meet all of the system requirements.* We hope the discussion and examples given in the following sections will help in your selection of stock lenses that will do the job. However, as noted in the following sections, your requirements may be sufficiently demanding that you will require custom lenses and the aid of a lens design professional. If that is the case, it would still be a good idea to try the design first yourself, with stock lenses, so that you can better discuss your requirements with the lens-maker or design professional.

Some Lens Basics

Consider the convex lens shown in figure 7. If we illuminate from the right with a collimated beam of light parallel to the optical axis, the beam will come to a focus a distance *f* to the left of the lens.

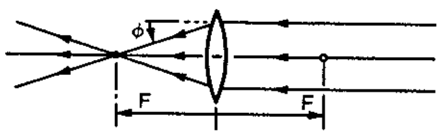


Figure 7.

For the ideal lens, all rays are turned through the correct angle ø upon passing through the lens so that they cross the optical axis at the same point *F.*

Of course, such an ideal convergence of all rays to a single point as shown does not occur exactly for a lens with spherical surfaces. The departures from this idealized picture are discussed later as **aberrations**. For now, however, we observe that if the angles ø are all so small that we may reasonably approximate sin ø by ø, then the lens behavior will be ideal as shown. This mathematical approximation is termed the **paraxial approximation** and gives rise to *Gaussian optics* or *First-order theory.* Rays which remain sufficiently near the optical axis to satisfy the approximation are termed **paraxial rays.** Snell's law in this same approximation becomes

$$n_1\phi_1 = n_2\phi_2$$

The figure shown above also usually raises the question: "From what point on the lens do I measure *F*?" In the thin lens approximation, a single plane can be used to define the position of the lens along the optical axis. Later we shall explore the more general case of the thick lens and then see under what conditions we are allowed to make this thin lens approximation.

If, instead of the situation illustrated in figure 7, we were to place a tiny point source of light at a distance *F* to the left of the lens (fig. 8), we would produce a collimated beam of light emerging to the right. Note that the rays are the same.

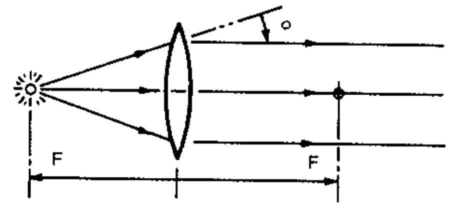


Figure 8.

The outermost ray is turned through the same angle ø as in the previous figure. Light rays are reversible in their transit through an optical system.

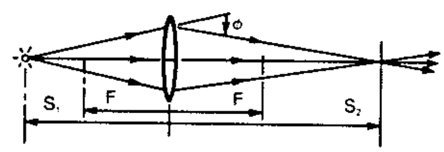


Figure 9.

If we were to move the point source of light to the left to a point s_1, farther from the lens than *F* as shown, then the outermost ray emerging from the lens would no longer be parallel to the axis, but instead cross the axis at a finite distance s_2 to the right of the lens (fig. 9). The same would be true of all the other rays. The relationship of these two points and the focal length of the lens is given by the famous **Gaussian Lens Equation:**

$$\frac{1}{s_1}+\frac{1}{s_2}=\frac{1}{F}$$

Finally, if we move the point source at s_1 a very long distance to the left, or infinity (fig. 10), then the light will arrive at the lens in parallel rays, and will come to a focus at a distance *F* to the right of the lens. For the thin lens approximation the distance *F* is the same in both figures 8 and 10 measured from the position of the thin lens.

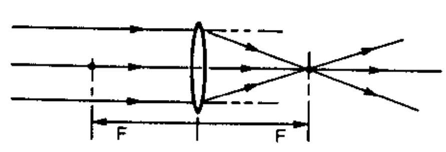


Figure 10.

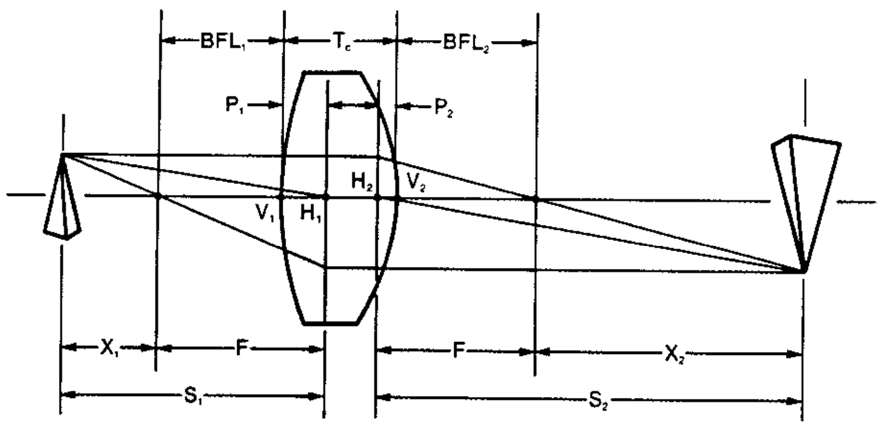


Figure 11.

When the lens is "too thick" to meet this approximation (or too curved, or is a compound lens made up of two or more thin lenses), we may still retain the concept of symmetrical focal distances *F*, but we now must measure them from two different positions in the lens, called the principal planes, H_1 and H_2 (fig. 11).

The principle planes are shown in figure 11 as falling inside the lens at distances P_1 and P_2 from the vertices, the points V_1 and V_2 where the optical axis intersects the two lens surfaces. The focal lengths *BFL1* and *BFL2* measured from the lens surfaces are not in general equal, while the effective focal length *F* measured from the principal planes is the same on both sides of the lens. The Gaussian lens equation relating the object and image distances, s_1 and s_1, to the focal length *F* holds as well for the thick lens.

The principal planes need not fall inside the lens itself. For some lenses one or even both of the principal planes may fall outside the lens as shown in figure 12.

Imaging is perhaps the most common use of a lens. In the figures above, we have shown lenses bringing all the rays from a point light source on the optical axis to a point focus, also on the optical axis. If we were to have an extended luminous object such as the light bulb shown in figure 13, then rays from each point on the object would be brought to focus so as to form an image of the object on the other side of the lens.

Our previous description of an ideal lens implied that it would faithfully transfer points in an object plane onto an image plane. However, figure 13 correctly shows that a lens actually images a three-dimensional object into a three-dimensional image. However, portions of the object that the lens cannot "see" cannot be part of the image. The figure also indicates several other things about the image-forming process with a *positive* lens.

— The object and image are on opposite sides of the lens. For this situation we say the image is "real." That is, an observer on the side of the lens opposite the object can reach out and "touch" the image without hitting the lens.

— The image is inverted. Though not obvious from the picture, the image is also inverted right to left.

— The image is *not* inverted front to back. That is, points on the front of the object as seen by an observer looking through the lens are also seen as being on the front of the image.

— The image is not necessarily the same size as the object. The lens

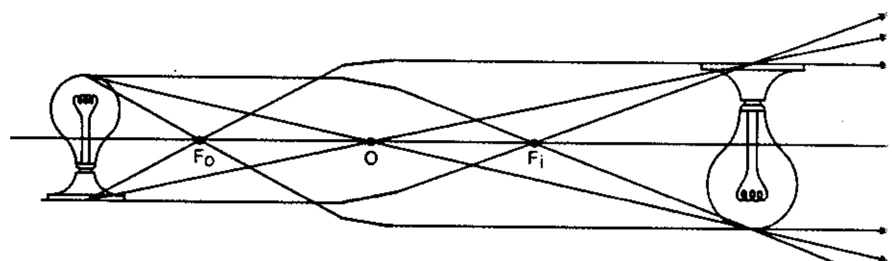


Figure 13.

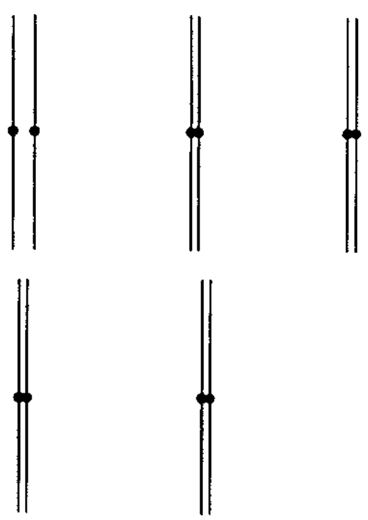


Figure 12.

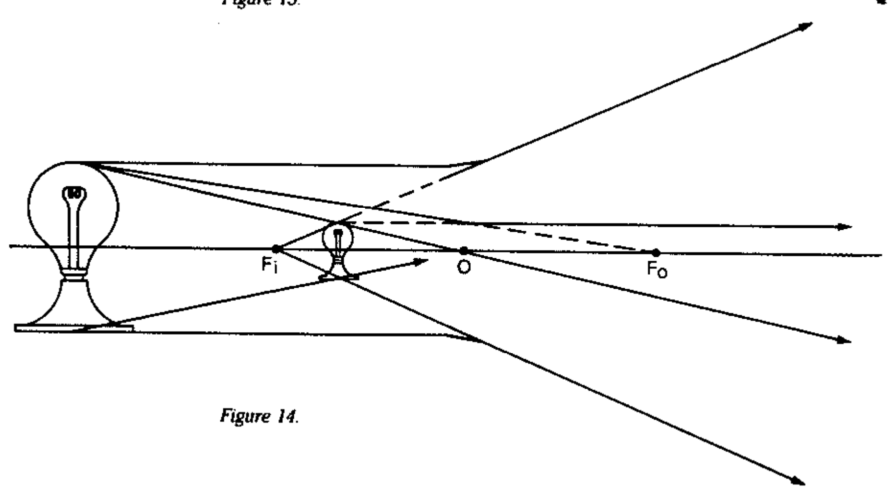


Figure 14.

can magnify or minify depending on how far the object and image are from the lens.

Only when the object and image distances are equal will the object and image sizes by equal.

We may draw similar conclusions about the image formation process for a *negative* lens (fig. 14).

The idea of a negative lens forming an "image" may seem a bit foreign the first time you think about it. By "image" we mean the locus of points where the rays from the object converge. Rays from an object point *diverge* after passing through a negative lens. If we extend these diverging rays in the reverse direction they intersect in a point, which we call an image point. With that definition of image we may then observe:

— The image is on the *same* side of the lens as the object. We say the image is "virtual" because a viewer on the side of the lens opposite the object can't get his hands on it.

— The image is erect and *not* inverted right to left.

— The image is *not* inverted front to back, identical to the positive lens.

— The image is always smaller than the object, no matter where the object is located, and is always located within the focal length, *F*, of the lens.

The concept of a virtual image is also required to fill in a gap in the description of the positive lens:

— If the object is within the focal length of the lens, the image will be erect, virtual and larger than the object. Drawing the ray diagram that shows this is left as a lunch napkin exercise for the reader.

Magnification of an image by a lens depends on where we place the object, as we've just seen. By using the Gaussian lens equation and a little geometry we can show that the magnification, M_T is

$$M_T = -\frac{s_2}{s_1}$$

— The ratio of the image and object distances to the lens. The negative sign simply reminds us that the image is inverted ("negative") if these distances are both positive. We note that this "magnification" can be greater or less than unity, depending on the ratio of the distances.

Another quantity that we should like is the **longitudinal** or **axial magnification**, M_L ,which is given by

$$M_L = -M_{T2}$$

Here the negative sign simply signifies that points on the object nearest the observer are also nearest the observer on the image; that is, the image is always orthoscopic. The important feature to note is that objects are not magnified equally in all three dimensions: If an object has a lateral magnification of two, the longitudinal magnification will be four. The light bulb image shown in the figures above should actually appear relatively fatter in the magnified image and relatively skinnier in the minified image.

While such distortion may strike you as not being very "ideal" for an ideal lens if you are a photographer, you should be thankful that nature works this way. How otherwise would you be able to focus on a distant scene with trees in the foreground and mountains in the distance? By minifying the mountains by the ratio of about 50 mm (a typical camera lens image distance) to many miles, you have also flattened the longitudinal extent of the image by the square of that ratio!

Lenses can be combined to perform a variety of tasks. The paraxial imaging properties of a sequence of thin lenses can be found by successive application of the Gaussian lens formula. For example, the image location and magnification of two positive lenses spaced by a distance *d* (fig. 15) is found by first calculating the image distance s_{i1} of the first lens, and then using this image as the object of the second lens, located at a distance of $s_{02} = d - s_{i1}$. The resulting distance from the second lens to the final image is

$$s_{i2} = \frac{\dfrac{F_2 d - F_2 s_{01} F_1}{(s_{01} - F_1)}}{d - F_2 - \dfrac{s_{01} F_1}{(s_{01} - F_1)}}$$

where F_1 and F_2 are the focal lengths of the two thin lenses.

The second lens can be thought of as simply magnifying the image from the first lens, which in turn is a magnified image of the object. The overall transverse magnification is then simply the product of these two magnifications:

$$M_T = M_{T1} M_{T2}$$

The actual size of the image and whether it is inverted or erect depends on the magnitudes and signs of both magnifications. In the situation illustrated in the figure, both M_{T1} and M_{T2} are negative, so M_T is positive and the final image is erect. The figure also has $|M_{T1}|<1$, $|M_{T2}|>1$, and $|M_T|>1$. If the same lenses were spaced by a distance *d* less than the sum of their focal lengths, then the overall magnification would be negative and the resulting image would be inverted.

The simple relation of transverse magnification to the object and image distances for a single thin lens no longer holds; that is,

$$M_T \neq -\frac{s_i}{s_o}$$

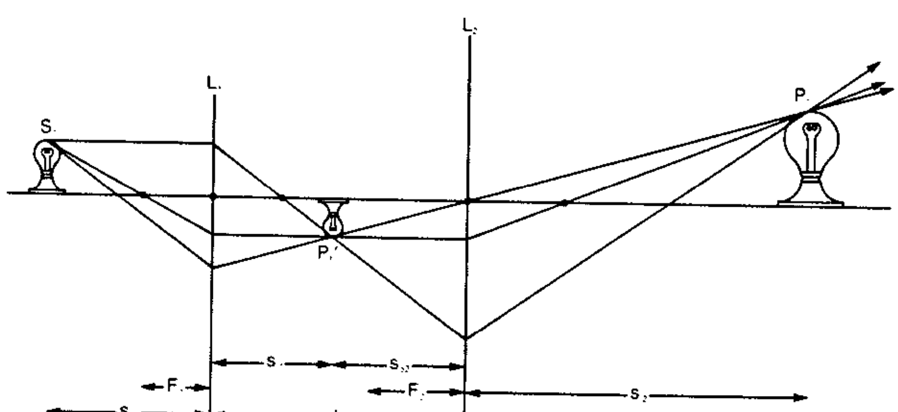


Figure 15

The reason is that the combination of two thin lenses with a space d in between is no longer a thin lens. In fact,

$$M_T = \frac{F_1 s_{12}}{d(s_{01} - F_1) - s_{01}F_1}$$

If we put the two lenses in contact, d=0, then this expression reverts to the previous thin lens formula for the magnification.

Another consequence of the thick lens nature of the two lens combination, usually termed a *spaced doublet*, is that the focal lengths of the combination are no longer the same when measured from the surfaces of the lenses. This is because the complete lens is now "thick." As in any thick lens, we can find principal planes H_1 and H_2 for the combination so that we can speak of an overall focal length F for the doublet that is the same on both sides of the lens when measured from these principal planes. The focal length F can be expressed in terms of the focal lengths of the individual thin lenses by:

$$\frac{1}{F} = \frac{1}{F_1} + \frac{1}{F_2} - \frac{d}{F_1 F_2}$$

The exact principal plane locations are given in Chapter 5 of Hecht and Zajac. However, when simple lenses are used in common laboratory mounts the above equation is adequate for initial trial designs.

A special case commonly encountered is two thin lenses spaced by the sum of their focal lengths, typically used to change the diameter of a collimated laser beam (fig. 16).

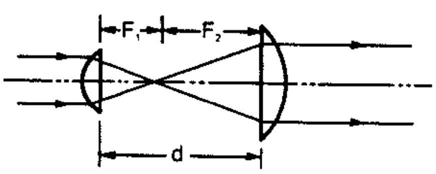


Figure 16.

