Projects in Optics

Applications Workbook

Created by the technical staff of Newport Corporation
with the assistance of Dr. Donald C. O'Shea of the School of Physics
at the Georgia Institute of Technology.

We gratefully acknowledge J. Wiley and Sons, publishers of
The Elements of Modern Optical Design by Donald C. O'Shea
for use of copyrighted material in the Optics Primer section.

Table of Contents

Projects In Optics

Preface

The Projects in Optics Kit is a set of laboratory equipment containing all of the optics and optomechanical components needed to complete a series of experiments that will provide students with a basic background in optics and practical hands-on experience in laboratory techniques. The projects cover a wide range of topics from basic lens theory through interferometry and the theory of imaging. The Project in Optics Handbook has been developed by the technical staff of Newport Corporation and Prof. D. C. O'Shea, in order to provide educators with a convenient means of stimulating their students' interest and creativity.

This handbook begins with a description of several mechanical assemblies that will be used in various combinations for each experiment. In addition, these components can be assembled in many other configurations that will allow more complex experiments to be designed and executed. One of the benefits from constructing these experiments using an optical bench (sometimes called an optical breadboard) plus standard components, is that the student can see that the components are used in a variety of different circumstances to solve the particular experimental problem, rather than being presented with an item that will perform only one task in one way.

A short Optics Primer relates a number of optical phenomena to the ten projects in this handbook. Each project description contains a statement of purpose that outlines the phenomena to be measured, the optical principle is being studied, a brief look at the relevant equations governing the experiment or references to the appropriate section of the Primer, a list of all necessary equipment, and a complete step-by-step instruction set which will guide the student through the laboratory exercise. After the detailed experiment description is a list of somewhat more elaborate experiments that will extend the basic concepts explored in the experiment. The ease with which these additional experiments can be done will depend both on the resources at hand and the inventiveness of the instructor and the student.

The equipment list for the individual experiments is given in terms of the components assemblies, plus items that are part of the project kits. There are a certain number of required items that are to be supplied by the instructor. Items such as metersticks and tape measures are easily obtainable. Others, for the more elaborate experiments, may be somewhat more difficult, but many are found in most undergraduate programs. Note that along with lasers and adjustable mirror mounts, index cards and tape is used to acquire the data. The student should understand that the purpose of the equipment is to get reliable data, using whatever is required. The student should be allowed some ingenuity in solving some of the problems, but if his or her choices will materially affect their data an instructor should be prepared to intervene.

These experiments are intended to be used by instructors at the sophomore/junior level for college engineering and physical science students or in an advanced high school physics laboratory course. The projects follow the general study outline found in most optical text books, although some of the material on lasers and imaging departs from the standard curriculum at the present time. They should find their greatest applicability as instructional aids to reinforcing the concepts taught in these texts.

Acknowledgement: A large part of the text and many of the figures of "An Optics Primer" are based on Chapter One of *Elements of Modern Optical Design* by Donald C. O'Shea, published by J. Wiley and Sons, Inc., New York ©1985. They are reprinted with permission of John Wiley & Sons, Inc.

0.0 An Optics Primer

The field of optics is a fascinating area of study. In many areas of science and engineering, the understanding of the concepts and effects in that field can be difficult because the results are not easy to display. But in optics, you can see exactly what is happening and you can vary the conditions and see the results. This primer is intended to provide an introduction to the 10 optics demonstrations and projects contained in this **Projects in Optics** manual. A list of references that can provide additional background is given at the end of this primer.

0.1 Geometrical Optics

There is no need to convince anyone that light travels in straight lines. When we see rays of sunlight pouring between the leaves of a tree in a light morning fog, we trust our sight. The idea of light rays traveling in straight lines through space is accurate as long as the wavelength of the radiation is much smaller than the windows, passages, and holes that can restrict the path of the light. When this is not true, the phenomenon of diffraction must be considered, and its effect upon the direction and pattern of the radiation must be calculated. However, to a first approximation, when diffraction can be ignored, we can consider that the progress of light through an optical system may be traced by following the straight line paths or rays of light through the system. This is the domain of geometrical optics.

Part of the beauty of optics, as it is for any good game, is that the rules are so simple, yet the consequences so varied and, at times, elaborate, that one never tires of playing. Geometrical optics can be expressed as a set of three laws:

1. **The Law of Transmission.**

   In a region of constant refractive index, light travels in a straight line.

2. **Law of Reflection.**

   Light incident on a plane surface at an angle θ_i with respect to the normal to the surface is reflected through an angle θ_r equal to the incident angle (**Fig. 0.1**).

$$\theta_i = \theta_r \qquad (0.1)$$

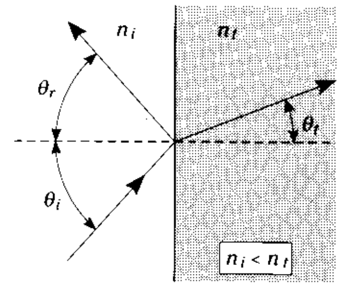


Figure 0.1 Reflection and refraction of light at an interface.

3. **Law of Refraction (Snell's Law).**

   Light in a medium of refractive index n_i incident on a plane surface at an angle θ_i with respect to the normal is refracted at an angle θ_t in a medium of refractive index n_t as (**Fig. 0.1**),

$$n_i \sin\theta_i = n_t \sin\theta_t \qquad (0\text{-}2)$$

A corollary to these three rules is that the incident, reflected, and transmitted rays, and the normal to the interface all lie in the same plane, called the **plane of incidence**, which is defined as the plane containing the surface normal and the direction of the incident ray.

Note that the third of these equations is not written as a ratio of sines, as you may have seen it from your earlier studies, but rather as a product of $n \sin\theta$. This is because the equation is unambiguous as to which refractive index corresponds to which angle. If you remember it in this form, you will never have any difficulty trying to determine which index goes where in solving for angles. **Project #1** will permit you to verify the laws of reflection and refraction.

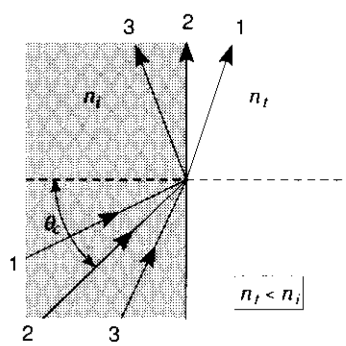


Figure 0.2. Three rays incident at angles near or at the critical angle.

A special case must be considered if the refractive index of the incident medium is greater than that of the transmitting medium, ($n_i > n_t$). Solving for θ_t, we get

$$\sin\theta_t = (n_i / n_t) \sin\theta_i \qquad (0\text{-}3)$$

In this case, $n_i / n_t > 1$, and $\sin\theta_i$ can range from 0 to 1. Thus, for large angles of θ_i it would seem that we could have $\sin\theta_t > 1$. But $\sin\theta_t$ must also be less than one, so there is a **critical angle** $\theta_i = \theta_c$, where $\sin\theta_c = n_t / n_i$ and $\sin\theta_t = 1$. This means the transmitted ray is traveling perpendicular to the normal (i.e., parallel to the interface), as shown by ray #2 in **Fig. 0.2**. For incident angles θ_i greater than θ_c no light is transmitted. Instead the light is totally reflected back into the incident medium (see ray #3, **Fig. 0.2**). This effect is called **total internal reflection** and will be used in **Project #1** to measure the refractive index of a prism.

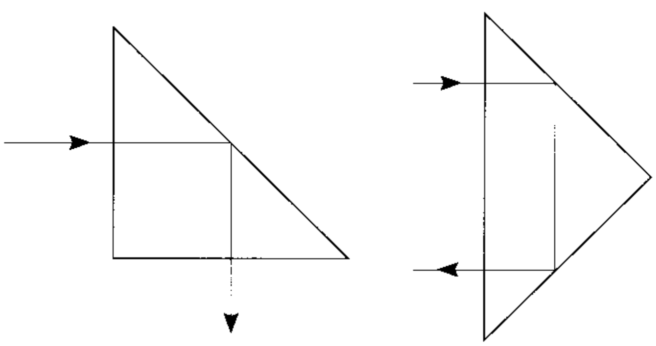


Figure 0.3. Total internal reflection from prisms.

As illustrated in **Fig. 0.3**, prisms can provide highly reflecting non-absorbing mirrors by exploiting total internal reflection.

Total internal reflection is the basis for the transmission of light through many optical fibers. We do not cover the design of optical fiber systems in this manual because the application has become highly specialized and more closely linked with modern communications theory than geometrical optics. A separate manual and series of experiments on fiber optics is available from Newport in our **Projects in Fiber Optics**.

0.1.1. Lenses

In most optical designs, the imaging components — the lenses and curved mirrors — are symmetric about a line, called the **optical axis**. The curved surfaces of a lens each have a center of curvature. A line drawn between the centers of curvatures of the two surfaces of the lens establishes the optical axis of the lens, as shown in **Fig.0.4**. In most cases, it is assumed that the optical axes of all components are the same. This line establishes a reference line for the optical system.

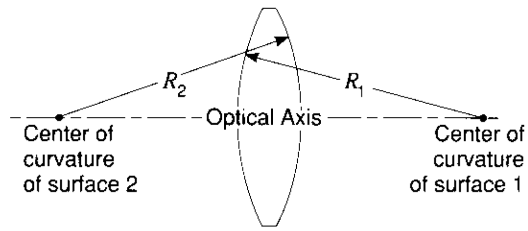


Figure 0.4 Optical axis of a lens.

By drawing rays through a series of lenses, one can determine how and where images occur. There are conventions for tracing rays; although not universally accepted, these conventions have sufficient usage that it is convenient to adopt them for sketches and calculations.

1. An object is placed to the left of the optical system. Light is traced through the system from left to right until a reflective component alters the general direction.

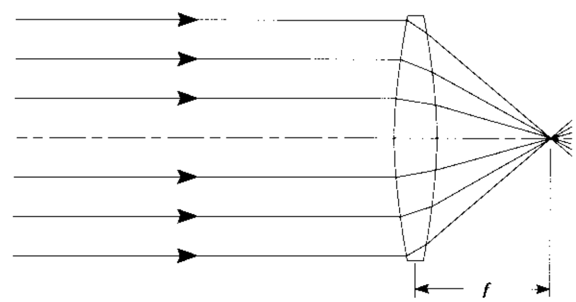
   Although one could draw some recognizable object to be imaged by the system, the simplest object is a vertical arrow. (The arrow, imaged by the optical system, indicates if subsequent images are erect or inverted with respect to the original object and other images.) If we assume light from the object is sent in all directions, we can draw a sunburst of rays from any point on the arrow. An image is formed where all the rays from the point, that are redirected by the optical system, again converge to a point.

   A positive lens is one of the simplest image-forming devices. If the object is placed very far away ("at infinity" is the usual term), the rays from the object are parallel to the optic axis and produce an image at the focal point of the lens, a distance f from the lens (the distance f is the **focal length** of the lens), as shown in **Fig. 0.5(a)**. A negative lens also has a focal point, as shown in **Fig. 0.5(b)**. However, in this case, the parallel rays do not converge to a point, but instead appear to diverge from a point a distance f in front of the lens.

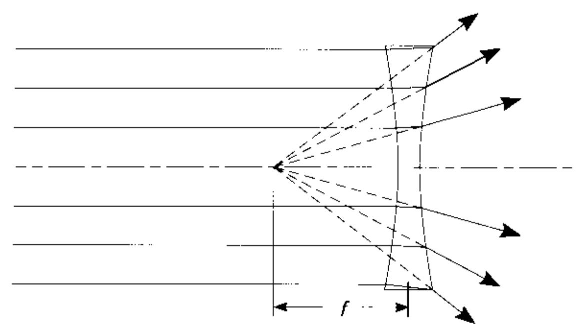
2. A light ray parallel to the optic axis of a lens will pass, after refraction, through the focal point, a distance f from the vertex of the lens.

3. Light rays which pass through the focal point of a lens will be refracted parallel to the optic axis.

4. A light ray directed through the center of the lens is undeviated.



a.



b.

Figure 0.5. Focusing of parallel light by positive and negative lenses.

The formation of an image by a positive lens is shown in **Fig. 0.6**. Notice that the rays cross at a point in space. If you were to put a screen at that point you would see the image in focus there. Because the image can be found at an accessible plane in space, it is called a **real image**. For a negative lens, the rays from an object do not cross after transmission, as shown in **Fig. 0.7**, but appear to come from some point behind the lens. This image, which cannot be observed on a screen at some point in space, is called a **virtual image**. Another example of a virtual image is the image you see in the bathroom mirror in the morning. One can also produce a virtual image with a positive lens, if the object is located between the vertex and focal point. The labels, "real" and "virtual", do not imply that one type of image is useful and the other is not. They simply indicate whether or not the rays redirected by the optical system actually cross.

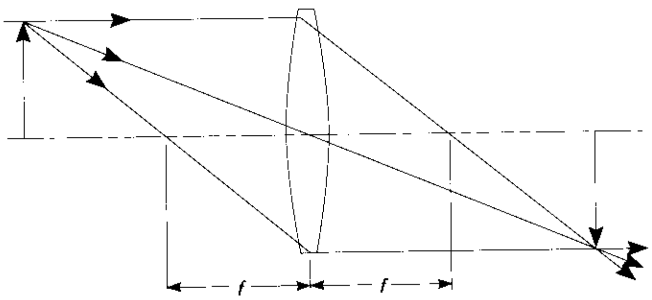


Figure 0.6. Imaging of an object point by a positive lens. A real inverted image with respect to the object is formed by the lens.

Most optical systems contain more than one lens or mirror. Combinations of elements are not difficult to handle according to the following rule:

5. The image of the original object produced by the first element becomes the object for the second element. The object of each additional element is the image from the previous element.

