JIWAJI UNIVERSITY GWALIOR

SELF LEARNING MATERIAL

FOR

B.SC. 2 YEAR PHYSICS

PAPER 201: OPTICS

PAPER CODE: 201

Published By:
Registrar,
Jiwaji University,Gwalior

Distance Education, Jiwaji University, Gwalior

JIWAJI UNIVERSITY GWALIOR

SELF LEARNING MATERIAL

FOR

B.SC. 2 YEAR PHYSICS

PAPER 201: OPTICS

PAPER CODE: 201

WRITER

Miss KARUNA MAHOR

Master of Computer Application

Physical Constants in CGS Units

speed of light $c = 3 \cdot 10^{10}$ cm/sec = 1000 ft /⌠sec = 1 ft /nanosecond

acceleration due to gravity at the surface of the earth $g = 980$ cm/sec^2 = 32 ft/sec^2

gravitational constant $G = 6.67 \cdot 10^{-8}$ cm^3/ (gm sec^2)

charge on an electron $e = 4.8 \cdot 10^{-10}$ esu

Planck's constant $h = 6.62 \cdot 10^{-27}$ erg sec (gm cm^2/sec)

Planck constant /2π $h = 1.06 \cdot 10^{-27}$ erg sec (gm cm^2 / sec)

Bohr radius $a_0 = .529 \cdot 10^{-8}$ cm

rest mass of electron $m_e = 0.911 \cdot 10^{-27}$ gm

rest mass of proton $M_p = 1.67 \cdot 10^{-24}$ gm

rest energy of electron $m_e c^2 = 0.51$ MeV (H 1 / 2 MeV)

rest energy of proton $M_p c^2 = 0.938$ BeV (H 1 BeV)

proton radius $r_p = 1.0 \cdot 10^{-13}$ cm

Boltzmann's constant $k = 1.38 \cdot 10^{-16}$ ergs/ kelvin

Avogadro's number $N_0 = 6.02 \cdot 10^{23}$ molecules/mole

absolute zero = 0° K = 273°C

density of mercury = 13 .6 gm / cm^3

mass of earth = $5.9\,8 \cdot 10^{27}$ gm

mass of the moon = $7.3\,5 \cdot 10^{25}$ gm

mass of the sun = $1.9\,7 \cdot 10^{33}$ gm

earth radius = $6.3\,8 \cdot 10^{8}$ cm = 3960mi

moon radius = $1.7\,4 \cdot 10^{8}$ cm = 1080mi

mean distance to moon = $3.8\,4 \cdot 10^{10}$ cm

mean distance to sun = $1.5\,0 \cdot 10^{13}$ cm

me an earth velocity in orbit about sun = 29.77 km / sec

UNIT-1

Geometrical Optics

For over 100 years, from the time of Newton and Huygens in the late 1600s, until 1801 when Thomas Young demonstrated the wave nature of light with his two slit experiment, it was not clear whether light consisted of beams of particles as proposed by Newton, or was a wave phenomenon as put forward by Huygens. The reason for the confusion is that almost all common optical phenomena can be explained by tracing light rays. The wavelength of light is so short compared to the size of most objects we are familiar with, that light rays produce sharp shadows and interference and diffraction effects are negligible.

To see how wave phenomena can be explained by ray tracing, consider the reflection of a light wave by a metal surface. When a wave strikes a very small object, an object much smaller than a wavelength, a circular scattered wave emerges as shown in the ripple tank photograph of Figure (36-1) reproduced here. But when a light wave impinges on a metal surface consisting of many small atoms, represented by the line of dots in Figure (36-2), the circular scattered waves all add up to produce a reflected wave that emerges at an angle of reflection θ_r equal to the angle of incidence θ_i. Rather than sketching the individual crests and troughs of the incident wave, and adding up all the scattered waves, it is much easier to treat the light as a ray that reflected from the surface. This ray is governed by the law of reflection, namely $\theta_r = \theta_i$.

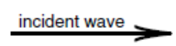
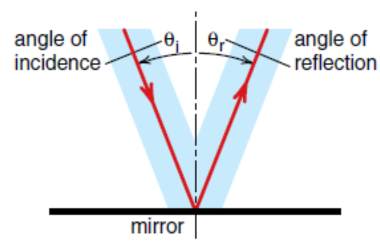


Figure 36-1
An incident wave passing over a small object produces a circular scattered wave.



Light ray reflected from a mirror.

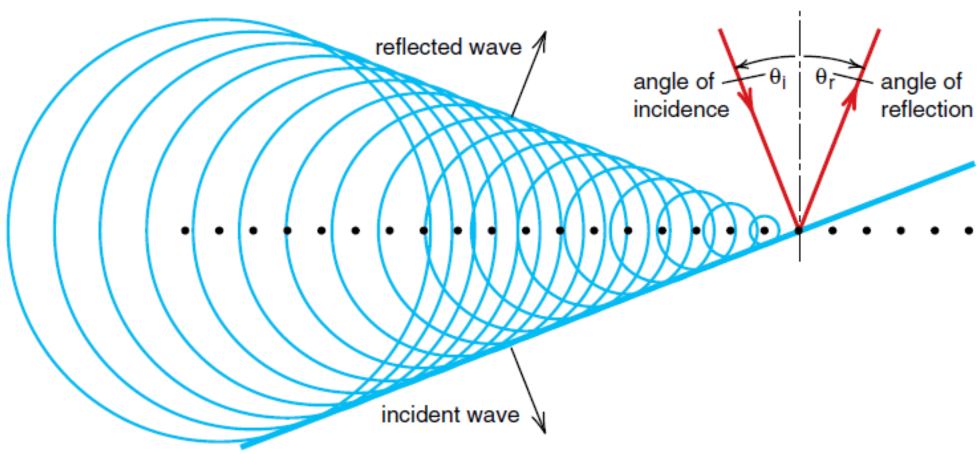


Figure 36-2
Reflection of light. In the photograph, we see an incoming plane wave scattered by a small object. If the object is smaller than a wavelength, the scattered waves are circular. When an incoming light wave strikes an array of atoms in the surface of a metal, the scattered waves add up to produce a reflected wave that comes out at an angle of reflection θ_r equal to the angle of incidence θ_i.

The subject of geometrical optics is the study of the behavior of light when the phenomena can be explained by ray tracing, where shadows are sharp and interference and diffraction effects can be neglected. The basic laws for ray tracing are extremely simple. At a reflecting surface $\boldsymbol{\theta_r = \theta_i}$*, as we have just seen. When a light ray passes between two media of different* ***indexes of refraction****, as in going from air into glass or air into water, the rule is* $\boldsymbol{n_1 \sin\theta_1 = n_2 \sin\theta_2}$ *, where* n_1 *and* n_2 *are constants called indices of refraction, and* θ_1 *and* θ_2 *are the angles that the rays made with the line perpendicular to the interface. This is known as* ***Snell's law****.*

This entire chapter is based on the two rules $\boldsymbol{\theta_r = \theta_i}$ *and* $\boldsymbol{n_1 \sin\theta_1 = n_2 \sin\theta_2}$*. These rules are all that are needed to understand the function of telescopes, microscopes, cameras, fiber optics, and the optical components of the human eye. You can understand the operation of these instruments without knowing anything about Newton's laws, kinetic and potential energy, electric or magnetic fields, or the particle and wave nature of matter. In other words, there is no prerequisite background needed for studying geometrical optics as long as you accept the two rules which are easily verified by experiment.*

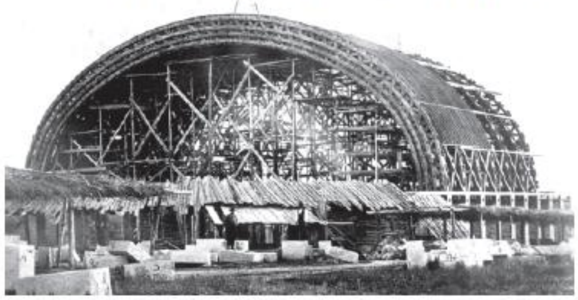
In most introductory texts, geometrical optics appears after Maxwell's equations and theory of light. There is a certain logic to this, first introducing a basic theory for light and then treating geometrical optics as a practical application of the theory. But this is clearly not an historical approach since geometrical optics was developed centuries before Maxwell's theory. Nor is it the only logical approach, because studying lens systems teaches you nothing more about Maxwell's equations than you can learn by deriving Snell's law. Geometrical optics is an interesting subject full of wonderful applications, a subject that can appear anywhere in an introductory physics course.

We have a preference not to introduce geometrical optics after Maxwell's equations. With Maxwell's theory, the student is introduced to the wave nature of one component of matter, namely light. If the focus is kept on the basic nature of matter, the next step is to look at the photoelectric effect and the particle nature of light. You then see that light has both a particle and a wave nature, which opens the door to the particle-wave nature of all matter and the subject of quantum mechanics. We have a strong preference not to interrupt this focus on the basic nature of matter with a long and possibly distracting chapter on geometrical optics.

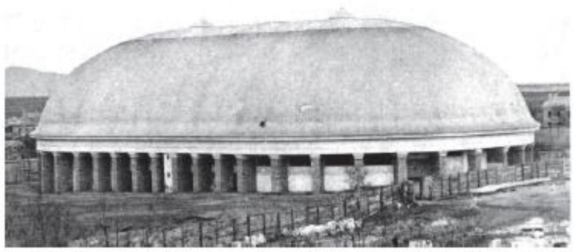
REFLECTION FROM CURVED SURFACES

The Mormon Tabernacle, shown in Figure (1), is constructed in the shape of an ellipse. If one stands at one of the focuses and drops a pin, the pin drop can be heard 120 feet away at the other focus. The reason why can be seen from Figure (2), which is similar to Figure (8-28) where we showed you how to draw an ellipse with a pencil, a piece of string, and two thumbtacks.

The thumbtacks are at the focuses, and the ellipse is drawn by holding the string taut as shown. As you move the pencil point along, the two sections of string always make equal angles θ_i and θ_r to a line perpendicular or normal to the part of the ellipse we are drawing. The best way to see that the angles θ_i and θ_r are always equal is to construct your own ellipse and measure these angles at various points along the curve.



Mormon Tabernacle under construction, 1866.



Mormon Tabernacle finished, 1871.

Mormon Tabernacle today.
Figure 1

If a sound wave were emitted from focus 1 in Figure (2), the part of the wave that traveled over to point A on the ellipse would be reflected at an angle θ_r equal to the angle of incidence θ_i , and travel over to focus 2. The part of the sound wave that struck point B on the ellipse, would be reflected at an angle θ_r equal to it's angle of incidence θ_i , and also travel over to focus 2. If you think of the sound wave as traveling out in rays, then all the rays radiated from focus 1 end up at focus 2, and that is why you hear the whisper there. We say that the rays are ***focused*** at focus 2, and that is why these points are called focuses of the ellipse. (Note also that the path lengths are the same, so that all the waves arriving at focus 2 are in phase.)

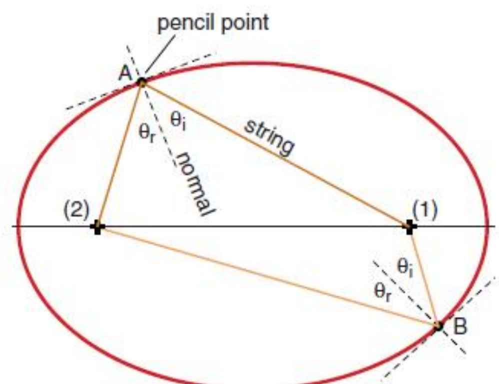


Figure 2
Drawing an ellipse using a string and two thumbtacks.

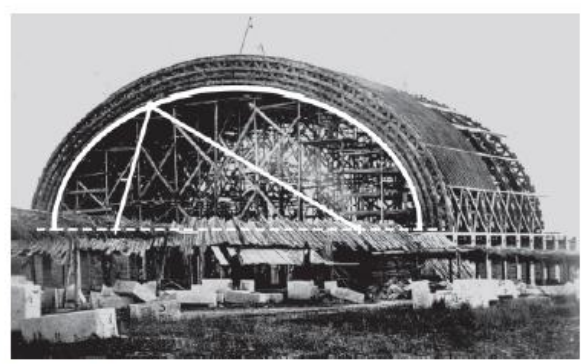


Figure 2a
A superposition of the top half of Figure 2 on Figure 1.

The Parabolic Reflection

You make a parabola out of an ellipse by moving one of the focuses very far away. The progression from a parabola to an ellipse is shown in Figure (3). For a true parabola, the second focus has to be infinitely far away.

Suppose a light wave were emitted from a star and traveled to a parabolic reflecting surface. We can think of the star as being out at the second, infinitely distant, focus of the parabola. Thus all the light rays coming in from the star would reflect from the parabolic surface and come to a point at the near focus. The rays from the star approach the reflector as a parallel beam of rays, thus a parabolic reflector has the property of focusing parallel rays to a point, as shown in Figure (4a).

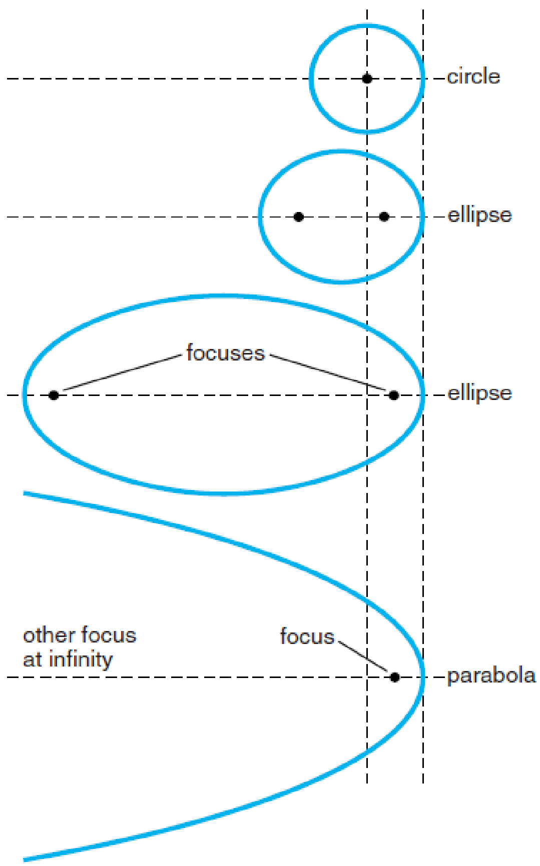


Figure 3
Evolution of an ellipse into a parabola. For a parabola, one of the focuses is out at infinity.

If parallel rays enter a deep dish parabolic mirror from an angle off axis as shown in Figure (4b), the rays do not focus to a point, with the result that an off axis star would appear as a blurry blob. (This figure corresponds to looking at a star 2.5° off axis, about 5 moon diameters from the center of the field of view.)

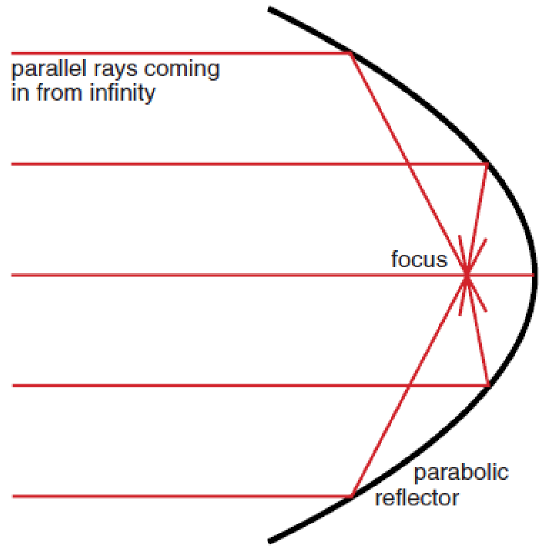


Figure 4a
Parallel rays, coming down the axis of the parabola, focus to a point.

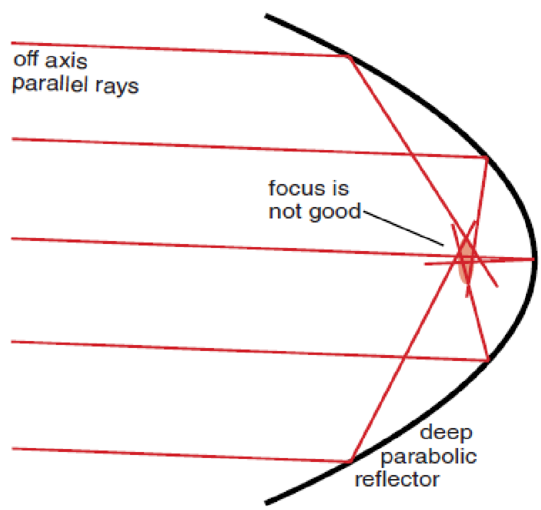


Figure 4b
For such a deep dish parabola, rays coming in at an angle of 2.5° do not focus well.

One way to get sharp images for parallel rays coming in at an angle is to use a shallower parabola as illustrated in Figure (4c). In that figure, the ***focal length*** (distance from the center of the mirror to the focus) is 2 times the mirror diameter, giving what is called an *f* 2 mirror. In Figure (4d), you can see that rays coming in at an angle of 2.5° (blue lines) almost focus to a point. Typical amateur telescopes are still shallower, around *f* 8, which gives a sharp focus for rays off angle by as much as 2° to 3°.

As we can see in Figure (4d), light coming from two different stars focus at two different points in what is called the ***focal plane*** of the mirror. If you placed a photographic film at the focal plane, light from each different star, entering as parallel beams from different angles, would focus at different points on the film, and you would end up with a photographic image of the stars. This is how distant objects like stars are photographed with what is called a ***reflecting telescope***.

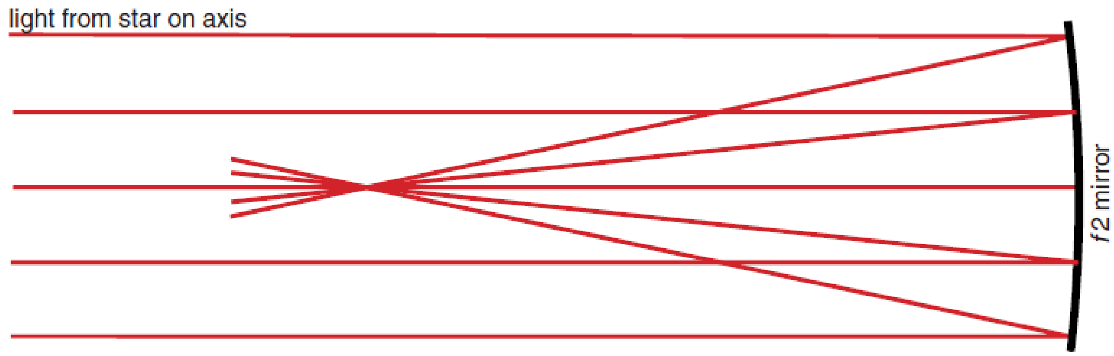


Figure 4c
A shallow dish is made by using only the shallow bottom of the parabola. Here the focal length is twice the diameter of the dish, giving us an f2 mirror. Typical amateur telescopes are still shallower, having a focal length around 8 times the mirror diameter (f8 mirrors). [The mirror in Figure 4b, that gave a bad focus, was f.125, having a focal length 1/8 the diameter of the mirror.]

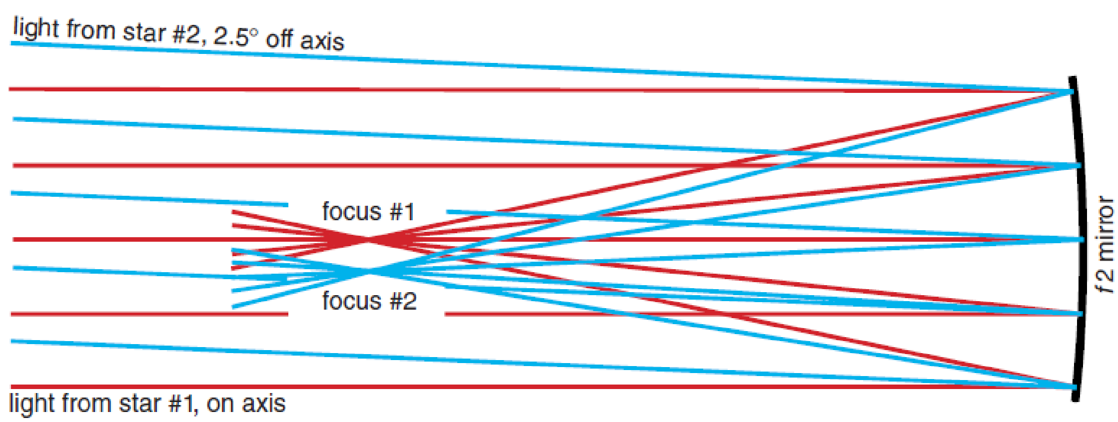


Figure 4d
We can think of this drawing as representing light coming in from a red star at the center of the field of view, and a blue star 2.5° (5 full moon diameters) away. Separate images are formed, which could be recorded on a photographic film. With this shallow dish, the off axis image is sharp (but not quite a point).

MIRROR IMAGES

The image you see in a mirror, although very familiar, is still quite remarkable in its reality. Why does it look so real? You do not need to know how your eye works to begin to see why.

Consider Figure (5a) where light from a point source reaches your eye. We have drawn two rays, one from the source to the top of the eye, and one to the bottom. In Figure (5b), we have placed a horizontal mirror as shown and moved the light source a distance h above the mirror equal to the distance it was below the mirror before the mirror was inserted. Using the rule that the angle of incidence equals the angle of reflection, we again drew two rays that went from the light source to

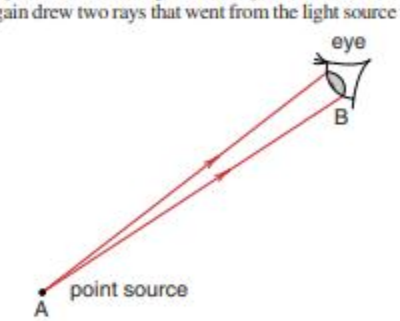


Figure 5a
Light from a point source reaching your eye.

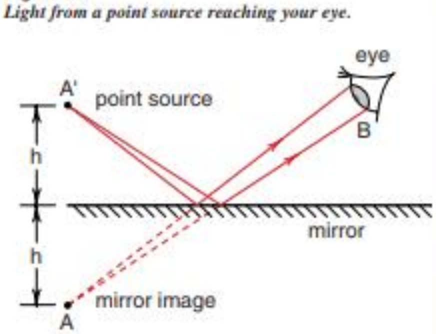


Figure 5b
There is no difference when the source is at point A, or at point A' and the light is reflected in a mirror.

the top and to the bottom of the eye. You can see that if you started at the eye and drew the rays back as straight lines, ignoring the mirror, the rays would intersect at the old source point A as shown by the dotted lines in Figure (5b).

To the eye (or a camera) at point B, there is no detectable difference between Figures (5a) and (5b). In both cases, the same rays of light, coming from the same directions enter the eye. Since the eye has no way of telling that the rays have been bent, we perceive that the light source is at the ***image point*** A rather than at the source point A′.

When we look at an extended object, its image in the mirror does not look identical to the object itself. In Figure (6), my granddaughter Julia is holding her right hand in front of a mirror and her left hand off to the side. The image of the right hand looks like the left hand. In particular, the fingers of the mirror image of the right hand curl in the opposite direction from those of the right hand itself. If she were using the right hand rule to find the direction of the angular momentum of a rotating object, the mirror image would look as if she were using a left hand rule.

It is fairly common knowledge that left and right are reversed in a mirror image. But if left and right are reversed, why aren't top and bottom reversed also? Think about that for a minute before you go on to the next paragraph.

Figure 6
The image of the right hand looks like a left hand.

To see what the image of an extended object should be, imagine that we place an arrow in front of a mirror as shown in Figure (7). We have constructed rays from the tip and the base of the arrow that reflect and enter the eye as shown. Extending these rays back to the image, we see that the image arrow has been reversed *front to back*. That is what a mirror does. The mirror image is reversed front to back, not left to right or top to bottom. It turns out that the right hand, when reversed front to back as in its image in Figure (6), has the symmetry properties of a left hand. If used to define angular momentum, you would get a left hand rule.

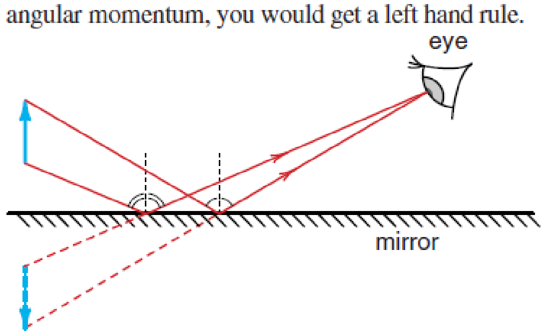


Figure 7
A mirror image changes front to back, not left to right.

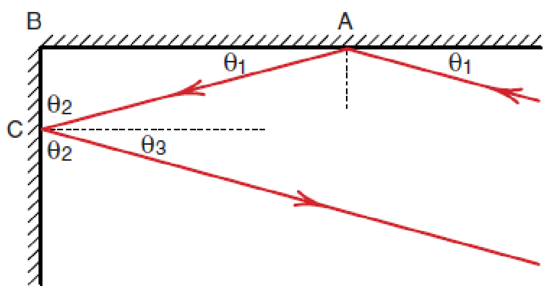


Figure 8a
With a corner reflector, the light is reflected back it the same direction from which it arrived.