We would have a bit of trouble applying the formulas given above for the spaced doublet, since the intermediate image has zero size (in the geometric optics limit), so that M_{T1}=0 and M_{T2}=∞, and their product is indeterminate. However, the paths of the rays are particularly simple, forming similar triangles on either side of the intermediate focus, so we may readily conclude that the overall transverse magnification is:

$$M_T = \frac{F_1}{F_2}$$

As a specific example, suppose we wished to expand the beam from a small helium-neon laser, with a Gaussian beam diameter of 0.63 mm to a diameter of 10 mm. We would then need to select a pair of lenses with a ratio F_1/F_2=16. A good pair of simple plano-convex lenses would be:

Newport Part #	F	Diameter
KPX019	25.4 mm	6.35 mm
KPX115	400 mm	25.4 mm

The lenses would be spaced by the sum of their focal lengths, 425.4 mm. We will want to orient the lenses with their convex sides outward to minimize spherical aberration, as discussed later. The diameters of these lenses are large enough to accommodate the laser beam as it enters and exits.

In this example we must be careful how we measure the lens separation when we actually construct the doublet. You will note in the catalog that the lenses given in these examples have slightly different values given for front and back focal distances, (listed as **BFL1** and **BFL2** in the table). This tells us that the lenses are not completely "thin." We will need to measure the effective focal length, F, from the *principal planes* of these lenses. The columns *P1* and *P2* in the catalog listing give the location of the principal planes with respect to the physical surfaces of the lens. The principal plane for the planar surface of the **KPX019** lies 2.23 mm inside the lens, measured from that surface. (The negative sign means inside the surface.) For the **KPX115** the principal plane lies 2.23 mm inside the planar surface. Thus, we would properly space the two planar surfaces by the sum of the two values of **BFL2** given for these lenses:

23.16 mm + 397.76 mm = 420.92 mm

You should also note that these distances are all measured on the *optical axis*. Lenses are usually mounted at their peripheries, so the offset of the periphery must be taken into account if you are designing a fabricated mount. If you are putting together a laboratory setup with individual lens holders then you will have enough freedom to compensate for this small difference.

Ray tracing is an alternative method of evaluating the performance of a lens or a system of lenses. Consider a sequence of optical surfaces separating regions of differing refractive index n_j by axial spacings t_j as shown in fig.17.

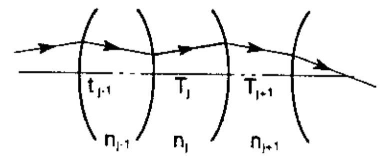


Figure 17

The surfaces may be spherical or aspherical, but their shapes must be known and described by given equations or numerical data. We may determine the path of any ray through this optical system with only two kinds of calculations: how a ray propagates between surfaces and how it changes passing through a surface. By repeating these two calculations, we can work our way through the system. We can do this with pencil and paper or an electronic computer; either method is called "ray tracing."

In free space, rays travel in straight lines. We can write the equations describing the transit between surfaces in "input-output" form as shown in fig. 17.1:

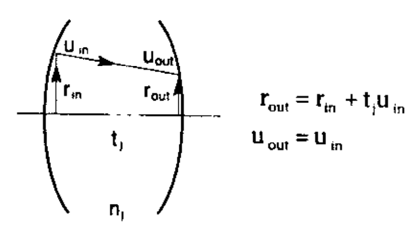
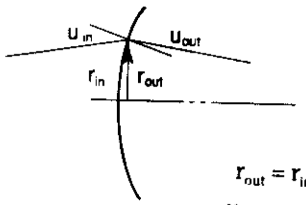


Figure 17.1

To be precise, we should use the actual separation of the two points of intersection rather than the axial spacing of the surfaces, t_j, but usually the error is small.

The calculation describing how rays change passing through a surface is simply the application of Snell's law to the particular surface. This is written schematically in fig 17.2:



$$r_{out} = r_{in} + t_j u_{in}$$
$$u_{out} = u_{in}$$

Figure 17.2

The function Sn represents the calculation of Snells law at the point the ray pierces the surface; it will depend on the radius and slope of the input ray, the relative refractive indices, and the shape of the surface.

To find the overall object and image distances for a complicated sequence of lens surfaces by ray tracing, simply originate a ray on axis (r=0) at the position of the object and trace it through the system until it exits. The image will be located where this exiting ray next crosses the axis. If the ray also crosses the axis inside the lens sequence, you will have found the intermediate images as well. Any slope u_{in} may be used for the calculation, but different slopes will give slightly different positions for the image; this is one measure of the aberration of your lens system.

We can also find the transverse magnification of the lens system by choosing an off-axis point on the object (finite r) and tracing it through the system until it reaches the image position (determined previously.) The transverse magnification is then the radius at the image divided by the initially chosen radius.

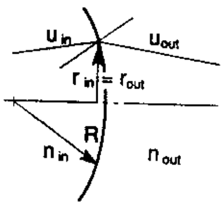
For simple spherical surfaces and paraxial rays, the input-output equations for a surface transit can be greatly simplified, and the sequential computation for ray tracing can be reduced to the multiplication of 2x2 matrices, the so-called "ABCD" matrices from radio-frequency transmission line theory:

$$\begin{pmatrix} r_{out} \\ u_{out} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_{in} \\ u_{in} \end{pmatrix}$$

Figure 17.3 gives the ABCD matrix for a spherical surface of curvature radius R.

The equations describing propagation in free space between surfaces is already in ABCD form:

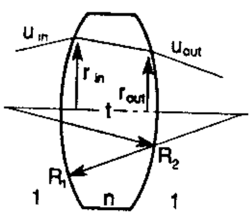
$$\begin{pmatrix} r_{out} \\ u_{out} \end{pmatrix} = \begin{pmatrix} 1 & t_j \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_{in} \\ u_{in} \end{pmatrix}$$



$$\begin{pmatrix} r_{out} \\ u_{out} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \left(\frac{1}{R}\right)\left(1 - \frac{n_{in}}{n_{out}}\right) & \left(\frac{n_{in}}{n_{out}}\right) \end{pmatrix} \begin{pmatrix} r_{in} \\ u_{in} \end{pmatrix}$$

Figure 17.3

We can do many interesting calculations with these simple 2x2 matricies. For example, we can combine the ABCD matricies for two spherical surfaces separated by a thickness t to find the ABCD matrix for a thick lens:



$$A = 1 - \frac{t(n-1)}{R_1 n}$$

$$B = \frac{t}{n}$$

$$C = -(n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + \frac{t(n-1)^2}{R_1 R_2 n}$$

$$D = 1 - \frac{t(n-1)}{R_2 n}$$

Figure 17.4

It is handy to note that all ABCD matricies, for individual elements or a complicated optical system, have the property:

$$AD - BC = n_{in}/n_{out}$$

This relation can be used as a check on your calculation after multiplying together a long chain of 2x2 matricies.

The simple recursive nature of the calculation is ideal for solution by available personal computer spreadsheet programs. For more information on ray tracing and ABCD matricies and their applications, see the texts by Hecht and Zajac, or O'Shea, or Chapter 15 of Lasers by A. E. Siegman, University Science Books, 1986.

In our discussion of lenses, we have said little about the diameters required, yet this is an important feature of the lens in determining its cost. In the example of the laser beam expanding doublet it was clear that the lenses had to be large enough to pass the desired input and output beam diameters. In other optical systems, the required diameters must be determined. The most general statement that can be made is that we need to keep the lens diameters large enough so that no desired light is lost by falling outside a lens as it traverses the system. The study of the light lost in a system is that of *stops, pupils,* and *windows.*

Stops come in two kinds: *aperture* stops and *field* stops. Aperture stops limit the amount of light that can pass through the lens system, while field stops limit the angular field of view that the system can transmit. ***Pupils*** are the image of aperture stops; an entrance pupil is the image of the aperture stop as it would be seen in object space (that is, the space on the object side of the first lens), while the exit pupil is the image of the aperture stop as seen from image space (that is, the space on the image side of the last lens of the system.) Likewise, ***entrance*** and ***exit windows*** are the images of the field stop in object and image space respectively.

Stops may be the lenses themselves, or intervening apertures placed intentionally or unintentionally between the lenses. A familiar example of an intentional aperture stop is the iris diaphragm in a camera used to change the amount of light allowed to expose the film. Cameras also contain field stops: the rectangular frame that the film rests on limits the field of view to just that which will fit on the film. Notice that this field stop keeps the sprocket hole part of 35 mm film from being exposed by the lens system.

In a simple combination of lenses, it may be obvious which diameter limits the amount of light and which limits the field of view. In more complicated lens systems it is necessary to trace rays to determine which dimensions are the critical elements. A simple procedure to determine the aperture stop is to trace rays that pass through the axis at the object plane, or *axial rays*, at ever increasing slope, until you find a ray that intersects the edge of a lens or aperture. That ray is the *marginal* *ray* of the system, and the object it hits is the aperture stop. Once the location and diameter of the aperture stop is determined, the exit and entrance pupils can be determined by tracing two more rays through the system, to find the images of the aperture stop from the two sides of the system.

Similarly, the field stop may be found once the aperture stop is located by tracing rays that cross the axis at the position of the aperture stop. Increase the slope of this ray until it intersects the edge of a lens or aperture. The object it hits is the field stop. The exit and entrance windows are found by tracing two more rays to find the images of the field stop in object and image space. For examples of this procedure, see Hecht and Zajac, or O'Shea.

Aberration

Aberration is the name given to various departures from the ideal performance of a lens. As we've seen above, a lens with spherical surfaces exhibits near-ideal behavior if we confine the rays through the lens to be very near the optical axis so that they make very shallow angles with respect to that axis. If we use such a lens with larger ray angles, we will find that the lens no longer behaves in an ideal fashion. For example, *marginal rays* that pass through the region near the edge of the lens will not cross the axis at the same distance from the lens as the central or paraxial rays cross.

We would like to have a theory that describes this nonideal behavior, so that we can understand it and learn how to minimize the "damage" done to the performance of our optical system. The simplest analytical theory that allows us to evaluate these departures from ideal imaging is *Third-order theory*, developed by Ludwig von Seidel in the nineteenth century. Seidel made the approximation $\sin\phi = \phi - \phi^3/3!$ for the angles used in tracing rays through a lens. Using this approximation he found that the rays do not exactly map one plane onto another to produce a faithful image of an object. He also found that he could separate the resulting departures from the ideal into five different aberrations. These are now called the Seidel aberrations in his honor.

Third-order calculations do not tell the whole story. More complete approximations to $\sin\phi$ have been used to calculate fifth-order and even seventh-order lens theories. The complexity increases astronomically with the order of the approximation. Unfortunately, the insight gained by these theories decreases at the same astronomical rate.

Modern optical design has abandoned high order approximations in favor of ray tracing, calculating $\sin\phi$ exactly with high-speed electronic computers. This can be done for a single lens, but more often the calculation is made for an entire optical system. Ray tracing procedures have resulted in very powerful lens design programs for mainframe and personal computers.

Although lens design programs can rapidly evaluate and refine lens systems, they cannot (as of this writing) independently determine a suitable starting design. A good design begins with insight and the intelligent use of trial and error. To minimize the number of trials, a lens-design professional will want to start with a reasonable first guess derived from a good understanding of the Seidel aberrations. At a more elementary level, it is helpful to understand the implications of the simplest aberrations so that you can optimize your system performance using stock spherical lenses. In many cases, a reasonable first guess is good enough to meet the system requirements.

Seidel Aberrations are usually discussed in this order: spherical aberration, coma, astigmatism, field curvature, and distortion. They must be corrected in the same order. That is, you cannot have pure coma until the spherical aberration is eliminated, and pure astigmatism requires that spherical aberration and coma both be eliminated, etc.

This order also tells how the magnitude of the particular aberration varies with the distance of the object point from the optical axis: spherical aberration is independent of this distance, while the amount of coma varies linearly with the distance off axis; astigmatism and field curvature vary quadratically with this distance and distortion varies cubically. The magnitudes of all aberrations increase as the distance between the optical axis and the ray entering the lens increases. This is as expected, since rays confined to the optical axis are ideally described by Gaussian first-order formulas. In the paragraphs below we present only brief physical descriptions of these aberrations and some of the techniques used to minimize their effects when designing an optical system.

Spherical Aberration is illustrated below (fig. 18). For a lens with spherical surfaces, the outer regions of the lens are too strong compared to the central region. Marginal rays are turned more than the paraxial rays and come to a focus closer to the lens.

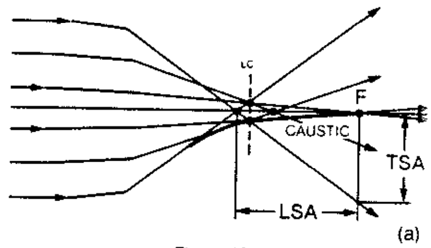


Figure 18.

The difference in distance between the paraxial focus and the marginal focus is a measure of the spherical aberration of the lens, or more properly, the *longitudinal spherical aberration, LSA*, to distinguish it from a closely related measure, the *transverse* spherical aberration, *TSA*. As shown in the figure, the TSA is the distance off axis that the marginal rays pierce the paraxial focal plane.

The envelope of rays passing through one side of the lens defines the *caustic curve*, and its intersection with the marginal rays defines the *circle of least confusion*, or the smallest spot to which a light source at infinity will be focused by such a spherical lens. We should point out that the effect is shown greatly exaggerated in figure 18. If the difference between the paraxial focus and the marginal focus were as large as shown, third order theory would not be adequate to accurately describe the situation.

Minimizing spherical aberration can be accomplished in several ways. A simple method is to use only the central portion of the lens. If the lens shown above were stopped down (larger *f/*) so that the marginal rays were the second from the axis rather than the third as shown, both LSA and TSA would be greatly reduced, and the circle of least confusion would be much smaller. Of course, you would likely do this by purchasing a smaller diameter lens in the first place. Either a smaller lens or an aperture stop reduce the aberration by the same amount.

The amount of spherical aberration exhibited by a lens also depends on how the lens is employed in the system. Figure 19 illustrates how the same plano-convex lens can exhibit different amounts of spherical aberration simply by reversing it when it is used to bring a collimated beam of light to a focus.

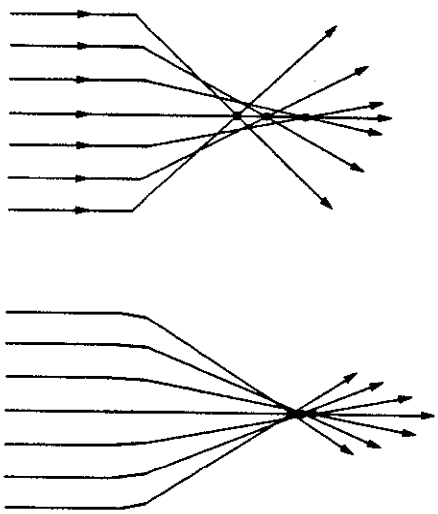


Figure 19.

Clearly, orienting the curved side of the lens toward the collimated beam reduces the spherical aberration. The same would be true if the lens were used to create a collimated beam from a point source.

There is a simple way to remember which way to orient a lens to minimize spherical aberration: Recall that spherical aberration results because the outer portions of the spherical lens are "too strong" compared to the central region. We should then orient the lens to give the *least possible bending of the outer rays.* If we recall that the minimum deviation of the ray passing through a prism occurs when the ray is deviated *equally* at both surfaces, then we can strive for this same condition in the outer region of the lens. In the case of the two orientations of the plano-convex lens shown in figure 19, orienting the flat side toward the parallel rays from infinity requires all the deviation to occur at the second surface, while reversing the orientation splits the deviation between the two surfaces, a better situation.

The amount of spherical aberration produced by a lens also depends on the ratio of the image distance to the object distance, s_2/s_1, often termed the *conjugate ratio,* since s_1 and s_2 are called the *conjugate points.* In the figure above, the conjugate ratio is zero (or infinity if object and image points are interchanged.) The lens with parameters that minimizes the spherical aberration for a particular conjugate ratio is said to be a *best form* lens. For glass with a refractive index of 1.5 at infinite conjugate ratio the best form lens is nearly plano-convex. We say nearly because the radius of curvature of the second side of the lens is six times that of the first and hence appears nearly flat. Note that multiple tools would be required to manufacture this lens, and even so, it would not completely eliminate spherical aberration. Plano-convex glass lenses are considerably less expensive to fabricate and they are so close to best form for infinite conjugate ratio that they will usually meet your system requirements for all but the most critical applications.

What do you do if you need a lens for a finite conjugate ratio? Application of our minimum deviation mnemonic tells us immediately that the best form lens for a conjugate ratio near unity (equal image and object distances) will be a ***symmetric*** lens, ***biconvex*** or ***biconcave***. For other conjugate ratios, the best form will be somewhere between symmetric biconvex and nearly plano-convex, with the most curved side toward the longer of the image and object distances. It is impractical for a lens supplier to stock all possible forms between these two extremes. You could consider ordering a custom lens designed for your particular conjugate ratio, or you could consider using two stock plano-convex lenses back to back, (fig. 20).