More elaborate systems can be handled in a similar manner. In many cases the elaborate systems can be broken down into simpler systems that can be handled separately, at first, then joined together later.

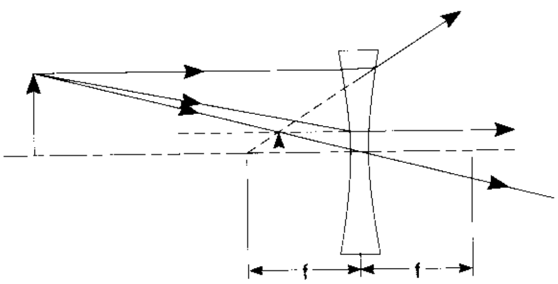


Figure 0.7. Imaging of an object point by a negative lens. A virtual erect image with respect to the object is formed by the lens.

0.2 Thin Lens Equation

Thus far we have not put any numbers with the examples we have shown. While there are graphical methods for assessing an optical system, sketching rays is only used as a design shorthand. It is through calculation that we can determine if the system will do what we want it to. And it is only through these calculations that we can specify the necessary components, modify the initial values, and understand the limitations of the design.

Rays traced close to the optical axis of a system, those that have a small angle with respect to the axis, are most easily calculated because some simple approximations can be made in this region. This approximation is called the **paraxial approximation**, and the rays are called **paraxial rays**.

Before proceeding, a set of sign conventions should be set down for the thin lens calculations to be considered next. The conventions used here are those used in most high school and college physics texts. They are not the conventions used by most optical engineers. This is unfortunate, but it is one of the difficulties that is found in many fields of technology. We use a standard right-handed coordinate system with light propagating generally along the z-axis.

1. Light initially travels from left to right in a positive direction.
2. Focal lengths of converging elements are positive; diverging elements have negative focal lengths.
3. Object distances are positive if the object is located to the left of a lens and negative if located to the right of a lens.
4. Image distances are positive if the image is found to the right of a lens and negative if located to the left of a lens.

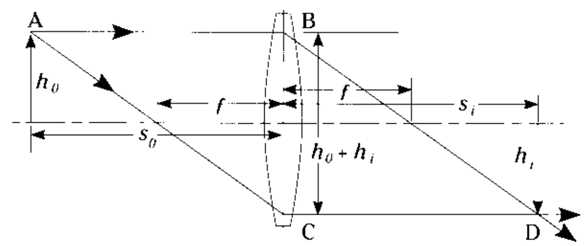


Figure 0.8. Geometry for a derivation of the thin lens equation.

We can derive the object-image relationship for a lens. With reference to **Fig. 0.8** let us use two rays from an off-axis object point, one parallel to the axis, and one through the front focal point. When the rays are traced, they form a set of similar triangles *ABC* and *BCD*. In ABC,

$$\frac{h_o + h_i}{s_o} = \frac{h_i}{f} \qquad \text{(0-4a)}$$

and in BCD

$$\frac{h_o + h_i}{s_i} = \frac{h_o}{f} \qquad \text{(0-4b)}$$

Adding these two equations and dividing through by $h_o + h_i$ we obtain the **thin lens equation**

$$\frac{1}{f} = \frac{1}{s_i} + \frac{1}{s_o} \qquad \text{(0-5)}$$

Solving equations 0-4a and 0-4b for $h_o + h_i$, you can show that the **transverse magnification** or lateral magnification, M, of a thin lens, the ratio of the image height h_i to the object height h_0, is simply the ratio of the image distance over the object distance:

$$M = \frac{h_i}{h_o} = \frac{-s_i}{s_o} \qquad \text{(0-6)}$$

With the inclusion of the negative sign in the equation, not only does this ratio give the size of the final image, its sign also indicates the orientation of the image relative to the object. A negative sign indicates that the image is inverted with respect to the object. The axial or longitudinal magnification, the magnification of a distance between two points on the axis, can be shown to be the square of the lateral or transverse magnification.

$$M_l = M^2 \tag{0-7}$$

In referring to transverse magnification, an unsubscripted M will be used.

The relationship of an image to an object for a positive focal length lens is the same for all lenses. If we start with an object at infinity we find from **Eq. 0-5** that for a positive lens a real image is located at the focal point of the lens ($1/s_o = 0$, therefore $s_i = f$), and as the object approaches the lens the image distance increases until it reaches a point $2f$ on the other side of the lens. At this point the object and images are the same size and the same distance from the lens. As the object is moved from $2f$ to f, the image moves from $2f$ to infinity. An object placed between a positive lens and its focal point forms a virtual, magnified image that decreases in magnification as the object approaches the lens. For a negative lens, the situation is simpler: starting with an object at infinity, a virtual image, demagnified, appears to be at the focal point on the same side of the lens as the object. As the object moves closer to the lens so does the image, until the image and object are equal in size at the lens. These relationships will be explored in detail in **Project #2**.

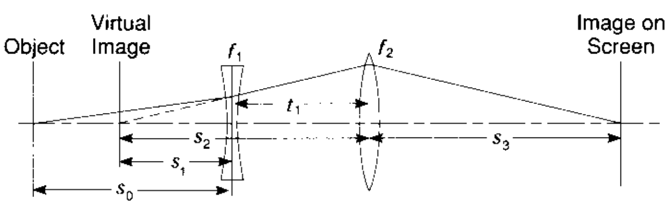


Figure 0.9 Determination of the focal length of a negative lens with the use of a positive lens of known focal length.

The calculation for a combination of lenses is not much harder than that for a single lens. As indicated earlier with ray sketching, the image of the preceding lens becomes the object of the succeeding lens.

One particular situation that is analyzed in **Project #2** is determining the focal length of a negative lens. The idea is to refocus the virtual image created by the negative lens with a positive lens to create a real image. In **Fig. 0.9** a virtual image created by a negative lens of unknown focal length f_1 is reimaged by a positive lens of known focal length f_2. The power of the positive lens is sufficient to create a real image at a distance s_3 from it. By determining what the object distance s_2 should be for this focal length and image distance, the location of the image distance for the negative lens can be found based upon rule 5 in Sec. 0.1: the image of one lens serves as the object for a succeeding lens. The image distance s_1 for the negative lens is the separation between lenses t_1 minus the object distance s_2 of the positive lens. Since the original object distance s_0 and the image distance s_1

have been found, the focal length of the negative lens can be found from the thin lens equation.

In many optical designs several lenses are used together to produce an improved image. The effective focal length of the combination of lenses can be calculated by ray tracing methods. In the case of two thin lenses in contact, the effective focal length of the combination is given by

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2}. \tag{0-8}$$

0.3 Diffraction

Although the previous two sections treated light as rays propagating in straight lines, this picture does not fully describe the range of optical phenomena that can be investigated within the experiments in **Projects in Optics**. There are a number of additional concepts that are needed to explain certain limitations of ray optics and to describe some of the techniques that allow us to analyze images and control the amplitude and direction of light. This section is a brief review of two important phenomena in physical optics, interference and diffraction. For a complete discussion of these and related subjects, the reader should consult one or more of the references.

0.3.1 Huygen's Principle

Light is an electromagnetic wave made up of many different wavelengths. Since light from any source (even a laser!) consists of fields of different wavelength, it would seem that it would be difficult to analyze their resultant effect. But the effects of light made up of many colors can be understood by determining what happens for a monochromatic wave (one of a single wavelength) then adding the fields of all the colors present. Thus by analysis of these effects for monochromatic light, we are able to calculate what would happen in non-monochromatic cases. Although it is possible to express an electromagnetic wave mathematically, we will describe light waves graphically and then use these graphic depictions to provide insight to several optical phenomena. In many cases it is all that is needed to get going.

An electromagnetic field can be pictured as a combination of electric (E) and magnetic (H) fields whose directions are perpendicular to the direction of propagation of the wave (k), as shown in **Fig. 0.10**. Because the electric and magnetic fields are proportional to each other, only one of the fields need to be described to understand what is happening in a light wave. In

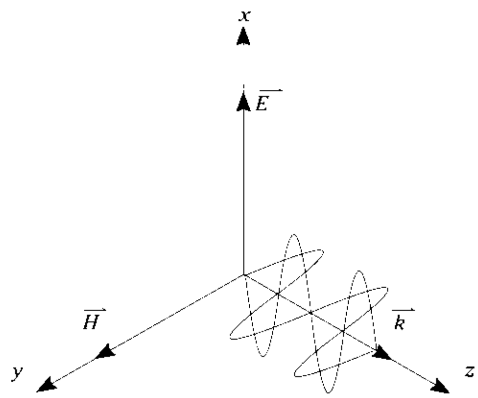


Figure 0.10. Monochromatic plane wave propagating along the z axis. For a plane wave, the electric field is constant in an x-y plane. The vector k is in the direction of propagation.

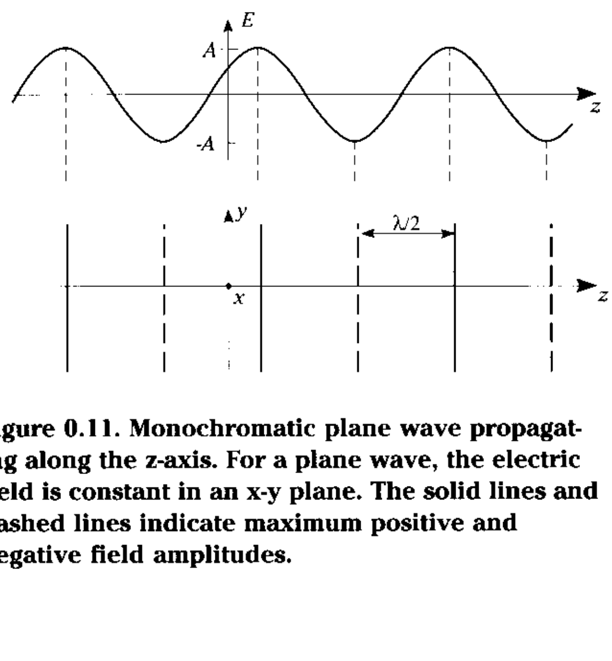


Figure 0.11. Monochromatic plane wave propagating along the z-axis. For a plane wave, the electric field is constant in an x-y plane. The solid lines and dashed lines indicate maximum positive and negative field amplitudes.

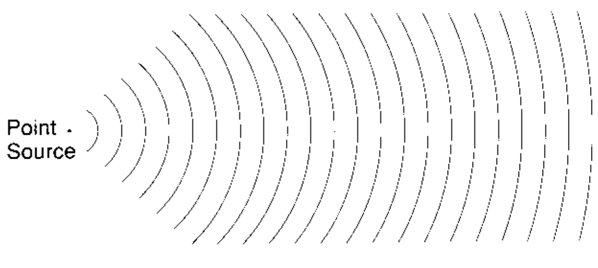


Figure 0.12. Spherical waves propagating outward from the point source. Far from the point source, the radius of the wavefront is large and the wavefronts approximate plane waves.

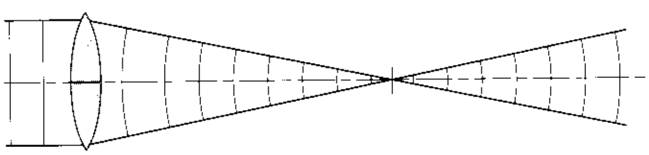


Figure 0.13. Generation of spherical waves by focusing plane waves to a point. Diffraction prevents the waves from focusing to a point.

most cases, a light wave is described in terms of the electric field. The diagram in **Fig 0.10** represents the field at one point in space and time. It is the arrangement of the electric and magnetic fields in space that determines how the light field progresses.

One way of thinking about light fields is to use the concept of wavefront. If we plot the electric fields as a function of time along the direction of propagation, there are places on the wave where the field is a maximum in one direction and other places where it is zero, and other places where the field is a maximum in the opposite direction, as shown in **Fig. 0.11**. These represent different **phases** of the wave. Of course, the phase of the wave changes continuously along the direction of propagation. To follow the progress of a wave, however, we will concentrate on one particular point on the phase, usually at a point where the electric field amplitude is a maximum. If all the points in the neighborhood have this same amplitude, they form a surface of constant phase, or **wavefront**. In general, the wavefronts from a light source can have any shape, but some of the simpler wavefront shapes are of use in describing a number of optical phenomena.

A **plane wave** is a light field made up of plane surfaces of constant phase perpendicular to the direction of propagation. In the direction of propagation, the electric field varies sinusoidally such that it repeats every wavelength. To represent this wave, we have drawn the planes of maximum electric field strength, as shown in **Fig. 0.11**, where the solid lines represent planes in which the electric field vector is pointing in the positive y-direction and the dashed lines represent plane in which the electric field vector is pointing in the negative y-direction. The solid planes are separated by one wavelength, as are the dashed planes.

Another useful waveform for the analysis of light waves is the spherical wave. A **point source**, a fictitious source of infinitely small dimensions, emits a wavefront that travels outward in all directions producing wavefronts consisting of spherical shells centered about the point source. These **spherical waves** propagate outward from the point source with radii equal to the distance between the wavefront and the point source, as shown schematically in **Fig. 0.12**. Far away from the point source, the radius of the wavefront is so large that the wavefronts approximate plane waves. Another way to create spherical waves is to focus a plane wave. **Figure 0.13** shows the spherical waves collapsing to a point and then expanding. The waves never collapse to a true point because of diffraction (next Section). There are many other possible forms of wave fields, but these two are all that is needed for our discussion of interference.