The Corner Reflector

When two vertical mirrors are placed at right angles as shown in Figure (8a), a horizontal ray approaching the mirrors is reflected back in the direction from which it came. It is a little exercise in trigonometry to see that this is so. Since the angle of incidence equals the angle of reflection at each mirror surface, we see that the angles labeled θ_1 must be equal to each other and the same for the angles θ_2. From the right triangle ABC, we see that $\theta_1 + \theta_2 = 90°$. We also see that the angles $\theta_2 + \theta_3$ also add up to 90°, thus $\theta_3 = \theta_1$, which implies the exiting ray is parallel to the entering one.

If you mount three mirrors perpendicular to each other to form the corner of a cube, then light entering this so called ***corner reflector*** from any angle goes back in the direction from which it came. The Apollo II astronauts placed the array of corner reflectors shown in Figure (8b) on the surface of the moon, so that a laser beam from the earth would be reflected back from a precisely known point on the surface of the moon. By measuring the time it took a laser pulse to be reflected back from the array, the distance to the moon could be measured to an accuracy of centimeters. With the distance to the moon known with such precision, other distances in the solar system could then be determined accurately.

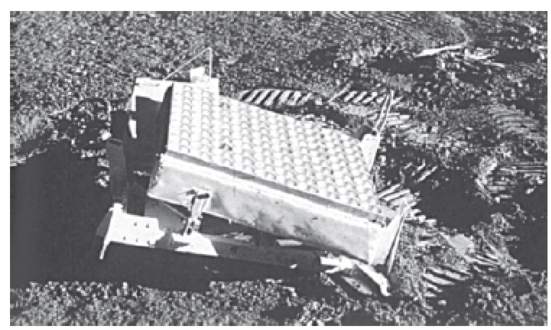


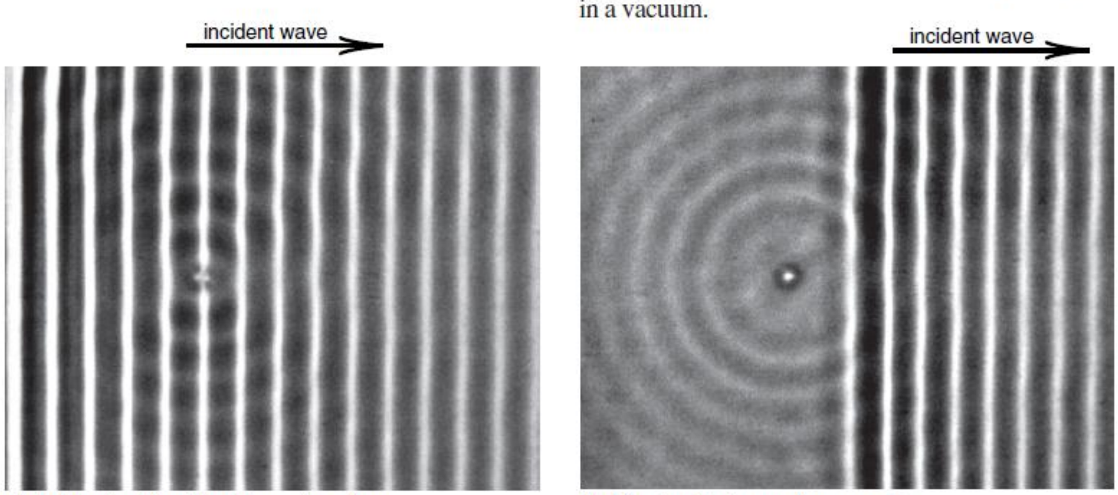
Figure 8b
Array of corner reflectors left on the moon by the Apollo astronauts. A laser pulse from the earth, aimed at the reflectors, returns straight back to the laser. By measuring the time the pulse takes to go to the reflectors and back, the distance to that point on the moon and back can be accurately measured.

MOTION OF LIGHT THROUGH A MEDIUM

We are all familiar with the fact that light can travel through clear water or clear glass. With some of the new glasses developed for fiber optics communication, light signals can travel for miles without serious distortion. If you made a mile thick pane from this glass you could see objects through it.

From an atomic point of view, it is perhaps surprising that light can travel any distance at all through water or glass. A reasonable picture of what happens when a light wave passes over an atom is provided by the ripple tank photograph shown in Figure (36-1) reproduced here. The wave scatters from the atom, and since atoms are considerably smaller than a wavelength of visible light, the scattered waves are circular like those in the ripple tank photograph. The final wave is the sum of the incident and the scattered waves as shown in Figure (36-1a).

When light passes through a medium like glass or water, the wave is being scattered by a huge number of atoms. The final wave pattern is the sum of the incident wave and all of the many billions of scattered waves. You might suspect that this sum would be very complex, but that is not the case. At the surface some of the incident wave is reflected. Inside the medium, the ***incident and scattered waves add up to a new wave*** of the same frequency as the incident wave but which travels ***at a reduced speed***. The speed of a light wave in water for example is 25% less than the speed of light in a vacuum.



a) Incident and scattered wave together.

b) After incident wave has passed.

Figure 36-1
If the scattering object is smaller than a wavelength, we get circular scattered waves.

The optical properties of lenses are a consequence of this effective reduction in the speed of light in the lens. Figure (9) is a rather remarkable photograph of individual short pulses of laser light as they pass through and around a glass lens. You can see that the part of the wave front that passed through the lens is delayed by its motion through the glass. The thicker the glass, the greater the delay. You can also see that the delay changed the shape and direction of motion of the wave front, so that the light passing through the lens focuses to a point behind the lens. This is how a lens really works.

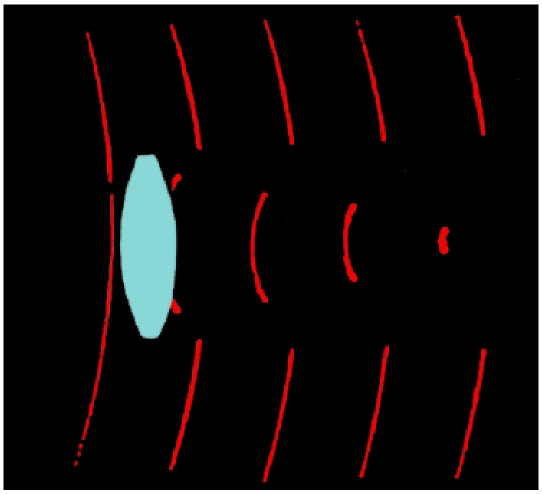


Figure 9
Motion of a wave front through a glass lens. The delay in the motion of the wave front as it passes through the glass changes the shape and direction of motion of the wave front, resulting in the focusing of light. (This photograph should not be confused with ripple tank photographs where wavelengths are comparable to the size of the objects. Here the wavelength of the light is about one hundred thousand times smaller than the diameter of the lens, with the result we get sharp shadows and do not see diffraction effects.)

In the 18/February/1999 issue of Nature ***it was announced that a laser pulse travelled through a gas of supercooled sodium atoms at a speed of 17 meters per second! (You can ride a bicycle faster than that.) This means that the sodium atoms had an index of refraction of about 18 million, 7.3 million times greater than that of diamond!***

Index of Refraction

The amount by which the effective speed of light is reduced as the light passes through a medium depends both upon the medium and the wavelength of the light. There is very little slowing of the speed of light in air, about a 25% reduction in speed in water, and nearly a 59% reduction in speed in diamond. In general, blue light travels somewhat slower than red light in nearly all media.

It is traditional to describe the slowing of the speed of light in terms of what is called the ***index of refraction*** of the medium. The index of refraction n is defined by the equation

$$\boxed{\left.\begin{array}{l}\text{speed of light}\\ \text{in a medium}\end{array}\right\} \quad v_{light} = \frac{c}{n}} \qquad (1)$$

The index n has to equal 1 in a vacuum because light always travels at the speed 3×10^8 meters in a vacuum. The index n can never be less than 1, because nothing can travel faster than the speed c. For yellow sodium light of wavelength $\lambda = 5.89 \times 10^{-5}$ cm (589 nanometers), the index of refraction of water at 20° C is $n = 1.333$, which implies a 25% reduction in speed. For diamond, $n = 2.417$ for this yellow light. Table 1 gives the indices of refraction for various transparent substances for the sodium light.

Vacuum	1.00000	exactly
Air (STP)	1.00029	
Ice	1.309	
Water (20° C)	1.333	
Ethyl alcohol	1.36	
Fuzed quartz	1.46	
Sugar solution (80%)	1.49	
Typical crown glass	1.52	
Sodium Chloride	1.54	
Polystyrene	1.55	
Heavy flint glass	1.65	
Sapphire	1.77	
Zircon	1.923	
Diamond	2.417	
Rutile	2.907	
Gallium phosphide	3.50	
Very cold sodium atoms	18000000	for laser pulse

Table 1
Some indices of refraction for yellow sodium light at a wavelength of 589 nanometers.

Exercise 1a

What is the speed of light in air, water, crown glass, and diamond. Express your answer in feet/nanosecond. (Take c to be exactly 1 ft/nanosecond.)

Exercise 1b

In one of the experiments announced in *Nature,* a laser pulse took 7.05 microseconds to travel .229 millimeters through the gas of supercooled sodium atoms. What was the index of refraction of the gas for this particular experiment? (The index quoted on the previous page was for the slowest observed pulse. The pulse we are now considering went a bit faster.)

CERENKOV RADIATION

In our discussion, in Chapter 1, of the motion of light through empty space, we saw that nothing, not even information, could travel faster than the speed of light. If it did, we could, for example, get answers to questions that had not yet been thought of.

When moving through a medium, the speed of a light wave is slowed by repeated scattering and it is no longer true that nothing can move faster than the speed of light in that medium. We saw for example that the speed of light in water is only 3/4 the speed c in vacuum. Many elementary particles, like the muons in the muon lifetime experiment, travel at speeds much closer to c. When a charged particle moves faster than the speed of light in a medium, we get an effect not unlike the sonic boom produced by a supersonic jet. We get a ***shock wave of light*** that is similar to a sound shock wave (sonic boom), or to the water shock wave shown in Figure (33-30) reproduced here. The light shock wave is called ***Cerenkov radiation*** after the Russian physicist Pavel Cerenkov who received the 1958 Nobel prize for discovering the effect.

In the muon lifetime picture, one observed how long muons lived when stopped in a block of plastic. The experiment was made possible by Cerenkov radiation. The muons that stopped in the plastic, entered moving faster than the speed of light in plastic, and as a result emitted a flash of light in the form of Cerenkov radiation. When the muon decayed, a charged positron and a neutral neutrino were emitted. In most cases the charged positron emerged faster than the speed of light in the plastic, and also emitted Cerenkov radiation. The two flashes of light were detected by the phototube which converted the light flashes to voltage pulses. The voltage pulses were then displayed on an oscilloscope screen where the time interval between the pulses could be measured. This interval represented the time that the muon lived, mostly at rest, in the plastic.

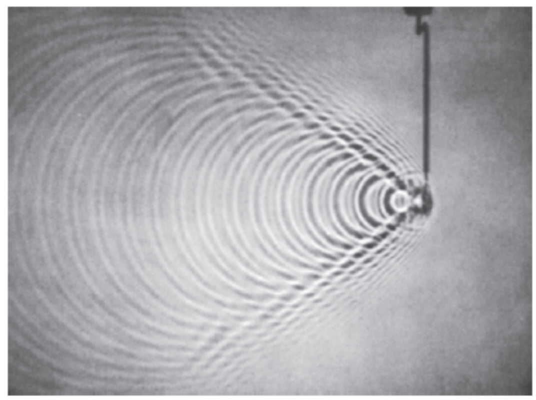


Figure 33-30
When the source of the waves moves faster than the speed of the waves, the wave fronts pile up to produce a shock wave as shown. This shock wave is the sonic boom you hear when a jet plane flies overhead faster than the speed of sound.

SNELL'S LAW

When a wave enters a medium of higher index of refraction and travels more slowly, the wavelength of the wave changes. The wavelength is the distance the wave travels in one period, and if the speed of the wave is reduced, the distance the wave travels in one period is reduced. (In most cases, the frequency or period of the wave is not changed. The exceptions are in fluorescence and nonlinear optics where the frequency or color of light can change.)

We can calculate how the wavelength changes with wave speed from the relationship

$$\lambda \frac{cm}{cycle} = \frac{v_{wave}\frac{cm}{sec}}{T\frac{sec}{cycle}}$$

Setting $v_{wave} = c/n$ for the speed of light in the medium, gives for the corresponding wavelength λ_n

$$\lambda_n = \frac{v_{wave}}{T} = \frac{c/n}{T} = \frac{1}{n}\frac{c}{T} = \frac{\lambda_0}{n} \qquad (2)$$

where $\lambda_0 = c/T$ is the wavelength in a vacuum. Thus, for example, the wavelength of light entering a diamond from air will be shortened by a factor of 1/2.42.

What happens when a set of periodic plane waves goes from one medium to another is illustrated in the ripple tank photograph of Figure (10). In this photograph, the water has two depths, deeper on the upper part where the waves travel faster, and shallower in the lower part where the waves travel more slowly. You can see that the wavelengths are shorter in the lower part, but there are the same number of waves. (We do not gain or loose waves at the boundary.) The frequency, the number of waves that pass you per second, is the same on the top and bottom.

The only way that the wavelength can be shorter and still have the same number of waves is for the wave to bend at the boundary as shown. We have drawn arrows showing the direction of the wave in the deep water (the incident wave) and in the shallow water (what we will call the ***transmitted*** or ***refracted*** wave), and we see that the change in wavelength causes a sudden change in direction of motion of the wave. If you look carefully you will also see reflected waves which emerge at an angle of reflection equal to the angle of incidence.

Figure (11) shows a beam of yellow light entering a piece of glass. The index of refraction of the glass is 1.55, thus the wavelength of the light in the glass is only .65 times as long as that in air ($n \approx 1$ for air). You can see both the bending of the ray as it enters the glass and also the reflected ray. (You also see internal reflection and the ray emerging from the bottom surface.) You cannot see the individual wave crests, but otherwise Figures (10) and (11) show similar phenomena.

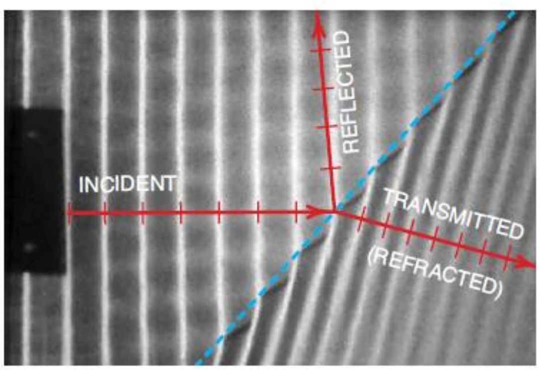


Figure 10
Refraction at surface of water. When the waves enter shallower water, they travel more slowly and have a shorter wavelength. The waves must travel in a different direction in order for the crests to match up.

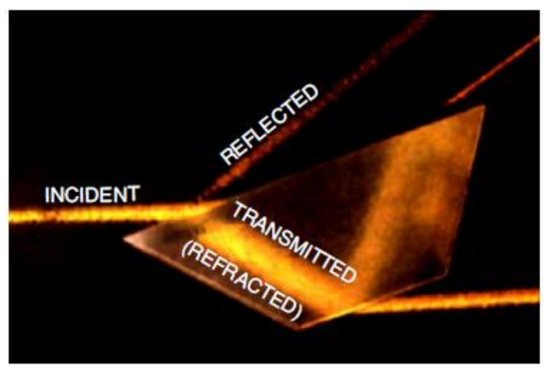


Figure 11
Refraction at surface of glass. When the light waves enter the glass, they travel more slowly and have a shorter wavelength. Like the water waves, the light waves must travel in a different direction in order for the crests to match up.

Derivation of Snell's Law

To calculate the angle by which a light ray is bent when it enters another medium, consider the diagram in Figure (12). The drawing represents a light wave, traveling in a medium of index n_1, incident on a boundary at an angle θ_1. We have sketched successive incident wave crests separated by the wavelength λ_1. Assuming that the index n_2 in the lower medium is greater than n_1, the wavelength λ_2 will be shorter than λ_1 and the beam will emerge at the smaller angle θ_2.

To calculate the angle θ_2 at which the transmitted or refracted wave emerges, consider the detailed section of Figure (12) redrawn in Figure (13a). Notice that we have labeled two apparently different angles by the same label θ_1. Why these angles are equal is seen in the construction of Figure (13b) where we see that the angles α and θ_1 are equal.

Exercise 2

Show that the two angles labeled θ_2 in Figure (13a) must also be equal.

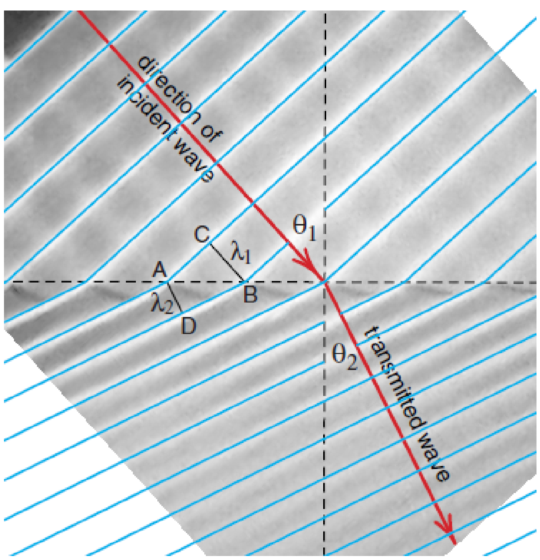


Figure 12
Analysis of refraction. The crests must match at the boundary between the different wavelength waves.

Since the triangles ACB and ADB are right triangles in Figure (13a), we have

$$\lambda_1 = AB \sin(\theta_1) = \lambda_0/n_1 \qquad (3)$$

$$\lambda_2 = AB \sin(\theta_2) = \lambda_0/n_2 \qquad (4)$$

where AB is the hypotenuse of both triangles and λ_0 is the wavelength when $n_0 = 1$. When we divide Equation 4 by Equation 5, the distances AB and λ_0 cancel, and we are left with

$$\frac{\sin(\theta_1)}{\sin(\theta_2)} = \frac{n_2}{n_1}$$

or

$$\boxed{n_1 \sin(\theta_1) = n_2 \sin(\theta_2)} \qquad \textit{Snell's law} \qquad (5)$$

Equation 5, known as Snell's law, allows us to calculate the change in direction when a beam of light goes from one medium to another.

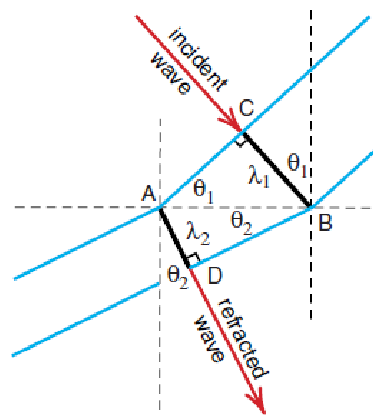


Figure 13a
The angles involved in the analysis.

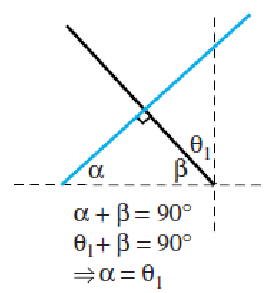


Figure 13b
Detail.

INTERNAL REFLECTION

Because of the way rays bend at the interface of two media, there is a rather interesting effect when light goes from a material of higher to a material of lower index of refraction, as in the case of light going from water into air. The effect is seen clearly in Figure (14). Here we have a multiple exposure showing a laser beam entering a tank of water, being reflected by a mirror, and coming out at different angles. The outgoing ray is bent farther away from the normal as it emerges from the water. We reach the point where the outgoing ray bends and runs parallel to the surface of the water. This is a critical angle, for if the mirror is turned farther, the ray can no longer get out and is completely reflected inside the surface.

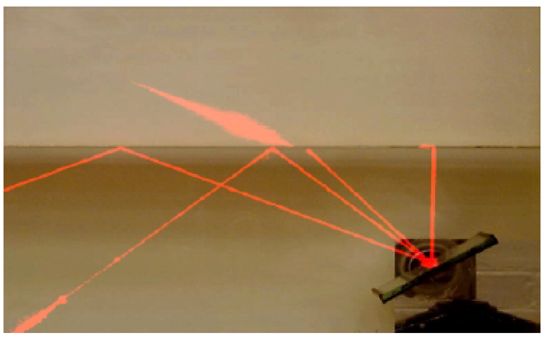


Figure 14
Internal reflection. We took three exposures of a laser beam reflecting off an underwater mirror set at different angles. In the first case the laser beam makes it back out of the water and strikes a white cardboard behind the water tank. In the other two cases, there is total internal reflection at the under side of the water surface. In the final exposure we used a flash to make the mirror visible.

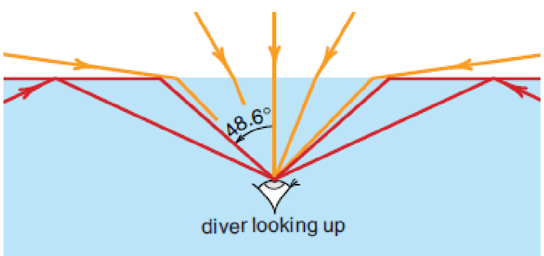


Figure 14a
When you are swimming under water and look up, you see the outside world through a round hole. Outside that hole, the surface is a silver mirror.

It is easy to calculate the critical angle θ_c at which this complete internal reflection begins. Set the angle of refraction, θ_2 in Figure (14), equal to 90° and we get from Snell's law

$$n_1 \sin\theta_c = n_2 \sin\theta_2 = n_2 \sin 90^\circ = n_2$$

$$\sin\theta_c = \frac{n_2}{n_1} \; ; \quad \boxed{\theta_c = \sin^{-1}\frac{n_2}{n_1}} \tag{6}$$

For light emerging from water, we have $n_2 \approx 1$ for air and $n_1 = 1.33$ for water giving

$$\sin^{-1}\theta_2 = \frac{1}{1.33} = .75$$

$$\theta_c = 48.6^\circ \tag{7}$$

Anyone who swims underwater, scuba divers especially, are quite familiar with the phenomenon of internal reflection. When you look up at the surface of the water, you can see the entire outside world through a circular region directly overhead, as shown in Figure (14a). Beyond this circle the surface looks like a silver mirror.

Exercise 3

A glass prism can be used as shown in Figure (15) to reflect light at right angles. The index of refraction n_g of the glass must be high enough so that there is total internal reflection at the back surface. What is the least value n_g one can have to make such a prism work? (Assume the prism is in the air where $n \approx 1$.)

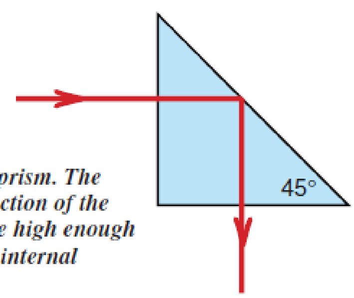


Figure 15
Right angled prism. The index of refraction of the glass has to be high enough to cause total internal reflection.

Fiber Optics

Internal reflection plays a critical role in modern communications and modern medicine through fiber optics. When light is sent down through a glass rod or fiber so that it strikes the surface at an angle greater than the critical angle, as shown in Figure (16a), the light will be completely reflected and continue to bounce down the rod with no loss out through the surface. By using modern very clear glass, a fiber can carry a light signal for miles without serious attenuation.

The reason it is more effective to use light in glass fibers than electrons in copper wire for transmitting signals, is that the glass fiber can carry information at a much higher rate than a copper wire, as indicated in Figure (16b). This is because laser pulses traveling through glass, can be turned on and off much more rapidly than electrical pulses in a wire. The practical limit for copper wire is on the order of a million pulses or bits of information per second (corresponding to a ***baud rate*** of one ***megabit***). Typically the information rate is much slower over commercial telephone lines, not much in excess of 30 to 50 thousand bits of information per second (corresponding to 30 to 50 ***kilobaud***). These rates are fast enough to carry telephone conversations or transmit text to a printer, but painfully slow for sending pictures and much too slow for digital television signals. High definition digital television will require that information be sent at a rate of about 3 million bits or pulses every 1/30 of a second for a baud rate of 90 million baud. (Compare that with the baud rate on your computer modem.) In contrast, fiber optics cables are capable of carrying pulses or bits at a rate of about a billion (10^9) per second, and are thus well suited for transmitting pictures or many phone conversations at once.

By bundling many fine fibers together, as indicated in Figure (17), one can transmit a complete image along the bundle. One end of the bundle is placed up against the object to be observed, and if the fibers are not mixed up, the image appears at the other end.

To transmit a high resolution image, one needs a bundle of about a million fibers. The tiny fibers needed for this are constructed by making a rather large bundle of small glass strands, heating the bundle to soften the glass, and then stretching the bundle until the individual strands are very fine. (If you have heated a glass rod over a Bunsen burner and pulled out the ends, you have seen how fine a glass fiber can be made this way.)