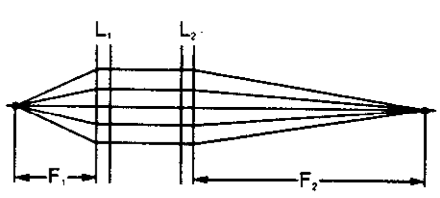


Figure 20.

Each lens operates at infinite conjugate ratio, while the pair operates at a finite conjugate ratio. Of course, you have introduced two extra surfaces, which will exhibit reflection unless they have antireflection coatings. However, recall that the art of optical design is to meet the system requirements at a minimum cost.

Coma is lens aberration that appears when light is brought to a focus at points off the optical axis. The name coma is Latin for comet, and that is the shape of the aberrated image of an off-axis point (fig. 21).

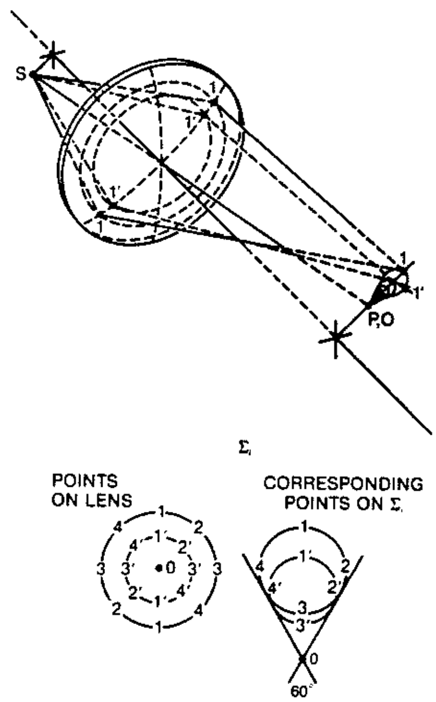


Figure 21.

Rays that pass very near the center of the lens *0* come to a paraxial focus. In the figure the object point *S* has a paraxial focus at *P.* However, rays that pass through the outer regions of the lens come to a different focus, both in position and in shape: The rays that come from *S* and pass through the lens at points *1-2-3-4* at some radius from *0* are imaged as a circle *1-2-3-4* which will be tangent to two lines in the focal plane through the point *P.* Rays that pass through the lens at a smaller radius *1'-2'-3'-4'* will be imaged in a smaller circle. Note that it is the tail of the comet that is the paraxial focal point, not the head.

Unlike spherical aberration, which always brings marginal rays to a closer focus than paraxial rays, coma may be positive or negative, depending on the shape of the lens. For negative coma, the tail points away from the optical axis and for positive coma, the tail points toward the axis. This means that for a particular lens shape, the coma can be made exactly zero.

The optimum lens shape to minimize or eliminate coma depends on the conjugate ratio of the imaging. Fortunately, the shape of the lens that produces zero coma is very nearly the same shape that produces the minimum spherical aberration. Techniques used to reduce coma are similar to those used in reducing spherical aberration. For example, a plano-convex lens with the curved side facing the collimated light nearly eliminates coma at infinite conjugate ratio. For conjugate ratios near unity, a biconvex lens is near optimal. And, at intermediate ratios, the most economical approach to reducing coma is to use two stock plano-convex lenses back to back as previously shown. It is also possible when lenses are used in combination to introduce a stop at a proper location to introduce positive or negative coma into the system to cancel out the net coma from the rest of the system. Hecht and Zajac give further details on this technique.

Astigmatism is the next aberration to be considered after spherical aberration and coma are corrected. Figure 22 illustrates this aberration.

We distinguish two planes in this system: the meridional plane, containing the optical axis and the object point, also called the tangential plane, and the sagittal plane, perpendicular to the tangential and containing the object point. Rays in the meridional plane come to a focus closer to the lens than rays in the sagittal plane. Thus, the circular cross section of a conical bundle of rays from an object point will become elliptical after passing through the lens, coming to a line focus first perpendicular to the meridional plane, then farther from the lens to a second line focus lying in the meridional plane. The nearer focus is called the primary or tangential focus and the farther focus is called the secondary or sagittal focus. Between these two foci the bundle of rays will have a minimum overall diameter, called the circle of least confusion.

The Seidel aberration astigmatism should not be confused with visual astigmatism. Visual astigmatism occurs when the eye's lens does not have spherical surfaces. The curvature of the lens surface in one meridional plane is not the same as the curvature in another. Images of orthogonal sets of lines (for example, window screen) will not come to focus at the same distance. Seidel astigmatism, on the other hand, occurs with perfectly spherical lens surfaces, and produces a different kind of image defect, as illustrated in the image of the wagon wheel (fig. 23).

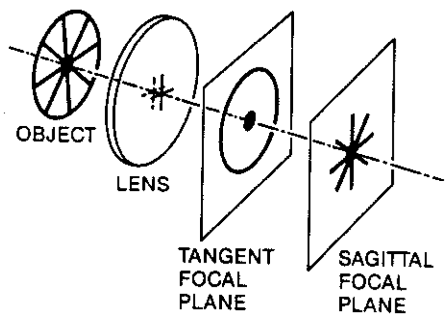


Figure 23.

Points will be blurred out only in a circular direction in the tangential focal plane; thus, circles will appear sharply focused in radius while the spokes will be blurred out. In the sagittal focal plane the blurring is only in the radial direction, so that the spokes will appear sharply focused while the rim blurs out.

Astigmatism can be reduced by judicious use of stops with simple lenses. Overall correction techniques are beyond the scope of this discussion, and fully corrected designs require the fabrication of custom multi-element lenses. If a high degree of astigmatism correction is required the best solution is to adapt commercially available lens systems (termed *anastigmats, copy lenses, process lenses*) to your application.

Field Curvature remains after spherical aberration, coma, and astigmatism are corrected. A lens corrected for all aberration except field curvature would image off-axis points in the object as off-axis points in the image. However, if the object points lie in a plane, the image points will not lie in a plane, but rather on a paraboloidal surface, termed the *Petzval surface*, after the nineteenth century mathematician, Josef Petzval. This surface curves *toward the lens for positive lenses* and *away from the lens for negative lenses*, which suggests that a combination of positive and negative lenses might cancel the effect. The *Petzval condition* does just this. If both lenses are made out of the same glass, then the magnitude of the focal lengths must be equal, so that the lenses must be spaced some finite distance *d* if the combination is to have a finite focal length. The overall focal length is then

$$F = \frac{(F_1)^2}{d}$$

This simple correction method is readily applicable with only common lenses but is limited in flexibility. Again, total correction requires specialized multi-element lenses. Fortunately, lenses corrected for astigmatism are usually corrected for field curvature.

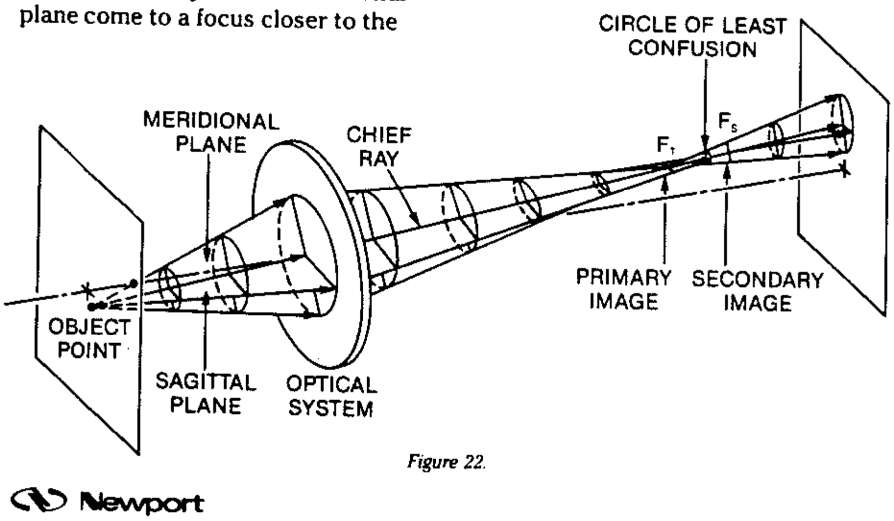


Figure 22.

Distortion is the last of the Seidel aberrations and manifests itself as a displacement of the points in the image as a whole rather than a blurring of the individual points. Positive or *pincushion* and negative or *barrel* distortion are illustrated below.

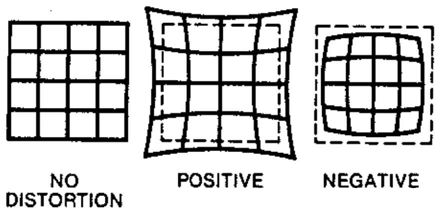


Figure 24.

Distortion may also be thought of as a change in the transverse magnification with the distance of the image point off axis. With positive distortion that magnification increases, while with negative distortion it decreases. Adding a stop to a lens system will often introduce distortion of one sign or the other. If the stop is added some distance in front of the lens, barrel distortion will result. If it is added some distance behind the lens, pincushion distortion will result. No added distortion results if the stop is placed in the plane of the lens or at the optical center of a system of lenses. The judicious introduction of a stop can be used to create distortion of the opposite sign to that of the lens and thus produce a system with no net distortion.

Chromatic aberration

Chromatic aberration, unlike the five Seidel aberrations which occur with monochromatic light, occurs when a lens or optical system must use many wavelengths or a continuum of wavelengths. Because the refractive index of optical materials varies with wavelength, the focal properties of a simple lens will vary as well. Figure 25 illustrates this.

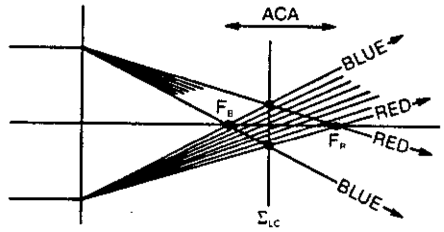


Figure 25.

Glasses exhibit normal dispersion in the visible spectrum, so the refractive index is higher for blue light than red. The ability of the lens to bend rays is thus stronger in the blue, and the focal length of a convex glass lens is shorter for blue light than for red, as shown in the figure. The distance between the red and blue foci on the optical axis is the *axial chromatic aberration*, or *ACA*. There is also a *lateral chromatic aberration*, or *LCA*, which is the lateral separation between the position of the blue and red foci of an off-axis point (fig. 26).

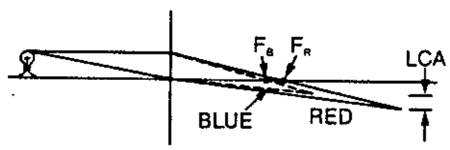


Figure 26.

A little thought will show that the positions of the red and blue foci with respect to the plane of the lens are reversed for a negative lens. This suggests that a combination of positive and negative lenses could be found to eliminate chromatic aberration. The simplest combination is called an *achromatic doublet*. To produce a doublet with finite focal length we must do one of two things: (1) put two lenses in contact, but make them of glasses with different dispersion; or (2) use two lenses made of the same glass but with a particular distance between the lenses.

The achromatic doublet illustrated below consists of two lenses of different optical glasses placed in close contact. The two lenses are shown as having a common radius of curvature (fig. 27), so that they fit intimately over their entire surface. An index-matched cement is used to eliminate the individual reflections from the two interior surfaces, and this form is called a *Fraunhofer cemented achromat*.

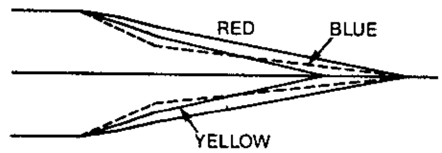


Figure 27.

By a proper choice of glasses, this doublet can be made to have the same focal lengths at two wavelengths, red and blue in the above figure. If the lens can be considered a thin lens, then the red and blue rays will cross the axis at the same point, and the lens will have zero ACA for those wavelengths. The red and blue rays will also exit parallel to each other so it will also be free from lateral chromatic aberration (LCA). However, if the lens must be treated as a thick lens, the ACA may not be zero because the positions of the principal planes may not be the same for the two wavelengths, even though the focal lengths are. The LCA will remain corrected.

Typically, rays at intermediate wavelengths (yellow) will not cross the optical axis at the same point as the red and blue rays, so that the lens will not be completely free from chromatic aberration. How much the yellow focal length differs from the red/blue focal length depends on the exact form of the dispersion curves for the two glasses. In demanding applications this small residual aberration must also be reduced. In visual optical instruments, the residual LCA produces magenta and green halos around white objects, termed the *secondary spectrum*. To reduce the secondary spectrum, more than two lenses are utilized, and the net focal length can be made the same at three or four wavelengths. Such lenses are custom designed and considerably more expensive. However, for a large percentage of applications, achromatic doublets listed in this catalog will be satisfactory.

In designing achromatic optical systems the dispersive property of glass is usually characterized by its *dispersive index*, also known as its *V-number* or *Abbe number*, defined by:

$$V = \frac{(n_{YELLOW} - 1)}{(n_{BLUE} - n_{RED})}$$

Industry standards specify exact wavelengths for red, yellow, and blue. The subscripts *C*, *d*, and *F* are used for these particular wavelengths, following Fraunhofer's historical designations for lines in the solar spectrum:

C (red) 656.3 nm
d (yellow) 589.3 nm
F (blue) 486.1 nm

The less dispersive the glass, the larger the V-number. Values of V and n_d vary from about 85 and 1.49 for fluoride crown glasses to 20 and 1.95 for dense flint glasses. Glasses typically used in simple achromats are borosilicate crown glass, e.g. BK-7 with V = 64 and n_d = 1.52; and flint glass e.g. SF-2 with V = 36 and n_d = 1.62.

The condition that a doublet be an achromat is then stated simply by:

$$F_{1d}V_1 + F_{2d}\ V_2 = 0$$

where the subscript d denotes the focal lengths of the two lenses at the yellow wavelength. The derivation of this relation is given in Chapter 6 of Hecht and Zajac. This condition specifies the ratio of the two lengths. The desired net focal length of the doublet places a second condition on the two focal lengths:

$$\frac{1}{F} = \frac{1}{F_{1d}} + \frac{1}{F_{2d}}$$

In the cemented achromatic doublet, there are three radii of curvature to be specified. The above conditions on focal lengths determine two of the radii. The remaining radius can be adjusted to provide the best form factor in order to correct for one other aberration. The achromatic doublets listed in this catalog have been carefully optimized by exact numerical ray tracing to eliminate spherical aberration at specific conjugate ratios.

It is not required that an achromatic doublet be in the cemented Fraunhofer form. Figure 28 illustrates other doublets of two different glasses.

The extra degree of freedom gained by having two different radii of curvature where the lenses are in contact can be used to make additional corrections for the Seidel aberrations. The disadvantages are that these two surfaces must be individually coated to reduce reflections, and the two lenses must be individually mounted, rather than mounted as a single unit.

We may also correct chromatic aberration using two lenses made of the same glass if we are willing to space the lenses some distance apart. As shown in Hecht and Zajac, for example, two lenses spaced a distance d given by

$$d = \frac{(F_{1d} + F_{2d})}{2}$$

will have the same focal lengths for red and blue wavelengths, provided the refractive index for yellow light is the arithmetic average of the indices for red and blue, a reasonable approximation for most glasses. Again, though the focal lengths are the same for red and blue, the location of the principal planes may not be, so such a lens is well corrected for LCA, but not ACA. The *Huygens ocular* or eyepiece is a good example of such a spaced achromatic doublet.

A final comment on achromatizing a system of lenses: Minimizing LCA requires that red and blue rays exit the system parallel to each other, and minimizing ACA requires that they cross the optical axis at the same point. This means that that the red and blue rays must actually be as collinear as possible when exiting the lens system. *When lenses are spaced some distance in a multi-element optical system, they should generally be individually achromatized, to keep the red and blue rays from separating significantly,* (fig. 29).

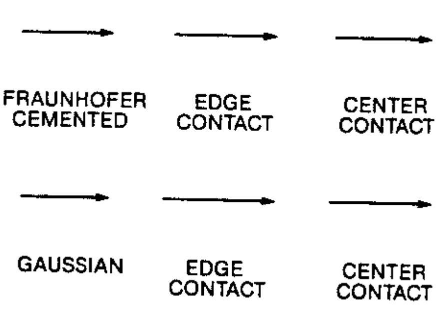


Figure 28.

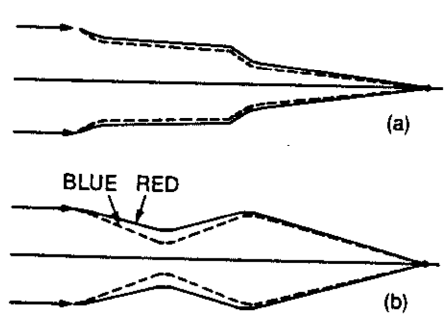


Figure 29.