What we have described are single wavefronts. What happens when two or more wavefronts are present in the same region? Electromagnetic theory shows that we can apply the **principle of superposition**: where waves overlap in the same region of space, the resultant field at that point in space and time is found by adding the electric fields of the individual waves at a point. For the present we are assuming that the electric fields of all the waves have the same polarization (direction of the electric field) and they can be added as scalars. If the directions of the fields are not the same, then the fields must be added as vectors. Neither our eyes nor any light detector "sees" the electric field of a light wave. All detectors measure the **square** of the time averaged electric field over some area. This is the **irradiance** of the light given in terms of watts/square meter (w/m^2) or similar units of power per unit area.

Given some resultant wavefront in space, how do we predict its behavior as it propagates? This is done by invoking **Huygen's Principle**. Or, in terms of the graphical descriptions we have just defined, Huygen's Construction (see **Fig. 0.14**): Given a wavefront of arbitrary shape, locate an array of point sources on the wavefront, so that the strength of each point source is proportional to the amplitude of the wave at that point. Allow the point sources to propagate for a time t, so that their radii are equal to ct (c is the speed of light) and add the resulting sources. The resultant envelope of the point sources is the wavefront at a time t after the initial wavefront. This principle can be used to analyze wave phenomena of considerable complexity.

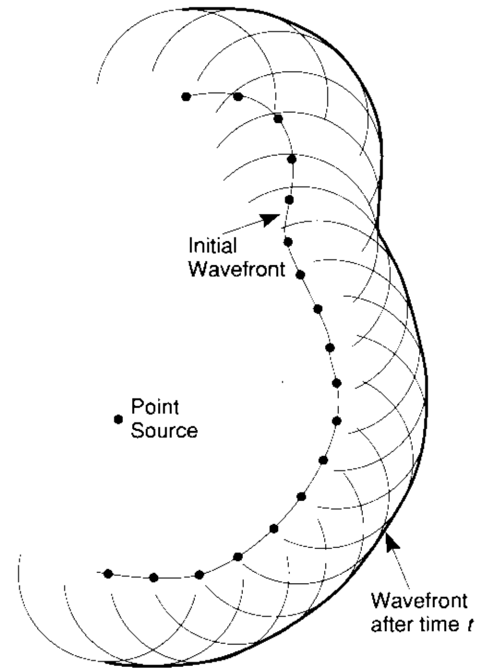
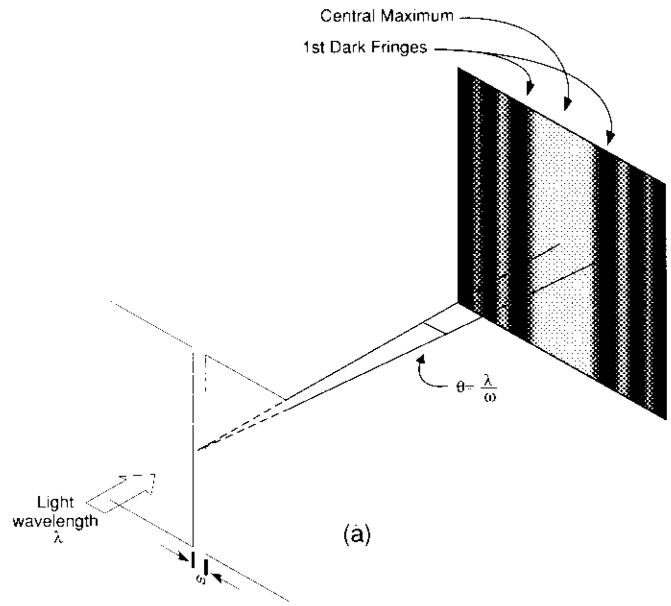


Figure. 0.14. Huygen's Construction of a propagating wavefront of arbitrary shape.

0.3.2 Fresnel and Fraunhofer Diffraction

Diffraction of light arises from the effects of apertures and interface boundaries on the propagation of light. In its simplest form, edges of lenses, apertures, and other optical components cause the light passing through the optical system to be directed out of the paths indicated by ray optics. While certain diffraction effects prove useful, ultimately all optical performance is limited by diffraction, if there is sufficient signal, and by electrical or optical "noise", if the signal is small.

When a plane wave illuminates a slit, the resulting wave pattern that passes the slit can be constructed using Huygens' Principle by representing the wavefront in the slit as a collection of point sources all emitting in phase. The form of the irradiance pattern that is observed depends on the distance from the diffraction aperture, the size of the aperture and the wavelength of the illumination. If the diffracted light is examined close to the aperture, the pattern will resemble the aperture with a few surprising variations (such as



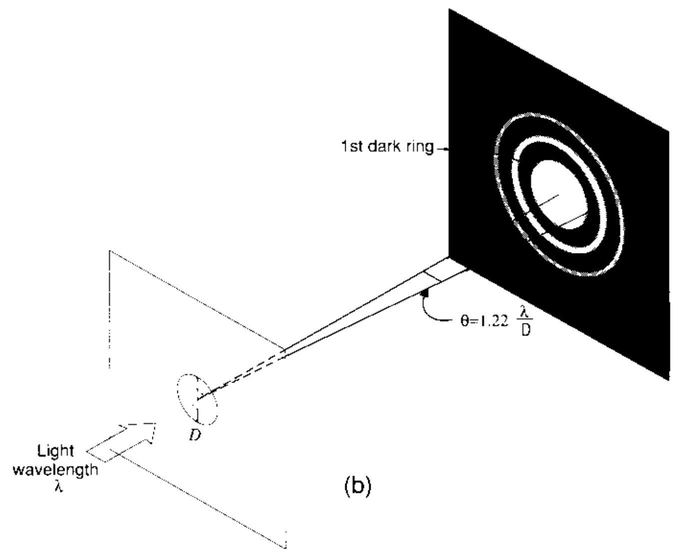


Figure 0.15. Diffraction of light by apertures. (a) Single slit. (b) Circular aperture.

finding a point of light in the shadow of circular mask!). This form of diffraction is called **Fresnel** (Freh-nell) **diffraction** and is somewhat difficult to calculate.

At a distance from the aperture the pattern changes into a **Fraunhofer diffraction** pattern. This type of diffraction is easy to calculate and determines in most cases, the optical limitations of most precision optical systems. The simplest diffraction pattern is that due to a long slit aperture. Because of the length of the slit relative to its width, the strongest effect is that due to the narrowest width. The resulting diffraction pattern of a slit on a distant screen contains maxima and minima, as shown in **Fig. 0.15(a)**. The light is diffracted strongly in the direction perpendicular to the slit edges. A measure of the amount of diffraction is the spacing between the strong central maximum and the first dark fringe in the diffraction pattern. The differences in Fraunhofer and Fresnel diffraction patterns will be explored in **Project #4**.

At distances far from the slit, the Fraunhofer diffraction pattern does not change in shape, but only in size. The fringe separation is expressed in terms of the sine of the angular separation between the central maximum and the center of the first dark fringe,

$$\sin\theta = \frac{\lambda}{w} \tag{0-9}$$

where w is the slit width and λ is the wavelength of the light illuminating the slit. Note that as the width of the slit becomes smaller, the diffraction angle becomes larger. If the slit width is not too small, the sine can be replaced by its argument,

$$\theta = \frac{\lambda}{w} \tag{0-10}$$

If the wavelength of the light illuminating the slit is known, the diffraction angle can be measured and the width of the diffracting slit determined. In **Project #5** you will be able to do exactly this.

In the case of circular apertures, the diffraction pattern is also circular, as indicated in **Fig. 0.15(b)**, and the angular separation between the central maximum and the first dark ring is given by

$$\sin\theta = 1.22\frac{\lambda}{D}$$

or for large D,

$$\theta = 1.22\frac{\lambda}{D} \tag{0-11}$$

where D is the diameter of the aperture. As in the case of the slit, for small values of λ/D, the sine can be replaced by its angle. The measurement of the diameter of different size pinholes is part of **Project #4**.

One good approximation of a point source is a bright star. A pair of stars close to one another can give a measure of the diffraction limits of a system. If the stars have the same brightness, the resolution of the system can be determined by the smallest angular separation between such sources that would still allow them to be resolved. This is provided that the aberrations of the optical system are sufficiently small and diffraction is the only limitation to resolving the images of these two point sources. Although it is somewhat artificial, a limit of resolution that has been used in many instances is that two point sources are just resolvable if the maximum of the diffraction pattern of one point source falls on the first dark ring of the pattern of the second point source, as illustrated in **Fig. 0.16**, then

$$\theta_R = 1.22\frac{\lambda}{D} \tag{0-12}$$

This condition for resolution is called the **Rayleigh criterion**. It is used in other fields of optical design, such as specifying the resolution of a optical systems.

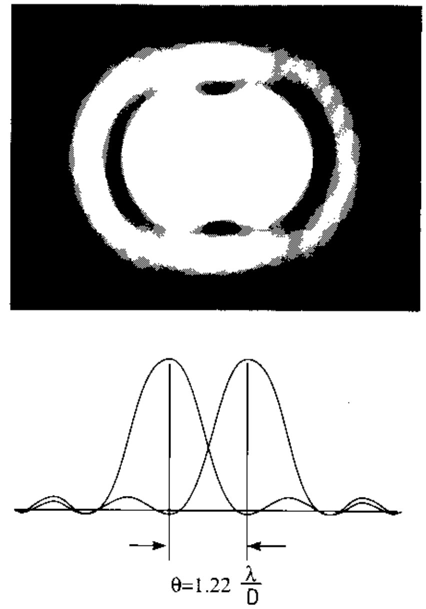


Figure 0.16. Rayleigh criterion. The plot of the intensity along a line between the centers of the two diffraction patterns is shown below a photo of two sources just resolved as specified by the Rayleigh criterion. (Photo by Vincent Mallette)

0.4 Interference

While diffraction provides the limits that tells us how far an optical technique can be extended, interference is responsible for some of the most useful effects in the field of optics — from diffraction gratings to holography. As we shall see, an interference pattern is often connected with some simple geometry. Once the geometry is discovered, the interference is easily understood and analyzed.

0.4.1. Young's Experiment

In **Fig. 0.17** the geometry and wave pattern for one of the simplest interference experiments, Young's experiment, is shown. Two small pinholes, separated by a distance d, are illuminated by a plane wave, producing two point sources that create overlapping spherical waves. The figure shows a cross-sectional view of the wavefronts from both sources in a plane containing the pinholes. Notice that at points along a line equidistant from both pinholes, the waves from the two sources are always in phase. Thus, along the line marked C the electric fields always add in phase to give a field that is twice that of a single field; the irradiance at a point

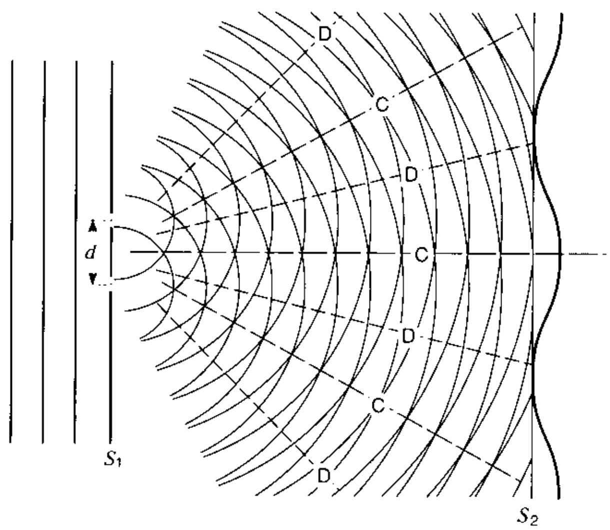


Figure 0.17. Young's Experiment. Light diffracted through two pinholes in screen S_1 spreads out toward screen S_2. Interference of the two spherical waves produces a variation in irradiance (interference fringes) on S_2 that is plotted to the right of the screen.

along the line, which is proportional to the square of the electric field, will be four times that due to a single pinhole. When electric fields add together to give a larger value it is referred to as **constructive interference**. There are other directions, such as those along the dotted lines marked D, in which the waves from the two sources are always 180° out of phase. That is, when one source has a maximum positive electric field, the other has the same negative value so the fields always cancel and no light is detected along these lines marked D, as long as both sources are present. This condition of canceling electric fields is called **destructive interference**. Between the two extremes of maximum constructive and destructive interference, the irradiance varies between four times the single pinhole irradiance and zero. It can be shown that the total energy falling on the surface of a screen placed in the interference pattern is neither more nor less than twice that of a single point source; it is just that interference causes the light distribution to be arranged differently!

Some simple calculations will show that the difference in distances traveled from pinholes to a point on the screen is

$$\Delta r = d \sin\theta. \qquad (0\text{-}13)$$

In the case of constructive interference, the wavefronts arrive at the screen in phase. This means that there is either one or two or some integral number of wavelength difference between the two paths traveled by the light to the point of a bright fringe. Thus, the angles at which the bright fringes occur are given by

$$\Delta r = d \sin\theta = n\lambda \quad (n = 1, 2, 3, \ldots). \qquad (0\text{-}14)$$

If the above equation is solved for the angles θ_n at which the bright fringes are found and one applies the approximation that for small angles the sine can be replaced by its angle in radians, one obtains:

$$\theta_n \cong n\lambda/d \qquad (n = 1, 2, 3, \ldots). \qquad (0\text{-}15)$$

The angular separation by neighboring fringes is then the difference between θ_{n+1} and θ_n:

$$\Delta\theta = \lambda/d. \qquad (0\text{-}16)$$

It is this angular separation between fringes that will be measured in **Project #5** to determine the separation between two slits.