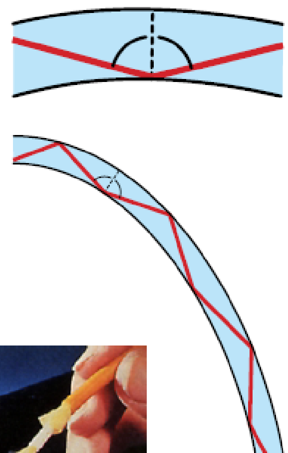


Figure 16a
Because of internal reflections, light can travel down a glass fiber, even when the fiber is bent.

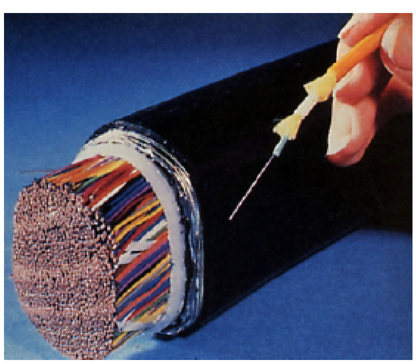


Figure 16b
A single glass fiber can carry the same amount of information as a fat cable of copper wires.

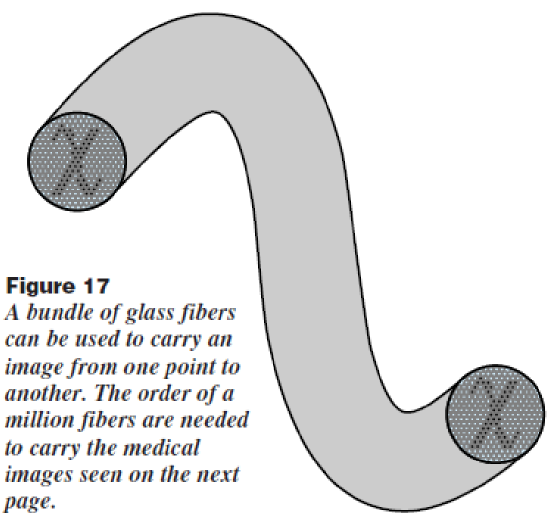


Figure 17
A bundle of glass fibers can be used to carry an image from one point to another. The order of a million fibers are needed to carry the medical images seen on the next page.

Medical Imaging

The use of fiber optics has revolutionized many aspects of medicine. It is an amazing experience to go down and look inside your own stomach and beyond, as the author did a few years ago. This is done with a flexible fiber optics instrument called a retroflexion, producing the results shown in Figure (18). An operation, such as the removal of a gallbladder, which used to require opening the abdomen and a long recovery period, can now be performed through a small hole near the navel, using fiber optics to view the procedure. You can see the viewing instrument and such an operation in progress in Figure (19).

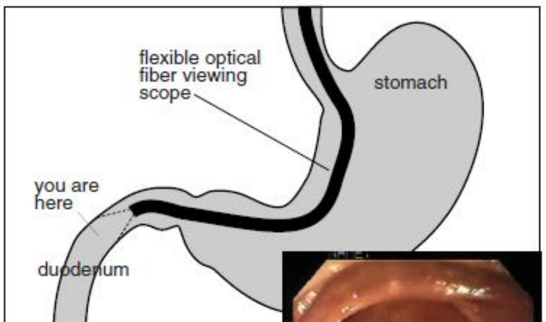


Figure 18
Close-up view of the author taken by photographer Dr. Richard Rothstein.

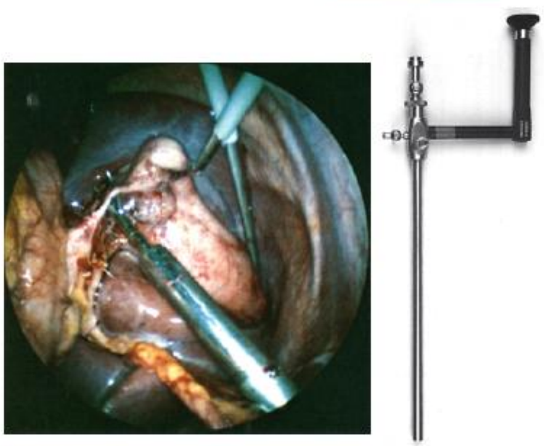


Figure 19
Gallbladder operation in progress, being viewed by the rigid laparoscope shown on the right. Such views are now recorded by high resolution television.

PRISMS

So far in our discussion of refraction, we have considered only beams of light of one color, one wavelength. Because the index of refraction generally changes with wavelength, rays of different wavelength will be bent at different angles when passing the interface of two media. Usually the index of refraction of visible light increases as the wavelength becomes shorter. Thus when white light, which is a mixture of all the visible colors, is sent through a prism as shown in Figure (20), the short wavelength blue light will be deflected by a greater angle than the red light, and the beam of light is separated into a rainbow of colors.

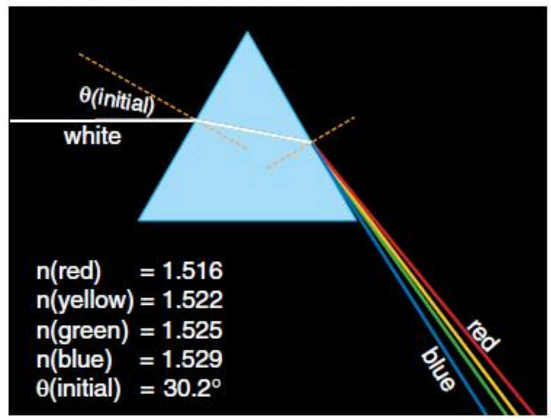


Figure 20
When light is sent through a prism, it is separated into a rainbow of colors. In this scale drawing, we find that almost all the separation of colors occurs at the second surface where the light emerges from the glass.

Rainbows

Rainbows in the sky are formed by the reflection and refraction of sunlight by raindrops. It is not, however, particularly easy to see why a rainbow is formed. René Descartes figured this out by tracing rays that enter and leave a spherical raindrop.

In Figure (21a) we have used Snell's law to trace the path of a ray of yellow light that enters a spherical drop of water (of index $n = 1.33$), is reflected on the back side, and emerges again on the front side. (Only a fraction of the light is reflected at the back, thus the reflected beam is rather weak.) In this drawing, the angle θ_2 is determined by $\sin(\theta_1) = 1.33 \sin(\theta_2)$. At the back, the angles of incidence and reflection are equal, and at the front we have $1.33 \sin(\theta_2) = \sin(\theta_1)$ (taking the index of refraction of air = 1). Nothing is hard about this construction, it is fairly easy to do with a good drafting program like Adobe Illustrator and a hand calculator.

In Figure (21b) we see what happens when a number of parallel rays enter a spherical drop of water. (This is similar to the construction that was done by Descartes in 1633.) When you look at the outgoing rays, it is not immediately obvious that there is any special direction for the reflected rays. But if you look closely you will see that the ray we have labeled #11 is the one that comes back at the widest angle from the incident ray.

Ray #1, through the center, comes straight back out. Ray #2 comes out at a small angle. The angles increase up to Ray #11, and then start to decrease again for Rays #12 and #13. In our construction the maximum angle, that of Ray #11, was 41.6°, close to the theoretical value of 42° for yellow light.

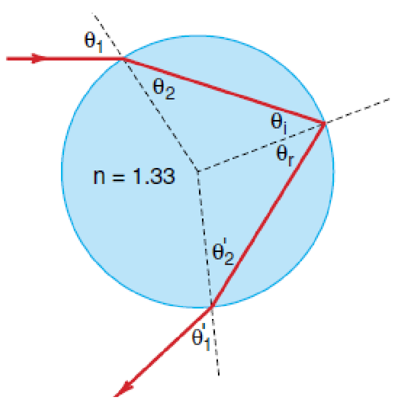


Figure 21a
Light ray reflecting from a raindrop.

What is more important than the fact that the maximum angle of deviation is 42° is the fact that the rays close to #11 emerge as more or less parallel to each other. The other rays, like those near #3 for example come out at diverging angles. That light is spread out. But the light emerging at 42° comes out as a parallel beam. When you have sunlight striking many raindrops, ***more yellow light is reflected back at this angle of 42° than any other angle.***

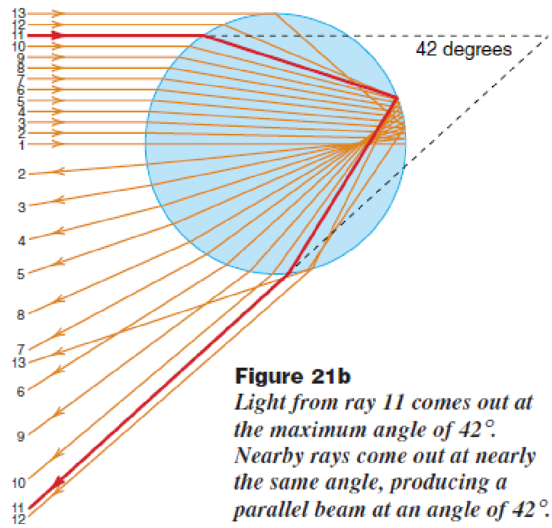


Figure 21b
Light from ray 11 comes out at the maximum angle of 42°. Nearby rays come out at nearly the same angle, producing a parallel beam at an angle of 42°.

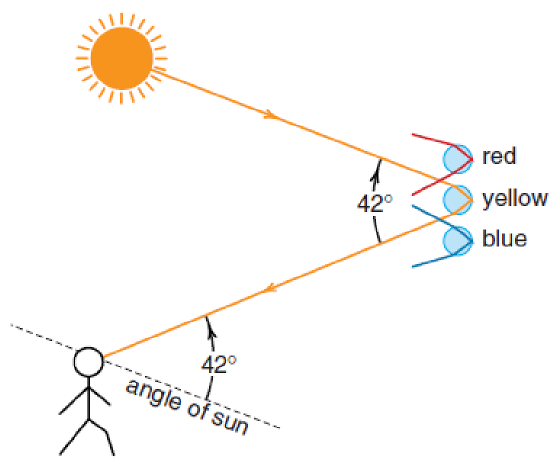


Figure 21c
You will see the yellow part of the rainbow at an angle of 42° as shown above. Red will be seen at a greater angle, blue at a lesser one.

If the atmosphere is not so clear, if there is a bit of haze or moisture as one often gets in the summer, the blue light is absorbed by the haze, and the last image we see setting is the green image. This is the origin of the green flash. With still more haze you get a red sunset, all the other colors having been absorbed by the haze.

Usually it requires binoculars to see the green or blue colors at the instant of sunset. But sometimes the atmospheric conditions are right so that this final light of the sun is reflected on clouds and can be seen without binoculars. If the clouds are there, there is probably enough moisture to absorb the blue image, and the resulting flash on the clouds is green.

Halos and Sun Dogs

Another phenomenon often seen is the reflection of light from hexagonal ice crystals in the atmosphere. The reflection is seen at an angle of 22° from the sun. If the ice crystals are randomly oriented then we get a complete halo as seen in Figure (23a). If the crystals are falling with their flat planes predominately horizontal, we only see the two pieces of the halo at each side of the sun, seen in Figure (24). These little pieces of rainbow are known as "sun dogs".

Figure 23
Halo caused by reflection by randomly oriented hexagonal ice crystals.

Figure 24
Sun dogs caused by ice crystals falling flat.

LENSES

The main impact geometrical optics has had on mankind is through the use of lenses in microscopes, telescopes, eyeglasses, and of course, the human eye. The basic idea behind the construction of a lens is Snell's law, but as our analysis of light reflected from a spherical raindrop indicated, we can get complex results from even simple geometries like a sphere.

Modern optical systems like the zoom lens shown in Figure (25) are designed by computer. Lens design is an ideal problem for the computer, for tracing light rays through a lens system requires many repeated applications of Snell's law. When we analyzed the spherical raindrop, we followed the paths of 12 rays for an index of refraction for only yellow light. A much better analysis would have resulted from tracing at least 100 rays for the yellow index of refraction, and then repeating the whole process for different indices of refraction, corresponding to different wavelengths or colors of light. This kind of analysis, while extremely tedious to do by hand, can be done in seconds on a modern desktop computer.

In this chapter we will restrict our discussion to the simplest of lens systems in order to see how basic instruments, like the microscope, telescope and eye, function. You will not learn here how to design a color corrected zoom lens like the Nikon lens shown below.

Figure 25
Nikon zoom lens.

If the atmosphere is not so clear, if there is a bit of haze or moisture as one often gets in the summer, the blue light is absorbed by the haze, and the last image we see setting is the green image. This is the origin of the green flash. With still more haze you get a red sunset, all the other colors having been absorbed by the haze.

Usually it requires binoculars to see the green or blue colors at the instant of sunset. But sometimes the atmospheric conditions are right so that this final light of the sun is reflected on clouds and can be seen without binoculars. If the clouds are there, there is probably enough moisture to absorb the blue image, and the resulting flash on the clouds is green.

Halos and Sun Dogs

Another phenomenon often seen is the reflection of light from hexagonal ice crystals in the atmosphere. The reflection is seen at an angle of 22° from the sun. If the ice crystals are randomly oriented then we get a complete halo as seen in Figure (23a). If the crystals are falling with their flat planes predominately horizontal, we only see the two pieces of the halo at each side of the sun, seen in Figure (24). These little pieces of rainbow are known as "sun dogs".

Figure 23
Halo caused by reflection by randomly oriented hexagonal ice crystals.

Figure 24
Sun dogs caused by ice crystals falling flat.

LENSES

The main impact geometrical optics has had on mankind is through the use of lenses in microscopes, telescopes, eyeglasses, and of course, the human eye. The basic idea behind the construction of a lens is Snell's law, but as our analysis of light reflected from a spherical raindrop indicated, we can get complex results from even simple geometries like a sphere.

Modern optical systems like the zoom lens shown in Figure (25) are designed by computer. Lens design is an ideal problem for the computer, for tracing light rays through a lens system requires many repeated applications of Snell's law. When we analyzed the spherical raindrop, we followed the paths of 12 rays for an index of refraction for only yellow light. A much better analysis would have resulted from tracing at least 100 rays for the yellow index of refraction, and then repeating the whole process for different indices of refraction, corresponding to different wavelengths or colors of light. This kind of analysis, while extremely tedious to do by hand, can be done in seconds on a modern desktop computer.

In this chapter we will restrict our discussion to the simplest of lens systems in order to see how basic instruments, like the microscope, telescope and eye, function. You will not learn here how to design a color corrected zoom lens like the Nikon lens shown below.

Figure 25
Nikon zoom lens.

Spherical Lens Surface

A very accurate spherical surface on a piece of glass is surprisingly easy to make. Take two pieces of glass, put a mixture of grinding powder and water between them, rub them together in a somewhat regular, somewhat irregular, pattern that one can learn in less than 5 minutes. The result is a spherical surface on the two pieces of glass, one being concave and the other being convex. The reason you get a spherical surface from this somewhat random rubbing is that only spherical surfaces fit together perfectly for all angles and rotations. Once the spheres have the desired radius of curvature, you use finer and finer grits to smooth out the scratches, and then jeweler's rouge to polish the surfaces. With any skill at all, one ends up with a polished surface that is perfectly spherical to within a fraction of a wavelength of light.

To see the optical properties of a spherical surface, we can start with the ray diagram we used for the spherical raindrop, and remove the reflections by extending the refracting medium back as shown in Figure (26a). The result is not encouraging. The parallel rays entering near the center of the surface come together—***focus***—quite a bit farther back than rays entering near the outer edge. This range of focal distances is not useful in optical instruments.

In Figure (26b) we have restricted the area where the rays are allowed to enter to a small region around the center of the surface. To a very good approximation all these parallel rays come together, focus, at one point. This is the characteristic we want in a simple lens, to bring parallel incoming rays together at one point as the parabolic reflector did.

Figure (26b) shows us that the way to make a good lens using spherical surfaces is to use only the central part of the surface. Rays entering near the axis as in Figure (26b) are deflected only by small angles, angles where we can approximate $\sin(\theta)$ by θ itself. When the angles of deflection are small enough to use small angle approximations, a spherical surface provides sharp focusing. As a result, in analyzing the small angle spherical lenses, we can replace the exact form of Snell's law

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \qquad \text{(5 repeated)}$$

by the approximate equation

$$n_1 \theta_1 = n_2 \theta_2 \qquad (8)$$

Snell's law for small angles

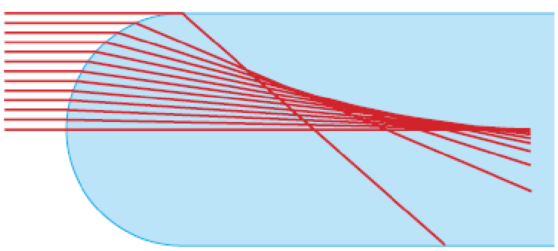


Figure 26a
Focusing properties of a spherical surface. (Not good!)

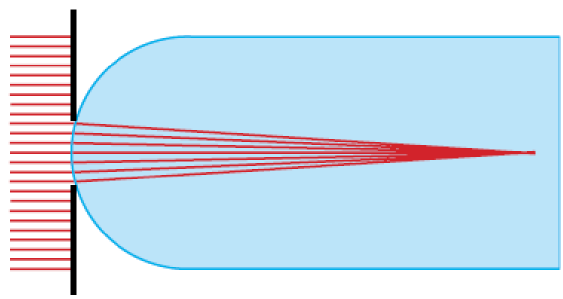


Figure 26b
We get a much better focus if we use only a small part of the spherical surface.

Focal Length of a Spherical Surface

Let us now use the simplified form of Snell's law to calculate the focal length f of a spherical surface, i.e., the distance behind the surface where entering parallel rays come to a point. Unless you plan to start making your own lenses, you do not really need this result, but the exercise provides an introduction to how focal lengths are related to the curvature of lenses.

Consider two parallel rays entering a spherical surface as shown in Figure (27). One enters along the axis of the surface, the other a distance h above it. The angle labeled θ_1 is the angle of incidence for the upper ray, while θ_2 is the refracted angle. These angles are related by Snell's law

$$n_1\theta_1 = n_2\theta_2$$

or

$$\theta_2 = \frac{n_1}{n_2}\theta_1 \tag{9}$$

If you recall your high school trigonometry you will remember that the outside angle of a triangle, θ_1 in Figure (27a), is equal to the sum of the opposite angles, θ_2 and α in this case. Thus

$$\theta_1 = \theta_2 + \alpha$$

or using Equation 9 for θ_2

$$\theta_1 = \frac{n_1}{n_2}\theta_1 + \alpha \tag{10}$$

Now consider the two triangles reproduced in Figures (27b) and (27c). Using the small angle approximation $\tan(\theta) \approx \sin(\theta) \approx \theta$, we have for Figure (27b)

$$\theta_1 \approx \frac{h}{r}\ ; \quad \alpha \approx \frac{h}{f} \tag{11}$$

Substituting these values for θ_1 and α into Equation 10 gives

$$\frac{h}{r} = \frac{n_1}{n_2}\frac{h}{r} + \frac{h}{f} \tag{12}$$

The height h cancels, and we are left with

$$\boxed{\frac{1}{f} = \frac{1}{r}\left(1 - \frac{n_1}{n_2}\right)} \tag{13}$$

The fact that the height h cancels means that parallel rays entering at any height h (as long as the small angle approximation holds) will focus at the same point a distance f behind the surface. This is what we saw in Figure (26b).

Figure (26b) was drawn for $n_1 = 1$ (air) and $n_2 = 1.33$ (water) so that $n_1/n_2 = 1/1.33 = .75$. Thus for that drawing we should have had

$$\frac{1}{f} = \frac{1}{r}(1 - .75) = \frac{1}{r}(.25) = \frac{1}{r}\left(\frac{1}{4}\right)$$

or

$$f = 4r \tag{14}$$

as the predicted focal length of that surface.

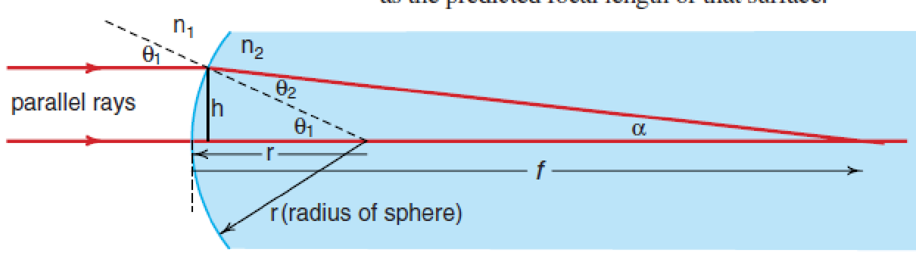


Figure 27
Calculating the focal length f of a spherical surface.

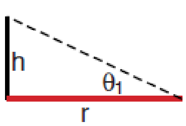


Figure 27b
$\theta_1 \approx h/r$

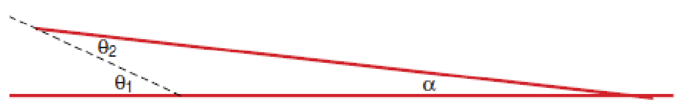


Figure 27a
$\theta_1 = \theta_2 + \alpha$

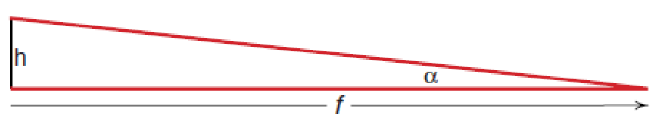


Figure 27c
$\alpha \approx h/f$

Exercise 5

Compare the prediction of Equation 14 with the results we got in Figure (26b). That is, what do you measure for the relationship between *f* and r in that figure?

Exercise 6

The index of refraction for red light in water is slightly less than the index of refraction for blue light. Will the focal length of the surface in Figure (26b) be longer or shorter than the focal length for red light?

Exercise 7

The simplest model for a fixed focus eye is a sphere of index of refraction n_2. The index n_2 is chosen so that parallel light entering the front surface of the sphere focuses on the back surface as shown in Figure (27d). What value of n_2 is required for this model to work when $n_1 = 1$? Looking at the table of indexes of refraction, Table 1, explain why such a model would be hard to achieve.

Aberrations

When parallel rays entering a lens do not come to focus at a point, we say that the lens has an ***aberration***. We saw in Figure (26a) that if light enters too large a region of a spherical surface, the focal points are spread out in back. This is called ***spherical aberration***. One cure for spherical aberration is to make sure that the diameter of any spherical lens you use is small in comparison to the radius of curvature of the lens surface.

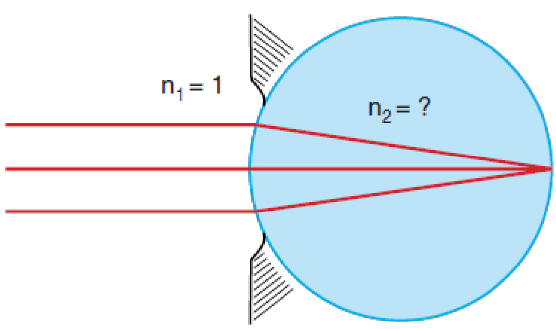


Figure 27d
A simple, but hard to achieve, model for an eye.

We get rainbows from raindrops and prisms because the index of refraction for most transparent substances changes with wavelength. As we saw in Exercise 6, this causes red light to focus at a different point than yellow or blue light, (resulting in colored bands around the edges of images). This problem is called ***chromatic aberration***. The cure for chromatic aberration is to construct complex lenses out of materials of different indices of refraction. With careful design, you can bring the focal points of the various colors back together. Some of the complexity in the design of the zoom lens in Figure (25) is to correct for chromatic aberration.

Astigmatism is a common problem for the lens of the human eye. You get astigmatism when the lens is not perfectly spherical, but is a bit cylindrical. If, for example, the cylindrical axis is horizontal, then light from a horizontal line will focus farther back than light from a vertical line. Either the vertical lines in the image are in focus, or the horizontal lines, but not both at the same time. (In the eye, the cylindrical axis does not have to be horizontal or vertical, but can be at any angle.)

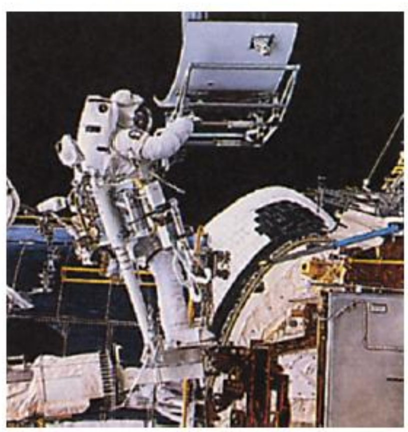
There can be many other aberrations depending upon what distortions are present in the lens surface. We once built a small telescope using a shaving mirror instead of a carefully ground parabolic mirror. The image of a single star stretched out in a line that covered an angle of about 30 degrees. This was an extreme example of an aberration called ***coma***. That telescope provided a good example of why optical lenses and mirrors need to be ground very accurately.

What, surprisingly, does not usually cause a serious problem is a small scratch on a lens. You do not get an image of the scratch because the scratch is completely out of focus. Instead the main effect of a scratch is to scatter light and fog the image a bit.