If the red and blue rays have separated significantly within the system, it may prove very difficult to make them collinear again with the final lenses of the system.

Diffraction

Diffraction phenomena are a consequence of the wave nature of light. The scalar theory of diffraction can be used to relate the phase and amplitude distributions of the optical fields in planes transverse to the direction of propagation of a beam of light as it propagates from its source. Figure 30 illustrates the arrangement usually considered: The source electric field $E_s\ (y, z)$ is known in both amplitude and phase over the y-z plane at $x = 0$. The actual source of light is somewhere to the left of $x = 0$ and is illuminating an aperture in a mask located at $x = 0$, so that the source in the y-z plane is zero except in the aperture.

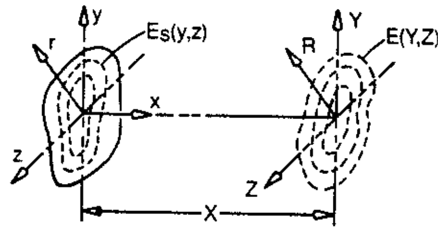


Figure 30.

We can then express the electric field $E(Y, Z)$ at a distance $x = X$ away from the source plane as an appropriate integral containing the source field, E_s. The resulting famous **Huygens-Fresnel Integral** has received much attention, both in optics and in the theory of antennas; it can be solved to varying degrees of approximation, depending on the particular variation of E_s and the dimensions of the problem. Chapter 10 of Hecht and Zajac gives a good introduction to applications of this integral in optics.

In many optical systems the Huygens-Fresnel integral can be simplified to:

$$E(Y,Z) \approx \iint_0^\infty E_S(y,z)\ e^{\frac{i2\pi Zz}{\lambda x}}\ e^{\frac{i2\pi Yy}{\lambda x}}\ dy\ dz$$

Note that this equation is in the form of a Fourier transform pair. The electric field distribution in the (Y, Z) observing plane is just the Fourier transform of the source electric field in the (y, z) plane. The fact that E and

E_s form a Fourier transform pair is certainly worth remembering, since we can use our knowledge of Fourier transforms from other branches of engineering and science to visualize how light diffracts in an optical system. However, the source and observing plane electric fields are Fourier transform pairs only if:

(1) The source aperture and its resulting field at the observing plane are very much smaller than the distance between source and observing planes.

(2) The observation plane is in the **far-field**. The **Fraunhofer condition** defines the far field as the observing distance, X, such that the maximum dimension of the source distribution is much smaller than $\sqrt{X\lambda}$.

— *Since many optical systems possess axial symmetry, it is often convenient to use the distribution of the light with the radius from the axis r or R* rather than the Cartesian variables *y, z* and *Y, Z*. In that case the two dimensional Fourier transform becomes a one dimensional Hankel transform relationship:

$$E(R) = \int_0^\infty E_S(r)\, J_0\left(\frac{2\pi Rr}{\lambda x}\right) r\,dr$$

where J_0 is the zero-order Bessel function. Hankel transforms may be less familiar, but their properties are quite analogous to Fourier transforms.

Both equations given above express the Fraunhofer approximation to the Fresnel integral, or *Fraunhofer diffraction*. Note that the relationship is between the source and observed electric fields, expressed as complex quantities, with both magnitude and phase. Ideally, this requires that the light waves be completely monochromatic and sinusoidal. Diffraction of polychromatic light is more complicated since the integral relationships must be used to relate each spectral component, wavelength by wavelength for all wavelengths in the source.

The magnetic fields can be found from the electric fields or vice versa using **Maxwell's equations**. However, in optical systems, we usually observe the **intensity**, I, rather than the electric or magnetic fields. The intensity is related to these fields by:

$$I = \frac{EE^*}{2\eta} = \frac{EH^*}{2} = \frac{E^*H}{2} = \frac{\eta HH^*}{2}$$

$$\eta = \frac{377\text{ ohms}}{n\ (\text{index})}$$

*denotes complex conjugate

with I measured in watts per square meter, E in volts per meter and H in amperes per meter. The factor of 1/2 in the above equation is the result of using peak values of E and H; it should be omitted if rms values are used.

Knowing only the intensity pattern in one plane does not enable you to find the intensity in the other plane unless you know the *phase distribution* as well. Diffraction patterns calculated by Fourier and Hankel transforms are extremely sensitive to phase in either plane. The integrals are simplified when the source plane is uniformly illuminated with monochromatic light of constant phase. This is known as *plane-wave illumination*, and is closely approximated by laser light.

A rectangular aperture under uniform plane-wave illumination is the simplest source distribution to illustrate the Fourier transform relationship. A slit illuminated by a laser beam is a common example. The resulting far-field distribution is:

$$I(Y,Z) = \left[\frac{\sin\left[\frac{\pi aZ}{\lambda X}\right]}{\left[\frac{\pi aZ}{\lambda X}\right]}\right]^2 \times \left[\frac{\sin\left[\frac{\pi bY}{\lambda X}\right]}{\left[\frac{\pi bY}{\lambda X}\right]}\right]^2$$

This distribution is the square of the F.T. of the rectangular impulse function, familiar in many fields of science and technology. The function is shown in perspective (fig. 31) for a rectangular aperture with $b = a/2$.

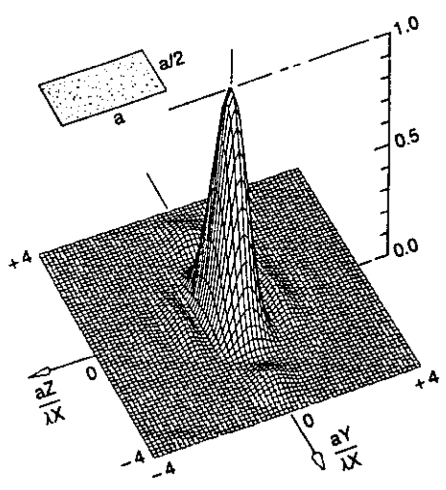


Figure 31.

It is characterized by a large central lobe which goes to zero in a rectangle bounded by $Y = \pm\lambda X/b$ and $Z = \pm\lambda X/a$. The full width at half maximum of this lobe is 0.886 $\lambda X/b$ in the Y dimension by 0.886 $\lambda X/a$ in the Z dimension. The smaller width occurs in the direction of the larger dimension of the aperture, illustrating the familiar inverse relation between widths of Fourier transform pairs.

The central lobe is surrounded by minor lobes that go to zero in rectangles. These are more easily seen if the intensity is plotted on a base 10 logarithmic scale (fig. 32).

The peak values of the minor lobes falls off least rapidly along the axes. They fall off most rapidly along the diagonals perpendicular to the diagonals of the source rectangle. The rectangles defining the zeros of the minor lobes have dimensions of $2\lambda X/b$ by $\lambda X/a$ along the Z axis, $\lambda X/b$ by $2\lambda X/a$ along the Y axis, and $\lambda X/b$ by $\lambda X/a$ off the axes.

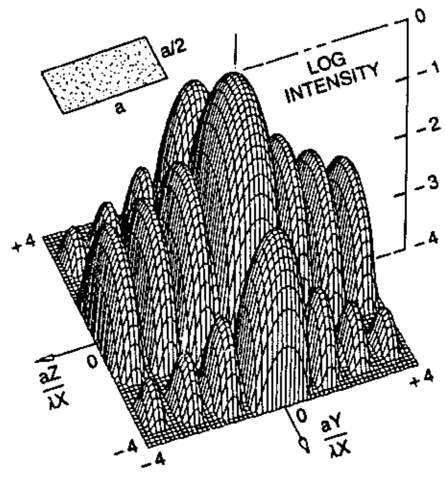


Figure 32.

Small variations in the phase distribution over the aperture can produce vastly different far-field patterns than those shown above. For example, a phase step of π in the center of the y-axis will produce a null on the optical axis, with a major lobe on either side of the Y-axis. Phase steps of π in both y and z will produce four large lobes, etc. A linear phase variation across the rectangular source aperture will move the pattern off the optical axis. More complicated phase variations produce correspondingly complicated far-field patterns. Thus you do not see the patterns shown unless the illumination is uniform and in phase.

A Circular Aperture with

uniform plane-wave illumination is perhaps the most commonly encountered source distribution in optical systems. If the light is monochromatic then the intensity distribution of the diffraction pattern at infinity will be:

$$I(R) = \left[\frac{2J_1\left[\frac{\pi dR}{\lambda X}\right]}{\left[\frac{\pi dR}{\lambda X}\right]} \right]^2$$

where d is the diameter of the aperture and $R = \sqrt{Y^2 + Z^2}$, the radius from the X axis. This function is shown in figure 33. Again, the log of the intensity has been plotted to make the minor lobes more visible.

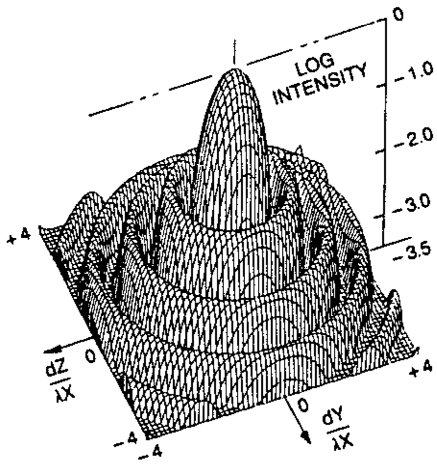


Figure 33.

A large circular major lobe is surrounded by annular minor lobes that decrease rapidly in intensity away from the optical axis. The central lobe is known as Airy's disk when projected on a screen, in honor of astronomer Sir George Airy, who first derived the equation for this diffraction pattern. The diameter of the first zero surrounding the central lobe is $R = 3.83(2/\pi)(l\,X/d)$, 3.83 being the first zero of the J_1 Bessel function. The diameters of the remaining zeros in the pattern can be found by substituting the other zeros of J_1. Unlike the rectangular diffraction pattern, the zeros are not uniformly spaced. The diameter of the central peak at its half maximum is $1.03(lX/d)$. The amplitudes of the rings fall off more rapidly than the side lobes from a square aperture along the Y and Z axes, but less rapidly than along the diagonals. Again, the far field distribution will be changed dramatically by departures from phase uniformity across the source aperture.

Fourier optics is the field that studies the science and application of the relationships described above. Practical applications would be severely limited if the transform relations were restricted to situations in which the source and observing planes had to be far distant from one another. However, recall that a simple ideal lens images a point at infinity at its focus. This property holds for diffraction as well. If the source distribution occurs in a collimated beam of light, the two-dimensional Fourier transform of that distribution will occur at the focus of the lens (fig. 34).

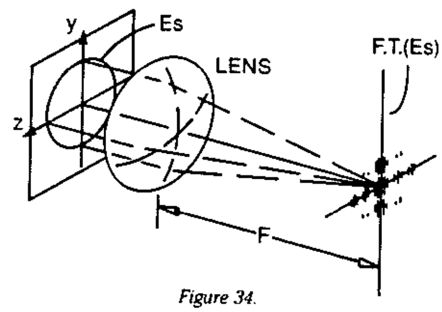


Figure 34.

Of course, the lens must be "ideal" or free from all aberrations. A plane wave must be brought to a single focal point. We've already mentioned that phase variations of the order of π (1/2) over the source aperture can totally change the pattern of the F.T. Such a restriction applies to the lens as well. The path lengths through the various regions of the lens must not depart from their ideal values by more than a very small fraction of the wavelength. The fraction that is allowable depends on how critical the application is, but varies typically from 1/10 to 1/100. Lenses this perfect are termed diffraction limited, and will convert an ideal plane wave into an ideal spherical wave. If your application involves such considerations, it is important to know just how departures from the ideal in the lens will affect your result. The cost of a **diffraction-limited** lens increases very rapidly with the phase or wavelength tolerances you place on it.

The Fourier transform relationship also results in some other interesting and readily understood consequences. Suppose the source distribution contains a particular frequency component, for example, the light intensity varies sinusoidally in a grating pattern with a spacing *d*: (fig. 35).

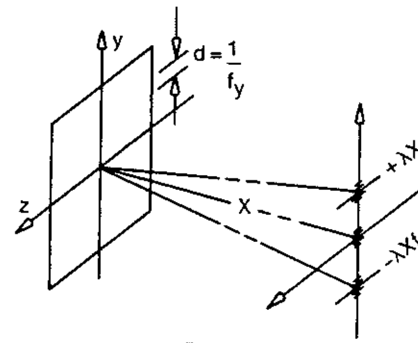


Figure 35.

We may characterize the grating as a single frequency component f_Y= 1/d (cycles/meter). The F.T. relationship tells us that the intensity in the far field pattern (or at the focal distance of a lens) will be concentrated at a distance Y from the optical axis proportional to f_Y. Actually, a concentration of energy will also occur at $-f_Y$ and at zero, since the simple real grating transmission function will contain both positive and negative frequency components and an average value as well. And if the grating is other than sinusoidal, harmonics will appear as well at $\pm 2f_Y$, $\pm 3f_Y$, ...

Many of the other features of the Fourier transform relationship can be used to see what will happen in this simple picture if we change the source distribution:

If we translate the grating in y or z, the intensity distribution in the Y,Z plane will be unaffected since we have introduced no new frequency components.

Rotating the grating around the optical axis will rotate the far-field pattern by the same angle. The two-dimensional transform is not rotationally invariant.

If we decrease the period of the grating, the size of the far-field pattern will increase. *The product of scale size in the source and scale size in the far-field is constant.*

If a feature in the source distribution is largely periodic, then it will be represented in the far field by a localized concentration of intensity at that frequency. If we amplitude, phase, or frequency modulate the grating at some lower frequency, additional frequency components would appear in the observing plane. Finally, if a feature is localized in the source (say a single line) then the intensity in the far field will be spread out.

Another interesting feature of the lens arrangement used to take an F.T. is that it makes no difference in which direction light propagates through the system. That is, if the distribution at the focal point of the lens is the source of light, then the distribution just on the far side of the lens will be the F.T. of that distribution. *Taking the transform of a transform gives us back the original function.* You can see this easily if you think of placing a mirror at the focal point of the lens. The source distribution will exactly be reflected on itself from the F.T. on the mirror.

A very important consequence of this idea is illustrated below. If, instead of a mirror, we use a second lens to restore the original source distribution, we can place masks in the F.T. plane to block certain components of the F.T. (fig. 36).

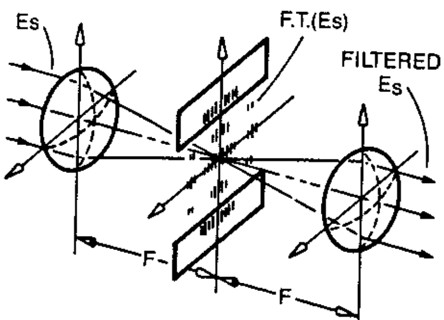


Figure 36.

We usually refer to these components as **spatial frequencies** to distinguish them from time signals that may occur elsewhere in an optical system, and we refer to the process of blocking certain spatial frequencies as **spatial filtering**. In its simplest form, spatial filtering can restore the quality of a collimated laser beam by blocking all spatial frequencies due to dust particles in a system. In more involved applications, the technique provides a powerful method of **optical signal processing**. For an introduction to this topic and some very striking photographic examples, see Chapter 14 of Hecht & Zajac,

Another obvious consequence of the Fourier transform is that the intensity distribution of light changes as it propagates through an optical system. This effect, if unwanted, may be detrimental in a particular system and quite baffling to the optics user. Recall that the F.T. only tells the relationship between source and observation planes, but the amplitude and phase must change continuously between these two points.

For example, if you were to pass a collimated laser beam through a 35 mm black and white transparency, you would see an exact shadow of the transparency on a screen very near the film. As you moved the screen farther away, you would notice a fuzziness develop on the edges of the image, then at still farther distances, a periodic "ringing" would appear, which grows in extent until the original shadow image is all but unrecognizable. Finally, at very large distances this distribution would evolve into the Fourier transform of the original pattern on the black and white transparency.

To obtain all the patterns at intermediate distance, we would have to solve the Fresnel integral without the mathematical approximations that gave us the simple Fourier transform. This is generally not easy to do except in trivial cases (slit, circular aperture, etc.).