0.4.2 The Michelson Interferometer

Another interference geometry that will be investigated in **Project #6** and used to measure an important parameter for a laser in **Project #7** is shown in **Fig. 0.18**. This is a **Michelson interferometer**, which is constructed from a beamsplitter and two mirrors. (This

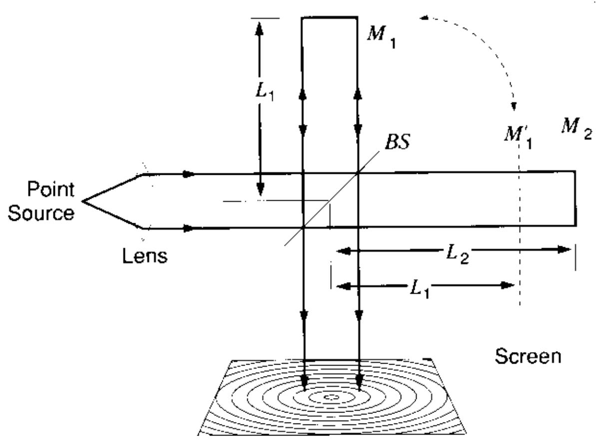


Figure 0.18. Michelson interferometer. By reflecting the mirror M_1 about the plane of the beamsplitter BS to location M'_1, one can see that a ray reflecting off mirror M_2 travels an additional distance $2(L_2 - L_1)$ over a ray reflecting off M_1.

device is sometimes called a **Twyman-Green** interferometer when it is used with a monochromatic source, such as a laser, to test optical components.) The beamsplitter is a partially reflecting mirror that separates the light incident upon it into two beams of equal strength. After reflecting off the mirrors, the two beams are recombined so that they both travel in the same direction when they reach the screen. If the two mirrors are the same distance ($L_1 = L_2$ in **Fig. 0.18**) from the beamsplitter, then the two beams are always in phase once they are recombined, just as is the case along the line of constructive interference in Young's experiment. Now the condition of constructive and destructive interference depends on the difference between the paths traveled by the two beams. Since each beam must travel the distance from the beamsplitter to its respective mirror and back, the distance traveled by the beam is $2L$. If the path-length difference, $2L_1 - 2L_2$, is equal to an integral number of wavelengths, $m\lambda$, where m is an integer, then the two waves are in phase and the interference at the screen will be constructive.

$$L_1 - L_2 = m\lambda/2 \quad (m = \ldots, -1, 0, 1, 2, \ldots). \qquad (0\text{-}17)$$

If the path-length difference is an integral number of wavelengths plus a half wavelength, the interference on the screen will be destructive. This can be expressed as

$$L_1 - L_2 = m\lambda/4 \quad (m = \text{odd integers}). \qquad (0\text{-}18)$$

In most cases the wavefronts of the two beams when they are recombined are not planar, but are spherical wavefronts with long radii of curvature. The interference pattern for two wavefronts of different curvature is a series of bright and dark rings. However, the above discussion still holds for any point on the screen. Usually, however, the center of the pattern is the point used for calculations.

In the above discussion, it was assumed that the medium between the beamsplitter and the mirrors is undisturbed air. If, however, we allow for the possibility that the refractive index in those regions could be different, then the equation for the bright fringes should be written as

$$n_1L_1 - n_2L_2 = m\lambda/2 \quad (m = \ldots -1, 0, 1, 2, \ldots). \qquad (0\text{-}17a)$$

Thus, any change in the refractive index in the regions can also contribute to the interference pattern as you will see in **Project #6**.

In optical system design, interferometers such as the Michelson interferometer can be used to measure very small distances. For example, a movement of one of the mirrors by only one quarter wavelength (correspond-

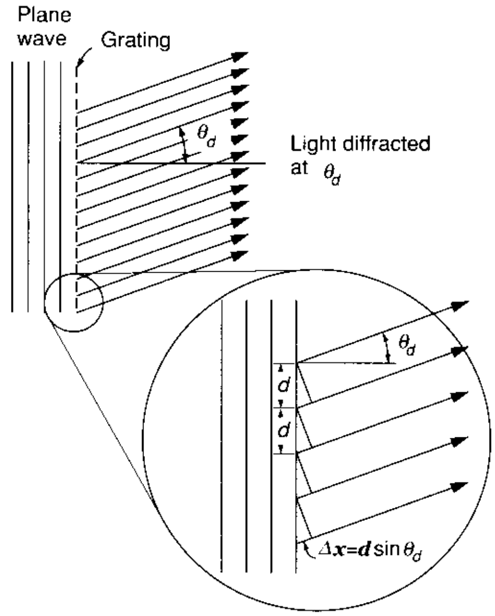


Figure 0.19. Diffraction of light by a diffraction grating.

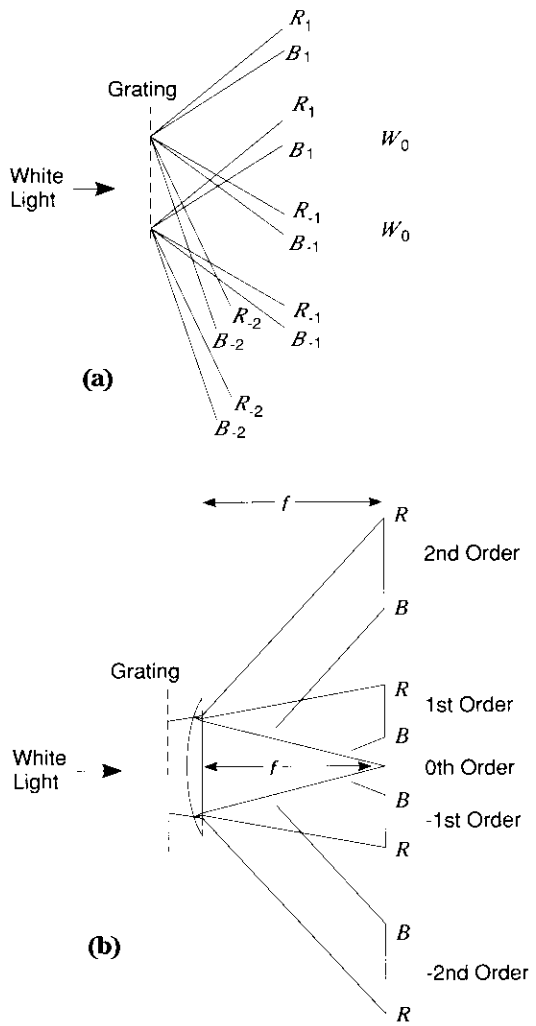


Figure 0.20. Orders of diffraction from a grating illuminated by white light. (a) Rays denoting the upper and lower bounds of diffracted beams for the red (R) and blue (B) ends of the spectrum; (b) spectra produced by focusing each collimated beam of wavelengths to a point in the focal plane.

ing to a path-length change of one half wavelength) changes the detected irradiance at the screen from a maximum to a minimum. Thus, devices containing interferometers can be used to measure movements of a fraction of a wavelength. One application of interference that has developed since the invention of the laser is holography. This fascinating subject is explored in a separate set of experiments in Newport's **Projects in Holography**.

0.4.3. The Diffraction Grating

It is a somewhat confusing use of the term to call the item under discussion a **diffraction grating**. Although diffraction does indeed create the spreading of light from a regular array of closely spaced narrow slits, it is the combined interference of multiple beams that permits a diffraction grating to deflect and separate the light. In **Fig. 0.19** a series of narrow slits, each separated from its neighboring slits by distance d, are illuminated by a plane wave. Each slit is then a point (actually a line) source in phase with all other slits. At some angle θ_d to the grating normal, the path-length difference between neighboring slits will be (see inset to **Fig. 0.19**)

$$\Delta x = d\sin(\theta_d),$$

Constructive interference will occur at that angle if the path-length difference Δx is equal to an integral number of wavelengths:

$$m\lambda = d\sin(\theta_d) \quad (m = \text{an integer}). \qquad (0\text{-}19)$$

This equation, called the grating equation, holds for any wavelength. Since any grating has a constant slit separation d, light of different wavelengths will be diffracted at different angles. This is why a diffraction grating can be used in place of a prism to separate light into its colors. Because a number of integers can satisfy the grating equation, there are a number of angles into which monochromatic light will be diffracted. This will be examined in **Project #5**. Therefore, when a grating is illuminated with white light, the light will be dispersed into a number of spectra corresponding to the integers $m = \ldots, \pm1, \pm2, \ldots$, as illustrated in **Fig. 0.20(a)**. By inserting a lens after the grating, the spectra can be displayed on a screen one focal length from the lens, **Fig. 0.20(b)**. These are called the orders of the grating and are labeled by the value of m.

0.5. Polarization

Since electric and magnetic fields are vector quantities, both their magnitude and direction must be specified. But, because these two field directions are always perpendicular to one another in non-absorbing media, the direction of the electric field of a light wave is used to specify the direction of polarization of the light. The kind and amount of polarization can be determined and modified to other types of polarization. If you understand the polarization properties of light, you can control the amount and direction of light through the use of its polarization properties.

0.5.1. Types of Polarization

The form of polarization of light can be quite complex. However, for most design situations there are a limited number of types that are needed to describe the polarization of light in an optical system. **Fig. 0.21** shows the path traced by the electric field during one full cycle of oscillation of the wave ($T = 1/\nu$) for a number of different types of polarization, where ν is the frequency of the light. **Fig. 0.21(a)** shows **linear polarization**, where orientation of the electric field vector of the wave does not change with time as the field amplitude oscillates from a maximum value in one direction to a maximum value in the opposite direction. The orientation of the electric field is referenced to some axis perpendicular to the direction of propagation. In some cases, it may be a direction in the laboratory or optical system, and it is specified as horizontally or vertically polarized or polarized at some angle to a coordinate axis.

Because the electric field is a vector quantity, electric fields add as vectors. For example, two fields, E_x and E_y, linearly polarized at right angles to each other and oscillating in phase (maxima for both waves occur at the same time), will combine to give another linearly polarized wave, shown in **Fig. 0.21(b)**, whose direction ($\tan\theta = E_y/E_x$) and amplitude ($\sqrt{E_x^2+E_y^2}$) are found by addition of the two components. If these fields are 90° out of phase (the maximum in one field occurs when the other field is zero), the electric field of the combined fields traces out an ellipse during one cycle, as shown in **Fig.0.21(c)**. The result is called **elliptically polarized** light. The eccentricity of the ellipse is the ratio of the amplitudes of the two components. If the two components are equal, the trace is a circle. This polarization is called **circularly polarized**. Since the direction of rotation of the vector depends on the relative phases of the two components, this type of polarization has a handedness to be specified. If the electric field coming from a source toward the observer rotates counterclockwise, the polarization is said to be **left handed**. **Right-handed polarization** has the opposite sense, clockwise. This nomenclature applies to elliptical as well as circular polarization. Light whose direction of polarization does not follow a simple pattern such as the ones described here is sometimes referred to as

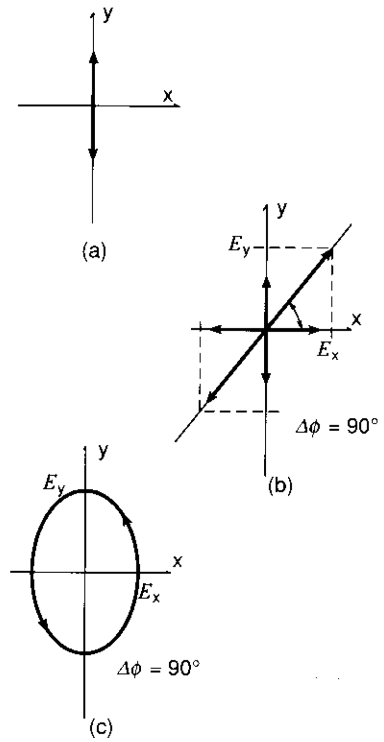


Figure 0.21. Three special polarization orientations: (a) linear, along a coordinate axis; (b) linear, components along coordinate axes are in phase ($\Delta\Phi = 0$) and thus produce linear polarization; (c) same components, 90° out of phase, produce elliptical polarization.

unpolarized light. This can be somewhat misleading because the field has an instantaneous direction of polarization at all times, but it may not be easy to discover what the pattern is. A more descriptive term is **randomly polarized light**.

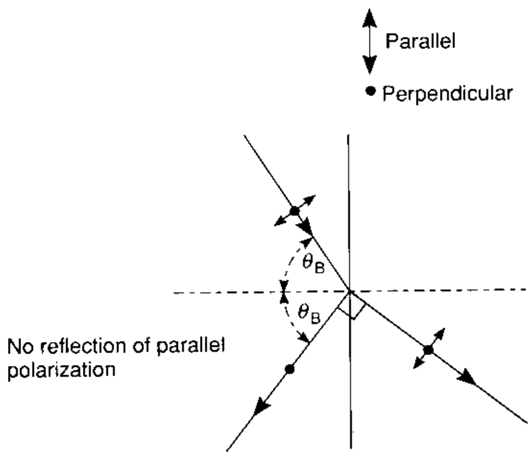


Figure 0.22. Geometry for the Brewster angle.

Light from most natural sources tends to be randomly polarized. While there are a number of methods of converting it to linear polarization, only those that are commonly used in optical design will be covered. One method is reflection, since the amount of light reflected off a tilted surface is dependent on the orientation of the incident polarization and the normal to the surface. A geometry of particular interest is one in which the propagation direction of reflected and refracted rays at an interface are perpendicular to each other, as shown in **Fig. 0.22**. In this orientation the component of light polarized parallel to the plane of incidence (the plane containing the incident propagation vector and the surface normal, i.e., the plane of the page for **Fig. 0.22**) is 100% transmitted. There is no reflection for this polarization in this geometry. For the component of light perpendicular to the plane of incidence, there is some light reflected and the rest is transmitted. The angle of incidence at which this occurs is called **Brewster's angle**, θ_B, and is given by:

$$\tan\theta_B = n_{trans}/n_{incident} \qquad (0\text{-}20)$$

As an example, for a crown glass, n = 1.523, and the Brewster angle is 56.7°. Measurement of Brewster's angle is part of **Project #8**.