Perhaps the most famous aberration in history is the spherical aberration in the primary mirror of the orbiting Hubble telescope. The aberration was caused by an undetected error in the complex apparatus used to test the surface of the mirror while the mirror was being ground and polished. The ironic part of the story is that the aberration could have easily been detected using the same simple apparatus all amateur telescope makers use to test their mirrors (the so called Foucault test), but such a simple minded test was not deemed necessary.

What saved the Hubble telescope is that the engineers found the problem with the testing apparatus, and could therefore precisely determine the error in the shape of the lens. A small mirror, only a few centimeters in diameter, was designed to correct for the aberration in the Hubble image. When this correcting mirror was inserted near the focus of the main mirror, the aberration was eliminated and we started getting the many fantastic pictures from that telescope.

Another case of historical importance is the fact that Issac Newton invented the reflecting telescope to avoid the chromatic aberration present in all lenses at that time. With a parabolic reflecting mirror, all parallel rays entering the mirror focus at a point. The location of the focal point does not depend on the wavelength of the light (as long as the mirror surface is reflecting at that wavelength). You also do not get spherical aberration either because a parabolic surface is the correct shape for focusing, no matter how big the diameter of the mirror is compared to the radius of curvature of the surface.



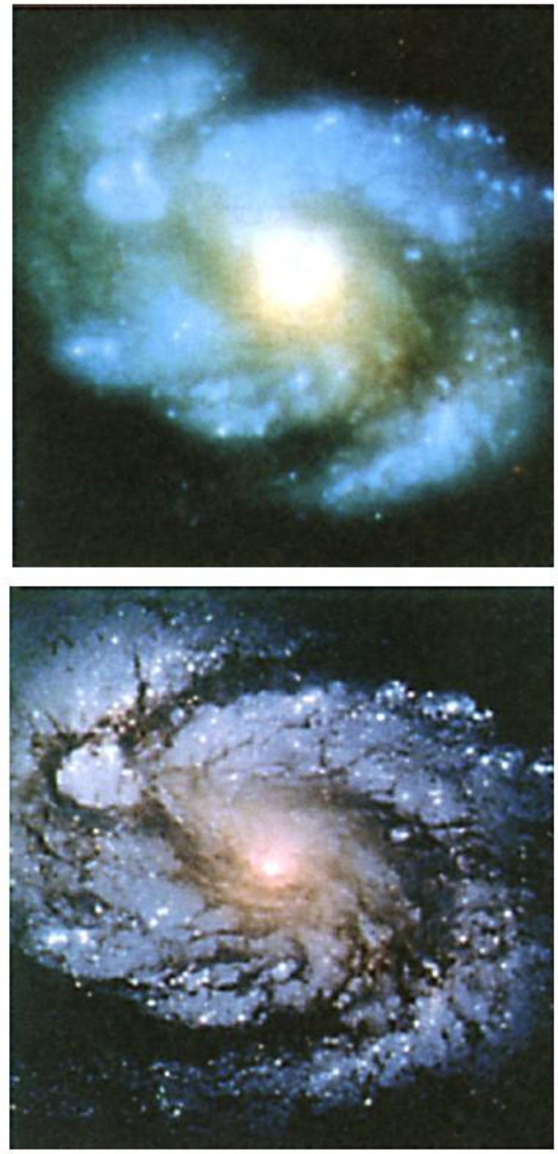


Figure 28
Correction of the Hubble telescope mirror. Top: before the correction. Bottom: same galaxy after correction. Left: astronauts installing correction mirror.

THIN LENSES

In Figure (29), we look at what happens when parallel rays pass through the two spherical surfaces of a lens. The top diagram (a) is a reproduction of Figure (26b) where a narrow bundle of parallel rays enters a new medium through a single spherical surface. By making the diameter of the bundle of rays much less than the radius of curvature of the surface, the parallel rays all focus to a single point. We were able to calculate where this point was located using small angle approximations.

In Figure (29b), we added a second spherical surface. The diagram is drawn to scale for indices of refraction n = 1 outside the gray region and n = 1.33 inside, and using Snell's law at each interface of each ray. (The drawing program Adobe Illustrator allows you to do this quite accurately.) The important point to note is that the parallel rays still focus to a point. The difference is that the focal point has moved inward.

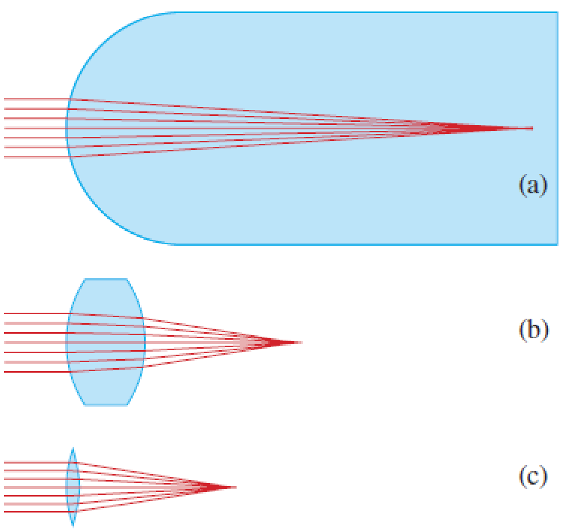


Figure 29
A two surface lens. Adding a second surface still leaves the light focused to a point, as long as the diameter of the light bundle is small compared to the radii of the lens surfaces.

In Figure (29c), we have moved the two spherical surfaces close together to form what is called a ***thin lens***. We have essentially eliminated the distance the light travels between surfaces. If the index of refraction outside the lens is 1 and has a value n inside, and surfaces have radii of curvature r_1 and r_2, then the focal length *f* of the lens given by the equation

$$\boxed{\frac{1}{f} = (n-1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right)} \qquad \textit{lens maker's equation} \qquad (15)$$

Equation 15, which is known as the ***lens maker's equation***, can be derived in a somewhat lengthy exercise involving similar triangles.

Unless you are planning to grind your own lenses, the lens maker's equation is not something you will need to use. When you buy a lens, you specify what focal length you want, what diameter the lens should be, and whether or not it needs to be corrected for color aberration. You are generally not concerned with how the particular focal length was achieved—what combination of radii of curvatures and index of refraction were used.

Exercise 8

(a) See how well the lens maker's equation applies to our scale drawing of Figure (29c). Our drawing was done to a scale where the spherical surfaces each had a radius of $r_1 = r_2 = 37\,\text{mm}$, and the distance *f* from the center of the lens to the focal point was 55 mm.

(b) What would be the focal length *f* of the lens if it had been made from diamond with an index of refraction n = 2.42?

The Lens Equation

What is important in the design of a simple lens system is where images are formed for objects that are different distances from the lens. Light from a very distant object enters a lens as parallel rays and focuses at a distance equal to the focal length f behind the lens. To locate the image when the object is not so far away, you can either use a simple graphical method which involves a tracing of two or three rays, or use what is called the ***lens equation*** which we will derive shortly from the graphical approach.

For our graphical work, we will use an arrow for the object, and trace out rays coming from the tip of the arrow. Where the rays come back together is where the image is formed. We will use the notation that the object is at a distance *(o)* from the lens, and that the image is at a distance *(i)* as shown in Figure (30).

In Figure (30) we have located the image by tracing three rays from the tip of the object. The top ray is parallel to the axis of the lens, and therefore must cross the axis at the focal point behind the lens. The middle ray, which goes through the center of the lens, is undeflected if the lens is thin. The bottom ray goes through the focal point in front of the lens, and therefore must come out parallel to the axis behind the lens. (Lenses are symmetric in that parallel light from either side focuses at the same distance f from the lens.) The image is formed where the three rays from the tip merge. To locate the image, you only need to draw two of these three special rays.

Exercise 9

(a) Graphically locate the image of the object in Figure (31).

(b) A ray starts out from the tip of the object in the direction of the dotted line shown. Trace out this ray through the lens and show where it goes on the back side of the lens.

In Exercise 9, you found that, once you have located the image, you can trace out any other ray from the tip of the object that passes through the lens, because these rays must all pass through the tip of the image.

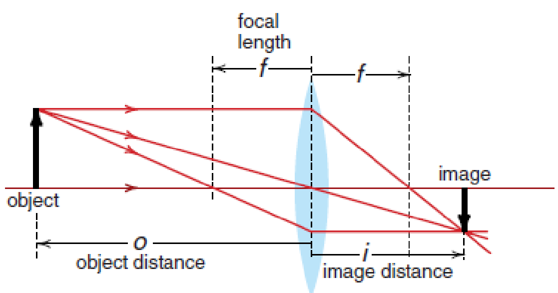


Figure 30
Locating the image using ray tracing. Three rays are easy to draw. One ray goes straight through the center of the lens. The top ray, parallel to the axis, intersects the axis where parallel rays would focus. A ray going through the left focus, comes out parallel to the axis. The image of the arrow tip is located where these rays intersect.

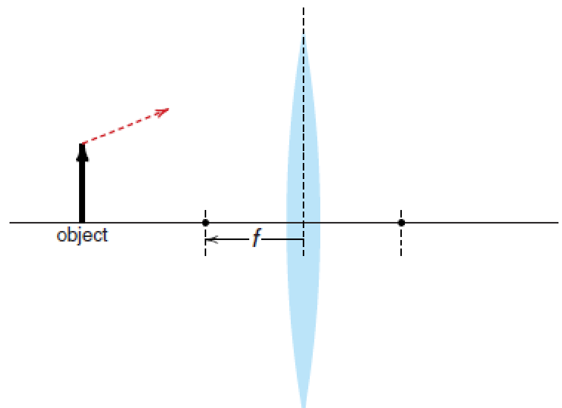


Figure 31
Locate the image of the arrow, and then trace the ray starting out in the direction of the dotted line.

Negative Image Distance

The lens equation is more general than you might expect, for it works equally well for positive and negative distances and focal lengths. Let us start by seeing what we mean by a negative image distance. Writing Equation 15 in the form

$$\frac{1}{i} = \frac{1}{f} - \frac{1}{o} \qquad (16a)$$

let us see what happens if $1/o$ is bigger than $1/f$ so that i turns out to be negative. If $1/o$ is bigger than $1/f$, that means that o is less than f and we have placed the object within the focal length as shown in Figure (33).

When we trace out two rays from the tip of the image, we find that the rays diverge after they pass through the lens. They diverge as if they were coming from a point behind the object, a point shown by the dotted lines. In this case we have what is called a ***virtual image***, which is located at a ***negative image distance*** (i). This negative image distance is correctly given by the lens equation (16a).

(We will not drag you through another geometrical proof of the lens equation for negative image distances. It should be fairly convincing that just when the image distance becomes negative in the lens equation, the geometry shows that we switch from a real image on the right side of the lens to a virtual image on the left.)

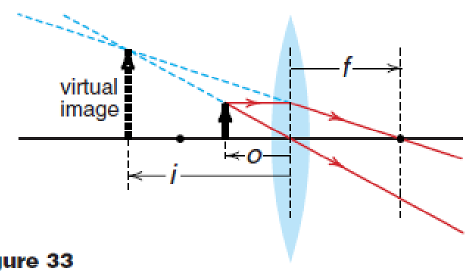


Figure 33
When the object is located within the focal length, we get a virtual image behind the object.

Negative Focal Length and Diverging Lenses

In Figure (33) we got a virtual image by moving the object inside the focal length. Another way to get a virtual image is to use a diverging lens as shown in Figure (34). Here we have drawn the three special rays, but the role of the focal point is reversed. The ray through the center of the lens goes through the center as before. The top ray parallel to the axis of the lens diverges outward as if it came from the focal point on the left side of the lens. The ray from the tip of the object headed for the right focal point, comes out parallel to the axis. Extending the diverging rays on the right, back to the left side, we find a virtual image on the left side.

You get diverging lenses by using concave surfaces as shown in Figure (34). In the lens maker's equation,

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \qquad \textit{lensmaker's equation} \qquad (15)$$

you replace 1/r by – 1/r for any concave surface. If $1/f$ turns out negative, then you have a diverging lens. Using this negative value of f in the lens equation (with $f = -|f|$) we get

$$\frac{1}{i} = -\left(\frac{1}{|f|} + \frac{1}{o}\right) \qquad (16b)$$

This always gives a negative image distance i, which means that diverging lenses only give virtual images.

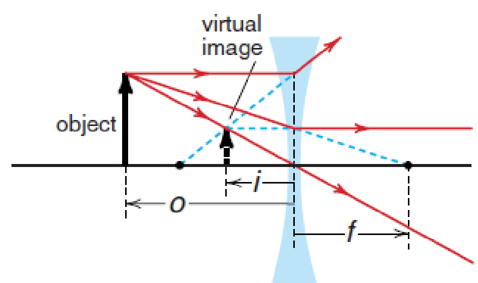


Figure 34
A diverging lens always gives a virtual image.

There is a very, very simple relationship between the object distance o, the image distance i and the lens focal length f. It is

$$\boxed{\frac{1}{o}+\frac{1}{i}=\frac{1}{f}} \qquad \textit{the lens equation} \qquad (16)$$

Equation 16 is worth memorizing if you are going to do any work with lenses. It is the equation you will use all the time, it is easy to remember, and as you will see now, the derivation requires some trigonometry you are not likely to remember. We will take you through the derivation anyway, because of the importance of the result.

In Figure (32a), we have an object of height A that forms an inverted image of height B. We located the image by tracing the top ray parallel to the axis that passes through the focal point behind the lens, and by tracing the ray that goes through the center of the lens.

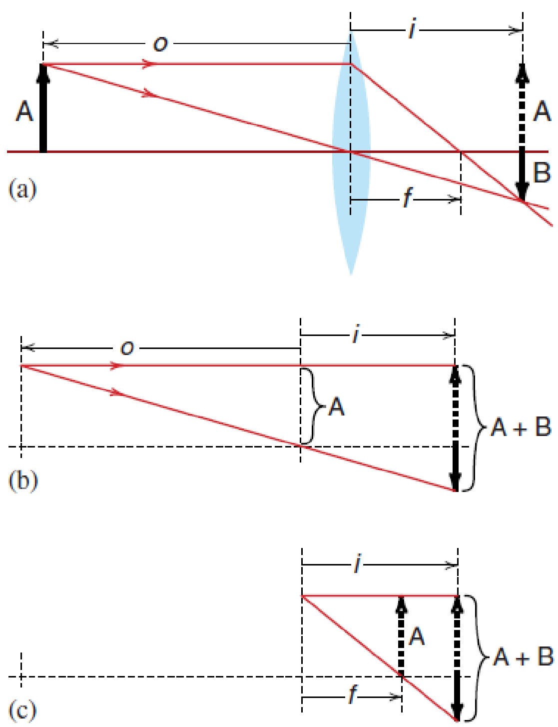


Figure 32
Derivation of the lens equation.

In Figure (32b) we have selected one of the triangles that appears in Figure (32a). The triangle starts at the tip of the object, goes parallel to the axis over to the image, and then down to the tip of the image. The length of the triangle is $(o + i)$ and the height of the base is (A+B). The lens cuts this triangle to form a smaller similar triangle whose length is o and base is (A). The ratio of the base to length of these similar triangles must be equal, giving

$$\frac{A}{o}=\frac{A+B}{(o+i)} \Rightarrow \frac{(A+B)}{A}=\frac{(o+i)}{o} \qquad (17)$$

In Figure (32c) we have selected another triangle which starts where the top ray hits the lens, goes parallel to the axis over to the image, and down to the tip of the image. This triangle has a length i and a base of height (A+B) as shown. This triangle is cut by a vertical line at the focal plane, giving a smaller similar triangle of length f and base (A) as shown. The ratio of the length to base of these similar triangles must be equal, giving

$$\frac{A}{f}=\frac{A+B}{i} \Rightarrow \frac{(A+B)}{A}=\frac{i}{f} \qquad (18)$$

Combining Equations 17 and 18 gives

$$\frac{i}{f}=\frac{o+i}{o}=1+\frac{i}{o} \qquad (19)$$

Finally, divide both sides by i and we get

$$\frac{1}{f}=\frac{1}{i}+\frac{1}{o} \qquad \textit{lens equation} \qquad (16)$$

which is the lens equation, as advertised.

Note that the lens equation is an exact consequence of the geometrical construction shown back in Figure (30). There is no restriction about small angles. However if you are using spherical lenses, you have to stick to small angles or the light will not focus to a point.

Exercise 10

You have a lens making machine that can grind surfaces, either convex or concave, with radius of curvatures of either 20 cm or 40 cm, or a flat surface. How many different kinds of lenses can you make? What is the focal length and the name of the lens type for each lens? Figure (35) shows the names given to the various lens types.

Negative Object Distance

With the lens equation, we can have negative image distances and negative focal lengths, and also negative object distances as well.

In all our drawings so far, we have drawn rays coming out of the tip of an object located at a positive object distance. A negative object distance means we have a virtual object where rays are converging toward the tip of the virtual object but don't get there. A comparison of the rays emerging from a real object and converging toward a virtual object is shown in Figure (36). The converging rays (which were usually created by some other lens) can be handled with the lens equation by assuming that the distance from the lens to the virtual object is negative.

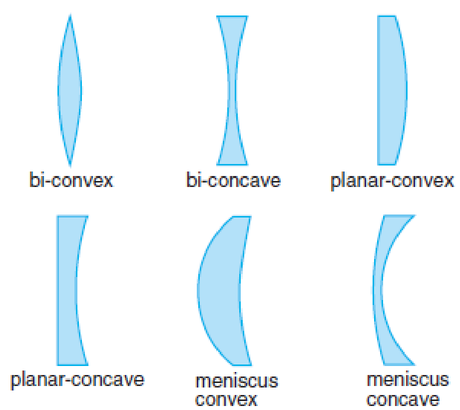


Figure 35
Various lens types. Note that eyeglasses are usually meniscus convex or meniscus concave.

As an example, suppose we have rays converging to a point, and we insert a diverging lens whose negative focal length $f=-|f|$ is equal to the negative object distance $o=-|o|$ as shown in Figure (37). The lens equation gives

$$\frac{1}{i}=\frac{1}{f}-\frac{1}{o}=\frac{1}{-|f|}-\frac{1}{-|o|}=\frac{1}{|o|}-\frac{1}{|f|} \quad (20)$$

If $|f|=|o|$, then $1/i=0$ and the image is infinitely far away. This means that the light emerges as a parallel beam as we showed in Figure (37).

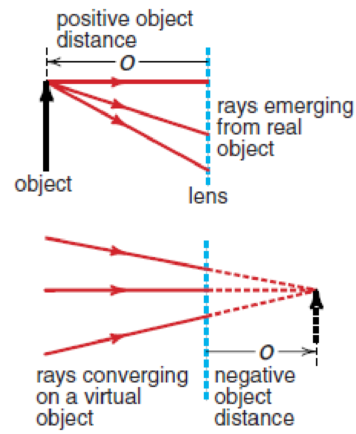


Figure 36
Positive and negative object distances.

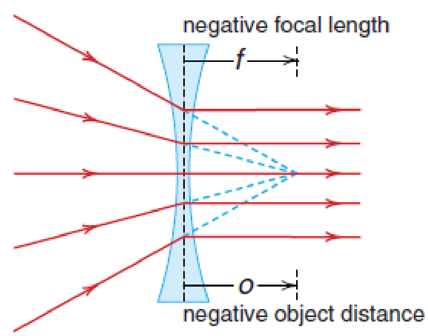


Figure 37
Negative focal length.

Multiple Lens Systems

Using the lens equation, and knowing how to handle both positive and negative distances and focal lengths, you can design almost any simple lens system you want. The idea is to work your way through the system, one lens at a time, where the image from one lens becomes the object for the next. We will illustrate this process with a few examples.

As our first example, consider Figure (38a) where we have two lenses of focal lengths $f_1 = 10$ cm and $f_2 = 12$ cm separated by a distance D = 40 cm. An object placed at a distance $o_1 = 17.5$ cm from the first lens creates an image a distance i_1 behind the first lens. Using the lens equation, we get

$$i_1 = \frac{1}{f_1} - \frac{1}{o_1} = \frac{1}{10} - \frac{1}{17.5} = \frac{1}{23.33} \qquad (21)$$

$$i_1 = 23.33 \text{ cm}$$

the same distance we got graphically in Figure (38a).

This image, which acts as the object for the second lens has an object distance

$$o_2 = \text{D} - i_1 = 40 \text{ cm} - 23.33 \text{ cm} = 16.67 \text{ cm}$$

This gives us a final upright image at a distance i_2 given by

$$\frac{1}{i_2} = \frac{1}{f_2} - \frac{1}{o_2} = \frac{1}{12} - \frac{1}{16.67} = \frac{1}{42.86} \qquad (22)$$

$$\boxed{i_2 = 42.86 \text{ cm}} \qquad (23)$$

which also accurately agrees with the geometrical construction.

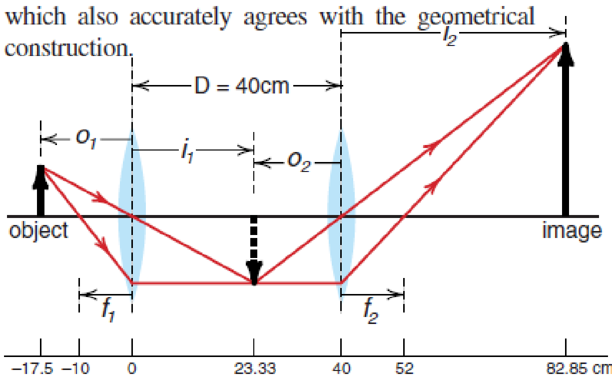


Figure 38a
Locating the image in a two lens system.

In Figure (38b), we moved the second lens up to within 8 cm of the first lens, so that the first image now falls behind the second lens. We now have a negative object distance

$$o_2 = \text{D} - i_1 = 8 \text{ cm} - 23.33 \text{ cm} = -15.33 \text{ cm}$$

Using this negative object distance in the lens equation gives

$$\frac{1}{i_2} = \frac{1}{f_2} - \frac{1}{o_2} = \frac{1}{12} - \frac{1}{-15.33}$$

$$= \frac{1}{12} + \frac{1}{15.33} = \frac{1}{6.73}$$

$$\boxed{i_2 = 6.73 \text{ cm}} \qquad (24)$$

In the geometrical construction we find that the still inverted image is in fact located 6.73 cm behind the second image.

While it is much faster to use the lens equation than trace rays, it is instructive to apply both approaches for a few examples to see that they both give the same result. In drawing Figure (38b) an important ray was the one that went from the tip of the original object, down through the first focal point. This ray emerges from the first lens traveling parallel to the optical axis. The ray then enters the second lens, and since it was parallel to the axis, it goes up through the focal point of the second lens as shown. The second image is located by drawing the ray that passes straight through the second lens, heading for the tip of the first image. Where these two rays cross is where the tip of the final image is located.

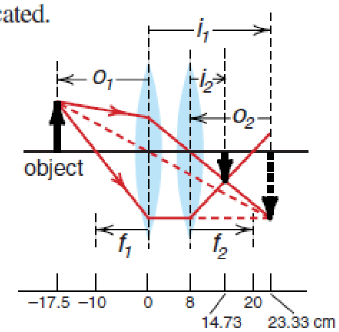


Figure 38b
We moved the second lens in so that the second object distance is negative. We now get an inverted image 6.73 cm from the second lens.

In Figure (38c) we sketched a number of rays passing through the first lens, heading for the first image. These rays are converging on the second lens, which we point out in Figure (36b) was the condition for a negative object distance.

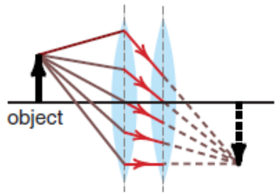


Figure 38c

Two Lenses Together

If you put two thin lenses together, as shown in Figure (39), you effectively create a new thin lens with a different focal length. To find out what the focal length of the combination is, you use the lens equation twice, setting the second object distance o_2 equal to minus the first image distance $-i_1$.

$$o_2 = -i_1 \qquad \textit{for two lenses together} \tag{25}$$

From the lens equations we have

$$\frac{1}{i_1} = \frac{1}{f_1} - \frac{1}{o_1} \tag{26}$$

$$\frac{1}{i_2} = \frac{1}{f_2} - \frac{1}{o_2} \tag{27}$$

Setting $o_2 = -i_1$ in Equation 27 gives

$$\frac{1}{i_2} = \frac{1}{f_2} - \frac{1}{(-i_1)} = \frac{1}{f_2} + \frac{1}{i_1}$$

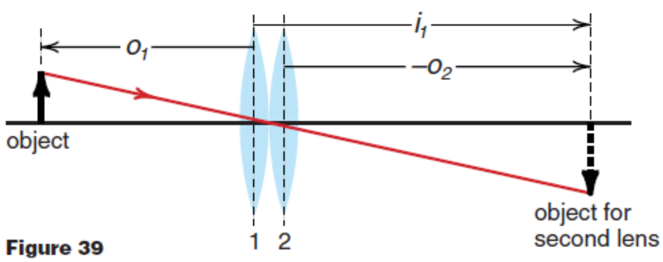


Figure 39
Two lenses together. Since the object for the second lens is on the wrong side of the lens, the object distance o_2 is negative in this diagram. If the lenses are close together, i_1 and $-o_2$ are essentially the same.