Gaussian Beam Optics

The Gaussian intensity distribution is the one important exception to the comments in the previous paragraph. This radially symmetric distribution whose electric field variation is given by

$$E_S = E_0 \exp\left(-\frac{r^2}{w_0^2}\right)$$

has the interesting mathematical property that its Fourier transform is also a Gaussian distribution. We have implied by the expression given above that the *phase* is uniform in *r*. And, if we were to solve the Fresnel integral itself rather than the Fraunhofer approximation, we would find that a *Gaussian source distribution remains Gaussian at every point along its path of propagation through the optical system*. Of course, its size will change as it is focused by lenses or mirrors, but the intensity remains Gaussian. This makes it particularly easy to visualize the distribution of the fields at any point in the optical system. We note that the intensity is also Gaussian:

$$I_S = \eta E_S E_S^* = \eta E_0 E_0^* \exp\left(-\frac{2r^2}{w_0^2}\right)$$

This relationship is much more than a mathematical curiosity, since it is now very easy to find a light source with a Gaussian distribution of intensity: *a laser*. Most lasers automatically oscillate with a Gaussian distribution of electric field. This is also due to the fact that the Gaussian is the only function whose F.T. is itself; As light bounces back and forth between the mirrors in a laser cavity, it must adopt some distribution that will remain unchanged no matter how many bounces it makes. The Gaussian is the only distribution that satisfies this requirement. It is possible for the basic Gaussian to also take on some particular polynomial multipliers and still remain its own transform. These field distributions are known as **higher order transverse modes** and are usually avoided by design in practical lasers.

The Gaussian has no obvious boundaries to give it a characteristic dimension like the diameter of the circular aperture, for example. The definition of the size of a Gaussian is somewhat arbitrary. One could define the radius of a Gaussian as the distance from the axis at which the intensity had decreased to some fraction of the value on axis. Figure 37 shows the Gaussian intensity distribution

$$I(r) = I_0 \exp\left(-\frac{2r^2}{w_0^2}\right)$$

which might be observed at the output of a typical He-Ne laser.

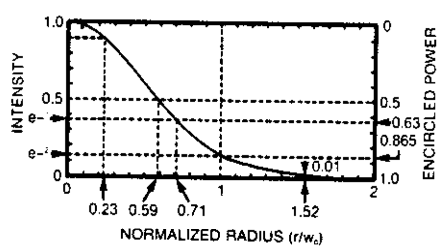


Figure 37.

The parameter w_0, usually called the Gaussian beam radius, is the radius at which the intensity has decreased to $1/e^2$ or 0.135 of its value on the axis. Another point to note is the radius of half maximum or 50% intensity, which is 0.59 w_0. At $2w_0$ or twice the Gaussian radius, the intensity is 0.0003 of its value on axis, usually completely negligible. We might also note that the region near the axis over which the Gaussian is reasonably constant is rather small. If we could tolerate only a 10% decrease in intensity, we could use the Gaussian only out to a radius of 0.23 w_0.

The power contained within a radius *r*, *P(r)*, is easily obtained by integrating the intensity distribution from 0 to r:

$$P(r) = P(\infty)\left[1 - \exp\left(-\frac{2r^2}{w_0^2}\right)\right]$$

When normalized to the total power in the beam, *P*(∞) in watts, the curve is the same curve as the intensity, but with the ordinate inverted so it may be included in the same graph. Nearly 100% of the power is contained in a radius $r = 2w_0$. One-half the power is contained within 0.59 w_0, and only 10% of the power is contained within 0.23 w_0, the radius at which the intensity has only decreased by 10%. The total power, *P*(∞) (watts), is the quantity typically advertised by manufacturers. It is related to the on axis intensity, *I*(0) (watts/m²), by:

$$P(\infty) = \left(\frac{\pi w_0^2}{2}\right) I(0)$$

$$I(0) = P(\infty)\left(\frac{2}{\pi w_0^2}\right)$$

A little thought will indicate that the on axis intensity can be very high due to the small area of the beam.

Care should be taken in cutting off the Gaussian distribution with a very small aperture to make the beam more uniform over its extent. The source distribution would no longer be Gaussian, and the far-field intensity distribution would develop zeros and other non-Gaussian features. However, if the aperture is at least 3 or $4w_0$ in diameter, then these effects will be negligible.

Propagation of Gaussian beams through an optical system can be treated almost as simply as geometric optics. Because of the unique self-Fourier Transform characteristic of the Gaussian, we do not need an integral to describe the evolution of the intensity profile with distance; the transverse distribution of intensity remains Gaussian at every point in the system; only the radius of the Gaussian and the radius of curvature of the wavefront change. Imagine that we somehow create a coherent light beam with a Gaussian distribution and a plane wavefront at a position x=0. The beam size and wavefront curvature will then vary with x as shown in figure 38.

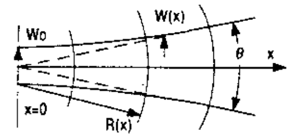


Figure 38.

The beam size will increase, slowly at first, then faster, eventually increasing proportionally to x. The wavefront radius of curvature, which was infinite at x=0, will become finite and initially decrease with x. At some point it will reach a minimum value, then increase with larger x, eventually becoming proportional to x. The equations describing the Gaussian beam radius w(x) and wavefront radius of curvature R(x) are:

$$w^2(x) = w_0^2\left[1 + \left(\frac{\lambda x}{\pi w_0^2}\right)^2\right]$$

$$R(x) = x\left[1 + \left(\frac{\pi w_0^2}{\lambda x}\right)^2\right]$$

where w_0 is the beam radius at x=0 and λ is the wavelength. The entire beam behavior is specified by these two parameters, and because they occur in the same combination in both equations, they are often merged into a single parameter, x_R, the Rayleigh range:

$$x_R = \frac{\pi w_0^2}{\lambda}$$

In fact, it is at $x = x_R$ that R has its minimum value.

Note that these equations are also valid for negative values of x. We only imagined that the source of the beam was at x=0; we could have created the same beam by creating a larger Gaussian beam with a negative wavefront curvature at some x<0. This we can easily do with a lens, as illustrated in fig. 39:

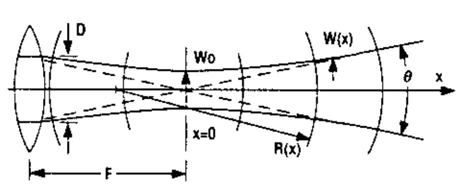


Figure 39.

The input to the lens is a Gaussian with diameter D and a wavefront radius of curvature which, when modified by the lens, will be R(x) given by the equation above with the lens located at -x from the beam waist at x=0. That input Gaussian will also have a beam waist position and size (or Rayleigh range) associated with it. Thus we can generallize the law of propagation of a Gaussian through even a complicated optical system:

In the free space between lenses, mirrors, etc., a Gaussian beam is specified in diameter and wavefront curvature by the equations given above; the position of the beam waist and the waist diameter (or Rayleigh range) completely determine the beam.

When a beam passes through a lens, mirror, or dielectric interface the diameter is unchanged, but the wavefront curvature is changed, resulting in new values of waist position and waist diameter (or Rayleigh range) on the output side of the interface.

These laws, with input values of w and R will allow you to trace a Gaussian beam through any optical system. Of course, some restrictions apply: optical surfaces need to be spherical, and with not too short a focal length, so that beams do not change diameter too fast. These are exactly the analog to the paraxial restrictions we applied to simplify geometrical optical propagation.

As stated above, the laws are not exactly suited for efficient calculation. Wouldn't it be nice if we could put them in a form as convenient as the ABCD matricies we introduced for geometric ray tracing? It turns out we can, and we even end up with the same values of A, B, C, and D for the elements of the optical system, or the system itself! There is a difference: w(x) and R(x) do not transform in matrix fashion as r and u did for ray tracing; rather they transform via a complex bilinear transformation:

$$q_{out} = \frac{[q_{in}A + B]}{[q_{in}C + D]}$$

where the quantity q is a complex composite of w and R:

$$\frac{1}{q(x)} = \frac{1}{R(x)} - \frac{j\lambda}{\pi w(x)^2}$$

We can see from the expression for q that at a beam waist (R= ∞ and $w = w_o$) that q is pure imaginary and equals jx_R. Thus if we know where one beam waist is and its size, we can calculate q there and then use the bilinear ABCD relation to find q anywhere else in the system. If you wanted to know the size and wavefront curvature of the beam everywhere in the system, you would have to use the ABCD values for each element of the system and trace q through them via successive bilinear transformations. On the other hand, if you only wanted the overall transformation of q, you could multiply the elemental ABCD values in matrix form, just as we did for geometric optics, to find the overall ABCD values for the system, then apply the bilinear transform. For more information on Gaussian beams, see Chapter 17 of Siegman's book Lasers.

There are some simple relations for Gaussian beams that can be stated easily, and are important enough to single out.

An interesting relationship follows from the two formulas for w and R. At large distances from a beam waist, the beam appears to diverge as a spherical wave from a point source located at the center of the waist. Note that "large" distances mean where $x >> x_R$ and are typically very manageable considering the small area of most laser beams. The diverging beam has a full angular width θ (again, defined to the $1/e^2$ points):

$$\theta = \frac{4\lambda}{2\pi w_0}$$

We have invoked the approximation tan $\theta \approx \theta$ since the angles are small. Since the origin can be approximated by a point source, θ is given by geometrical optics as the diameter illuminated on the lens, *D*, divided by the focal length of the lens.

$$\theta \approx \frac{D}{F} = (f/\#)^{-1}$$

where f/# is the photographic f-number of the lens.

Equating these two expressions allows us to find the beam waist diameter in terms of the input beam parameters

$$2w_0 = \left(\frac{4\lambda}{\pi}\right)\left(\frac{F}{D}\right)$$

The **focal spot diameter** of a Gaussian beam is about equal to the wavelength times the f-number of the focusing system. (There will be some restrictions on this that will be discussed later.)

We can also find the depth of focus for this same beam from the formulas above. If we define the depth of focus (somewhat arbitrarily) as the distance between the values of *x* where the beam is √2 times larger than it is at the beam waist, then using the equation for *w(x)* we can determine that

$$\mathrm{DOF} = \left(\frac{8\lambda}{\pi}\right)\left(\frac{F}{D}\right)^2$$

The depth of focus is about two and a half times the wavelength times the square of the f-number of the focusing system.

Using these relations, we can make simple calculations for optical systems employing Gaussian beams. For example, suppose that we use a 10 mm focal length lens to focus the collimated output of a helium-neon laser (632.8 nm) that has a 1 mm diameter beam. The diameter of the focal spot will be

$$\left(\frac{4}{\pi}\right) \times (632.8\,\mathrm{nm}) \times \left(\frac{10\,\mathrm{mm}}{1\,\mathrm{mm}}\right),$$

or about 8 microns. The depth of focus for this beam is then

$$\left(\frac{8}{\pi}\right) \times (632.8\,\mathrm{nm}) \times \left(\frac{10\,\mathrm{mm}}{1\,\mathrm{mm}}\right)^2,$$

or about 160 microns.

If we were to change the focal length of the lens in this example from 10 mm to 100 mm, the focal spot size would increase 10 times to 80 microns or 8% of the original beam diameter. The depth of focus would increase 100 times to 16 mm.

However, suppose we increase the focal length of the lens to 2,000 mm, for example. The "focal spot size" given by our simple equation would be 200 times larger, or 1.6 mm, 60% larger than the original beam! Obviously, something is wrong. The trouble lies not with the equations giving *w(x)* and *R(x)*, but rather with our assumption that the beam waist occurs at the focal distance from the lens. *For weakly focusing systems, the beam waist does not occur at the focal length.* In fact, the position of the beam waist changes contrary to what we would expect from geometric optics: For a weakly focusing system, the waist *moves toward* the lens as the focal length of the lens is increased, as we could demonstrate by using the ABCD transformation. However, we can easily believe the limiting case of this behavior by noting that a lens of infinite focal length such as a flat piece of glass, placed at the beam waist of a collimated beam will produce a new beam waist not at infinity but at the position of the glass itself. Fortunately, the simple approximations for spot size and depth of focus can still be used in most optical systems to select pinhole diameters, couple light into fibers, or compute laser intensities. Only when f-numbers are large should the full Gaussian equations be needed.

Waveplates

We have stated that the interaction of light with the atoms or molecules of a material is wavelength dependent. A consequence of this dependence is the resonant interactions discussed in regard to the material **dispersion** on page N-2. Another consequence of such **resonant interaction** is **birefringence**, the change in refractive index with the polarization of light. The orderly arrangement of atoms in some crystals results in different resonant frequencies for different orientations of the electric vector relative to the crystalline axes. This, in turn, results in different refractive indices for different polarizations. Unlike dispersion, birefringence is easy to avoid: use amorphous materials such as glass, or crystals that have simple symmetries, such as NaCl or GaAs. On the other hand we can "use" birefringence to modify the polarization state of light, a useful thing to do in many situations. The optical components that do this trick are called **birefringent waveplates** or **retardation plates** (or just waveplates or retarders for short).

By taking just the right slice of a crystal with respect to the crystalline axes, we can arrange it so that the minimum index of refraction is exhibited for one polarization of the electric vector of a plane-polarized wave, as shown in fig. 40.

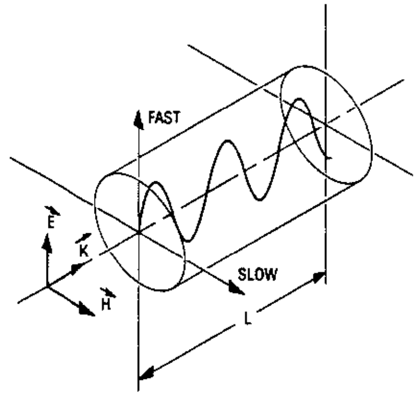


Figure 40.

We say that wave is polarized along the fast axis, since its phase velocity will be a maximum. A plane-polarized wave with its plane rotated 90° will propogate with the maximum index of refraction and minimum phase velocity, as shown in fig. 41.

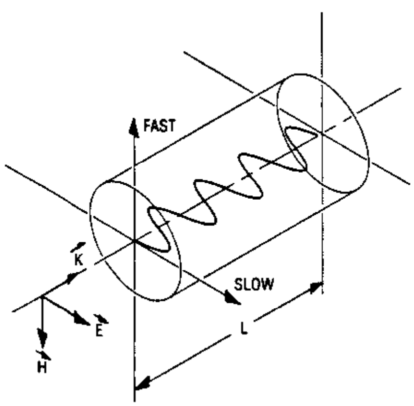


Figure 41.

We say it is polarized along the slow axis. The difference in the number of wavelengths shown in figs. 40 and 41 (2-2/3, and 4 respectively) would imply a ratio of the two indices of refraction $n_{fast}/n_{slow} = 2/3$, a much larger difference than in typical natural crystals; we have exaggerated the ratio for clarity.

The propagation phase constant k can be written as $2\pi fn/c$ radians per meter, so that a wave will experience a phase shift of $\varnothing=2\pi fnL/c$ radians in travelling a distance L through the crystal. Thus, the phase shift for the wave in fig. 40. will be $\varnothing_{fast}=2\pi fn_{fast}L/c$, and for the wave in fig. 41., $\varnothing_{slow}=2\pi n_{slow}fL/c$ (8π radians as shown.) The difference between these two phase shifts is termed the **retardation**, $\Gamma=2\pi f(n_{slow}-n_{fast})L/c$. The value of Γ in this formula is in radians, but is more common to express in "wavelengths" or "waves", with a "full wave" meaning $\Gamma=2\pi$, a "half wave" meaning $\Gamma=\pi$, a "quarter wave" meaning $\Gamma=\pi/2$, and so forth. Thus, we would term the crystal shown in the figs. a "4/3 wave plate"; that is, it retards the phase of the slow wave by 4/3 of a wave (cycle) relative to the fast wave.

Since waves repeat themselves every 2π radians, we could just as well subtract out an integral number of 2πs or waves and call the crystal shown a $2\pi/3$ radian or 1/3 wave plate. We would never know the difference, provided we only used it at exactly the optical frequency shown in the figs. However, if we change the frequency we will quickly note that the retardation will change at a rate faster than it would for a plate that had really only 1/3 wave retardation. We can note this difference by calling it a "**multiple order** 1/3 wave plate."

Half-wave Plates

By far the most commonly used waveplates are the half-wave plate ($\Gamma=\pi$) and the quarter-wave plate ($\Gamma=\pi/2$). The half-wave plate can be used to rotate the plane of plane polarized light as shown in fig. 42.

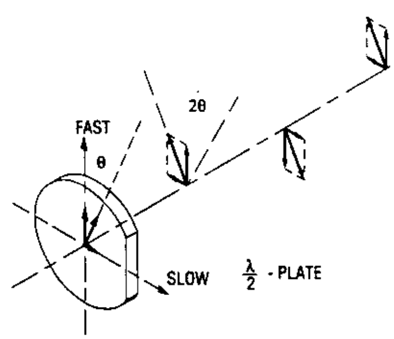


Figure 42.