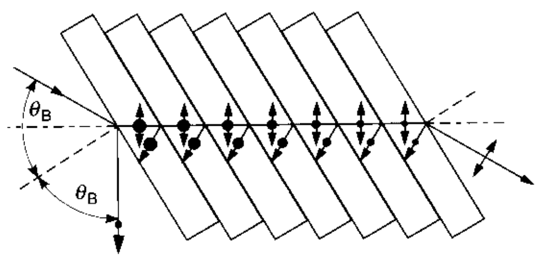


Figure 0.23. A "Pile of Plates" polarizer. This device working at Brewster angle, reflects some portion of the perpendicular polarization (here depicted as a dot, indicating an electric field vector perpendicular to the page) and transmits all parallel polarization. After a number of transmissions most of the perpendicular polarization has been reflected away leaving a highly polarized parallel component.

Sometimes only a small amount of polarized light is needed, and the light reflected off of a single surface tilted at Brewster's angle may be enough to do the job. If nearly complete polarization of a beam is needed, one can construct a linear polarizer by stacking a number of glass slides (e.g., clean microscope slides) at Brewster's angle to the beam direction. As indicated in **Fig. 0.23**, each interface rejects a small amount of light polarized perpendicular to the plane of incidence.

The "pile of plates" polarizer just described is somewhat bulky and tends to get dirty, reducing its efficiency. Plastic polarizing films are easier to use and mount. These films selectively absorb more of one polarization component and transmit more of the other. The source of this polarization selection is the aligned linear chains of a polymer to which light-absorbing iodine molecules are attached. Light that is polarized parallel to the chains is easily absorbed, whereas light polarized perpendicular to the chains is mostly transmitted. The sheet polarizers made by Polaroid Corporation are labeled by their type and transmission. Three common linear polarizers are

HN-22, HN-32, and HN-38, where the number following the HN indicates the percentage of incident unpolarized light that is transmitted through the polarizer as polarized light.

When you look through a crystal of calcite (calcium carbonate) at some writing on a page, you see a double image. If you rotate the calcite, keeping its surface on the page, one of the images rotates with the crystal while the other remains fixed. This phenomenon is known as double refraction. (Doubly refracting is the English equivalent for the Latin **birefringent.**) If we examine these images through a sheet polarizer, we find that each image has a definite polarization, and these polarizations are perpendicular to each other.

Calcite crystal is one of a whole class of birefringent crystals that exhibit double refraction. The physical basis for this phenomenon is described in detail in most optics texts. For our purposes it is sufficient to know that the crystal has a refractive index that varies with the direction of propagation in the crystal and the direction of polarization. The **optic axis** of the crystal (no connection to the optical axis of a lens or a system) is a direction in the crystal to which all other directions are referenced. Light whose component of the polarization is perpendicular to the optic axis travels through the crystal as if it were an ordinary piece of glass with a single refraction index, n_0. Light of this polarization is called an **ordinary ray**. Light polarized parallel to a plane containing the optic axis has a refractive index that varies between n_0 and a different value, n_e. The material exhibits a refractive index n_e where the field component is parallel to the optic axis and the direction of light propagation is perpendicular to the optic axis. Light of this polarization is called an **extraordinary ray**. The action of the crystal upon light of these two orthogonal polarization components provides the double images and the polarization of light by transmission through the crystals. If one of these components could be blocked or diverted while the other component is transmitted by the crystal, a high degree of polarization can be achieved.

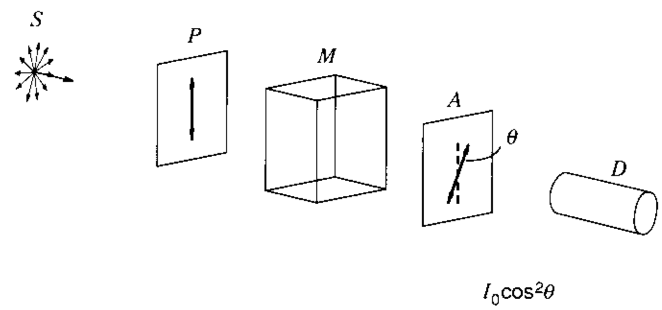


Figure 0.24. Analysis of polarized light. Randomly polarized light from source S is linearly polarized after passage through the polarizer P with irradiance I_0. After passage through optically active material M, the polarization vector has been rotated through an angle θ. (The dashed line of both polarizers A and P denote the transmission axes; the arrow indicates the polarization of the light.) The light is analyzed by polarizer A, transmitting an amount $I_0\cos^2\theta$ that is detected by detector D.

In many cases polarizers are used to provide information about a material that produces, in some manner, a change in the form of polarized light passing through it. The standard configuration, shown in **Fig. 0.24**, consists of a light source *S*, a polarizer *P*, the material M, another polarizer, called an analyzer *A*, and a detector *D*. Usually the polarizer is a linear polarizer, as is the analyzer. Sometimes, however, polarizers that produce other types of polarization are used.

The amount of light transmitted by a polarizer depends on the polarization of the incident beam and the quality of the polarizer. Let us take, for example, a perfect polarizer — one that transmits all of the light for one polarization and rejects (by absorption or reflection) all of the light of the other polarization. The direction of polarization of the transmitted light is the polarization axis, or simply the axis of the polarizer. Since randomly polarized light has no preferred polarization, there would be equal amounts of incident light for two orthogonal polarization directions. Thus, a perfect linear polarizer would have a Polaroid designation of HN-50, since it would pass half of the incident radiation and absorb the other half. The source in **Fig. 0.24** is randomly polarized, and the polarizer passes linearly polarized light of irradiance I_o. If the material M changes the incident polarization by rotating it through an angle θ, what is the amount of light transmitted through an analyzer whose transmission axis is oriented parallel to the axis of the first polarizer? Since the electric field is a vector, we can decompose it into two components, one parallel to the axis of the analyzer, the other perpendicular to this axis. That is

$$\tilde{E} = E_0 \cos\theta\, \hat{e}_{\parallel} + E_0 \sin\theta\, \hat{e}_{\perp} \tag{0-21}$$

(Note that the parallel and perpendicular components here refer to their orientation with respect to the axis of the analyzer and not to the plane of incidence as in the case of the Brewster angle.) The transmitted field is the parallel component, and the transmitted irradiance I_{trans} is the time average square of the electric field

$$I_{trans} = \langle E_0^2 \cos^2\theta \rangle = \langle E_0^2 \rangle \cos^2\theta$$

or

$$I_{trans} = I_0 \cos^2\theta \tag{0-22}$$

This equation, which relates the irradiance of polarized light transmitted through a perfect polarizer to the irradiance of incident polarized light, is called the **Law of Malus**, after its discoverer, Etienne Malus, an engineer in the French army. For a nonperfect polarizer, I_0 must be replaced by αI_0, where α is the fraction of the preferred polarization transmitted by the polarizer.

0.5.2. Polarization Modifiers

Besides serving as linear polarizers, birefringent crystals can be used to change the type of polarization of a light beam. We shall describe the effect that these polarization modifiers have on the beam and leave the explanation of their operation to a physical optics text.

In a birefringent crystal, light whose polarization is parallel to the optic axis travels at a speed of $c/n_{\parallel}$; for a polarization perpendicular to that, the speed is $c/n_{\perp}$. In calcite $n_{\perp} > n_{\parallel}$, and therefore the speed of light polarized parallel to the optic axis, $v_{\parallel}$, is greater than $v_{\perp}$.

Thus, for calcite, the optic axis is called the fast axis and a perpendicular axis is the slow axis. (In other crystals n may be greater than $n_{\perp}$ and the fast-slow designation would have to be reversed.)

The first device to be described is a **quarter-wave plate**. The plate consists of a birefringent crystal of a specific thickness d, cut so that the optic axis is parallel to the plane of the plate and perpendicular to the edge, as shown in **Fig.0.25**. The plate is oriented so that its plane is perpendicular to the beam direction and its fast and slow axes are at 45° to the polarized direction of the incident linearly polarized light. Because of this 45° geometry, the incident light is split into slow and fast components of equal amplitude traveling through the crystal. The plate is cut so that the components, which were in phase at the entrance to the crystal, travel at different speeds through it and exit at the point when they are 90°, or a quarter wave, out of phase. This output of equal amplitude components, 90° out of phase, is then circularly polarized. It can be shown that when circularly polarized light is passed through the same plate, linearly polarized light results. Also, it should be noted that if the 45° input geometry is not maintained, the output is elliptically polarized. The angle between the input polarization direction and the optic axis determines the eccentricity of the ellipse.

If a crystal is cut that has twice the thickness of the quarter-wave plate, one has a **half-wave plate**. In this case, linearly polarized light at any angle θ with respect to the optic axis provides two perpendicular components which end up 180° out of phase upon passage through the crystal. This means that relative to one of the polarizations, the other polarization is 180° from its original direction. These components can be combined, as shown in **Fig. 0.26**, to give a resultant whose direction has been rotated 2θ from the original polarization. Sometimes a half-wave plate is called a **polarization rotator**. It also changes the "handedness" of circular polarization from left to right or the reverse. This discussion of wave plates assumes that the crystal thickness d is correct only for the wavelength of the incident radiation. In practice, there is a range of wavelengths about the correct value for which these polarization modifiers work fairly well.

Waveplates provide good examples of the use of polarization to control light. One specific demonstration that you will perform as part of **Project #9** concerns reflection reduction. Randomly polarized light is sent through a polarizer and then through a quarter wave plate to create circularly polarized light, as noted above. When circularly polarized light is reflected off a

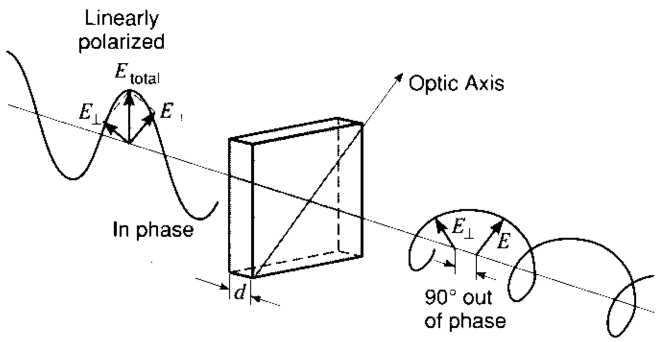


Figure 0.25. Quarter-wave plate. Incident linearly polarized light is oriented at 45° to the optic axis so that equal $E_{\parallel}$ and $E_{\perp}$ components are produced. The thickness of the plate is designed to produce a phase retardation of 90° of one component relative to the other. This produces circularly polarized light. At any other orientation elliptically polarized light is produced.

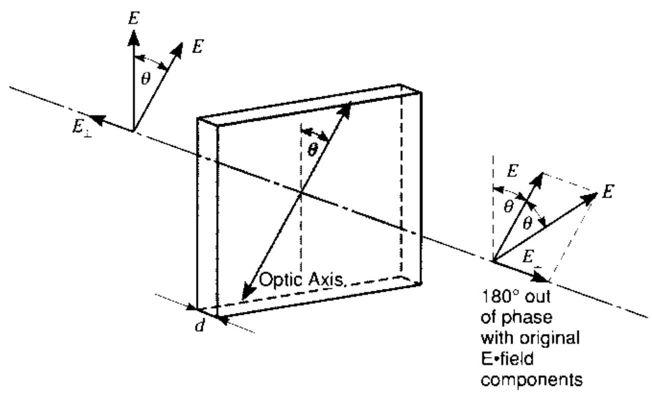


Figure 0.26. Half-wave plate. The plate produces a 180° phase lag between the $E_{\parallel}$ and $E_{\perp}$ components of the incident linearly polarized light. If the original polarization direction is at an angle θ to the optic axis, the transmitted polarization is rotated through 2θ from the original.

surface, its handedness is reversed (right to left or left to right). When the light passes through the quarter wave plate a second time, the circularly polarized light of the opposite handedness is turned into linearly polarized light, but rotated 90° with respect to the incident polarization. Upon passage through the linear polarizer a second time, the light is absorbed. However, light emanating from behind a reflective surface (computer monitor, for example) will not be subject to this absorption and a large portion will be transmitted by the polarizer. A computer anti-reflection screen is an application of these devices. Light from the room must undergo passage through the polarizer-waveplate combination twice and is, therefore suppressed, whereas light from the computer screen is transmitted through the combination but once and is only reduced in brightness. Thus, the contrast of the image on the computer screen is enhanced significantly using this polarization technique.

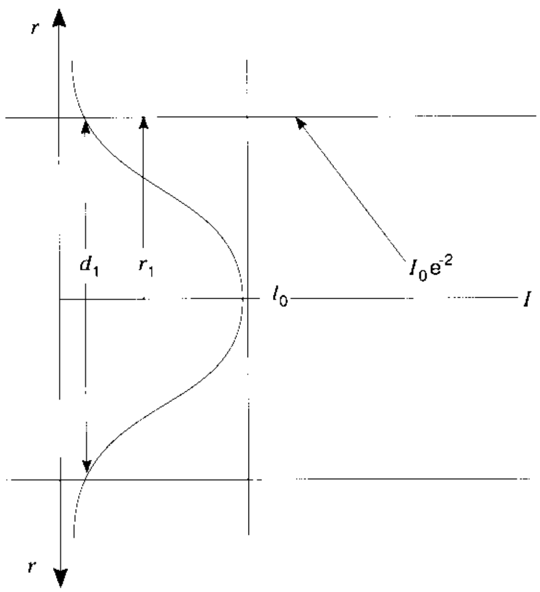


Figure 0.27. Gaussian beam profile. Plot of irradiance versus radial distance from the beam axis. [*Elements of Modern Optical Design*, Donald C. O'Shea, copyright ©, J. Wiley & Sons, 1985. Reprinted by permission of John Wiley & Sons, Inc.]