Using Equation 26 for $1/i_1$ gives

$$\frac{1}{i_2} = \frac{1}{f_2} + \frac{1}{f_1} - \frac{1}{o_1}$$

$$\frac{1}{o_1} + \frac{1}{i_2} = \frac{1}{f_1} + \frac{1}{f_2} \tag{28}$$

Now o_1 is the object distance and i_2 is the image distance for the pair of lenses. Treating the pair of lenses as a single lens, we should have

$$\frac{1}{o_1} + \frac{1}{i_2} = \frac{1}{f} \tag{29}$$

where f is the focal length of the combined lens.

Comparing Equations 28 and 29 we get

$$\boxed{\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}} \qquad \textit{focal length of two thin lenses together} \tag{30}$$

as the simple formula for the combined focal length.

Exercise 11

(a) Find the image distances i_2 for the geometry of Figures (38), but with the two lenses reversed, i.e., with $f_1 = 12$ cm, $f_2 = 10$ cm. Do this for both length D = 40 cm and D = 8 cm.

(b) If the two lenses are put together (D = 0) what is the focal length of the combination?

Magnification

It is natural to define the magnification created by a lens as the ratio of the height of the image to the height of the object. In Figure (40) we have reproduced Figure (38a) emphasizing the heights of the objects and images.

We see that the shaded triangles are similar, thus the ratio of the height B of the first image to the height A of the object is

$$\frac{\mathrm{B}}{\mathrm{A}} = \frac{i_1}{o_1} \tag{31}$$

We could define the magnification in the first lens as the ratio of B/A, but instead we will be a bit tricky and include a - (minus) sign to represent the fact that the image is inverted. With this convention we get

$$\boxed{m_1 = \frac{-\mathrm{B}}{\mathrm{A}} = \frac{-i_1}{o_1}} \qquad \textit{definition of magnification m} \tag{32}$$

Treating B as the object for the second lens gives

$$m_2 = \frac{-\mathrm{C}}{\mathrm{B}} = \frac{-i_2}{o_2} \tag{33}$$

The total magnification m_{12} in going from the object A to the final image C is

$$m_{12} = \frac{\mathrm{C}}{\mathrm{A}} \tag{34}$$

which has a + sign because the final image C is upright. But

$$\frac{\mathrm{C}}{\mathrm{A}} = \left(\frac{-\mathrm{C}}{\mathrm{B}}\right)\left(\frac{-\mathrm{B}}{\mathrm{A}}\right) \tag{35}$$

Thus we find that the final magnification is the product of the magnifications of each lens.

$$\boxed{m_{12} = m_1 m_2} \tag{36}$$

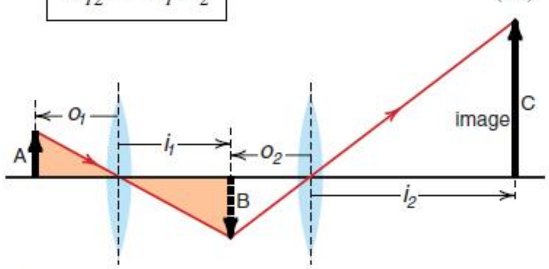


Figure 40
Magnification of two lenses.

Exercise 12

Figures (38) and (40) are scale drawings, so that the ratio of image to object sizes measured from these drawings should equal the calculated magnifications.

(a) Calculate the magnifications m_1, m_2 and m_{12} for Figure (38a) or (40) and compare your results with magnifications measured from the figure.

(b) Do the same for Figure (38b). In Figure (38b), the final image is inverted. Did your final magnification m_{12} come out negative?

Exercise 13

Figure (41a) shows a magnifying glass held 10 cm above the printed page. Since the object is inside the focal length we get a virtual image as seen in the geometrical construction of Figure (41b). Show that our formulas predict a positive magnification, and estimate the focal length of the lens. (Answer: about 17 cm.)

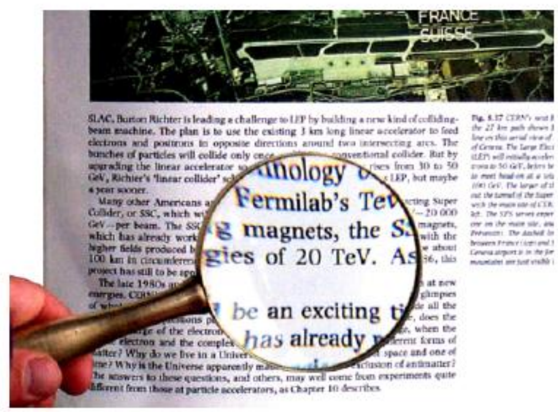


Figure 41a
Using a magnifying glass.

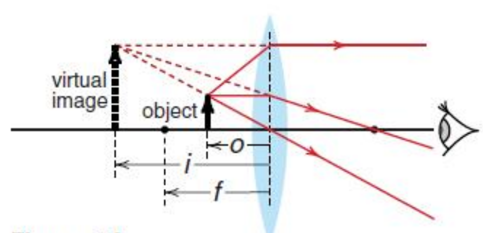


Figure 41b
When the magnifying glass is less than a focal length away from the object, we see an upright virtual image.

THE HUMAN EYE

A very good reason for studying geometrical optics is to understand how your own eye works, and how the situation is corrected when something goes wrong.

Back in Exercise 6 (p21), during our early discussion of spherical lens surfaces, we considered as a model of an eye a sphere of index of refraction n_2, where n_2 was chosen so that parallel rays which entered the front surface focused on the back surface as shown in Figure (27d). The value of n_2 turned out to be $n_2 = 2.0$. Since the only common substance with an index of refraction greater than zircon at $n = 1.923$ is diamond at $n = 2.417$, it would be difficult to construct such a model eye. Instead some extra focusing capability is required, both to bring the focus to the back surface of the eye, and to focus on objects located at various distances.

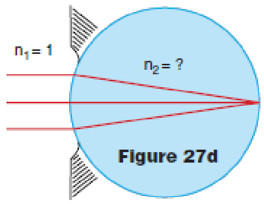
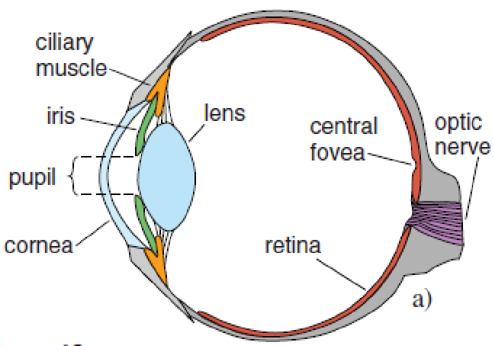


Figure 27d



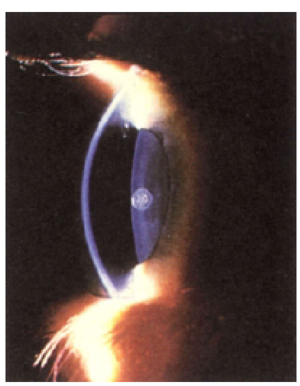


Figure 42
The human eye. The cornea and the lens together provide the extra focusing power required to focus light on the retina. (Photograph of the human eye by Lennart Nilsson.)

Figure (42a) is a sketch of the human eye and Figure 42b a remarkable photograph of the eye. As seen in (42a), light enters the ***cornea*** at the front of the eye. The amount of light allowed to enter is controlled by the opening of the ***iris***. Together the cornea and ***crystalline lens*** focuses light on the retina which is a film of ***nerve fibers*** on the back surface of the eye. Information from the new fibers is carried to the brain through the optic nerve at the back. In the retina there are two kinds of nerve fibers, called ***rods*** and ***cones***. Some of the roughly 120 million rods and 7 million cones are seen magnified about 5000 times in Figure (43). The slender ones, the rods, are more sensitive to dim light, while the shorter, fatter, cones, provide our color sensitivity.

In our discussion of the human ear, we saw how there was a mechanical system involving the basilar membrane that distinguished between the various frequencies of incoming sound waves. Information from nerves attached to the basilar membrane was then enhanced through processing in the local nerve fibers before being sent to the brain via the auditory nerve. In the eye, the nerve fibers behind the retina, some of which can be seen on the right side of Figure (43), also do a considerable amount of information processing before the signal travels to the brain via the optic nerve. The way that information from the rods and cones is processed by the nerve fibers is a field of research.

Returning to the front of the eye we have the surface of the cornea and the crystalline lens focusing light on the retina. Most of the focusing is done by the cornea. The shape, and therefore the focal length of the crystalline lens can be altered slightly by the ***ciliary muscle*** in order to bring into focus objects located at different distances.

Figure 43
Rods and cones in the retina. The thin ones are the rods, the fat ones the cones.

In a normal eye, when the ciliary muscle is in its resting position, light from infinity is focused on the retina as shown in Figure (44a). To see a closer object, the ciliary muscles contract to shorten the focal length of the cornea-lens system in order to continue to focus light on the cornea (44b). If the object is too close as in Figure (44c), the light is no longer focused and the object looks blurry. The shortest distance at which the light remains in focus is called the ***near point***. For children the near point is as short as 7 cm, but as one ages and the crystalline lens becomes less flexible, the near point recedes to something like 200 cm. This is why older people hold written material far away unless they have reading glasses.

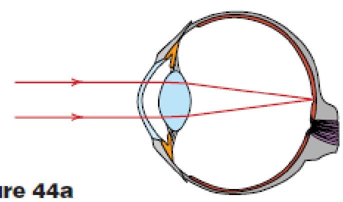


Figure 44a
Parallel light rays from a distant object are focuses on the retina when the ciliary muscles are in the resting position.

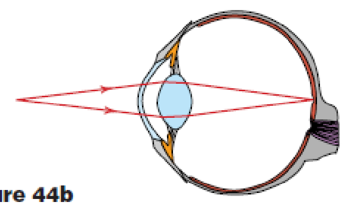


Figure 44b
The ciliary muscle contracts to shorten the focal length of the cornea-lens system in order to focus light from a more nearby object.

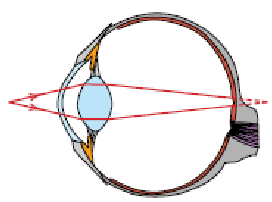


Figure 44c
When an object is to close, the light cannot be focused. The closest distance we. ....

Nearsightedness and Farsightedness

Not all of us have the so called ***normal*** eyes described by Figure (44). There is increasing evidence that those who do a lot of close work as children end up with a condition called ***nearsightedness*** or ***myopia*** where the eye is elongated and light from infinity focuses inside the eye as shown in Figure (45a). This can be corrected by placing a diverging lens in front of the eye to move the focus back to the retina as shown in Figure (45b).

The opposite problem, farsightedness, where light focuses behind the retina as shown in Figure (46a) is corrected by a converging lens as shown in Figure (46b).

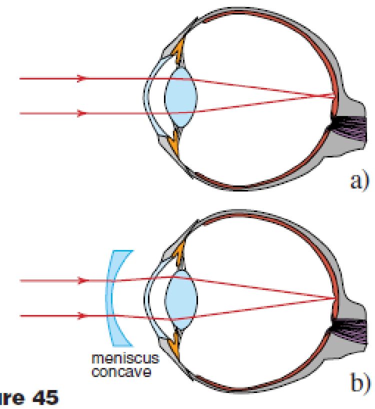


Figure 45
Nearsightedness can be corrected by a convex lens..

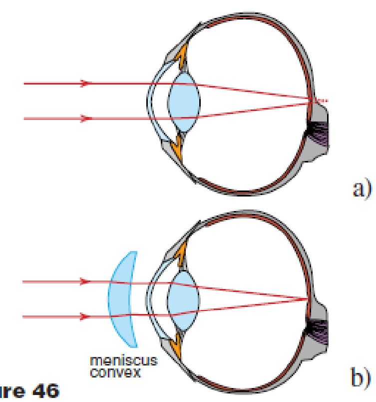


Figure 46
Farsightedness can be corrected by a convex lens

THE CAMERA

There are a number of similarities between the human eye and a simple camera. Both have an iris to control the amount of light entering, and both record an image at the focal plane of the lens. In a camera, the focus is adjusted, not by changing the shape of the lens as in the eye, but by moving the lens back and forth. The eye is somewhat like a TV camera in that both record images at a rate of about 30 per second, and the information is transmitted electronically to either the brain or a TV screen.

On many cameras you will find a series of numbers labeled by the letter *f*, called the ***f number*** or ***f stop***. Just as for the parabolic reflectors in figure 4 (p5), the *f* number is the ***ratio of the lens focal length to the lens diameter***. As you close down the iris of the camera to reduce the amount of light entering, you reduce the effective diameter of the lens and therefore increase the *f* number.

Exercise 14

The iris on the human eye can change the diameter of the opening to the lens from about 2 to 8 millimeters. The total distance from the cornea to the retina is typically about 2.3 cm. What is the range of *f* values for the human eye? How does this range compare with the range of *f* value on your camera? (If you have one of the automatic point and shoot cameras, the *f* number and the exposure time are controlled electronically and you do not get to see or control these yourself.)

Figure 47a
The Physics department's Minolta single lens reflex camera.

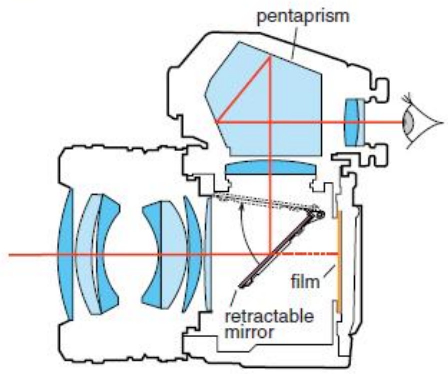


Figure 47b
The lens system for a Nikon single lens reflex camera. When you take the picture, the hinged mirror flips out of the way and the light reaches the film. Before that, the light is reflected through the prism to the eyepiece.

Depth of Field

There are three ways to control the exposure of the film in a camera. One is by the speed of the film, the second is the exposure time, and the third is the opening of the iris or *f* stop. In talking a picture you should first make sure the exposure is short enough so that motion of the camera and the subject do not cause blurring. If your film is fast enough, you can still choose between a shorter exposure time or a smaller *f* stop. This choice is determined by the ***depth of field*** that you want.

The concept of depth of field is illustrated in Figures (48a and b). In (48a), we have drawn the rays of light from an object to an image through an *f 2* lens, a lens with a focal length equal to twice its diameter. (The effective diameter can be controlled by a flexible diaphragm or iris like the one shown.) If you placed a film at the image distance, the point at the tip of the object arrow would focus to a point on the film. If you moved the film forward to position 1, or back to position 2, the image of the arrow tip would fill a circle about equal to the thickness of the three rays we drew in the diagram.

If the film were ideal, you could tell that the image at positions 1 or 2 was out of focus. But no film or recording medium is ideal. If you look closely enough there is always a graininess caused by the size of the basic medium like the silver halide crystals in black and white film, the width of the scan lines in an analog TV camera, or the size of the pixels in a digital camera. If the image of the arrow tip at position 1 is smaller than the grain or pixel size then you cannot tell that the picture is out of focus. You can place the recording medium anywhere between position 1 and 2 and the image will be as sharp as you can get.

In Figure (48b), we have drawn the rays from the same object passing through a smaller diameter *f 8* lens. Again we show by dotted lines positions 1 and 2 where the image of the arrow point would fill the same size circle as it did at positions 1 and 2 for the *f 2* lens above. Because the rays from the *f 8* lens fill a much narrower cone than those from the *f 2* lens, there is a much greater distance between positions 1 and 2 for the *f 8* lens.

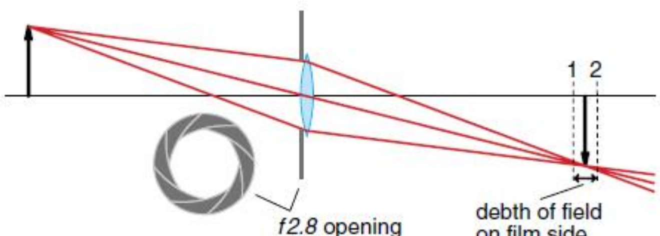


Figure 48a
A large diameter lens has a narrow depth of field.

Photograph taken at f 5.6.

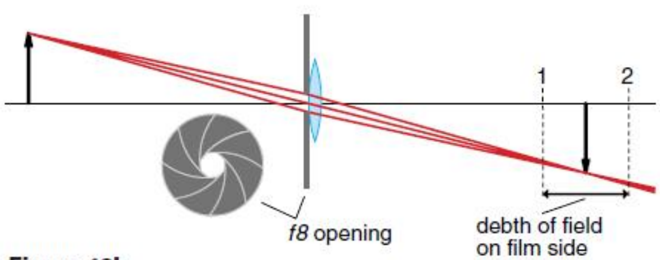


Figure 48b
Reducing the effective diameter of the lens increases the depth of field.

Photograph taken at f 22.

If Figures (48) represented a camera, you would not be concerned with moving the film back and forth. Instead you would be concerned with how far the image could be moved back and forth and still appear to be in focus. If the film were at the image position and you then moved the object in and out, you could not move it very far before it's image was noticeably out of focus with the *f2* lens. You could move it much farther for the *f8* lens.

This effect is illustrated by the photographs on the right side of Figures 48, showing a close-up tree and the distant tower on Baker Library at Dartmouth College. The upper picture taken at *f5.6* has a narrow depth of field, and the tower is well out of focus. In the bottom picture, taken at *f22*, has a much broader depth of field and the tower is more nearly in focus. (In both cases we focused on the nearby tree bark.)

Camera manufacturers decide how much blurring of the image is noticeable or tolerable, and then figure out the range of distances the object can be moved and still be acceptably in focus. This range of distance is called the ***depth of field***. It can be very short when the object is up close and you use a wide opening like *f2*. It can be quite long for a high *f* number like *f22*. The inexpensive fixed focus cameras use a small enough lens so that all objects are "in focus" from about 3 feet or 1 meter to infinity.

In the extreme limit when the lens is very small, the depth of field is so great that everything is in focus everywhere behind the lens. In this limit you do not even need a lens, a pinhole in a piece of cardboard will do. If enough light is available and the subject doesn't not move, you can get as good a picture with a pinhole camera as one with an expensive lens system. Our pinhole camera image in Figure (49) is a bit fuzzy because we used too big a pinhole.

(If you are nearsighted you can see how a pinhole camera works by making a tiny hole with your fingers and looking at a distant light at night without your glasses. Just looking at the light, it will look blurry. But look at the light through the hole made by your fingers and the light will be sharp. You can also see the eye chart better at the optometrists if you look through a small hole, but they don't let you do that.)

Figure 48c
Camera lens. This lens is set to f11, and adjusted to a focus of 3 meters or 10 ft. At this setting, the depth of field ranges from 2 to 5 meters.

Figure 49a
We made a pinhole camera by replacing the camera lens with a plastic film case that had a small hole poked into the end.

Figure 49b
Photograph of Baker library tower, taken with the pinhole camera above. If we had used a smaller hole we would have gotten a sharper focus.

Eye Glasses and a Home Lab Experiment

When you get a prescription for eyeglasses, the optometrist writes down number like -1.5, -1.8 to represent the ***power*** of the lenses you need. These cryptic number are the power of the lenses measured in ***diopters***. What a diopter is, is simply the reciprocal of the focal length $1/f$, where f is measured in meters. A lens with a power of 1 diopter is a converging lens with a focal length of 1 meter. Those of us who have lenses closer to –4 in power have lenses with a focal length of –25 cm, the minus sign indicating a diverging lens to correct for nearsightedness as shown back in Figure (45).

If you are nearsighted and want to measure the power of your own eyeglass lenses, you have the problem that it is harder to measure the focal length of a diverging lens than a converging lens. You can quickly measure the focal length of a converging lens like a simple magnifying glass by focusing sunlight on a piece of paper and measuring the distance from the lens to where the paper is starting to smoke. But you do not get a real image for a diverging lens, and cannot use this simple technique for measuring the focal length and power of diverging lenses used by the nearsighted.

As part of a project, some students used the following method to measure the focal length and then determine the power in diopters, of their and their friend's eyeglasses. They started by measuring the focal length f_0 of a simple magnifying glass by focusing the sun. Then they placed the magnifying glass and the eyeglass lens together, measured the focal length of the combination, and used the formula

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \qquad \text{(30 repeated)}$$

to calculate the focal length of the lens.

(Note that if you measure distances in meters, then $1/f_1$ is the power of lens 1 in diopters and $1/f_2$ that of lens 2. Equation 30 tells you that the power of the combination $1/f$ is the sum of the powers of the two lenses.

Exercise 15

Assume that you find a magnifying lens that focuses the sun at a distance of 10 cm from the lens. You then combine that with one of your (or a friends) eyeglass lenses, and discover that the combination focus at a distance of 15 cm. What is the power, in diopters, of

(a) The magnifying glass.

(b) The combination.

(c) The eyeglass lens.

Exercise 16 – Home Lab

Use the above technique to measure the power of your or your friend's glasses. If you have your prescription compare your results with what is written on the prescription. (The prescription will also contain information about axis and amount of astigmatism. That you cannot check as easily.

THE EYEPIECE

When the author was a young student, he wondered why you do not put your eye at the focal point of a telescope mirror. That is where the image of a distance object is, and that is where you put the film in order to record the image. You do not put your eye at the image because it would be like viewing an object by putting your eyeball next to it. The object would be hopelessly out of focus. Instead you look through an eyepiece.

The eyepiece is a magnifying glass that allows your eye to comfortably view an image or small object up close. For a normal eye, the least eyestrain occurs when looking at a distant object where the light from the object enters the eye as parallel rays. It is then that the ciliary muscles in the eye are in a resting position. If the image or small object is placed at the focal plane of a lens, as shown in Figure (50), light emerges from the lens as parallel rays. You can put your eye right up to that lens, and view the object or image as comfortably as you would view a distant scene.

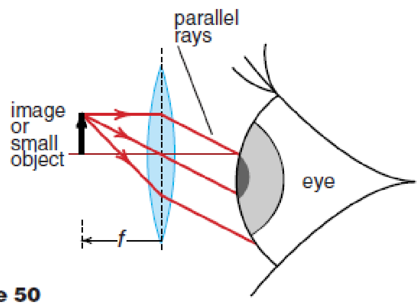


Figure 50
The eyepiece or magnifier. To look at small object, or to study the image produced by another lens or mirror, place the image or object at the focal plane of a lens, so that the light emerges as parallel rays that your eye can comfortably focus upon.

Exercise 17 - The Magnifying Glass

There are three distinct ways of viewing an object through a magnifying glass, which you should try for yourself. Get a magnifying glass and use the letters on this page as the object to be viewed.

(a) First measure the focal length of the lens by focusing the image of a distant object onto a piece of paper. A light bulb across the room or scene out the window will do.

(b) Draw some object on the paper, and place the paper at least several focal lengths from your eye. Then hold the lens about 1/2 a focal length above the object as shown in Figure (51a). You should now see an enlarged image of the object as indicated in Figure (51a). You are now looking at the virtual image of the object. Check that the magnification is roughly a factor of 2×.

(c) Keeping your eye in the same position, several focal lengths and at least 20 cm from the paper, pull the lens back toward your eye. The image goes out of focus when the lens is one focal length above the paper, and then comes back into focus upside down when the lens is farther out. You are now looking at the real image as indicated in Figure (51b). Keep your head far enough back that your eye can focus on this real image.

Hold the lens two focal lengths above the page and check that the inverted real image of the object looks about the same size as the object itself. (As you can see from Figure (51b), the inverted image should be the same size as the object, but 4 focal lengths closer.)

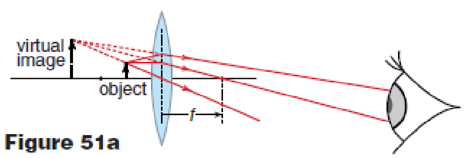


Figure 51a
Looking at the virtual image.

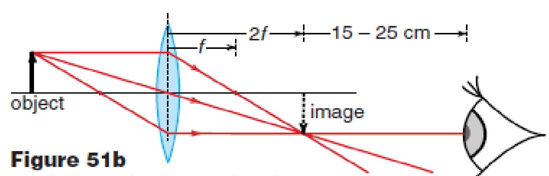


Figure 51b
Looking at the inverted real image.

(d) Now hold the lens one focal length above the page and put your eye right up to the lens. You are now using the lens as an eyepiece as shown in Figure (50). The letters will be large because your eye is close to them, and they will be comfortably in focus because the rays are entering your eye as parallel rays like the rays from a distant object. When you use the lens as an eyepiece you are not looking at an image as you did in parts (b) and (c) of this exercise, instead your eye is creating an image on your retina from the parallel rays.

(e) As a final exercise, hold the lens one focal length above a page of text, start with your eye next to the lens, and then move your head back. Since the light from the page is emerging from the lens as parallel rays, the size of the letters should not change as you move your head back. Instead what you should see is fewer and fewer letters in the magnifying glass as the magnifying glass itself looks smaller when farther away. This effect is seen in Figure (52).