Suppose a plane-polarized wave is normally incident on a waveplate, and the plane of polarization is at an angle θ with respect to the fast axis. To see what happens, resolve the incident field into components polarized along the fast and slow axes, as shown. After passing through the plate, pick a point in the wave where the fast component passes through a maximum. Since the slow component is retarded by one half wave, it will also be a maximum, but 180° out of phase, or pointing along the negative slow axis. If we follow the wave further, we see that the slow component remains exactly 180° out of phase with the original slow component, relative to the fast component. This describes a plane-polarized wave, but making an angle θ on the opposite side of the fast axis. Our original plane wave has been rotated through an angle 2θ. You can satisfy yourself that you will find the same result if the incident wave makes an angle θ with respect to the slow axis.

A half-wave plate is very handy in rotating the plane of polarization from a polarized laser to any other desired plane (especially if the laser is too large to rotate!). Most large ion lasers are vertically polarized, for example, so to obtain horizontal polarization, simply place a half-wave plate in the beam with its fast (or slow) axis 45° to the vertical. If it happens that your half-wave plate does not have marked axes (or if the markings are obscured by the mount), put a polarizer in the beam first and orient it for extinction (horizontally polarized), then interpose the half-wave plate normal to the beam and rotate it around the beam axis so that the beam remains extinct, you have found one of the axes. Then rotate the half-wave plate exactly 45° around the beam axis (in either direction) from this position, and you will have rotated the polarization of the beam by 90°. You may check this by rotating the polarizer 90° to see that extinction occurs again. If you need some other angle, instead of 90° polarization rotation, simply rotate the wave-plate by half the angle you desire. A convenient wave-plate mount calibrated in angle is the **RSA-1T**(page H-30).

Incidently, if the polarizer doesn't give you as good an extinction as you had before you inserted the wave plate, it likely means your wave-plate isn't exactly a half-wave plate at your operating wavelength. You can correct for small errors in retardation by rotating the wave-plate a small amount around its fast or slow axes. Rotation around the fast axis decreases the retardation while rotatation around the slow axis increases the retardation. Try it both ways and use your polarizer to check for improvement in extinction ratio.

Quarter-Wave Plates

Quarter-wave plates are used to turn plane-polarized light into circularly-polarized light and vice versa. To do this, we must orient the waveplate so that equal amounts of fast and slow waves are excited. We may do this by orienting an incident plane-polarized wave at 45° to the fast (or slow) axis, as shown in figure 43.

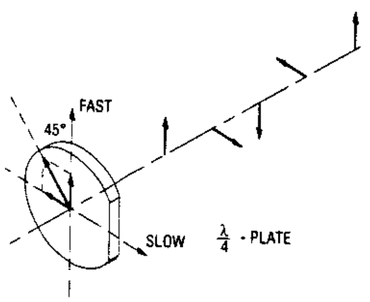


Figure 43.

On the other side of the plate, we again examine the wave at a point where the fast-polarized component is maximum. At this point, the slow-polarized component will be passing through zero, since it has been retarded by a quarter wave or 90° in phase. If we move an eighth wavelength farther, we will note that the two are the same magnitude, but the fast component is decreasing and the slow component is increasing. Moving another eighth wave, we find the slow component is maximum and the fast component is zero. If we trace the tip of the total electric vector, we find it traces out a **helix**, with a period of just one wavelength. This describes **circularly polarized light**. Right-hand light is shown in the figure; the helix wraps in the opposite sense for left-hand polarized light. You may produce left-hand polarized light by rotating either the waveplate or the plane of polarization of the incident light 90° in the figure.

Setting up a wave-plate to produce circularly polarized light proceeds exactly as we described for rotating 90° with a half-wave plate: first, cross a polarizer in the beam to find the plane of polarization. Next, insert the quarter-wave plate between the source and the polarizer and rotate the wave plate around the beam axis to find the orientation that **retains** the extinction. Then rotate the wave-plate 45° from this position. You should now have half the incident light passing through the polarizer (the other half being absorbed or deflected, depending on which kind of polarizer you are using). You can check the quality of the circularly polarized light by rotating the polarizer - the intensity of light passing through the polarizer should remain unchanged. If it varies somewhat, it means the light is actually **elliptically polarized**, and your wave plate isn't exactly a quarter-wave plate at your operating wavelenth. You may correct this as with the half-wave plate by tilting the wave-plate about its fast or slow axes slightly, while rotating the polarizer to check for constancy.

You may wonder what effect retardations other than a half wave or a quarter wave have on linearly polarized light. Figure 44 shows the effect of retardation on plane polarized light with the plane of polarization making an arbitrary angle with respect to the fast axis.

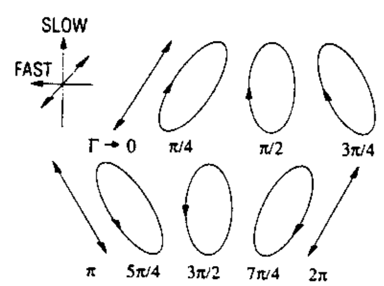


Figure 44.

The result is elliptically polarized light, with the amount of ellipticity and the tilt of the axis of the ellipse a function of the retardation and the tilt of the incident plane wave. The exception is a half wave retardation, in which case the ellipse degenerates into a plane wave making an angle of 2θ with the fast axis. Note that the quarter wave plate does not produce circularly polarized light here, because equal amounts of fast and slow wave components were not used; the incident tilt angle must be exactly 45° with respect to the fast (or slow) axis to make these components equal.

Waveplate Applications

We have already mentioned the two most common applications of wave plates: rotating the plane of polarization with a half-wave plate and creating circular polarization with a quarter-wave plate. Obviously, you can also also use a quarter wave plate to create plane polarization from circular polarization - just reverse the direction of light propagation in fig. 43!

Optical Isolation - We can use a quarter-wave plate as an optical isolator, that is, a device that eliminates undesired reflections. This application is discussed in detail in the introduction to optical isolators on page I-2.

Polarization Cleanup - Often an optical system will require several reflections from metal or dielectric mirrors. There is no change in the polarizations state of the reflection if the beam is incident normally on the mirrors, or if the plane of polarization lies in or normal to the plane of incidence. However, if the polarization direction makes some angle with the plane of incidence, then the reflection often makes a small phase shift between the parallel and perpendicular components. This is particularly true for metal mirrors, which always have some loss. The resulting reflected wave is no longer plane polarized, but will be slightly elliptically polarized, as you can easily determine by its degraded extinction when you insert a polarizer and rotate it. This small ellipticity can often be removed by inserting a **full wave** plate (which ordinarily does nothing) and tilting it slightly about either fast or slow axes to change the retardation slightly to just cancel the ellipticity.

Waveplate Material and Practice

Materials - Many natural occuring crystals exhibit birefringence, and could, in principle, be used for waveplates. Calcite and crystalline quartz are typical materials. They are durable and of high optical quality. However, the refractive index difference, n_{slow}-n_{fast} is so large that a true half-wave plate would be impractically thin to polish.

It is also possible to induce small amounts of birefringence into a normally isotropic material through stress. For example, most plastics exhibit birefringence from stress applied in the manufacture. Plastic waveplate material is available in half- or quarter-wave retardation values in very large sheets. It is inexpensive, but not of the highest optical quality or durability.

Multiple-order wave plates - One alternative to polishing or cleaving very thin plates is to use a practical thickness of a durable material such as crystalline quartz and obtain a high-order wave plate, say a 15.5 wave-plate for a 1 mm thickness. Such a plate will behave exactly the same as a half-wave plate at the design wavelength. However, as the optical wavelength is changed, the retardation will change much more rapidly than it would for a true half-wave plate. The formula for this change is easily derived from the definition of Γ:

$$\Gamma = (2m+1)\pi\left(\frac{\delta f}{f_0}\right)$$
$$\approx -(2m+1)\pi\left(\frac{\delta\lambda}{\lambda_0}\right)$$

where f_0 and λ_0 are the design frequency and wavelength, and m is the order of the waveplate. Thus, the rate of change of retardation with frequency $\delta\Gamma/\delta f$ will be 2m+1 times as large for an m^{th} order plate as a true half-wave plate, (m=0, or "zero order" plate). This would be 31 times larger for our 1 mm "15.5-wave" plate! You should calculate the frequency or wavelength range your system requires, and see if the error in retardation will be tolerable over that range with a multiple order waveplate.

By like reasoning, the sensitivity of the retardation to rotation about the fast and slow axes is found to be about (2m+1) times larger for a multiple order plate than a true zero-order half-wave plate. This means much smaller rotations are required to correct for retardation errors. But it also means that light rays not parallel to the optical axis will see a (2m+1) larger change in retardation. Multiple order waveplates are not recommended in strongly converging or diverging beam portions of your optical system. Similarly, the sensitivity of retardation to changes in length caused by changes in temperature are multiplied by (2m+1), so that tighter temperature control will be required. A typical temperature sensitivity is 0.0015 wave per degree C for a visible 1 mm thick half-wave plate.

Multiple order waveplates can be used to advantage if you require a waveplate that can be used at two discrete wavelengths, for example the 488 and 514 nm wavelengths of an argon-ion laser or the 532 and 1064 nm wavelengths from a Nd:YAG laser. By choosing the thickness to give a $(2m_1+1)$ plate at one wavelength and a $(2m_2+1)$ plate at the other, both wavelengths will see a "half-wave" plate (But not the wavelengths in between!) The integers have to be chosen by a computer program, since the dispersion in index has to be accounted for also, but it is usually possible to find a plate of reasonable thickness provided the two wavelengths are not too close together.

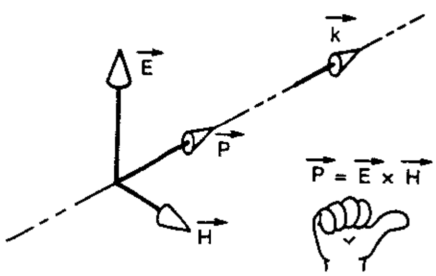
Zero-Order waveplates - Fortunately, a technique is available for realizing true half-wave plate performance, while retaining the high optical quality and rugged construction of crystalline quartz waveplates. By combining two waveplates whose retardations differ by exactly half a wave, a true half-wave plate results. The fast axis of one plate is aligned with the slow axis of the other, so that the net retardation is the difference of the two retardations. The change in retardation with frequency (or wavelength) is minimized. Temperature sensitivity is also reduced; a typical value is 0.0001-wave per degree C. The change in retardation with rotation is highly dependent on manufacturing conditions and may be equal to greater than that of a multiple order waveplate.

These waveplates are recommended for use in systems using tuneable radiation sources, such as a dye laser or white light sources.

Reference Guide

Light

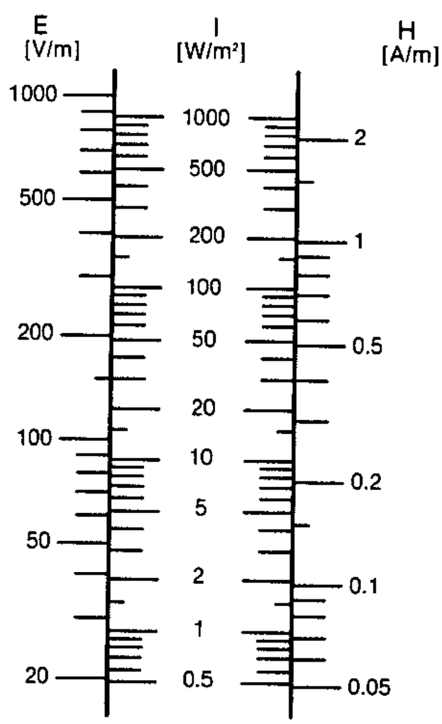
Light is a transverse electromagnetic wave. The electric and magnetic fields are perpendicular to each other and to the propagation vector *k*, as shown below.



Power density is given by **Poynting's vector,** *P*, the vector product of *E* and *H*. You can easily remember the directions if you "curl" *E* into *H* with the fingers of the right hand: your thumb points in the direction of propagation.

Intensity Nomogram

The nomogram below relates *E*, *H*, and *I* in vacuum. You may also use it for other area units, for example, [V/mm], [A/mm] and [W/mm^2]. If you change the electrical units, remember to change the units of *I* by the product of the units of *E* and *H*: for example [V/m], [mA/m], [mW/m^2] or [kV/m], [kA/m], [MW/m^2].



Light intensity

Light intensity, *I* is measured in Watts/m^2, *E* in Volts/m, and *H* in Amperes/m. The equations relating *I* to *E* and *H* are quite analogous to OHMS LAW. For peak values:

$$E = \eta H, \quad H = \frac{E}{\eta}, \quad \eta = \frac{E}{H}$$

$$I = \frac{EH}{2}, \quad I = \frac{E^2}{2\eta}, \quad I = \frac{\eta H^2}{2}$$

$$E = \sqrt{2\eta I}, \quad H = \sqrt{\frac{2I}{\eta}}$$

$$\eta_0 = 377 \text{ ohms } (\Omega)$$

$$\eta = \frac{\eta_0}{n}$$

The quantity η_0 is the wave impedance of vacuum, and η is the wave impedance of a medium with refractive index *n*.

Wave quantity relationships

$$k = \frac{2\pi}{\lambda} = \frac{2\pi n}{\lambda_0} = \frac{2\pi n \nu}{c} = \frac{n\omega}{c}$$

$$\nu = \frac{c}{\lambda_0} = \frac{c}{n\lambda} = \frac{kc}{2\pi n} = \frac{\omega}{2\pi}$$

$$\lambda = \frac{c}{n\nu} = \frac{\lambda_0}{n} = \frac{2\pi}{k} = \frac{2\pi c}{n\omega}$$

k: wave vector [radians/m]
ν: frequency [Hertz]
ω: angular frequency [radians/sec]
λ: wavelength [m]
λ_0: wavelength in vacuum [m]
n: refractive index

Energy conversions

Wavenumber $(\nu)[\text{cm}^{-1}]$

$$= \frac{10^7}{\lambda_0}[\text{nm}]$$

Electron volts (eV) per photon

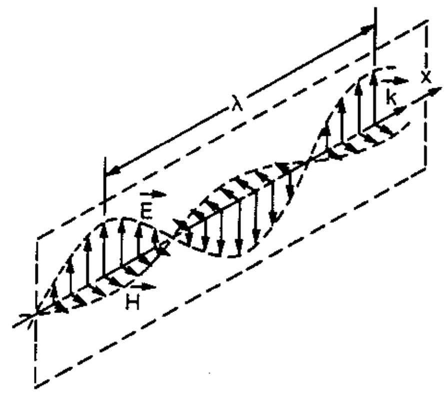
$$= \frac{1242}{\lambda_0}[\text{nm}]$$

Wavelength conversions

1 nm = 10 Angstroms(Å) = 10^{-9} m
= 10^{-7} cm = 10^{-3} micron

Plane polarized light

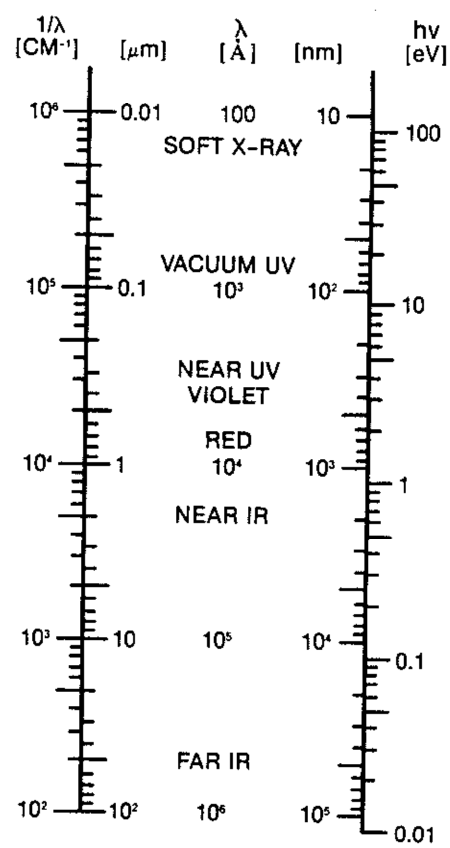
For **plane polarized light** the *E* and *H* fields remain in perpendicular planes parallel to the propagation vector *k* as shown below.