0.6 Lasers

The output of a laser is very different than most other light sources. After a description of the simplest type of beam, the TEM_{00} mode Gaussian beam and its parameters, we look at means of collimating the beam. The effect of a laser's construction on its output and a method by which this output can be measured will be discussed.

0.6.1. Characteristics of a Gaussian Beam

The term **Gaussian** describes the variation in the irradiance along a line perpendicular to the direction of propagation and through the center of the beam, as shown in **Fig. 0.27**. The irradiance is symmetric about the beam axis and varies radially outward from this axis with the form

$$I(r) = I_0 e^{-2r^2/r_1^2} \qquad (0\text{-}23)$$

or in terms of a beam diameter

$$I(d) = I_0 e^{-2d^2/d_1^2}$$

where r_1 and d_1 are the quantities that define the radial extent of the beam. These values are, by definition, the radius and diameter of the beam where the irradiance is $1/e^2$ of the value on the beam axis, I_0.

0.6.1.1. Beam Waist and Beam Divergence

Figure 0.27 shows a beam of parallel rays. In reality, a Gaussian beam either diverges from a region where the beam is smallest, called the **beam waist**, or converges to one, as shown in **Fig. 0.28**. The amount of divergence or convergence is measured by the **full angle beam divergence** θ, which is the angle subtended by the $1/e^2$ diameter points for distances far from the beam waist as shown in **Fig. 0.28**. In some laser texts and articles, the angle is measured from the beam axis to the $1/e^2$ asymptote, a *half* angle. However, it is the full angle divergence, as defined here, that is usually given in the specification sheets for most lasers. Because of symmetry on either side of the beam waist, the convergence angle is equal to the divergence angle. We will refer to the latter in both cases.

Under the laws of geometrical optics a Gaussian beam converging at an angle of θ should collapse to a point. Because of diffraction, this, does not occur. However, at the intersection of the asymptotes that define θ, the beam does reach a minimum value d_0 the **beam waist diameter**. It can be shown that for a TEM_{00} mode d_0 depends on the beam divergence angle as:

$$d_0 = \frac{4\lambda}{\pi\theta} \qquad (0\text{-}24)$$

where λ is the wavelength of the radiation. Note that for a Gaussian beam of a particular wavelength, the product $d_0\theta$ is constant. Therefore for a very small beam waist the divergence must be large, for a highly collimated beam (small θ), the beam waist must be large.

The variation of the beam diameter in the vicinity of the beam waist is shown in **Fig. 0.28** and given as

$$d^2 = d_0^2 + \theta^2 z^2 \qquad (0\text{-}25)$$

where d is the diameter at a distance $\pm z$ from the waist along the beam axis.

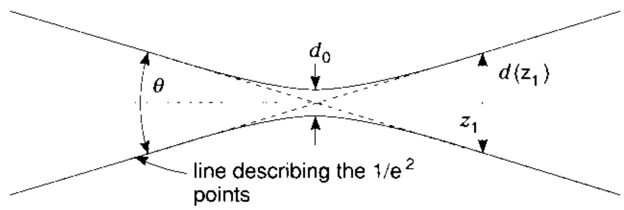


Figure 0.28. Variation of Gaussian beam diameter in the vicinity of the beam waist. The size of the beam at its smallest point is d_0; the full angle beam divergence, defined by the smallest asymptotes for the $1/e^2$ points at a large distance from the waist is θ.

0.6.1.2. The Rayleigh Range

It is useful to characterize the extent of the beam waist region with a parameter called the **Rayleigh range**. (In other descriptions of Gaussian beams this extent is sometimes characterized by the confocal beam parameter and is equal to twice the Rayleigh range.) Rewriting **Eq. 0.25** as

$$d = d_0\sqrt{1 + (\theta z / d_0)^2} \qquad (0\text{-}26)$$

we define the Rayleigh range as the distance from the beam waist where the diameter has increased to $d_0\sqrt{2}$. Obviously this occurs when the second term under the radical is unity, that is, when

$$z = z_R = d_0/\theta \qquad (0\text{-}27)$$

Although the definition of z_R might seem rather arbitrary, this particular choice offers more than just convenience. **Figure 0.29** shows a plot of the radius of curvature of the wavefronts in a Gaussian beam as a function of z. For large distances from the beam waist the wavefronts are nearly planar, giving values tending

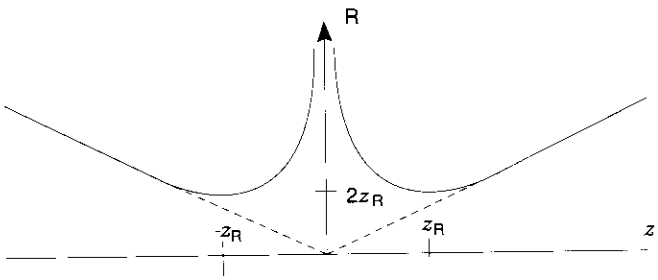


Figure 0.29. Plot of radius of curvature (*R*) versus distance (*z*) from the beam waist. The absolute value of the radius is a minimum at the Rayleigh range point, z_R. In the limit of geometrical optics, the radius of curvature of the wavefronts follows the dashed line.

toward infinity. At the beam waist the wavefronts are also planar, and, therefore, the absolute value of the radius of curvature of the wavefronts must go from infinity at large distances through a minimum and return to infinity at the beam waist. This is also true on the other side of the beam waist but with the opposite sign. It can be shown that the minimum in the absolute value of the radius of curvature occurs at $z = \pm z_R$, that is, at a distance one Rayleigh range either side of the beam waist. From **Fig. 0.29**, the "collimated" region of Gaussian beam waist can be taken as $2z_R$.

The Rayleigh range can be expressed in a number of ways:

$$z_R = \frac{d_0}{\theta} = \frac{4\lambda}{\pi\theta^2} = \frac{\pi d_0^{\ 2}}{4\lambda} \qquad (0\text{-}28)$$

From this we see that all three characteristics of a Gaussian beam are dependent on each other. Given any of the three quantities, d_0, θ, z_R, and the wavelength of the radiation, the behavior of the beam is completely described. Here, for example, if a helium-neon laser ($\lambda = 633$ nm) has a specified TEM_{00} beam diameter of 1 mm, then

$\theta = 4\lambda/\pi d_0 = (1.27 \times 6.33 \times 10^{-7}\text{m})/(1 \times 10^{-3}\text{m}) = 0.8$ mrad

and

$z_R = d_0/\theta = (1 \times 10^{-3}\text{m})/(0.8 \times 10^{-3}\text{rad}) = 1.25$ m.

The Rayleigh range of a typical helium-neon laser is considerable.

0.6.2 Collimation of a Laser Beam

Through the use of lenses the divergence, beam waist, and Rayleigh range of the Gaussian beam can be changed. However, from the above discussion it is clear that the relations between the various beam parameters cannot be changed. Thus, to increase the collimation of a beam by reducing the divergence requires that the beam waist diameter be increased, since the beam waist diameter-divergence product is constant. This is done by first creating a beam with a strong divergence and small beam waist and then putting the beam waist at the focal point of a long focal length lens. What this amounts to is putting the beam through a telescope — backward. The laser beam goes in the eyepiece lens and comes out the objective lens.

There are two ways of accomplishing this. One uses a Galilean telescope, which consists of a negative eyepiece lens and a positive objective lens, as shown in **Fig. 0.30(a)**. The light is diverged by the negative lens producing a virtual beam waist and the objective lens is positioned at a separation equal to the algebraic sum of the focal lengths of the lenses to produce a more collimated beam. It can be shown that the decrease in the divergence is equal to the original divergence divided by the magnification of the telescope. The magnification of the telescope is equal to the ratio of the focal lengths of the objective divided by the eyepiece. The second method uses a Keplerian telescope (**Fig. 0.30(b)**). The eyepiece lens is a positive lens so the beam comes to a focus and then diverges to be collimated by the objective lens.

Project #3 will demonstrate the design of these two types of laser beam expanders. Each has distinct advantages. The advantage of the Galilean type of beam expander occurs for high power or pulsed laser systems. Since the beam does not come to a focus anywhere inside of the beam expander's optical path, the power density of the beam decreases. Thus, if the lenses and environment can survive the initial beam, they can survive the beam anywhere in the optical path. Although the Keplerian beam expander can give similar ratios of beam expansion, the power density at the focus of the first lens is very large. In fact, with a high power, pulsed laser it is possible to cause a breakdown of the air in the space between the lenses. This breakdown is caused by the very strong electrical field that results from focusing the beam to a small diameter creating miniature lightning bolts. (Many researchers have been unpleasantly surprised when these "miniature" lightning bolts destroyed some very expensive optics!)

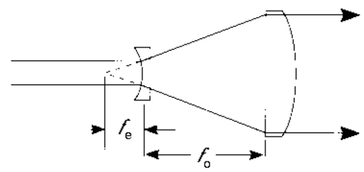
The primary advantage for the Keplerian beam expander is that a pinhole of an optimum diameter can be placed at the focus of the first lens to "clean" up the laser beam by rejecting the part of the laser energy that is outside of the pinhole diameter. This concept of "**spatial filtering**" will be explored in **Project #10**.

0.6.3 Axial Modes of a Laser

The properties of laser light include monochromaticity, low divergence (already explored in the previous sections), and a high degree of coherence, which encompasses both of these properties. This section is a discussion of the coherence of the laser and a historical experiment that illustrates one of the concepts using a modern device.

A complete discussion of the principle of laser action would take a substantial amount of space and reading time. For an explanation of the concept we refer you to the references. The basis of lasers is a physical process called stimulated emission. It appears as the third and fourth letters of the acronym, **LASER** (**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation). Amplification is only the beginning of the process in most lasers,

(a)



(b)

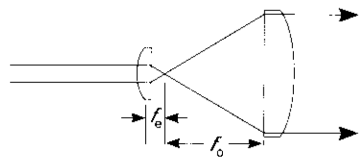


Figure 0.30. Gaussian beam collimation. (a) Galilean telescope. (b) Keplerian telescope. Eyepiece focal length, f_e; objective focal length, f_o.

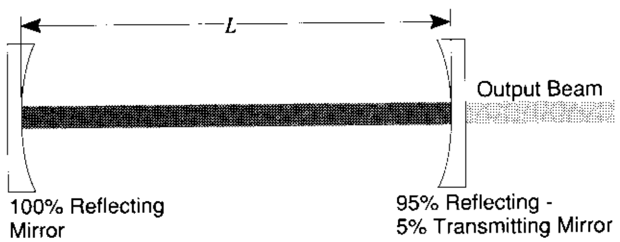


Figure 0.31. The laser cavity. The distance between mirrors is an important parameter in the output of a laser.

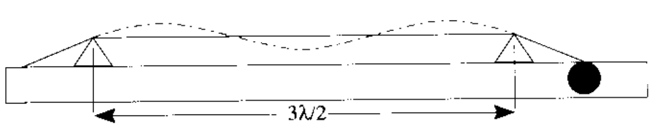


Figure 0.32. Standing wave picture.

since the increase in light as it passes through an amplifying volume is usually quite modest. If the radiation was only amplified during a single pass through the volume, it would be only marginally useful. However, when mirrors are placed at both ends of the amplifying medium, the light is returned to the medium for additional amplification **(Fig. 0.31)**. The useful output from the laser comes through one of the mirrors, which reflects most of the light, but transmits a small fraction of the light, usually on the order of 5% (up to 40% for high power lasers). The other mirror is totally reflecting. But the laser mirrors do more than confine most of the light. They also determine the distribution of wavelengths that can support amplification in the laser.

The mirrors serve as a simple, but effective, interferometer and for only certain wavelengths, just as in the case of the Michelson interferometer, will there be constructive interference. The mirrors form a resonant structure that stores or supports only certain frequencies. It is best compared to the resonances of a guitar string in which the note that the string produces when plucked is determined by the length of the string. By changing the location of the finger on a guitar string a different note is played. The note (really, notes) is determined by the amount of tension the guitarist has put on the string and the length of the string. Any fundamental physics book will show that the conditions imposed upon the string of length L will produce a note whose wavelength is such that an integral number of half wavelengths is equal L,

$$q\lambda/2 = L. \qquad (0\text{-}29)$$

In **Fig. 0.32** a **standing wave** with three half wavelengths is shown. In most lasers, unless special precautions are taken, a number of wavelengths will satisfy this resonance condition. These wavelengths are referred to as the **axial modes** of the laser. Since $L = q\lambda/2$, where q is an integer, the wavelengths supported by the laser are

$$\lambda_q = 2L/q \qquad (0\text{-}29a)$$

The frequencies of these modes are given by $\nu = c/\lambda$, where c is the speed of light.

Inserting the expression for the wavelengths, the frequencies of the resonant modes are given by

$$\nu_q = q\,(c/2L),$$

where q is an integer. The frequency separation between these axial modes equals the difference between modes whose integers differ by one:

$$\Delta n = \nu_{q+1} - \nu_q = (q+1)c/2L - qc/2L = c/2L, \qquad (0\text{-}30)$$

so the separation between neighboring modes of a laser is constant and dependent only on the distance between the mirrors in the laser, as shown in **Fig. 0.33**. Since the amount of power obtained from small helium-neon lasers, such as those used for the projects described in this manual, is related to the **length** of the laser, the separation between mirrors is set by the laser manufacturers to produce the required power for the laser. But the band of wavelengths that can maintain stimulated emission is determined by the atomic physics of the lasing medium, in this case, neon. That band does not change radically for most helium-neon laser tubes. Therefore, the **number** of axial modes is mainly dependent on the distance between the mirrors, L. The farther apart the mirrors are, the closer are the axial mode frequencies. Thus, long high power helium-neon lasers have a large number of axial modes, whereas, the modest power lasers used in this **Projects in Optics** kit produce only a small number (usually three) of axial modes.