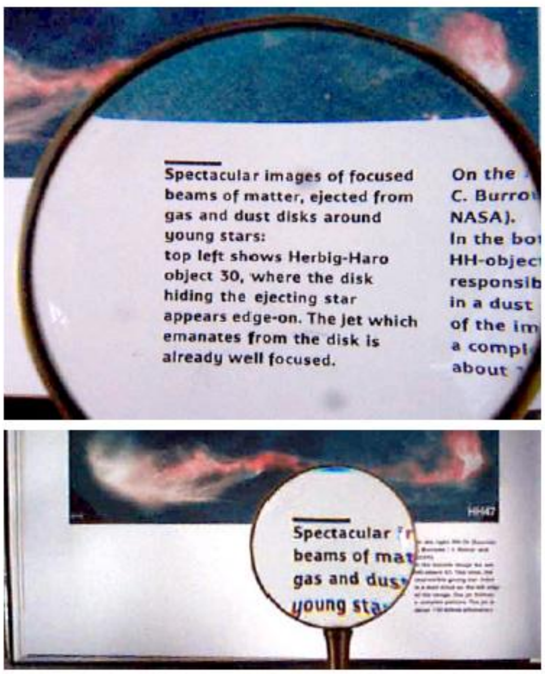


Figure 52
When the lens is one focal length from the page, the emerging rays are parallel. Thus the image letters do not change size as we move away. Instead the lens looks smaller, and we see fewer letters in the lens.

The Magnifier

When jewelers work on small objects like the innards of a watch, they use what they call a ***magnifier*** which can be a lens mounted at one end of a tube as shown in Figure (53). The length of the tube is equal to the focal length of the lens, so that if you put the other end of the tube up against an object, the lens acts as an eyepiece and light from the object emerges from the lens as parallel rays. By placing your eye close to the lens, you get a close up, comfortably seen view of the object. You may have seen jewelers wear magnifiers like that shown in Figure (54).

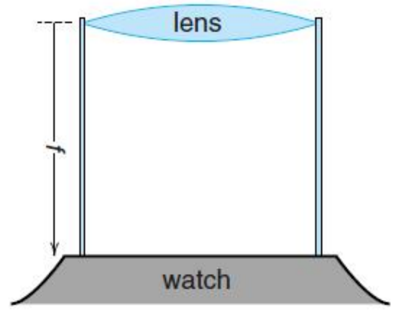


Figure 53
A magnifier.

Figure 54
Jeweler Paul Gross with magnifier lenses mounted in visor.

Angular Magnification

Basically all the magnifier does is to allow you to move the object close to your eye while keeping the object comfortably in focus. It is traditional to define the ***magnification*** of the magnifier as the ratio of the size of the object as seen through the lens to the size of the object as you would see it without a magnifier. By size, we mean the angle the object subtends at your eye. This is often called the ***angular magnification***.

The problem with this definition of magnification is that different people, would hold the object at different distances in order to look at it without a magnifier. For example, us nearsighted people would hold it a lot closer than a person with normal vision. To avoid this ambiguity, we can choose some standard distance like 25 cm, a standard near point, at which a person would normally hold an object when looking at it. Then the angular magnification of the magnifier is the ratio of the angle θ_m subtended by the object when using the magnifier, as shown in Figure (55a), to the angle θ_0 subtended by the object held at a distance of 25 cm, as shown in Figure (55b).

$$\text{angular magnification} = \frac{\theta_m}{\theta_0} \qquad \textit{angles defined in Figure 55} \tag{37}$$

To calculate the angular magnification we use the small angle approximation $\sin\theta \approx \theta$ to get

$$\theta_m = \frac{y}{f} \qquad \textit{from Figure 55a}$$

$$\theta_0 = \frac{y}{25\text{ cm}} \qquad \textit{from Figure 55b}$$

which gives

$$\text{angular magnification} = \frac{y/f}{y/25\text{ cm}} = \frac{25\text{ cm}}{f} \tag{38}$$

Thus if our magnifier lens has a focal length of 5 cm, the angular magnification is 5×. Supposedly the object will look five times bigger using the magnifier than without it.

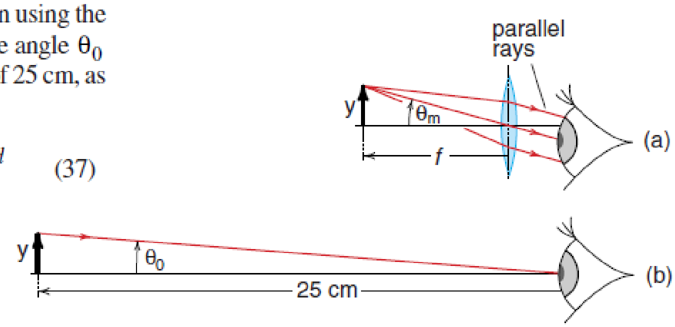


Figure 55
The angles used in defining angular magnification.

TELESCOPES

The basic design of a telescope is to have a large lens or parabolic mirror to create a bright real image, and then use an eyepiece to view the image. If we use a large lens, that lens is called an ***objective lens***, and the telescope is called a ***refracting telescope***. If we use a parabolic mirror, then we have a ***reflecting telescope***.

The basic design of a refracting telescope is shown in Figure (56). Suppose, as shown in Figure (56a), we are looking at a constellation of stars that subtend an angle θ_0 as viewed by the unaided eye. The eye is directed just below the bottom star and light from the top star enters at an angle θ_0. In Figure (56b), the lens system from the telescope is placed in front of the eye, and we are following the path of the light from the top star in the constellation.

The parallel rays from the top star are focused at the focal length f_0 of the objective lens. We adjust the eyepiece so that the image produced by the objective lens is at the focal point of the eyepiece lens, so that light from the image will emerge from the eyepiece as parallel rays that the eye can easily focus.

As with the magnifier, we define the magnification of the telescope as the ratio of the size of (angle subtended by) the object as seen through the object to the size of (angle subtended by) the object seen by the unaided eye. In Figure (56) we see that the constellation subtends an angle θ_0 as viewed by the unaided eye, and an angle θ_i when seen through the telescope. Thus we define the magnification of the telescope as

$$m = \frac{\theta_i}{\theta_0} \qquad \textit{magnification of telescope} \tag{39}$$

To calculate this ratio, we note from Figure (56c) that, using the small angle approximation $\sin\theta \approx \theta$, we have

$$\theta_0 = \frac{y_i}{f_0} \ ; \quad \theta_i = \frac{y_i}{f_e} \tag{40}$$

where f_0 and f_e are the focal lengths of the objective and eyepiece lens respectively. In the ratio, the image height y_i cancels and we get

$$m = \frac{\theta_i}{\theta_0} = \frac{y_i/f_e}{y_i/f_0}$$

$$\boxed{m = \frac{f_0}{f_e}} \tag{41}$$

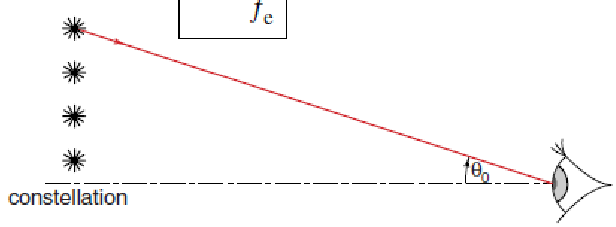


Figure 56a
The unaided eye looking at a constellation of stars that subtend an angle θ_o.

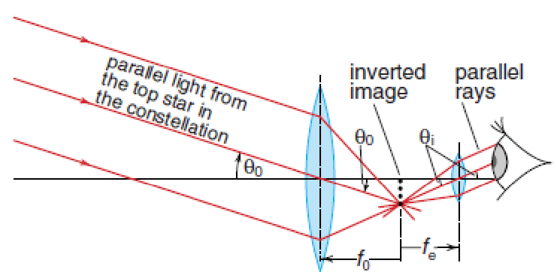


Figure 56b
Looking at the same constellation through a simple refracting telescope. The objective lens produces an inverted image which is viewed by the eyepiece acting as a magnifier. Note that the parallel light from the star focuses at the focal point of the objective lens. With the image at the focal point of the eyepiece lens, light from the image emerges as parallel rays that are easily focused by the eye.

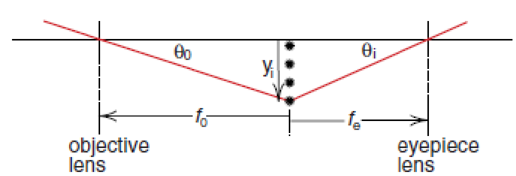


Figure 56c
Relationship between the angles $\boldsymbol{\theta_0}$, $\boldsymbol{\theta_i}$, and the focal lengths.

The same formula also applies to a reflecting telescope with f_0 the focal length of the parabolic mirror. Note that there is no arbitrary number like 25 cm in the formula for the magnification of a telescope because telescopes are designed to look at distant objects where the angle θ_0 the object subtends to the unaided eye is the same for everyone.

The first and the last of the important refracting telescopes are shown in Figures (57). The telescope was invented in Holland in 1608 by Hans Lippershy. Shortly after that, Galileo constructed a more powerful instrument and was the first to use it effectively in astronomy. With a telescope like the one shown in Figure (57a), he discovered the moons of Jupiter, a result that provided an explicit demonstration that heavenly bodies could orbit around something other than the earth. This countered the long held idea that the earth was at the center of everything and provided support for the Copernican sun centered picture of the solar system.

When it comes to building large refracting telescopes, the huge amount of glass in the objective lens becomes a problem. The 1 meter diameter refracting telescope at the Yerkes Observatory, shown in Figure (57b), is the largest refracting telescope ever constructed. That was built back in 1897. The largest reflecting telescope is the new 10 meter telescope at the Keck Observatory at the summit of the inactive volcano Mauna Kea in Hawaii. Since the area and light gathering power of a telescope is proportional to the area or the square of the diameter of the mirror or objective lens, the 10 meter Keck telescope is 100 times more powerful than the 1 meter Yerkes telescope.

Figure 57a
Galileo's telescope. With such an instrument Galileo discovered the moons of Jupiter.

Exercise 8

To build your own refracting telescope, you purchase a 3 inch diameter objective lens with a focal length of 50 cm. You want the telescope to have a magnification $m = 25\times$.

(a) What will be the *f* number of your telescope? (1 inch = 2.54 cm).

(b) What should the focal length of your eyepiece lens be?

(c) How far behind the objective lens should the eyepiece lens be located?

(d) Someone give you an eyepiece with a focal length of 10 mm. Using this eyepiece, what magnification do you get with your telescope?

(e) You notice that your new eyepiece is not in focus at the same place as your old eyepiece. Did you have to move the new eyepiece toward or away from the objective lens, and by how much?

(f) Still later, you decide to take pictures with your telescope. To do this you replace the eyepiece with a film holder. Where do you place the film, and why did you remove the eyepiece?

Figure 57b
The Yerkes telescope is the world's largest refracting telescope, was finished in 1897. Since then all larger telescopes have been reflectors.

Reflecting telescopes

In several ways, the reflecting telescope is similar to the refracting telescope. As we saw back in our discussion of parabolic mirrors, the mirror produces an image in the focal plane when the light comes from a distant object. This is shown in Figure (58a) which is similar to our old Figure (4). If you want to look at the image with an eyepiece, you have the problem that the image is in front of the mirror where, for a small telescope, your head would block the light coming into the scope. Issac Newton, who invented the reflecting telescope, solved that problem by placing a small, flat, 45° reflecting surface inside the telescope tube to deflect the image outside the tube as shown in Figure (58b). There the image can easily be viewed using an eyepiece. Newton's own telescope is shown in Figure (58d). Another technique, used in larger telescopes, is to reflect the beam back through a hole in the mirror as shown in Figure (58c).

The reason Newton invented the reflecting telescope was to avoid an effect called ***chromatic aberration***. When white light passes through a simple lens, different wavelengths or colors focus at different distances behind the lens. For example if the yellow light is in focus the red and blue images will be out of focus. In contrast, all wavelengths focus at the same point using a parabolic mirror.

Figure 58d
Issac Newton's reflecting telescope.

However, problems with keeping the reflecting surface shinny, and the development of lens combinations that eliminated chromatic aberration, made refracting telescopes more popular until the late 1800's. The invention of the durable silver and aluminum coatings on glass brought reflecting telescopes into prominence in the twentieth century.

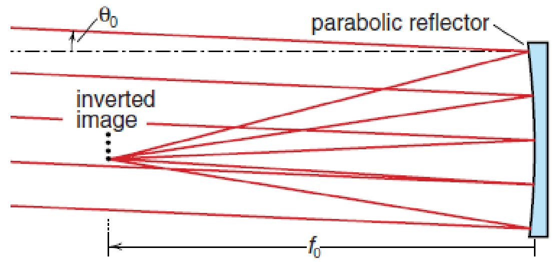


Figure 58a
A parabolic reflector focuses the parallel rays from a distant object, forming an image a distance f_0 in front of the mirror.

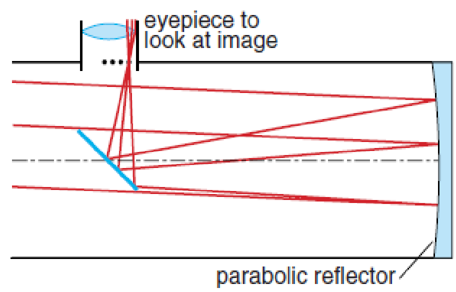


Figure 58b
Issac Newton's solution to viewing the image was to deflect the beam using a 45° reflecting surface so that the eyepiece could be outside the telescope tube.

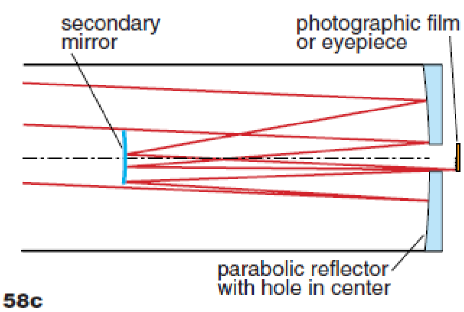


Figure 58c
For large telescopes, it is common to reflect the beam back through a hole in the center of the primary mirror. This arrangement is known as the Cassegrain *design.*

Large Reflecting Telescopes.

The first person to build a really large reflecting telescope was William Hershel, who started with a two inch reflector in 1774 and by 1789 had constructed the four foot diameter telescope shown in Figure (59a). Among Hershel's accomplishments was the discovery of the planet Uranus, and the first observation a distant nebula. It would be another 130 years before Edwin Hubble, using the 100 inch telescope on Mt. Wilson would conclusively demonstrate that such nebula were in fact galaxies like our own milky way. This also led hubble to discover the expansion of the universe.

During most of the second half of the twentieth century, the largest telescope has been the 200 inch (5 meter) telescope on Mt. Palomar, shown in Figure (59b). This was the first telescope large enough that a person could work at the prime focus, without using a secondary mirror. Hubbel himself is seen in the observing cage at the prime focus in Figure (59c).

Recently it has become possible to construct mirrors larger than 5 meters in diameter. One of the tricks is to cast the molten glass in a rotating container and keep the container rotating while the glass cools. A rotating liquid has a parabolic surface. The faster the rotation the deeper the parabola. Thus by choosing the right rotation speed, one can cast a mirror blank that has the correct parabola built in. The surface is still a bit rough, and has to be polished smooth, but the grinding out oh large amounts of glass is avoided. The 6.5 meter mirror, shown in Figures (59d and e), being installed on top of Mt. Hopkins in Arizona, was built this way. Seventeen tons of glass would have to have been ground out if the parabola had not been cast into the mirror blank.

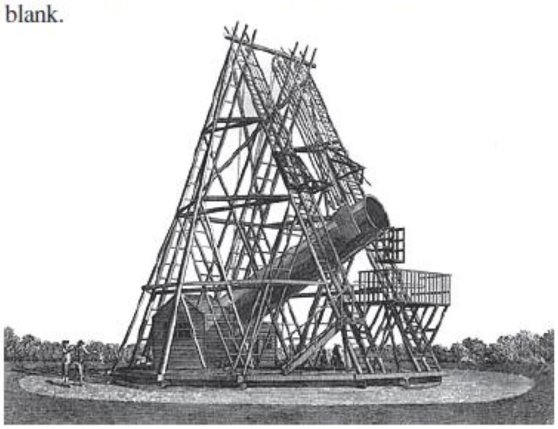
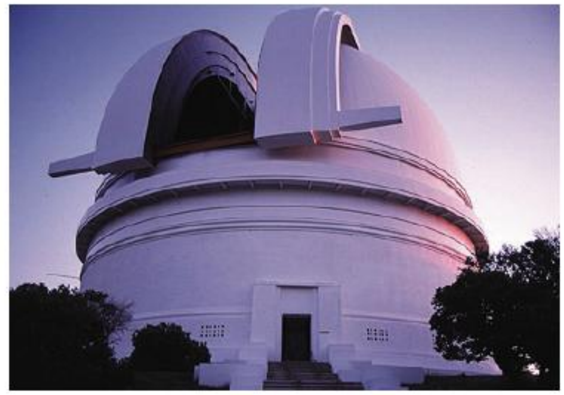
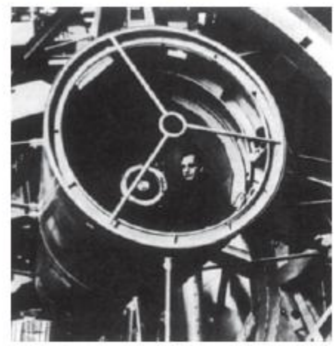
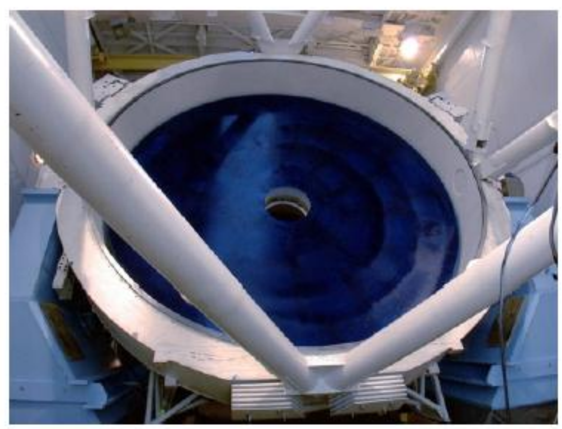


Figure 59a
William Hershel's 4 ft diameter, 40 ft long reflecting telescope which he completed in 1789.





Figures 59b,c
The Mt. Palomar 200 inch telescope. Below is Edwin Hubble in the observing cage.



Figures 59d,e
The 6.5 meter MMT telescope atop Mt. Hopkins. Above, the mirror has not been silvered yet. The blue is a temporary protective coating. Below, the mirror is being hoisted into the telescope frame.

Hubbel Space Telescope

An important limit to telescopes on earth, in their ability to distinguish fine detail, is turbulence in the atmosphere. Blobs of air above the telescope move around causing the star image to move, blurring the picture. This motion, on a time scale of about 1/60 second, is what causes stars to appear to twinkle.

The effects of turbulance, and any distortion caused by the atmosphere, are eliminated by placing the telescope in orbit above the atmosphere. The largest telescope in orbit is the famous Hubble telescope with its 1.5 meter diameter mirror, seen in Figure (60). After initial problems with its optics were fixed, the Hubble telescope has produced fantastic images like that of the Eagle nebula seen in Figure (7-17) reproduced here.

With a modern telescope like the Keck (see next page), the effects of atmospheric turbulance can mostly be eliminated by having a computer can track the image of a bright star. The telescopes mirror is flexible enough that the shape of the mirror can then be be modified rapidly and by a tiny amount to keep the image steady.

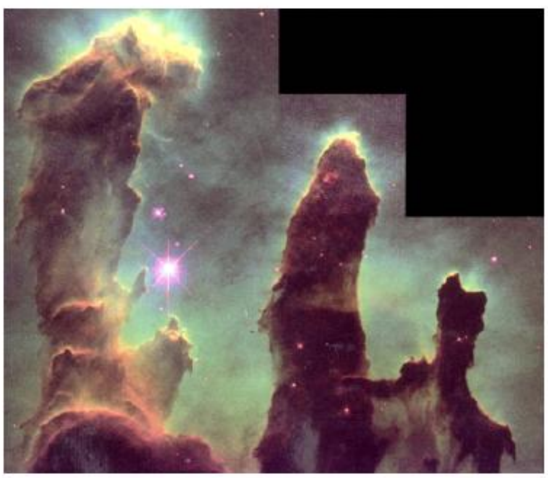


Figure 7-17
The eagle nebula, birthplace of stars. This Hubble photograph, which apeared on the cover of Time magazine, is perhaps the most famous.

Figure 60a
The Hubble telescope mirror. How is that for a shaving mirror?

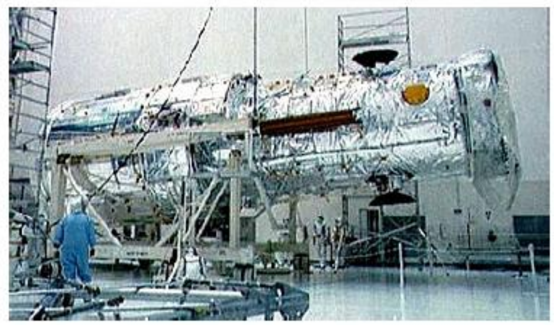


Figure 60b
Hubble telescope before launch.

Figure 60c
Hubble telescope being deployed.

World's Largest Optical Telescope

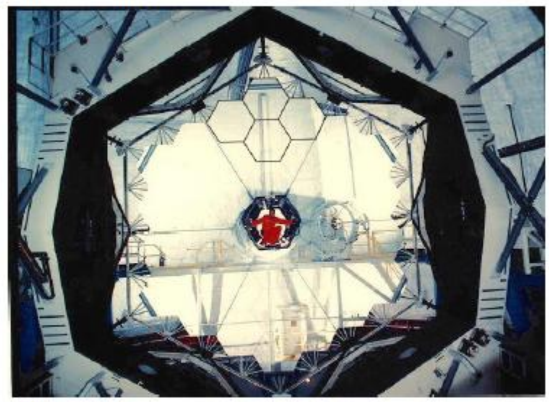
As of 1999, the largest optical telescope in the world is the Keck telescope located atop the Mauna Kea volcano in Hawaii, seen in Figure (61a). Actually there are two identical Keck telescopes as seen in the close-up, Figure (61b). The primary mirror in each telescope consists of 36 hexagonal mirrors fitted together as seen in Figure (61c) to form a mirror 10 meters in diameter. This is twice the diameter of the Mt. palomar mirrorwe discussed earlier.

The reason for building two Keck telescopes has to do with the wave nature of light. As we mentioned in the introduction to this chapter, geometrical optics works well when the objects we are studying are large compared to the wavelength of light. This is illustrated by the ripple tank photographs of Figures (33-3) and (33-8) reproduced here. In the left hand figure, we see we see a wave passing through a gap that is considerably wider than the wave's wavelength. On the other side of the gap there is a well defined beam with a distinct shadow. This is what we assume light waves do in geometrical optics.

In contrast, when the water waves encounter a gap whose width is comparable to a wavelength,as in the right hand figure, the waves spread out on the far side. This is a phenomenon called ***diffraction***. We can even see some diffraction at the edges of the beam emerging from the wide gap.

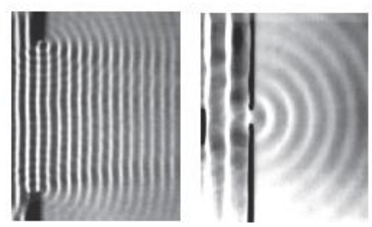
Figures 61 a,b
The Keck telescopes atop Mauna Kea volcano in Hawaii

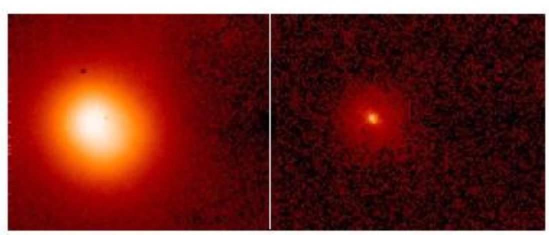
Diffraction also affects the ability of telescopes to form sharp images. The bigger the diameter of the telescope, compared to the light wavelength, the less important diffraction is and the sharper the image that can be formed. By combining the output from the two Keck telescopes, one creates a telescope whose effective diameter, for handling diffraction effects, is equal to the 90 meter separation of the telescopes rather than just the 10 meter diameter of one telescope. The great improvement in the image sharpness that results is seen in Figure (61d). On the left is the best possible image of a star, taken using one telescope alone. When the two telescopes are combined, they get the much sharper image on the right.



Figures 61 c
The 36 mirrors forming Keck's primary mirror. We have emphasized the outline of the upper 4 mirrors.

Figures 33-3,8
Unless the gap is wide in comparison to a wavelength, diffraction effects are important.





Figures 62
Same star, photographed on the left using one scope, on the right with the two Keck telescopes combined.

You might wonder why you have to cool an infrared camera and not a visible light camera. The answer is that warm bodies emit infrared radiation. The hotter the object, the shorter the wavelength of the radiation. If an object is hot enough, it begins to glow in visible light, and we say that the object is red hot, or white hot. Since you do not want the infrared detector in the camera seeing camera walls glowing "infrared hot", the camera has to be cooled.