Both *E* and *H* oscillate in time and space as:

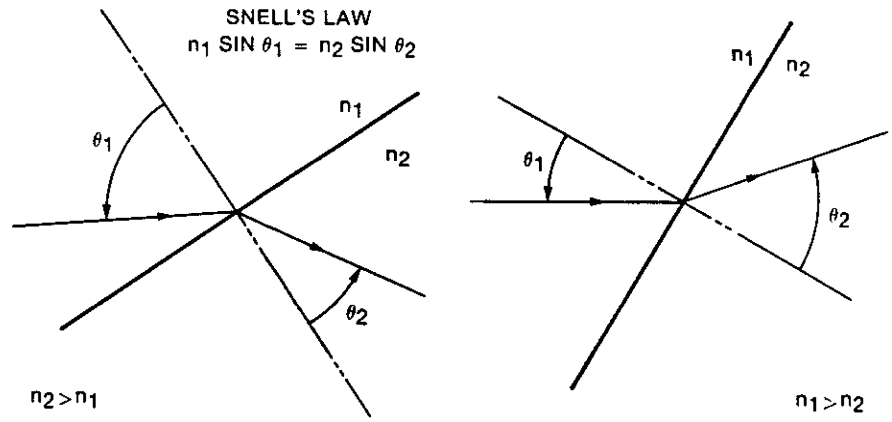
sin (ωt-kx)

The nomogram relates wavenumber, photon energy and wavelength.



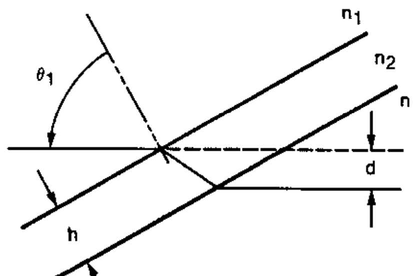
Snell's law

Snell's law tells how a light ray changes direction at a single surface between two media with different refractive indices. The angle of incidence, θ, is measured from the normal to the surface. A ray passing from low to high index is bent *toward* the normal; passing from high to low index it is bent *away from* the normal.

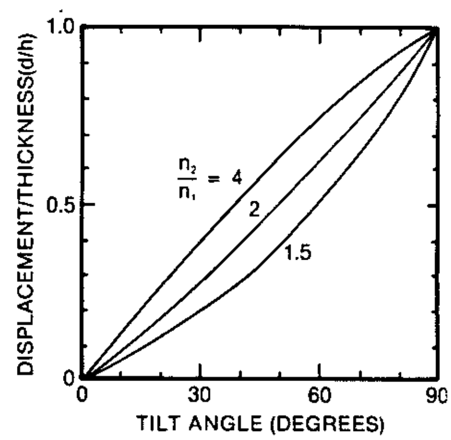


Displacement

A flat piece of glass can be used to *displace* a light ray laterally without changing its direction. The displacement varies with the angle of incidence; it is zero at normal incidence and equals the thickness of the flat at grazing incidence. The shape of the curve depends on the refractive index of the glass, as shown in the next column.

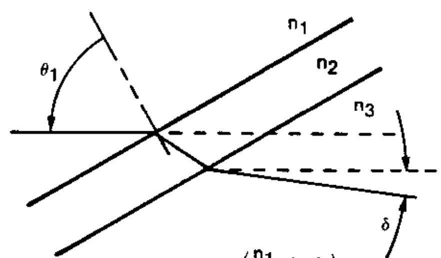


$$d = h \sin\theta_1 \left[1 - \frac{\cos\theta_1}{\sqrt{\left(\frac{n_2}{n_1}\right)^2 - \sin^2\theta_1}}\right]$$



Deviation

Both **displacement** and **deviation** occur if the media on the two sides of the tilted flat are different - for example, a tilted window in a fish tank. The displacement is the same, but the angular deviation δ is given by the formula. Note that δ is independent of the index of the flat; it is the same as if a single boundary existed between media 1 and 3.



$$\delta = \theta_1 - \sin^{-1}\left(\frac{n_1}{n_3}\sin\theta_1\right)$$

$$\cong (n_3 - n_1)\ \tan\theta_1, \text{ if } n_3 \approx n_1$$

Example: The refractive index of air at STP is about 1.0003. The deviation of a light ray passing through a glass Brewster's angle window on a HeNe laser is then:

$\delta = (n_3 - n_1) \tan\theta$

At Brewster's angle, $\tan\theta = n_2$

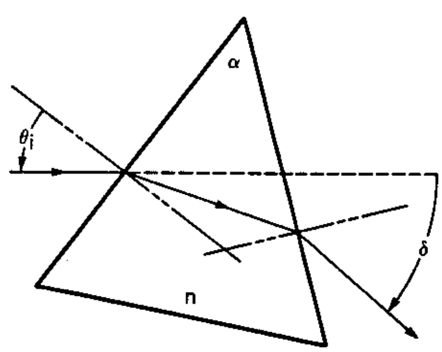
$\delta = (0.0003) \times 1.5 = 0.45$ mrad

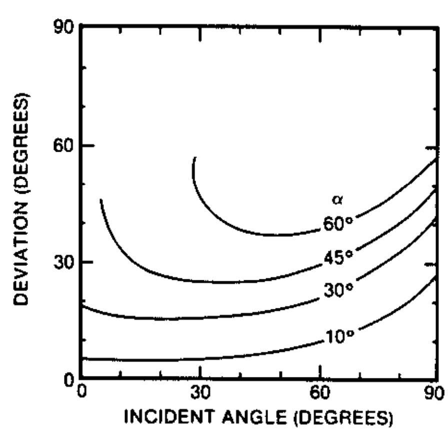
At 10,000 ft. altitude, air pressure is 2/3 that at sea level; the deviation is 0.30 mrad. This change may misalign the laser if its two windows are symmetrical rather than parallel.

Angular deviation of a prism

Angular deviation of a prism depends on the prism angle α, the refractive index, and the angle of incidence θ_i. Minimum deviation occurs when the ray within the prism is normal to the bisector of the prism angle. For small prism angles (optical wedges), the deviation is constant over a fairly wide range of angles around normal incidence. For such wedges the deviation is:

$$\delta \approx (n - 1)\ \alpha$$





Reference Guide

Field reflection

The **field reflection** and transmission coefficients are given by:

$r = E_r/E_i \qquad t = E_t/E_i$

Non-normal incidence:

$r_s = (n_i\cos\theta_i - n_t\cos\theta_t)/(n_i\cos\theta_i + n_t\cos\theta_t)$

$r_p = (n_t\cos\theta_i - n_i\cos\theta_t)/(n_t\cos\theta_i + n_i\cos\theta_t)$

$t_s = 2n_i\cos\theta_i/(n_i\cos\theta_i + n_t\cos\theta_t)$

$t_p = 2n_i\cos\theta_i/(n_t\cos\theta_i + n_i\cos\theta_t)$

Conservation of energy:

$R + T = 1$

This relation holds for *p* and *s* components individually and for total power.

Power reflection

The **power reflection** and transmission coefficients are denoted by capital letters:

$R = r^2 \quad T = t^2(n_t\cos\theta_t)/n_i\cos\theta_i)$

The refractive indices account for the different light velocities in the two media; the cosine ratio corrects for the different cross sectional areas of the beams on the two sides of the boundary.

The **intensities** [watts/area] must also be corrected by this geometric obliquity factor:

$I_t = T \times I_i\,(\cos\theta_i/\cos\theta_t)$

Fresnel Equations:

i - incident medium
t - transmitted medium
use Snell's law to find θ_t

Normal incidence:

$r = (n_i - n_t)/(n_i + n_t)$

$t = 2n_i/(n_i + n_t)$

Brewster's Angle

$\theta_\beta = \arctan(n_t/n_i)$

Only s-polarized light reflected.

Total Internal Reflection (TIR)

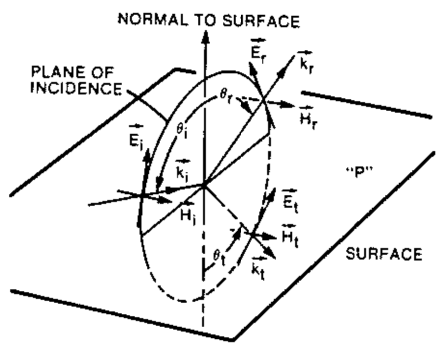
$\theta_{TIR} > \arcsin(n_t/n_i)$

$n_t < n_i$ is required for TIR

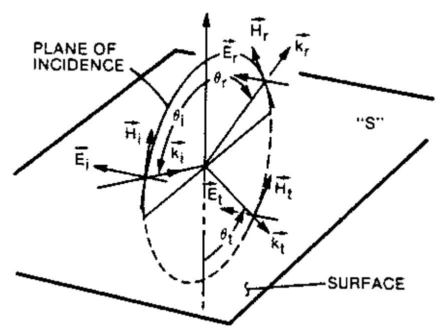
Polarization

Polarization. To simplify reflection and transmission calculations, the incident electric field is broken into two plane polarized components. The **plane of incidence** is denoted by the "wheel" in the pictures below. The normal to the surface and all propagation vectors (k_i, k_r, k_t) lie in this plane.

E parallel to the plane; **p-polarized.**

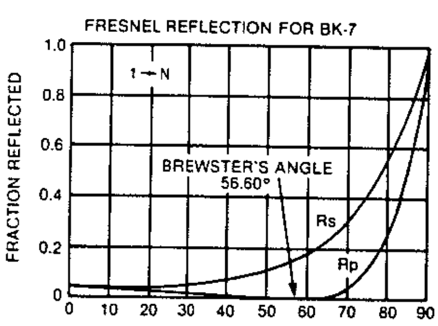


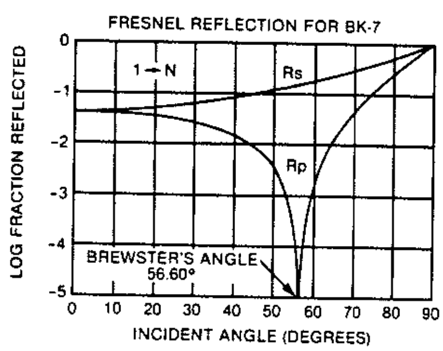
E normal to the plane; **s-polarized.**



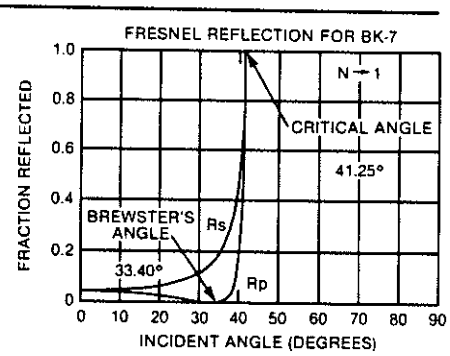
Power reflection coefficients

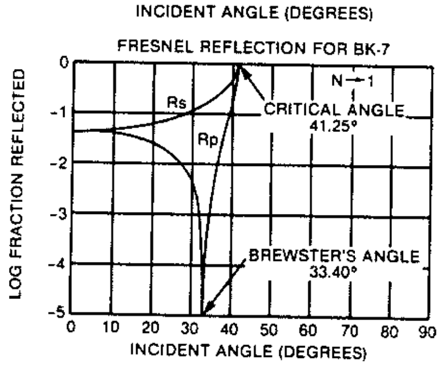
Power reflection coefficients R_s and R_p are plotted linearly and logarithmically for light traveling from air (n_i = 1) into BK-7 glass (n_t = 1.51673). **Brewster's angle** = 56.60°.





The corresponding reflection coefficients are shown below for light traveling from BK-7 glass into air **Brewster's angle** = 33.40°. **Critical angle (TIR angle)** = 41.25°.





Thin lens

If a lens can be characterized by a single plane then the lens is "thin." Various relations hold among the quantities shown in the figure.

Gaussian: $\frac{1}{s_1}+\frac{1}{s_2}=\frac{1}{F}$

Newtonian: $x_1x_2 = -F_2$

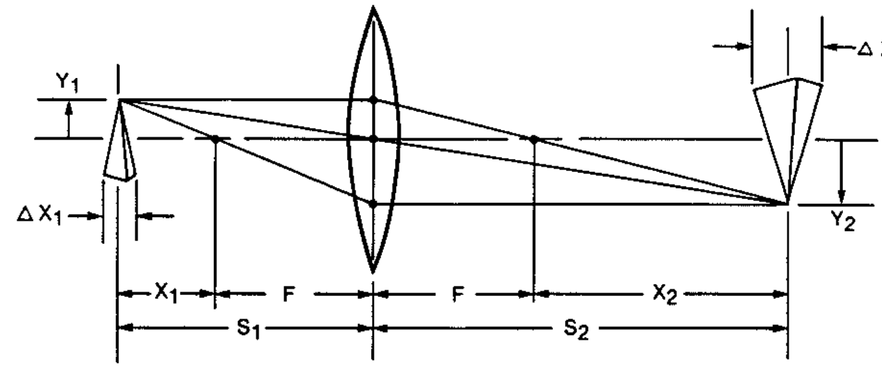
Magnification:

Transverse: $M_T=\frac{y_2}{y_1}=-\frac{s_2}{s_1}$

$M_T<0$ - Image inverted

Longitudinal: $M_L=\frac{\Delta x_2}{\Delta x_1}=-M_T^2$

$M_L<0$ - No front to back inversion



Sign conventions for images and lenses

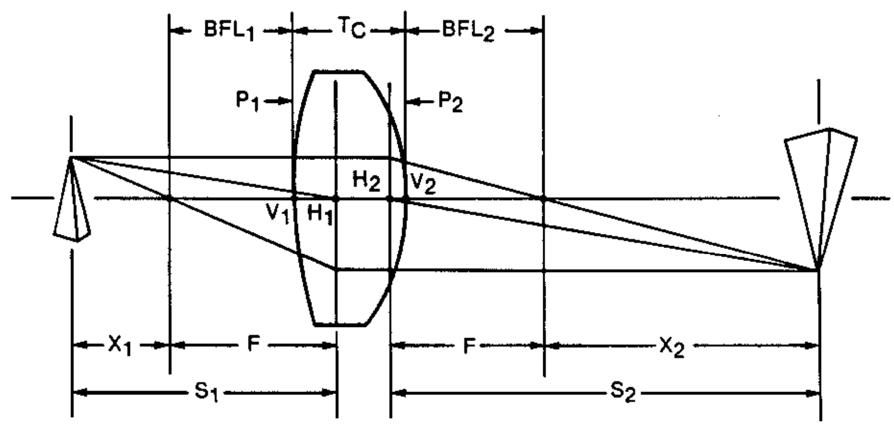
Quantity	+	-
s_1	real	virtual
s_2	real	virtual
F	convex lens	concave lens

Lens types for minimum aberration

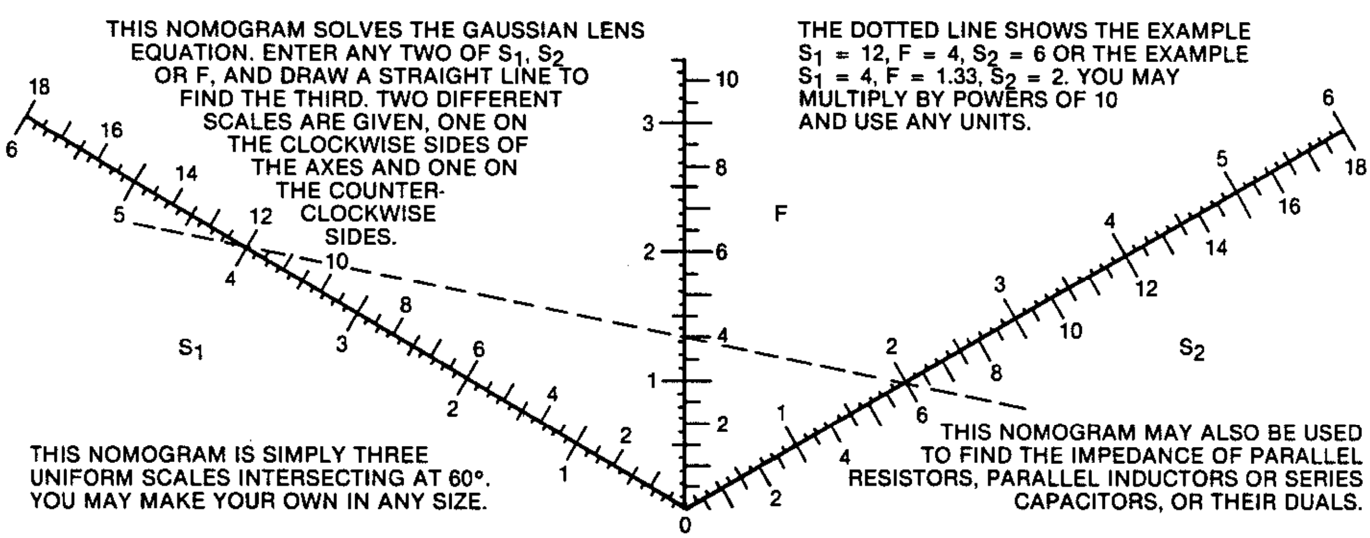
$\lvert s_2/s_1 \rvert$	Best lens
<0.2	plano-convex/concave
>5	plano-convex/concave
>0.2 or <5	bi-convex/concave

Thick lens

A **thick lens** cannot be characterized by a single focal length measured from a single plane. A single focal length *F* may be retained if it is measured from two planes, H_1, H_2, at distances *P1*, *P2* from the vertices of the lens, *V1*, *V2*. The two back focal lengths, *BFL* and *BFL2*, are measured from the vertices. The thin lens equations may be used, provided all quantities are measured form the principal planes.