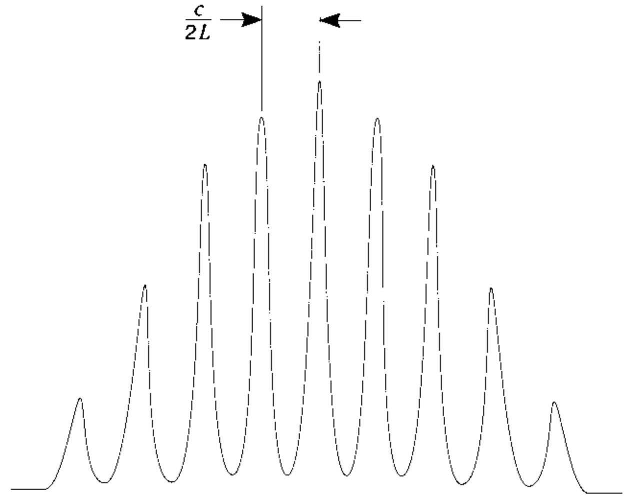


Figure 0.33. Laser mode distribution. Plot of power in laser output as a function of frequency.

One of the other relations between neighboring laser modes, beside their separation, is that their polarization is orthogonal (crossed) to that of their neighbors (**Fig. 0.34**). Thus, if we examined a three-mode laser with the appropriate tools, we would expect to find that two of the modes would have one polarization and the other would have a perpendicular polarization. This means that, while axial modes are separated in frequency by $c/2L$, modes of the **same polarization** are separated by c/L.

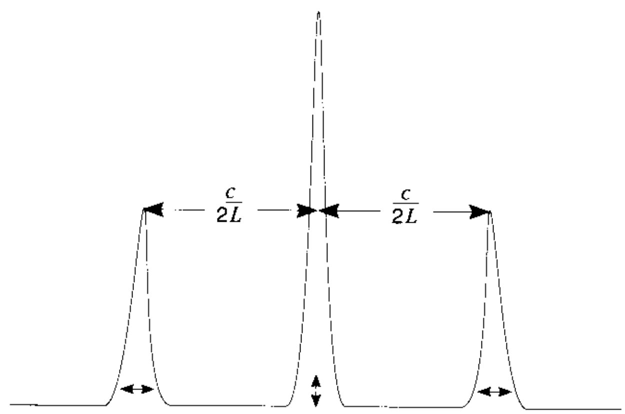


Figure 0.34. Output from a three mode laser. The relative polarization of each mode is indicated at its base.

Looking through a diffraction grating at the output of a three-mode laser, we see a single color. High resolution interferometers must be employed to display the axial modes of a laser. However, it is also possible to use a Michelson interferometer to investigate the modes without resorting to high resolution devices. This technique has special applications in the infrared region of the spectrum.

0.6.4 Coherence of a Laser

If we speak of something as being "coherent" in everyday life, we usually mean that it, a painting, a work of music, a plan of action, "makes sense." It "hangs together." There is in this concept the idea of consistency and predictability. The judgement of what is coherent, however, is one of individual taste. What one person may find coherent in heavy metal rock music, another person would hear in rhythm and blues ... or elevator music, perhaps. This concept of coherence as a predictable, consistent form of some idea or work of art has much the same meaning when applied to light sources. How consistent is a light field from one point

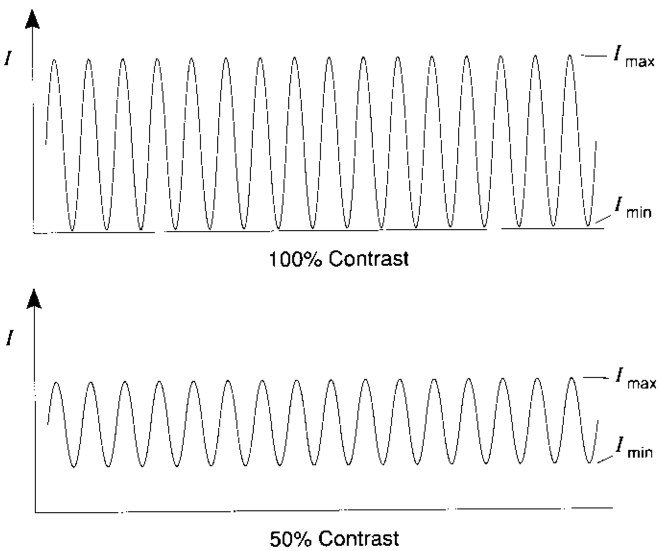


Figure 0.35. Contrast.

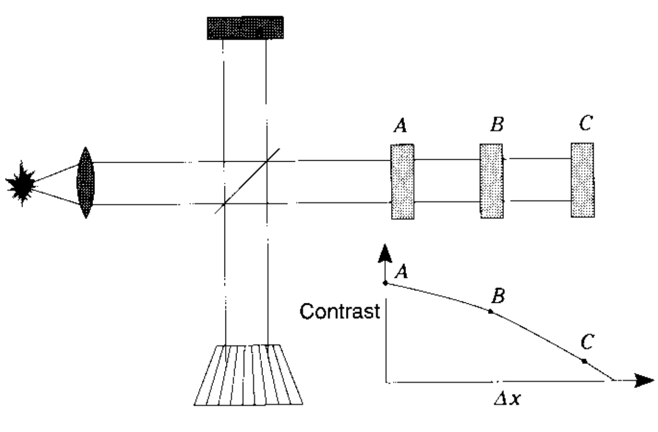


Figure 0.36. Visibility function.

to another? How do you make the comparison? The interference of the light beam with itself does the comparing. If there is a constant relation between one point on a laser beam and another point, then the interference of waves separated by that distance should produce a stable interference pattern. If, however, the amplitude or phase or wavelength changes between these two points, the interference, while it is still there at all times, will constantly vary with time. This unstable interference pattern may still exhibit fringes, but the fringes will be washed out. This loss of visibility of fringes as a function of the distance between the points of comparison is measure of the coherence of the light. This visibility can be measured by the **contrast** of the interference fringes. The contrast is defined by

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \qquad \text{(0-31)}$$

where I_{max} is the irradiance of the bright interference fringes and I_{min} is the irradiance of the dark interference fringes (**Fig. 0.35**). This contrast is determined by passing the light from the source through a Michelson interferometer with unequal arms. By changing the path length difference between the arm in the interferometer, the visibility of the fringes as a function of this difference can be recorded. From these observations, the measurement of the coherence of a source can be done using a Michelson interferometer.

If a source were absolutely monochromatic, there would be no frequency spread in its spectrum. That is, its frequency bandwidth would be zero. For this to be true, all parts of the wave exhibit the same sinusoidal dependence from one end of the wave to the other. Thus, a truly monochromatic wave would never show any lack of contrast in the fringes, no matter how large of a path length difference was made. But all sources, even laser sources contain a distribution of wavelengths. Therefore, as the path length difference is increased, the wavefront at one point on the beam gets out of phase with another point on the beam. A measure of the distance at which this occurs is the coherence length l_c of a laser. It is related to the frequency bandwidth of a laser by

$$\Delta \nu = c / l_c \qquad \text{(0-32)}$$

Any measurement of the coherence length of a light source by observation of the visibility of fringes from a Michelson interferometer will yield information on the bandwidth of that source and, therefore, its coherence. For example, suppose the source is a laser with some broadening. As the length of the one of the arms in a Michelson interferometer, as shown in **Fig. 0.36**, becomes unequal (mirror moved from A to B), the one part of a wave will interfere with another part that is delayed by a time equal to the difference in path length divided by the speed of light. Eventually the waves begin to get out of step and the fringe contrast begins to fall because the phase relations between the two waves is varying slightly due to the spread in frequencies in the light. The greater the broadening, the more rapidly the visibility of the fringes will go to zero.

One particularly interesting case consists of a source with only a few modes present as is the case for the three-mode helium-neon laser discussed above. Because only light of the same polarization can interfere, there will be two modes (λ_1, λ_3) in the laser that can interfere with each other. The third mode (λ_2) with orthogonal polarization is usually eliminated by passing the output of the laser through a polarizer. With the interferometer mirrors set at equal path length there are two sets of fringes, one from each mode. Since the path length difference is zero, these two sets of high contrast fringes overlap each other. But as the path length increases, the fringes begin to get out of step. Until, finally, the interference maximum of one set of fringes overlaps the interference minimum of the other set of fringes and the fringe contrast goes to zero. The calculation of this condition is fairly simple. The condition for an interference maximum is given by

$$L_1 - L_2 = m\lambda/2 \quad m = \text{an integer} \tag{0-33}$$

and for an interference minimum by

$$L_1 - L_2 = m\lambda/4 \quad m = \text{odd integers} \tag{0-34}$$

If we assume that the change in path length is from zero path length to the point where the visibility first goes to zero, then for one wavelength, λ_1,

$$L_1 - L_2 = m\lambda_1/2 \quad m = \text{an integer} \qquad (0\text{-}35)$$

and for the other mode with the same polarization, there is a minimum.

$$L_1 - L_2 = m\lambda_3/2 + \lambda_3/4 \qquad (0\text{-}35)$$

Equating these two expressions and rearranging terms, gives

$$m\lambda_1/2 - m\lambda_3/2 = m(\lambda_1 - \lambda_3)/2 = \lambda_3/4. \qquad (0\text{-}35)$$

or

$$m\Delta\lambda = \lambda_3/2$$

Wavelength separation can be expressed as a frequency separation by Δv

$$\Delta\lambda = \lambda\Delta v/v \tag{0-36}$$

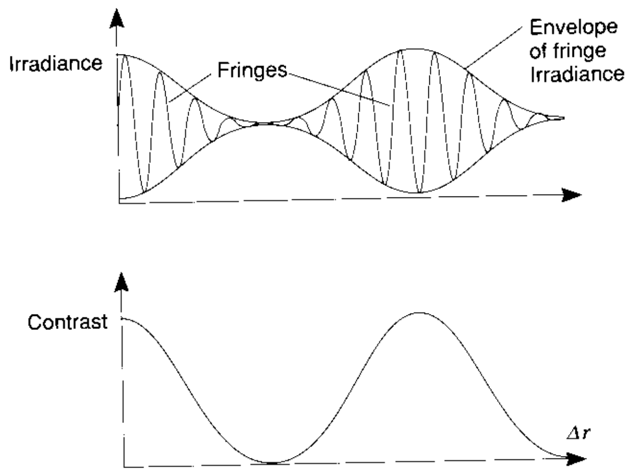


Figure 0.37. Visibility function for two mode system.

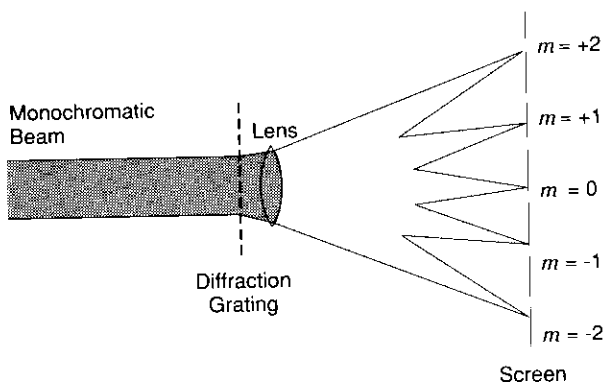


Figure 0.38. Diffraction orders.

where λ and v are the average values in the intervals $\Delta\lambda$ and Δv. Inserting this expression for $\Delta\lambda$, we obtain

$$\Delta v = v/2m\,. \tag{0-37}$$

The integer m is an extremely large number in most cases and is not easily determined, but it is related to the average wavelength of the source by $L_1 - L_2 = m\lambda/2$. If we set $\Delta L = L_1 - L_2$, solve for m and insert in the expression for Δv,

$$\Delta v = v/2m = \lambda v/4\Delta L = c/4\Delta L, \tag{0-38}$$

since $\lambda v = c$.

Thus the frequency separation between modes can be measured by determining the path length difference when the two interference fringe patterns are out of step with one another, causing the visibility to go to zero, as depicted in **Fig. 0.37**. It can also be demonstrated that there are additional minima in the visibility at $\Delta v = 3c/4\Delta L$, $5c/4\Delta L$, etc. Visibility maxima occur halfway between these minima as the two fringes patterns get back into step. In **Project#7**, this effect will enable you to determine the mode separations for the laser used in these projects. What has been derived here is a simple case of a much more involved application of this technique. It is possible to measure the fringe contrast as a function of mirror position (called an interferogram) and store it in the memory of a computer. It has been shown that a mathematical transformation (the same Fourier Transform that will be discussed in the next section) of the visibility function yields the frequency spectrum of the source.

While it might be considered difficult, the advent of powerful computers has reduced the cost and enhanced the utility of this technique, particularly in the far infrared part of the spectrum. These devices are known as Fourier transform spectrometers.

0.7 Abbe Theory of Imaging

The earlier discussion of imaging depended upon tracing a series of rays to determine the location and size of the image. It was shown that only a few rays were needed. This approach ignores the possibilities that the source could be monochromatic and sufficiently coherent that diffraction and interference effects could play a part in the formation of an image. What we will describe and then demonstrate in **Project #10**, is that after the light that will form an image has traversed the lens, we can intervene and change the image in very special ways. This approach to imaging has found use in a number of applications in modern optics. To begin to understand this concept, we need to review briefly the diffraction grating discussed in

Section 0.4.3, since the grating is one of the simplest illustrations of this new way of thinking about imaging. Consider a diffraction grating consisting of a series of equally spaced, narrow absorbing and transmitting (black and white) bands. It is possible to determine mathematically not only the directions of the diffracted orders

$$\sin\theta_m = m\,\lambda/d \qquad m = \text{an integer,} \tag{0-39}$$

but also the relative irradiances of the diffracted spots to one another. If we insert a lens after the diffraction grating, we can relocate the orders of the diffraction grating from infinity to the back focal plane of the lens **(Fig. 0.38)**. We will see how this can be used to understand imaging.