Not all infrared radiation can make it down through the earth's atmosphere. Water vapor, for example is very good at absorbing certain infrared wavelengths. To observe the wavelengths that do not make it through, infrared telescopes have been placed in orbit. Figure 65 is an artist's drawing of the Infrared Astronomical Satellite (IRAS) which was used to make the infrared map of the entire sky seen in Figure (66). The map is oriented so that the Milky Way, our own galaxy, lies along the center horizontal plane. In visible light photographs, most of the stars in our own galaxy are obscured by the immense amount of dust in the plane of the galaxy. But in an infrared photograph, the huge concentration of stars in the plane of the galaxy show up clearly.

At the center of our galaxy is a gigantic black hole, with a mass of millions of suns. For a visible light telescope, the galactic center is completely obscured by dust. But the center can be clearly seen in the infrared photograph of Figure (67), taken by the Mt. Hopkins telescope of Figure 64. This is not a single exposure, instead it is a composite of thousands of images in that region of the sky. Three different infrared wavelengths were recorded, and the color photograph was created by displaying the longest wavelength image as red, the middle wavelength as green, and the shortest wavelength as blue. In this photograph, you not only see the intense radiation from the region of the black hole at the center, but also the enormous density of stars at the center of our galaxy. (You do not see radiation from the black hole itself, but from nearby stars that may be in the process of being captured by the black hole.)

Figure 65
Artist's drawing of the infrared telescope IRAS in orbit.

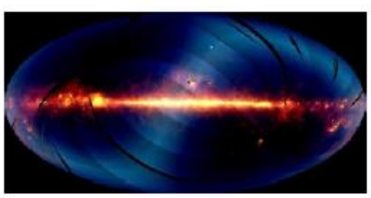


Figure 66
Map of the entire sky made by IRAS. The center of the Milky Way is in the center of the map. This is essentially a view of our galaxy seen from the inside.

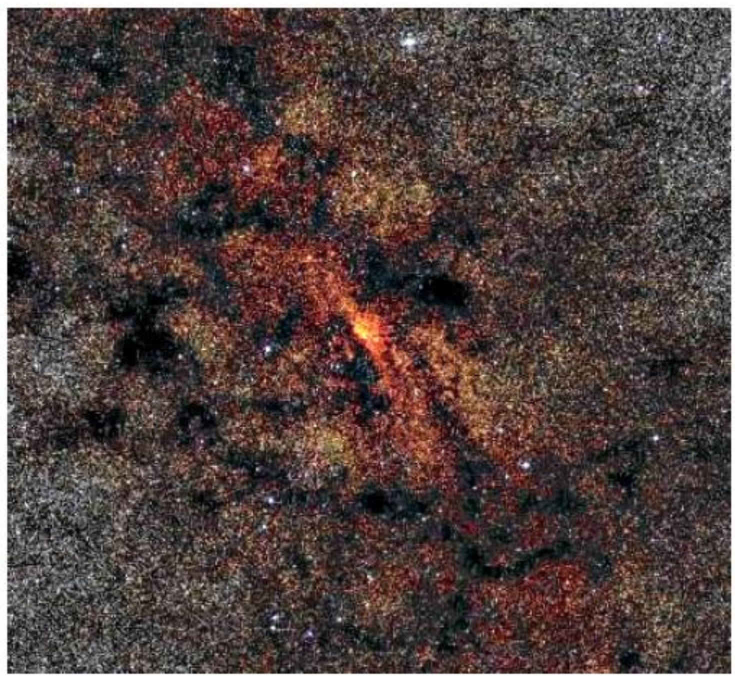


Figure 67
Center of our galaxy, where an enormous black hole resides. Not only is the galactic center rich in stars, but also in dust which prevents viewing this region in visible light.

Radio Telescopes

The earth's atmosphere allows not only visible and some infrared light from stars to pass through, but also radio waves in the wavelength range from a few millimeters to a good fraction of a meter. To study the radio waves emitted by stars and galaxies, a number of ***radio telescopes*** have been constructed.

For a telescope reflector to produce a sharp image, the surface of the reflector should be smooth and accurate to within about a fifth of a wavelength of the radiation being studied. For example, the surface of a mirror for a visible wavelength telescope should be accurate to within about 10^{-4} millimeters since the wavelength of visible light is centered around 5×10^{-4} millimeters. Radio telescopes that are to work with 5 millimeter wavelength radio waves, need surfaces accurate only to about a millimeter. Telescopes designed to study the important 21 cm wavelength radiation emitted by hydrogen, can have a rougher surface yet. As a result, radio telescopes can use sheet metal or even wire mesh rather than polished glass for the reflecting surface.

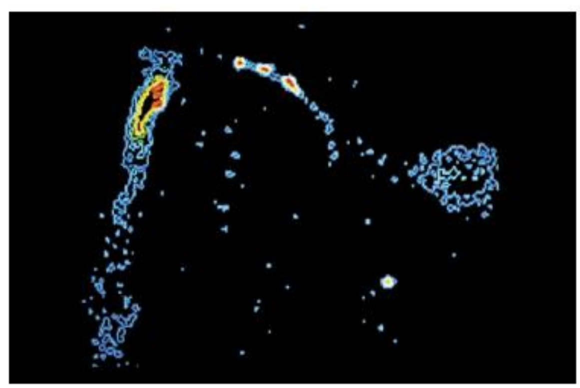
This is a good thing, because radio telescopes have to be much bigger than optical telescopes to order to achieve comparable images. The sharpness of an image, due to diffraction effects, is related to the ratio of the reflector diameter to the radiation wavelength. Since the radio wavelengths are at least 10^4 times larger than those for visible light, a radio telescope has to be 10^4 times larger than an optical telescope to achieve the same resolution.

The worlds largest radio telescope dish, shown in Figure (68), is the 305 meter dish at the Arecibo Observatory in Perto Rico. While this dish can see faint objects because of it's enormous size, and has been used to make significant discoveries, it has the resolving ability of an optical telescope about 3 centimeters in diameter, or a good set of binoculars .

Figure 68
Arecibo radio telescope. While the world's largest telescope dish remains fixed in the earth, the focal point can be moved to track a star.

As we saw with the Keck telescope, there is a great improvement in resolving power if the images of two or more telescopes are combined. The effective resolving power is related to the separation of the telescopes rather than to the diameter of the individual telescopes. Figure (69) shows the ***Very Large Array (VLA)*** consisting of twenty seven 25 meter diameter radio telescopes located in southern New Mexico. The dishes are mounted on tracks, and can be spread out to cover an area 36 kilometers in diameter. At this spacing, the resolving power is nearly comparable to a 5 meter optical telescope at Mt. Palomar.

Figures 69
The "Very Large Array" (VLA) of radio telescopes. The twenty seven telescopes can be spread out to a diameter of 36 kilometers.



Figures 69b
Radio galaxy image from the VLA. Studying the radio waves emitted by a galaxy often gives a very different picture than visible light.

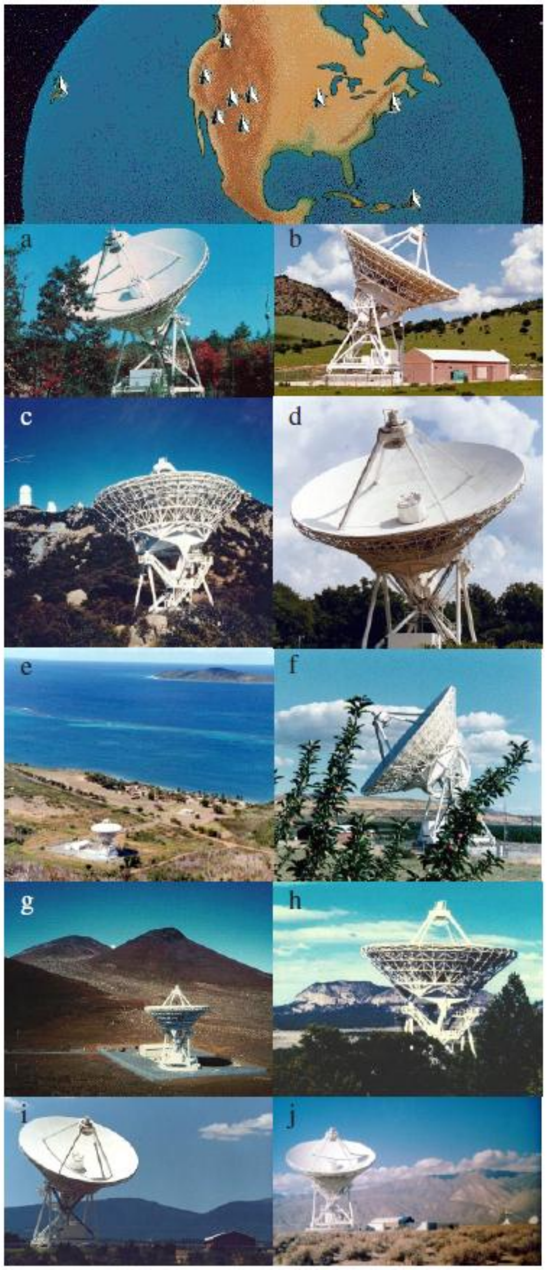


Figure 70
The Very Long Baseline Array of radio antennas. They are located at a) Hancock New Hampshire b) Ft. Davis Texas c) Kitt Peak Arizona d) North Liberty Iowa e) St. Croix Virgin Islands f) Brewster Washington g) Mauna Kea Hawaii h) Pie Town New Mexico i) Los Alamos New Mexico j) Owen's Valley California.

The Very Long Baseline Array (VLBA)

To obtain significantly greater resolving power, the ***Very Long Baseline Array (VLBA)*** was set up in the early 1990's. It consists of ten 25 meter diameter radio telescopes placed around the earth as shown in Figure (70). When the images of these telescopes are combined, the resolving power is comparable to an optical telescope 1000 meters in diameter (or an array of optical telescopes spread over an area one kilometer across).

The data from each telescope is recorded on a high speed digital tape with a time track created by a hydrogen maser atomic clock. The tapes are brought to a single location in Socorrow New Mexico where a high speed computer uses the accurate time tracks to combine the data from all the telescopes into a single image. To do this, the computer has to correct, for example, for the time difference of the arrival of the radio waves at the different telescope locations.

Because of it's high resolution, the VLBA can be used to study the structure of individual stars. In Figure 72 we see two time snapshots of the radio emission from the stellar atmosphere of a star 1000 light years away. With any of the current optical telescopes, the image of this star is only a point.

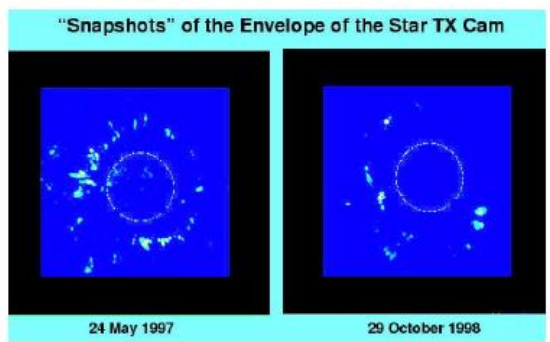


Figure 71
Very Long Baseline Array (VLBA) radio images of the variable star TX Cam which is located 1000 light years away. The approximate size of the star as it would be seen in visible light is indicated by the circle. The spots are silicon Monoxide (SiO) gas in the star's extended atmosphere. Motion of the these spots trace the periodic changes in the atmosphere of the star.
(Credit P.J. Diamond & A.J. Kembal, National Radio Astronomy, Associated Universities, Inc.)

MICROSCOPES

Optically, microscopes like the one seen in Figure (72), are telescopes designed to focus on nearby objects. Figure (73) shows the ray diagram for a simple microscope, where the objective lens forms an inverted image which is viewed by an eyepiece.

To calculate the magnification of a simple microscope, note that if an object of height y_0 were viewed unaided at a distance of 25 cm, it would subtend an angle θ_0 given by

$$\theta_0 = \frac{y_0}{25\text{ cm}} \qquad (42)$$

where throughout this discussion we will use the small angle approximation $\sin\theta \approx \tan\theta \approx \theta$.

A ray from the tip of the object (point A in Figure 73b), parallel to the axis, will cross the axis at point D, the focal point of the objective lens. Thus the height BC is equal to the height y_0 of the object, and the distance BD is the focal length f_0 of the objective, and the angle β is given by

$$\beta = \frac{y_0}{f_0} \qquad \textit{from triangle BCD} \qquad (43)$$

From triangle DEF, where the small angle at D is also β, we have

$$\beta = \frac{y_i}{L} \qquad \textit{from triangle DEF} \qquad (44)$$

where y_i is the height of the image and the distance L is called the ***tube length*** of the microscope.

Equating the values of β in Equations 29 and 30 and solving for y_i gives

$$\beta = \frac{y_0}{f_0} = \frac{y_i}{L}\ ; \qquad y_i = y_0\frac{L}{f_0} \qquad (45)$$

The eyepiece is placed so that the image of the objective is in the focal plane of the eyepiece lens, producing parallel rays that the eye can focus. Thus the distance EG equals the focal length f_0 of the eyepiece. From triangle EFG we find that the angle θ_i that image subtends as seen by the eye is

$$\theta_i = \frac{y_i}{f_e} \qquad \textit{angle subtended by image} \qquad (46)$$

Substituting Equation 45 for y_i in Equation 46 gives

$$\theta_i = \frac{L}{f_0}\frac{y_0}{f_e} \qquad (47)$$

Finally, the magnification m of the microscope is equal to the ratio of the angle θ_i subtended by the image in the microscope, to the angle θ_0 the object subtends at a distance of 25 cm from the unaided eye.

$$m = \frac{\theta_i}{\theta_0} = \frac{L}{f_0}\frac{y_0}{f_e} \times \frac{1}{y_0/25\text{ cm}} \qquad (48)$$

where we used Equation 47 for θ_i and Equation 42 for θ_0. The distance y_0 cancels in Equation 48 and we get

$$\boxed{m = \frac{L}{f_0} \times \frac{25\text{ cm}}{f_e}} \qquad \textit{magnification of a simple microscope} \qquad (49)$$

(We could have inserted a minus sign in the formula for magnification to indicate that the image is inverted.)

Figure 72
Standard optical microscope, which my grandfather purchased as a medical student in the 1890's. Compare this with a microscope constructed 100 years later, seen in Figure (69) on the next page.

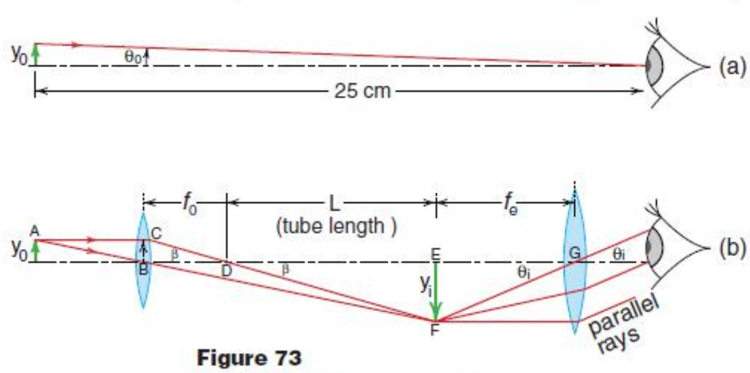
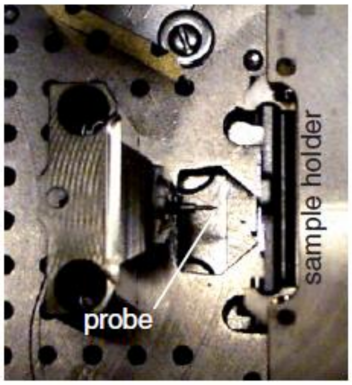


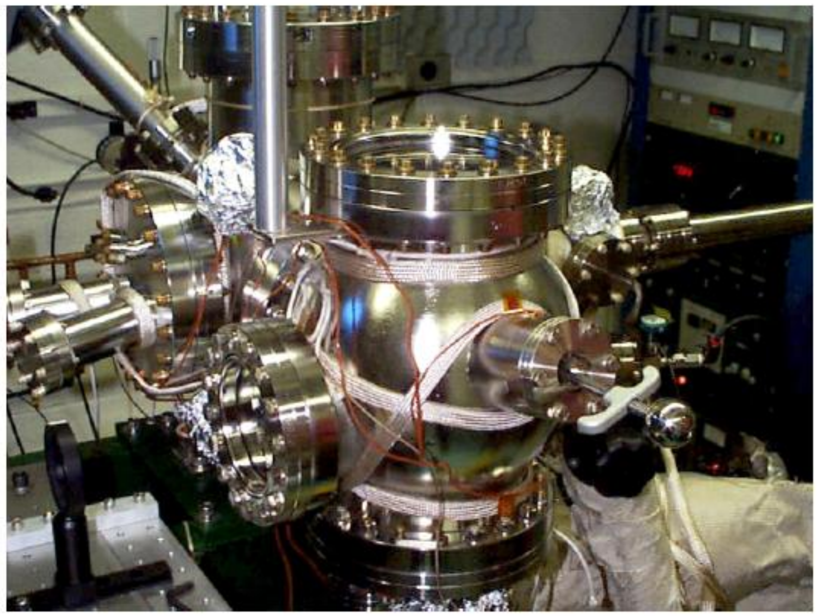
Figure 73
Optics of a simple microscope.

Scanning Tunneling Microscope

Modern research microscopes bear less resemblance to the simple microscope described above than the Hubble telescope does to Newton's first reflector telescope. In the research microscopes that can view and manipulate individual atoms, there are no lenses based on geometrical optics. Instead the surface to be studied is scanned, line by line, by a tiny probe whose operation is based on the particle-wave nature of electrons. An image of the surface is then reconstructed by computer and displayed on a computer screen. These microscopes work at a scale of distance much smaller than the wavelength of light, a distance scale where the approximations inherent in geometrical optics do not apply.

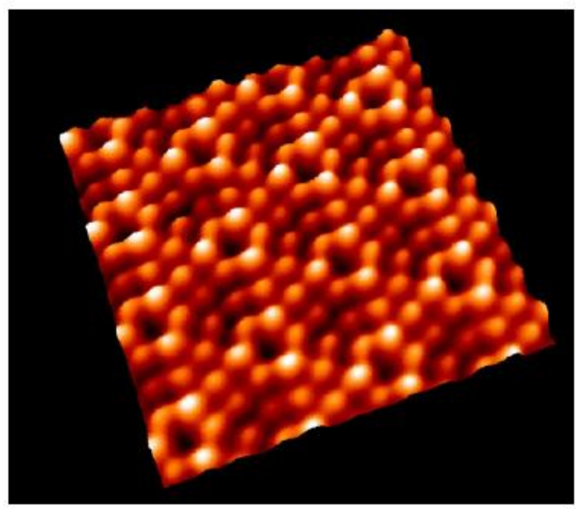


a) Probe and sample holder.



b) Vacuum chamber enclosing the probe and sample holder. Photograph taken in Geoff Nunes' lab at Dartmouth College.

Figure 74
Scanning Tunneling Microscope (STM). The tungsten probe seen in (a) has a very sharp point, about one atom across. With a couple of volts difference between the probe and the silicon crystal in the sample holder, an electric current begins to flow when the tip gets to within about fifteen angstroms (less than fifteen atomic diameters) of the surface. The current flows because the wave nature of the electrons allows them to "tunnel" through the few angstrom gap. The current increases rapidly as the probe is brought still closer. By moving the probe in a line sideways across the face of the silicon, while moving the probe in and out to keep the current constant, the tip of the probe travels at a constant height above the silicon atoms. By recording how much the probe was moved in and out, one gets a recording of the shape of the surface along that line. By scanning across many closely spaced lines, one gets a map of the entire surface. The fine motions of the tungsten probe are controlled by piezo crystals which expand or contract by tiny amounts when a voltage is applied to them. The final image you see was created by computer from the scanning data.



c) *Surface (111 plane) of a silicon crystal imaged by this microscope. We see the individual silicon atoms in the surface*

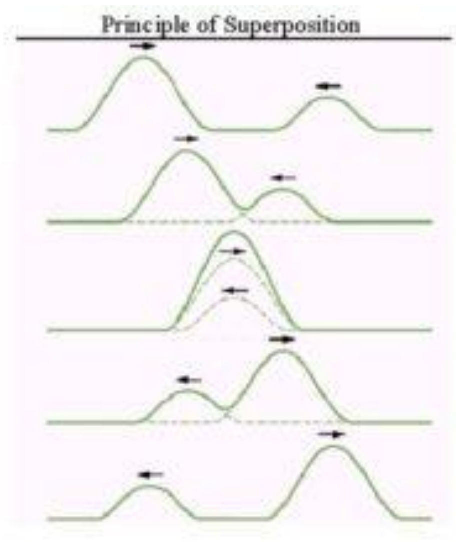
UNIT-2

Principle of Superposition of Waves

Waves surround us and their presence works to channelize a number of phenomena. Imagine you are in a boat and hear the siren of a ship. In this case, you'll be able to receive sound wave directly from the ship siren, as well as the sound wave that gets reflected by the seawater. In order to understand this concept, let us put our focus on the core concept of Superposition of Waves, together with the in-depth knowledge related to **superposition theorem.**

Introduction to Superposition of Waves

Let us take the example of a string wave to define the principle of superposition of wave that is based on the superposition theorem. And according to this, the net displacement of any component on the string for a given time is equal to the algebraic totality of the displacements caused due to each wave. Hence, this method of adding up individual waveforms for the evaluation of net waveform is termed as the **principle of superposition**.



The principle of superposition is expressed by affirming that overlapping waves add algebraically to create a resultant wave. Based on the principle, the overlapping waves ($f_1, f_2 \ldots, f_n$) do not hamper the motion or travel of each other. Therefore, the wave function (y) labelling the disturbance in the medium can be denoted as:

$$y = f_1(x - vt) + f_2(x - vt) + \ \ldots + f_n(x - vt)$$

$$= \sum_{i=1}^{n} f_i\,(x - vt)$$

Hence, the superposition of waves can lead to the following three effects:

1. Whenever two waves having the same frequency travel with the same speed along the same direction in a specific medium, then they superpose and create an effect termed as the interference of waves.

2. In a situation where two waves having similar frequencies move with the same speed along opposite directions in a specific medium, then they superpose to produce stationary waves.

3. Finally, when two waves having slightly varying frequencies travel with the same speed along the same direction in a specific medium, they superpose to produce beats.

Learn more about Transverse and Longitudinal Wave here.

Constructive & Destructive Interference

It is when two waves (similar wavelength, amplitude, and frequency) move in a specific or same direction. According to the superposition principle, the subsequent wave displacement can be written as:

$$y(x,t) = y\,m\,\sin(kx-\omega t) + y\,m\,\sin(kx-\omega t+\phi) = 2\,y\,m\,\cos(\phi/2)\,\sin(kx-\omega t+\phi/2)$$

This wave has an amplitude that depends on the phase (φ). Hence, when the two waves are believed to be in-phase (φ=0), then they interfere constructively. Furthermore, the resultant wave holds twice the amplitude as compared to the individual waves. On the other hand, when two waves possess opposite-phase (φ=180), then they interfere destructively; canceling each other out.

Two Sine Waves Moving in Opposite Directions (Standing Wave)

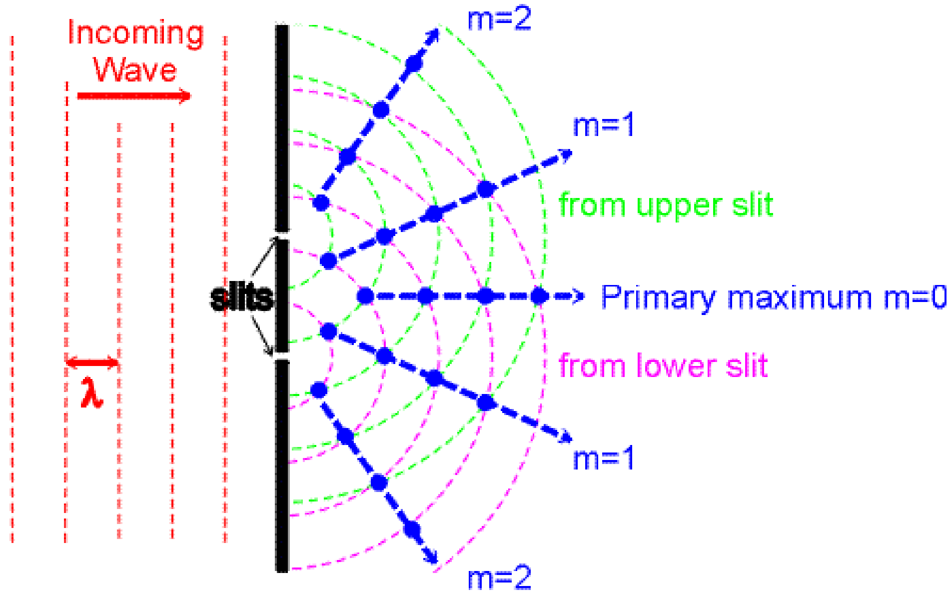
Do remember that, a traveling wave propagates from one place to another, however, a standing wave looks as if its still. Suppose two waves (having the same amplitude, wavelength, and frequency) move in opposite directions. Based on the principle of superposition, the final wave amplitude can be written as:

$$y\,(x\,,t\,) = y\,m\,\sin\,(kx-\omega t) + y\,m\,\sin\,(kx+\omega t) = 2\,y\,m\,\sin\,(kx)\,\cos\,(\omega t)$$

As per the **superposition theorem**, this wave is no longer termed as a traveling wave since the position and time dependency has been separated. In this, the wave amplitude as a function of point or location is 2ymsin(kx). To be precise, this amplitude wouldn't travel but will stand and oscillate up and down based on cos(ωt).