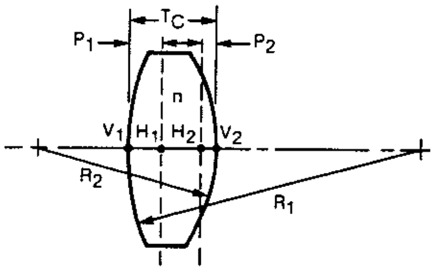


Lens nomogram



Reference Guide

The Lensmaker's Equation



$$\frac{1}{F}=(n-1)\left[\frac{1}{R_1}-\frac{1}{R_2}+\frac{(n-1)T_c}{nR_1R_2}\right]$$

$$P_1=-\frac{F(n-1)T_c}{nR_2}$$

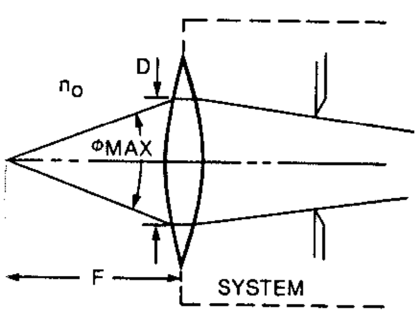
$$P_2=-\frac{F(n-1)T_c}{nR_1}$$

Convex surfaces facing left have positive radii. In the above $R_1>0$, $R_2<0$. Principal plane offsets, P, are positive to right. As illustrated, $P_1>0$, $P_2<0$. The thin lens focal length is given when $T_c=0$.

Numerical Aperture

$$NA=n_0\sin\left(\frac{\phi_{MAX}}{2}\right)$$

ϕ_{MAX} is the full angle of the cone of light rays that can pass through the system.



For small ϕ:

$$f/\#=\frac{F}{D}\approx\frac{1}{2\,NA}$$

Both f-number and NA refer to the *system* and not the exit lens.

Constants and Prefixes

Vacuum light vel.	$c=2.998\times10^{8}$ m/s
Planck's const.	$h=6.625\times10^{-34}$ J-s
Boltzmann's const.	$k=1.308\times10^{-23}$ J/°K
Stefan-Boltzmann	$s=5.67\times10^{8}$ W/m2°K⁴
1 electron volt	$eV=1.602\times10^{-19}$ J

exa (E)	10^{18}
peta (P)	10^{15}
tera (T)	10^{12}
giga (G)	10^{9}
mega (M)	10^{6}
Kilo (k)	10^{3}
milli (m)	10^{-3}
micro (μ)	10^{-6}
nano (n)	10^{-9}
pico (p)	10^{-12}
femto (f)	10^{-15}

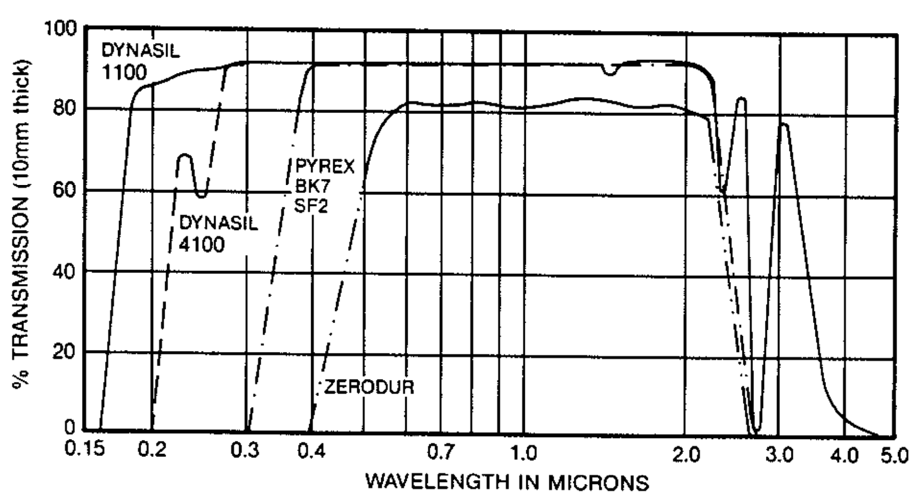
Wavelengths of common lasers

Source	(nm)
KrF	248
Nd:YAG(4)	266
XeCl	308
HeCd	325, 441.6
N_2	337
XeF	350
Nd:YAG(3)	354.7
Ar	488, 514.5, *351.1, 363.8*
Cu	510, 578
Nd:YAG(2)	532
HeNe	632.8, 1152, *543, 594,604*
Kr	647
Ruby	694
Nd:Glass	1060
Nd:YAG	1064, *1319*

Italics indicate secondary lines

Properties of optical materials

	Pyrex	Zerodur	BK-7	SF2	Fused Silica
λ(nm)			Index n (λ)		
1060			1.507	1.628	1.449
643.8	1.473		1.515	1.643	1.457
546.1	1.477	≈1.54	1.517	1.652	1.460
486	1.481		1.522	1.661	1.463
346.6					1.477
248.2					1.508
Abbe number V_d			64.2	33.8	67.8
Birefringence(nm/cm)	10	5	6	6	5
Expansivity (10^{-7}/°C)	32.5	-0.2	71	84	5.5
Conductivity (mW/cm°C)	11.3	16.3	11.3	7.3	13.8
Heat capacity (J/gm°C)	0.75	0.85	0.85	0.50	0.75
Max. Temp. (°C)	500	600	280	200	950
Density (gm/cm³)	2.23	2.52	2.51	3.86	2.20
Hardness	460	550	510	350	500
Young's Modulus (kN/mm²)	65.5	90.2	81.5	55	70.3



Gaussian intensity distribution

The **Gaussian intensity distribution:**

$$I(r) = I(0)\exp(-2r^2/w_0^2)$$

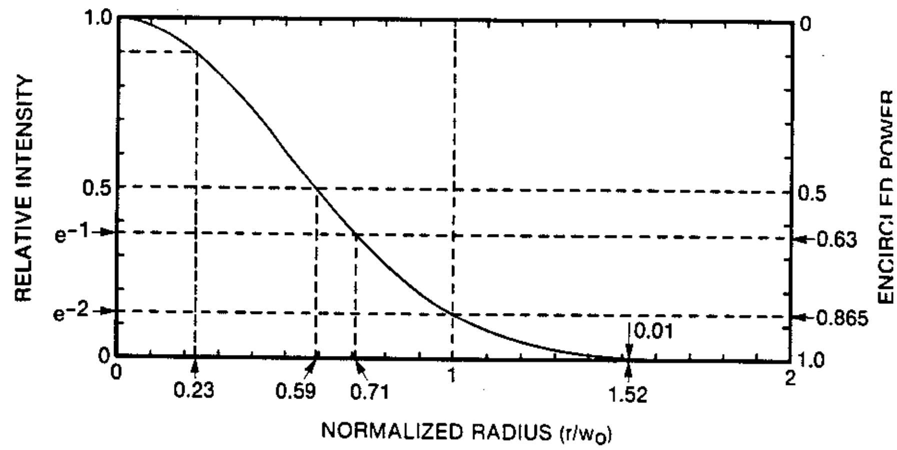
is shown at right. The right hand ordinate gives the fraction of the total power encircled at radius r:

$$P(r) = P(\infty)[1-\exp(-2r^2/w_0^2)]$$

The total beam power, $P(\infty)$ [watts], and the on-axis intensity I(0) [watts/area] are related by:

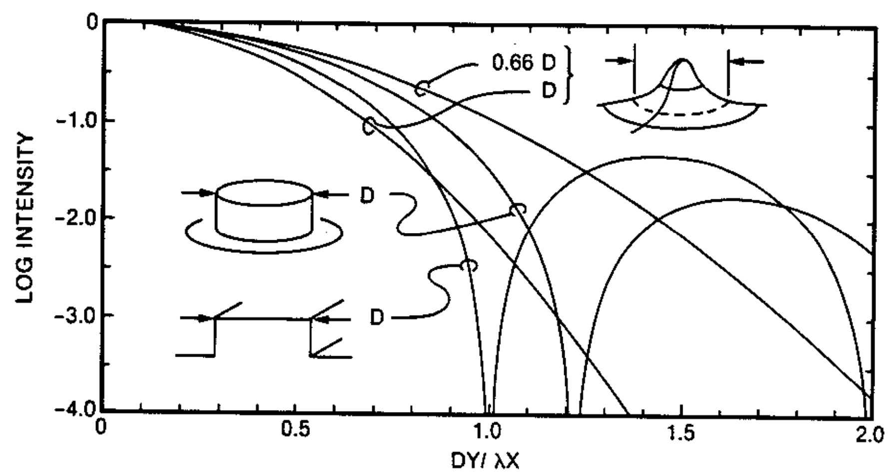
$$P(\infty) = (\pi w_0^2)\, I(0)$$

$$I(0) = (2/\pi w_0^2)\, P(\infty)$$



Diffraction

The second figure compares the far-field intensity distributions of a uniformly illuminated slit, a circular hole, and Gaussian distributions with e^{-2} diameters of D and 0.66D. (99% of a 0.66D Gaussian will pass through an aperture of diameter D.) The point of observation is Y off axis at a distance X>>Y from the source.



Focusing a collimated Gaussian beam

In the third figure the e^{-2} radius, $w(x)$, and the wavefront curvature, $R(x)$, change with x through a beam waist at $x = 0$. The governing equations are:

$$w^2(x) = w_0^2[1 + (\lambda x/\pi w_0^2)^2]$$

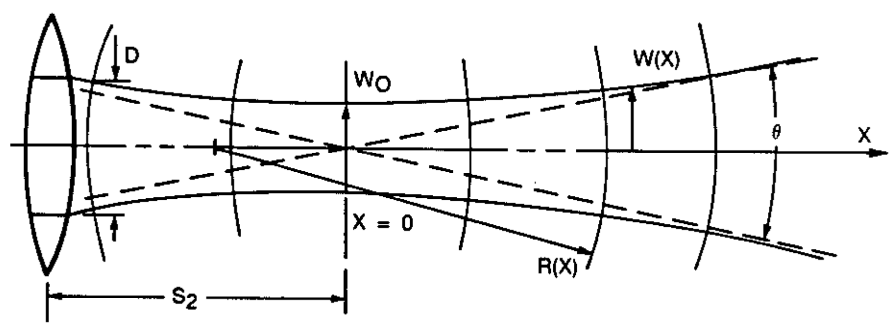
$$R(x) = x[1 + (\pi w_0^2/\lambda x)^2]$$

$2w_0$ is the waist diameter at the e^{-2} intensity points. The wavefronts are plane at the waist [$R(0) = \infty$].

At the waist, the distance from the lens will be approximately the focal length: $s_2 \approx F$.

D = collimated beam diameter or diameter illuminated on lens.

$$\text{f-number} \equiv f/\# = \frac{F}{D}$$



Depth of focus (DOF)

$$DOF = (8\lambda/\pi)(f/\#)^2$$

Only if DOF <F, then:

New waist diameter

$$2w_0 = \left(\frac{4\lambda}{\pi}\right)(f/\#)$$

Beam spread

$$\theta = (f/\#)$$

Optimal pinhole diameter for spatial filtering

$$D_{OPT} = 2\lambda(f/\#)$$

This aperture passes 99.3% of total beam energy and blocks spatial wavelengths smaller than the diameter of the initial beam. No diffraction effects will be caused by this aperture.

Reference Guide

Cleaning of any precision optic risks degrading the surface. The need for cleaning should be minimized by returning optics to their case or covering the optic and mount with a protective bag when not in use. If cleaning is required, we recommend one of the following procedures:

Cleaning Materials

Polyethylene lab gloves. Please wear them. Solvents are harsh to the skin.

Dust free tissue. Lens tissue or equivalent.

Dust free blower. Filtered dry nitrogen blown through an antistatic nozzle (Simco Inc., Hatfield, PA) is best. Bulb type blowers and brushes must be very clean to prevent redistribution of dirt.

Mild, neutral soap, 1% in water. Avoid perfumed, alkali or colored products. Several drops of green soap (available in any pharmacy) per 100 cc of distilled water is acceptable.

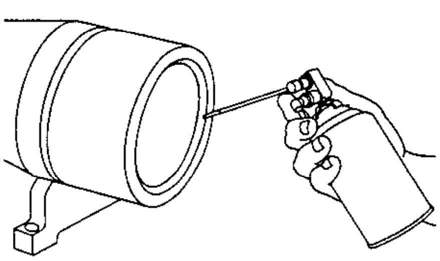
Spectroscopic grade ***isopropyl alcohol and acetone.***

Cotton swabs. Avoid plastic stems which can dissolve in alcohol or acetone.

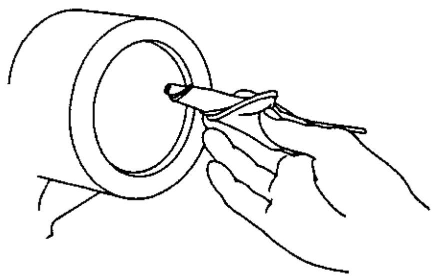
Cleaning Procedures

Dust on optics can be very tightly bound by static electricity. Blowing removes some dirt; the remainder can be collected by the surface tension of a wet alcohol swab. Acetone promotes rapid drying of the optic to eliminate streaks.

1) Blow off dust.



2) If any dust remains, twist tissue around a swab, soak in alcohol and wipe the optic in one direction with a gentle figure-eight motion. Repeat.

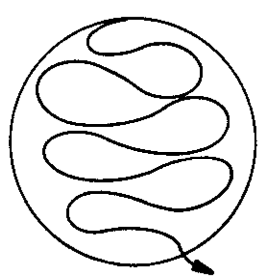


3) Repeat Step (2) with acetone soaked swabs.

Fingerprints, oil or water spots should be cleaned immediately. Skin acids attack coatings and glass. Cleaning with solvents alone tends to redistribute grime. These contaminants must be lifted from an optical surface with soap or other wetting agent. The part is then rinsed in water and the water removed with alcohol. Acetone speeds drying and eliminates streaks.

1) Blow off dust.

2) Using a soap saturated lens tissue around a swab, wipe the optic gently in the same figure 8 motion. Repeat.



3) Repeat (2) with distilled water only.

4) Repeat (2) with alcohol.

5) Repeat (2) with acetone.

Delicate optics such as UV aluminum mirrors are most safely cleaned by immersion. ***Do not immerse cemented optics.*** Washing solutions should be used only once to prevent recontamination.

1) Blow off dust.

2) Prepare petri dishes filled with soap solution, distilled water, alcohol, and acetone. Line the bottom of each with tissue to prevent blemishing an optic.

3) Immerse the optic in soap solution. Agitate gently.

4) Immerse in distilled water. Agitate.

5) Immerse in alcohol. Agitate.

6) Immerse in acetone. Agitate.

7) Blow dry.

Military specifications are used by Newport to communicate the durability of optical coatings in an industry consistent manner. The primary MILSPECS used are:

MIL-C-675 specifies that the coating will not show degradation to the naked eye after 20 strokes with a rubber pumice eraser. Coatings meeting MIL-C-675 can be cleaned repeatedly and survive moderate to severe handling.

MIL-M-13850 sets durability standards for metallic coatings. Coatings will not peel away from the substrate when pulled with cellophane tape. Further, no damage visible to the naked eye will appear after 50 strokes with a dry cheesecloth pad. Gentle, nonabrasive cleaning is advised.

MIL-C-14806 specifies durability of surfaces under environmental stress. Coatings are tested at high humidity, or in brine solutions to determine resistance to chemical attack. These coatings can survive in humid or vapor filled areas.

NOTE – IN APh 24 WE LET THE INSTRUCTOR & THE T.A.'S DO ALL THE CLEANING REQUIRED! WBB.