0.7.1 Spatial Frequencies

We are used to the idea of repetitions in time. Electrical and audio sources of signals with single frequencies, particularly as they relate to sound are used to test equipment for their response. A good high fidelity system will reproduce a wide range of frequencies ranging from the deep bass around 20 Hz (cycles per second), that is as much felt as it is heard, to the nearly impossible to hear 15,000 Hz, depending on how well you have treated your ears during life. As noted earlier, the frequency of the electromagnetic field determines whether the radiation is visible to the eye. Again, this periodic variation in the electric field takes place in time. Just as it is possible to speak about variation of electrical waves and sound with time, in optics, variations in space can be expressed as **spatial frequencies**. These are usually expressed in cycles/mm (or mm^{-1}). They indicate the rapidity with which an object or image varies in space instead of time. An example that shows a number of spatial frequencies is given in **Fig. 0.39.**

Figure 0.39. Spatial frequencies in an object.

As in the case of many sounds and electrical signals, most spatially repetitive patterns do not consist of a single frequency, but as a musical chord, are made up of some fundamental frequency plus its overtones, or higher harmonics. The discussion of spatial frequencies in optics is based on some interesting, but relatively complicated mathematics. You may want to read this section once to get the general ideas, then come back later after you have done **Project #10**. Certainly, here is a case where hands-on work will improve your understanding of the discussion of the subject.

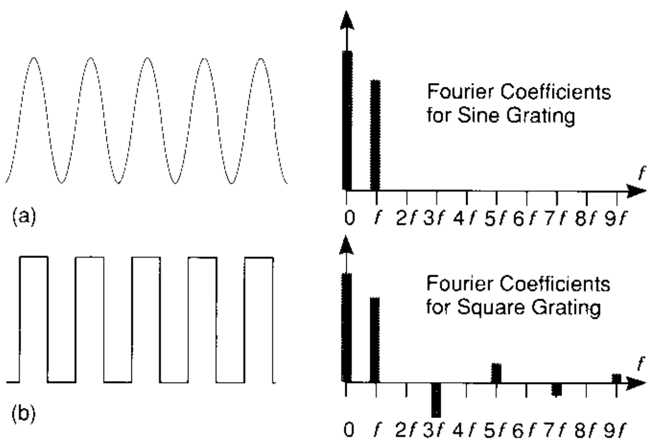


Figure 0.40. Sinusoidal grating versus black and white grating (Fourier analysis).

An example of an object with a few spatial frequencies is the diffraction grating. If the grating just discussed consisted of a sinusoidal variation, as shown in **Fig. 0.40(a)**, there would only be a zero order and the first orders ($m = \pm 1$). As repetitive patterns depart from sinusoidal, additional diffraction orders appear and in the case of the black and white grating, a whole series of diffraction orders are present **(Fig. 0.40(b))**.

All of this can be expressed mathematically in terms of **Fourier (Four-ee-ay) Theory**. We will not go into the mathematical expression of the theory, but only graphically express the result as simply as possible.

Any periodic (repeating) function can be expressed as a series of sine and cosine functions consisting of the fundamental periodic frequency (f) and its higher harmonics (those frequencies that are multiples of the fundamental frequency, f ($2f$, $3f$, $4f$, ...) The amount that each frequency contributes to the original function can be calculated using some standard integral calculus expressions. The decomposition of the periodic pattern into its harmonics is referred to as **Fourier Analysis**. This analysis determines the amplitude of each harmonic contribution to the original function and its phase relative to the fundamental (in phase or 180° out of phase).

The procedure can be, in a sense, reversed. If a pattern at the fundamental frequency is combined with the appropriate amounts of the higher harmonics, it is possible to approximate any function with a repetition frequency of the fundamental. This is referred to as **Fourier Synthesis**. To completely synthesize a function such as our example of an alternating black and white grating, an infinite number of harmonics would be needed. If only frequencies up to some specific value are used, the synthesized function will resemble the function, but it will have edges that are not as sharp as the original. A simple example **(Fig. 0.41)** using only a fundamental and two harmonics shows the beginning of the synthesis of a square wave function, similar to our black and white grating. What you will be investigating in **Project #10** are optical techniques that use Fourier analysis and synthesis in creating images.

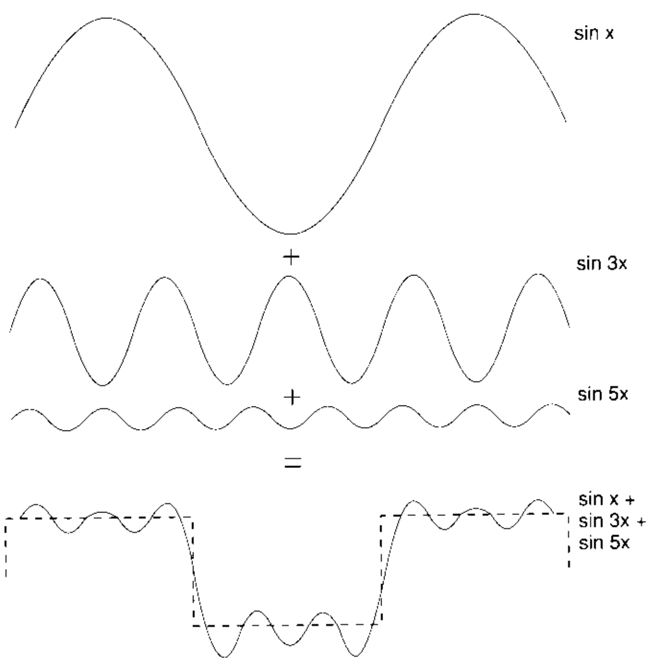


Figure 0.41. Fourier synthesis.

0.7.2 Image Formation

If the black and white grating is illuminated with plane waves of monochromatic light, a number of diffraction orders will be generated by the grating. These plane wave beams diffracted at different angles given by Eq. 0-39, can be focused with a lens located behind the diffraction grating, as shown in **Fig. 0.42**. The focused spots have intensities that are proportional to the square of the amplitudes that we could calculate for this diffraction grating. In effect, the laser plus lens combination serves as an optical Fourier analyzer for a diffractive object.

When a laser beam illuminates a grating, the light will diffract and the spectrum of spatial frequencies will be displayed in the back focal plane of the lens. It turns out that even if the object is not a grating or a series of lines with a number of repetitive spacings, the light pattern in the back focal plane still describes the content of spatial frequencies found in the object. Objects that are large and smoothly varying in their shading represent low spatial frequencies and will not diffract the beam much. Their contributions, therefore, lay close to the optical axis of the lens. Objects that are small or have fine detail and sharp edges will cause substantial diffraction and their contributions will be found further from the optical axis in the back focal plane of the lens.

Suppose as an object we use a fine, square mesh wire screen. This is a pair of crossed gratings only coarser than the diffraction grating just discussed earlier (**Fig. 0.42**). When this square grating is illuminated by a laser beam, the diffraction orders are focused to a series of spots at the back of the focal plane of the lens and the spatial frequencies form a two-dimensional grid of points. The separation between the points is determined by the distance between neighboring wires, representing the grating constant for this coarse grating.

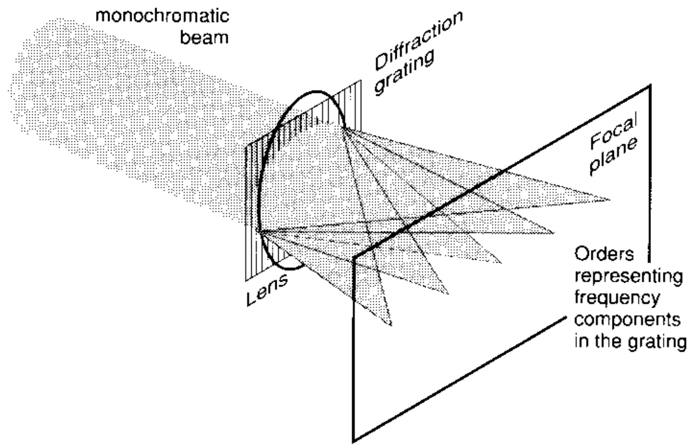


Figure 0.42. Optical Fourier analysis.

If the lens is more than one focal length from the grating, then somewhere beyond the Fourier transform plane, a real image of the grating will be formed. The geometry is shown in **Fig. 0.43**. The interesting point of this arrangement is that the image can be understood as a light distribution that arises out of the interference of the light from the spatial frequency components of the object located at the back focal plane of the lens. In other words, the image is a Fourier synthesis of the the spatial frequencies in the grating.

This imaging can be thought of as a two step process: Fourier analysis followed by Fourier synthesis. This approach to analyzing images was first proposed by Ernst Abbe, a lecturer at the University of Jena, who was hired by Carl Zeiss to understand and design microscope lenses. After some study one sees that the larger the collection angle of an imaging lens, the higher the possible resolution, since the higher spatial frequency components appear farther from the axis. Although very little light is collected near the edge of the lens, it is this light that contributes to the fine detail in high resolution images.

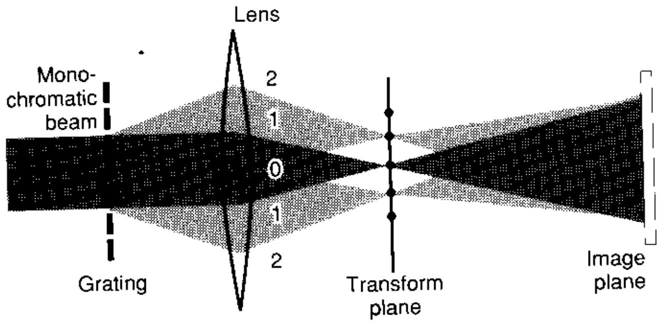


Figure 0.43. Abbe theory of imaging.

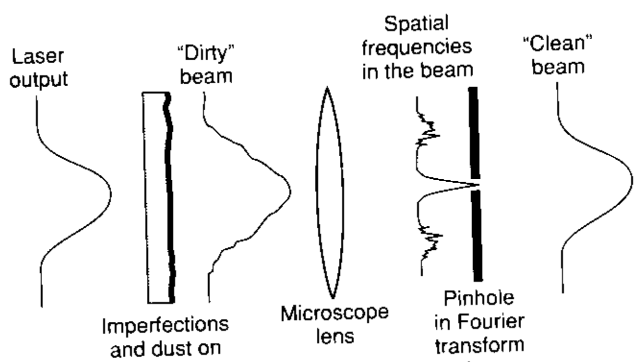


Figure 0.44. Spatial filtering.

0.7.3 Spatial Filtering

Since the light in the Fourier transform plane (**Fig 0.43**) is arranged according to increasing spatial frequency with radius, then any intervention in that plane in the form of a mask will change the distribution of spatial frequencies in the plane. It will also change the content of the image, but in a very predictable way.

The procedure of modifying an image by "changing" the spatial frequencies contained in it is called **spatial filtering**. One example of such a procedure is the spatial filtering of a transparency of a picture of a television screen. The Fourier transform of the picture is a rather raggedy-looking patch at the center of the picture and a series of equally spaced dots arranged in a vertical line. These dots represent a periodic, grating-like feature in the picture. This grating is due to the series of parallel lines, called a **raster**, that is used to build up an image on the TV screen. The electron beam that writes on the face of the tube in a TV set does so as a series of parallel lines. By turning the beam on and off as it is swept across the screen and dropping down a little on each sweep, the circuitry builds up a picture on the tube. If you look at a TV screen up close you can see the raster. If the dots represent the raster, where is the rest of the image? It resides in the raggedy-looking patch at the center of the beam. This analysis and synthesis process of imagery will be demonstrated as one of the experiments in **Project #10**.

There are a number of applications that are based on this approach to imaging. One of these "cleans up" a laser beam. The irradiance distribution of the beam in many lasers is Gaussian (Section 0.6.1) as it exits the output mirror of the laser. However, dust and small imperfections in the lenses, windows, and surfaces that it traverses or reflects from can produce irregularities in the irradiance pattern. The Gaussian distribution represents a low frequency spatial variation in the beam, whereas the irregularities contain higher spatial frequencies. When the laser beam is focused with a microscope objective, as shown in **Fig. 0.44**, these variations are arranged according to their spatial frequencies. If a small pinhole, whose diameter is sufficiently large to pass the low frequency Gaussian portion of the beam and block the high frequency part, the irregularities will be removed from emerging beam and a "clean" laser beam will result.

Another application involves the use of spatial frequencies for object recognition. In some areas of photographic surveys, the amount of data to be analyzed is enormous. Provided a feature of interest has some particular set of spatial frequencies connected with it (spacing between ties in the case of railroads, for example), the laser and lens combination can be used to recognize the possible presence of these features in the photograph. Other applications include inspection of products, such as the tips of hypodermic needles. The average spatial frequency pattern for a large number of good needles is stored in a computer memory. Then the pattern of each new tip is compared to it. Those needles that do not fit the stored pattern to within certain criteria are then rejected. Since the actual position of the tip does not affect the spatial frequency pattern, the test is insensitive to location errors, whereas a direct inspection of the image of the needle tip would have to locate it with a high degree of precision.

0.8 References

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

Optics (2nd Edition), E. Hecht
Addison Wesley,

Fundamentals of Optics, F. Jenkins and H. White
McGraw Hill, 1957,

Contemporary Optics for Scientists and Engineers,
Nussbaum
Prentice-Hall, inc. 1976

The Optics Problem Solver,
Research and Education Association, 1986