Two-slit interference

The prototypical example of interference is the two-slit experiment. Consider monochromatic (single wavelength) light incident on two narrow slits as shown.

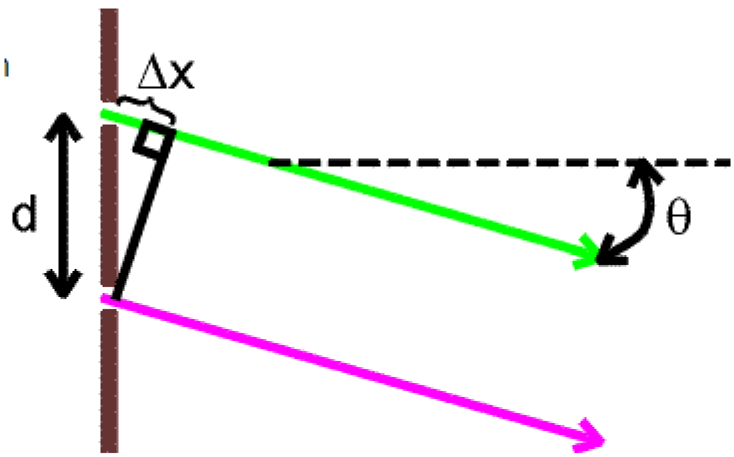


Wave crests leaving the upper and lower slits at a given instant of time are indicated by the violet and green lines. In the directions shown by the blue lines, the violet and green waves rise and fall together, giving **constructive interference** in those directions. Midway between those directions, the two contributions are 180 degrees out of phase, so they tend to cancel one another. The result is that if a screen is placed to the right, an **interference pattern** is seen: there are peaks in the brightness in the direction of the blue arrows, with dark bands in between.

To solve for the angular position of the maxima, consider two rays emited at an angle θ from each of the slits. The screen is assumed to be very far away (in comparison to the distance between the slits), so the rays from to the two slits to any given point on the screen are nearly parallel. If the distance **Δx** is an integer number ***m*** of wavelengths, a bright maximum will appear on the screen. The formula for the positions of those maxima is

$$d \sin\theta = m\lambda$$

where ***m*** = 0,1,2,3,...



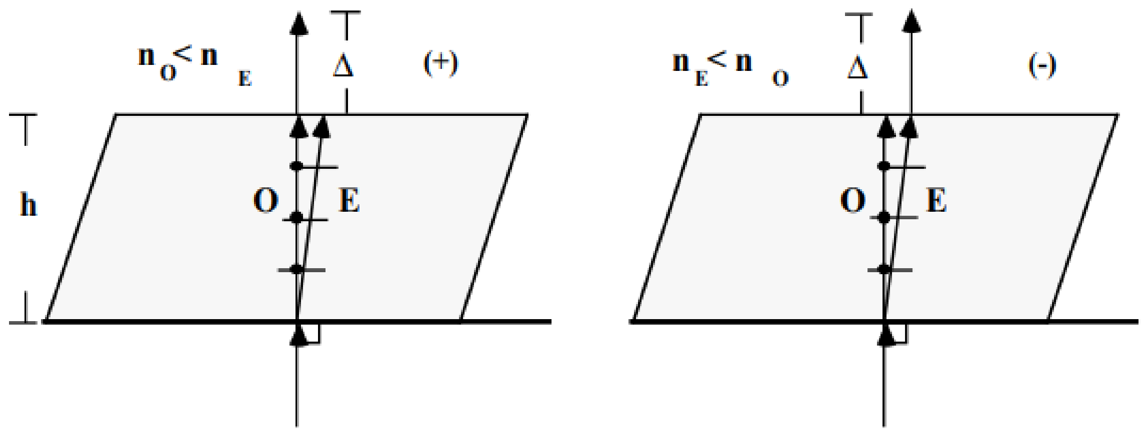
A second formula is needed to relate the positions on the screen to the angle θ. The screen, the center line, and the line to an arbitrary point on the screen form a right triangle, leading to the formula

$$\tan\theta = y / L$$

where ***L*** is the distance from the slits to the screen and ***y*** is the distance measured on the screen from the forward direction (the central maximum). For two-slit interference, the interference pattern is only easy to see at rather small angles, for which it is an excellent approximation to take tanθ and sinθ to be equal. For diffraction gratings, however, that approximation is not necessarily adequate because pattern can be observed also at large angles.

Retardation, Interference Colors

• In anisotropic crystals, the two rays of light produced by double refraction travel at different velocities through the crystal. It takes the slow ray longer to traverse the crystal than it takes the fast ray. The fast ray will have passed through the crystal and traveled some distance Δ beyond the crystal before the slow ray reaches the surface of the crystal. This distance Δ is called the retardation.



• The retardation Δ may be calculated as follows. If t_S is the time in seconds that it takes the slow ray to traverse the crystal and t_F is the time it takes the fast ray to traverse the crystal, then the distance Δ that the fast ray travels beyond the crystal before the slow ray emerges is

$$\Delta = c\,(t_S - t_F) \qquad \{\text{units: } m = (m/s)(s)\},$$

where c is the velocity of light in a vacuum, which is very close to the velocity of light in air. For a crystal of thickness h with velocities v_F and v_S , t F and t_S may be replaced by h/v_F and h/v_S {units: (m)/(m/s) = s}, respectively, to give

$$\Delta = c\left(\frac{h}{v_S} - \frac{h}{v_F}\right) = h\left(\frac{c}{v_S} - \frac{c}{v_F}\right)$$

Recalling the definition of the refractive index n, the equation for Δ becomes

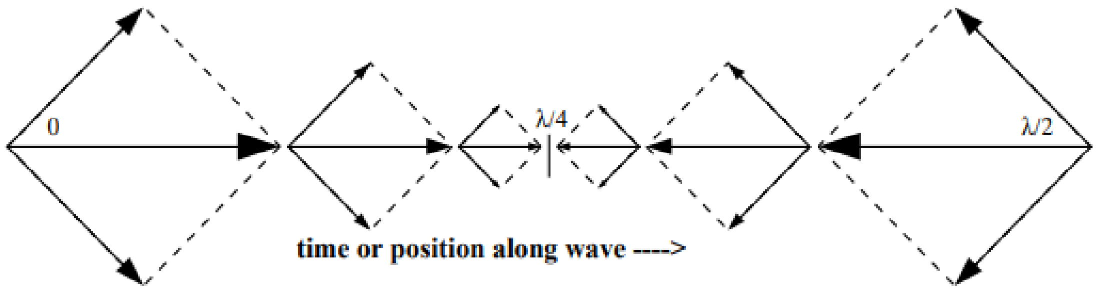
$$\Delta = h\,(n_S - n_F).$$

Because refractive indices are dimensionless, Δ will be in the same units as h, normally nanometers (nm). Note that the difference in path length for the O and E rays has been neglected in this calculation. In fact, for calcite the angle is only about 5°, so the path length difference is only about a factor of 0.005. For most other minerals the angle is much smaller.

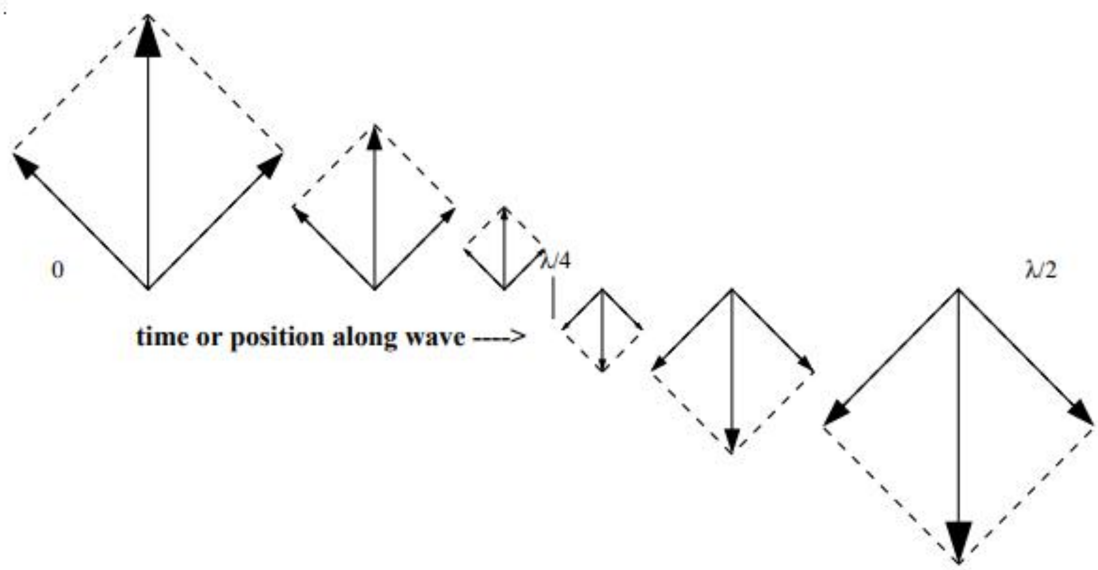
• The birefringence of a mineral grain is defined as the absolute value of the difference between the refractive indices of the two rays |nS - nF| for that grain. The maximum birefringence of a mineral is defined as the difference δ between the largest and smallest refractive indices for that mineral. Because thin sections are always the same thickness (h=3000 nm), the birefringence for a mineral in a particular orientation should be the same in all thin sections. Retardation for a particular mineral will be greatest when the mineral is oriented so that the two rays have the maximum and minimum refractive indices for the mineral.

• When the two rays of light emerge from an anisotropic crystal, they will recombine (following the rules of vector addition) to produce a resultant ray. If there were no retardation, the resultant ray would be identical to the incident ray. No light would pass the analyzer and the crystal would appear dark (extinct). However, retardation leads to a new resultant that does have an electric vector component that will pass the analyzer. If the light source is monochromatic, the crystal will appear lighter or darker, depending on the retardation. If the light source is polychomatic (white light), the crystal will exhibit interference colors.

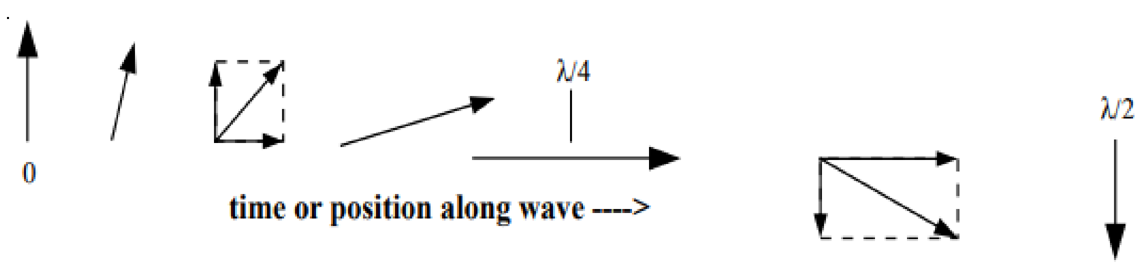
• To understand the origin of interference colors, we must examine the electric vectors at various points along a pair of light waves (emerging from an anisotropic crystal) and the resultant light wave. If the two rays of monochromatic light are in phase, the resultant wave will have the



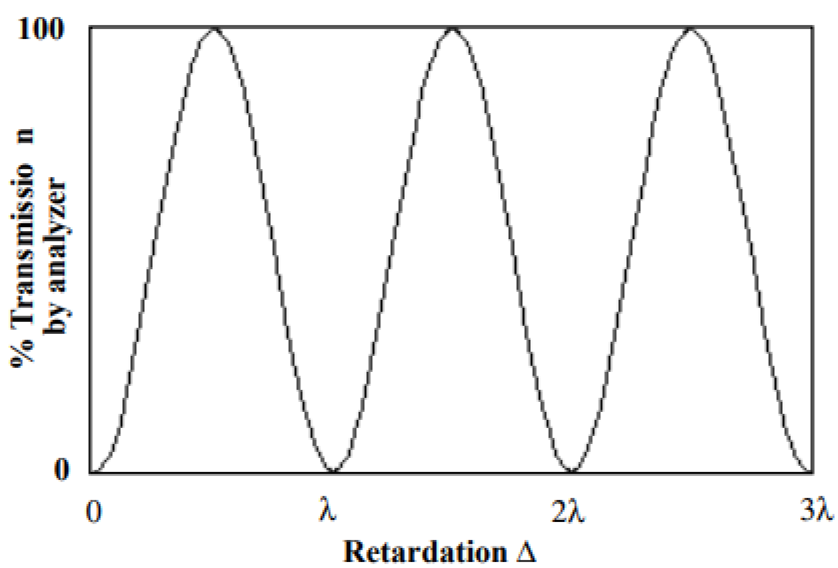
same plane of polarization as the incident wave: If the two rays of monochromatic light are out of phase due to retardation, then the resultant



wave will have a new orientation. If the two rays are λ/2 out of phase, the resultant will be:



If the two rays are λ/4 out of phase, the resultant will be circularly polarized: • Transmission of the resultant wave when the analyzer (the upper polarizing filter) is in place will depend on the orientation(s) of the resultant vibration directions with respect to the orientation of the analyzer. In most cases, some of the resultant wave is transmitted and interference colors are observed. However, if the one of the vibration directions of the crystal is parallel to that of the polarizer, then all of the light will pass through the crystal maintaining the analyzer's plane of polarization. Because there is in effect only one ray in this case, there is no interference when the light emerges from the crystal and, therefore, no interference color. Extinction is the dark appearance of a crystal between crossed polarizers when a vibration direction in the crystal is parallel to the vibration direction of the polarizer. Anisotropic crystals will become extinct four times as the stage of a polarizing microscope is rotated 360°. The maximum amount of light will be transmitted by the analyzer when stage is rotated 45° from an extinction position.



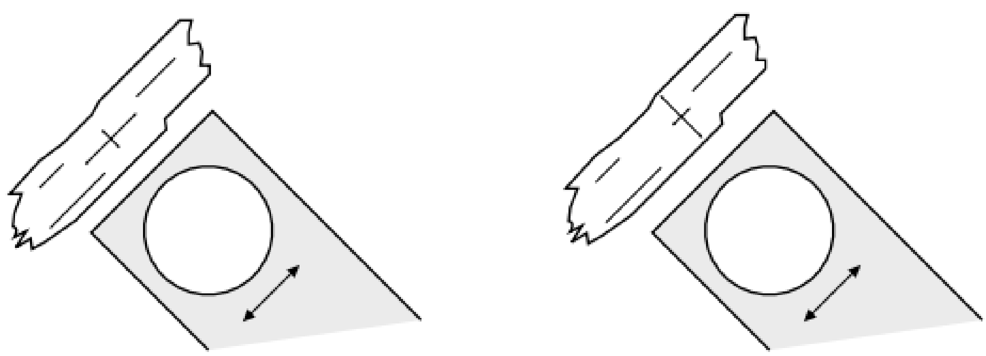
• For monochromatic light illuminating a crystal at 45° from extinction, the intensity of the light transmitted by the analyzer as a function of the retardation is given by this -> graph. Note that no light passes the analyzer when the retardation is an integral number of wavelengths for the wavelength of light used. This effect can be observed by viewing a quartz wedge between crossed polarizers in sodium light. Retardation for the quartz wedge increases with thickness so that a series of parallel dark bands (for $\Delta = \lambda$, 2λ, etc.) can be observed.

Because the light source in our microscopes is not monochromatic, the actual interference colors observed result from the summation of dark bands for all visible wavelengths. The characteristic sequence of colors as a function of retardation is shown as the chart of interference colors in Nesse and elsewhere. You will have seen these colors on soap bubbles and oil slicks, where they are produced by the interference of light waves reflected off the front and back surfaces of these films. However, in these cases no polarization or retardation is involved; the colors are due to destructive interference of the two (out of phase) reflected rays. Note that interference colors are not the same as the rainbow or spectrum produced from white light by a prism or a diffraction grating.

• Retardation is a function of the mineral, its orientation, and its thickness. If the thickness is doubled, so is the retardation. Similarly, if a second crystal of the same mineral with the same orientation is placed on top of the crystal being studied, the retardation will increase. In fact, if a second crystal of any mineral is placed on top with its slow vibration direction parallel to the slow vibration direction of the crystal being studied, the retardation will increase. This effect is called addition of retardation.

• Petrographic microscopes are equipped with a quartz plate designed to be placed in the light path above the crystal with the slow vibration direction of the quartz crystal oriented at 45° to the planes of the polarizing filters. The slow vibration direction of the plate is indicated on the plate by a double pointed arrow or similar mark. Use of this plate permits identification of the slow and fast vibration directions of a crystal by watching for addition or subtraction of the retardation. The thickness of the quartz plate is selected to add to (or subtract from) the retardation exactly 550 nm. Older microscopes were equipped with a similar plate made of gypsum (of a different thickness!) that caused the same amount (550 nm) of addition (or subtraction).

• Crystals that occur in a prismatic form may be characterized has having the slow direction either parallel to their long dimension or perpendicular to it. The former are called length slow and the latter length fast. These two cases may be easily distinguished by insertion of a quartz plate when the long direction of the mineral is 45° from the planes of the polarizer. If the crystal has no long direction, this test is not possible.



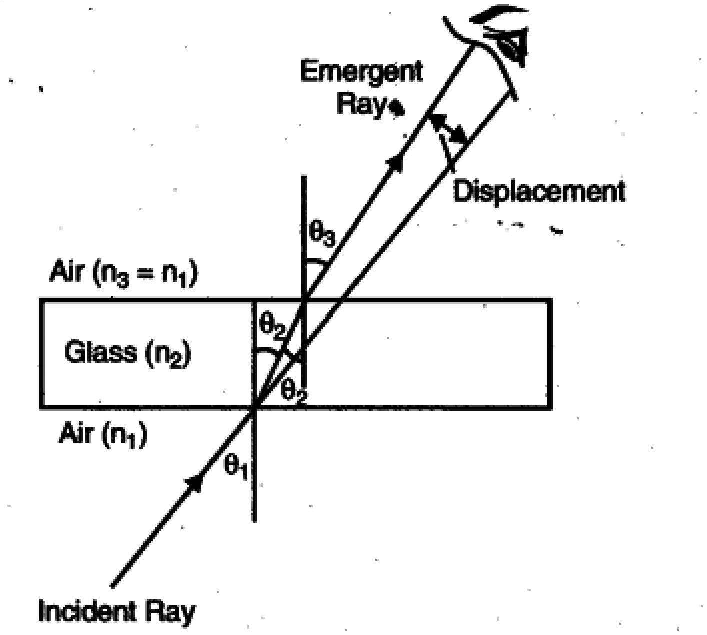
Length Slow
Retardation Increases
"Addition"

Length Fast
Retardation Decreases
"Subtraction"

Lateral Shift:-

When a ray of light travels through a glass slab from air, it bends towards the normal and when it comes out of the other side of the glass slab, it bends away from the normal. It is found that the incident ray and the emergent ray are not along the same straight line, but the emergent ray seems to be displaced with respect to the incident ray. This shift in the emergent ray with respect to the incident ray is called lateral shift or lateral displacement. The incident and the emergent rays, however, remain

parallel. The diagram is as shown:



Interference in thin films

The two rays BC and DE reflected from the top and the bottom of the air film have a varying path difference along the length of the film due to variation of the film thickness. Because ray DE travels more distance than BC.

Also ray DE undergoes a phase change of half wave length (π *change*) occurs at the air to glass boundary due to reflection .
The optical phase difference between the two rays BC and DE is given by :

$$\Delta = 2nt + \lambda/2$$

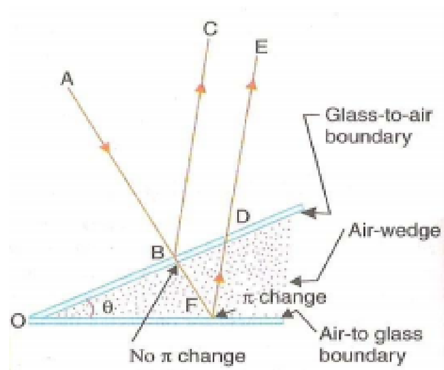


Figure.2

Minima occurs when the phase difference is an *odd* multiple of $\lambda/2$,the two waves arriving are 180^o out of phase and give rise to destructive interference .Therefore ,the condition for dark fringes, or **destructive interference** is:

$$\Delta = \left(m + \frac{1}{2}\right)\lambda$$

$$2nt = m\lambda$$

Because the film produced from air n=1

$$\boxed{2t = m\lambda}$$

The thickness of the spacer used to form the wedge shaped air film between the glass slides can be determined using a travelling microscope .

$$\boxed{t = \frac{L \times \frac{\lambda}{2}}{d}}$$

(t)is thickness of the spacer.
(L) is the length of the glass piece.
(λ)is the wave length of the used monochromatic light (sodium) in vacum.
(d) is the thickness of the fringe.

Procedure :

1. Set the apparatus as shown in (fig.1.a)
2. Fix the cross hair to one of the parallel fringes produced ,take the readings of the vernier at one of the dark fringes(d0)
3. And then take the reading again after counting 20 dark fringes from the previous one(d21)
4. Calculate d using $\boxed{d = \frac{d_{21} - d_0}{20}}$
5. Measure the length of the glass piece ,starting from the edge of the thin spacer to the end of the plate (L).
6. Calculate the thickness of the plate using the following relation :

$$\boxed{t = \frac{L \times \frac{\lambda}{2}}{d}}$$

7.explain the shape of the fringes produced from the air wedge experiment?

Newton's rings

EQUIPMENT NEEDED

–Travelling microscope.
– Convex lens.
–Sodium(monochromatic light) lamp with known wave length .
–Glass plate.
–Spherometer.
–Stand.

PURPOSE

1)Explain the formation of Newton's rings .
2)Measure the wave length of the monochromatic light (Sodium).

THEORY

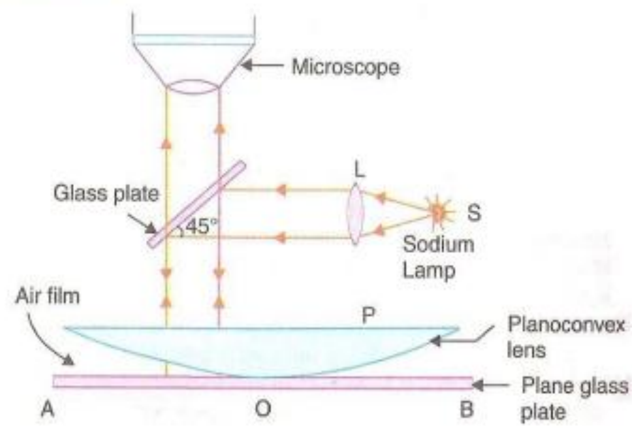


Figure.3a

Figure.3b

Newton's rings are formed when a plano-convex lens of large radius of curvature is placed on a plane glass sheet. The combination forms a thin circular air film of variable thickness in all directions around the point of contact of the lens and the glass plate at O.
If monochromatic light is allowed to fall normally (fig.3a)on the lens using the 45º inclined glass plate, and the film is viewed in reflected light, interference fringes are observed in the form of a series with concentric rings(fig.3b)

when the light is incident on the plano-convex lens part of the light incident on the system is reflected from glass-to-air boundary (say at point D).The reminder of the light is transmitted through the air film .it is again reflected from the air-to-glass boundary (say from point J)
The two rays are (1 and 2) reflected from the top and bottom of the air film interfere with each other to produce darkness and brightness .

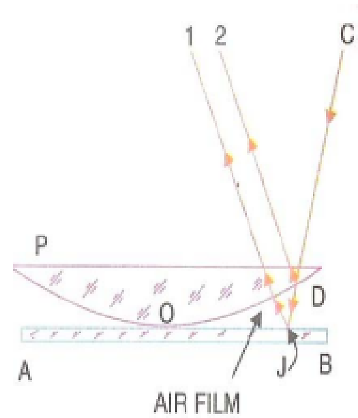


Figure.4

The condition For destructive interference is the same obtained from the air wedge experiment

$$\boxed{2t = m\lambda}$$

To determine the wave length of the sodium light we use the following relation

$$\boxed{\boxed{r^2 = m\lambda R}}$$

(r) is the radius of the fringe.
(m) is the order of the ring.
(R) is the radius of curvature of the plano convex lens.
(λ)is the wave length of the used monochromatic light (sodium) in vacum.

PROCEDURE

1. Turn on the sodium lamp and Adjust the apparatus so we have a parallel light falling in the lens and the rings are seen clearly in the eyepiece of the travelling microscope .
2. Fix the cross hair to the center ring and then move it to the L.H.S until you reach the 8^{th} dark fringe, take the reading of all the dark fringes until you reach the other end of the 8^{th} fringe from the R.H.S.
3. Calculate the radius of each ring r

$$\boxed{r = \frac{|L.H.S - R.H.S|}{2}}$$

The readings should be obtained as the following:

m	L.H.S (mm)	R.H.S (mm)	r	r^2

4. Use the spherometer to measure the radius of curvature of the lens R.

$$\boxed{R = \frac{l^2}{2h} + \frac{h}{2}}$$

(h) the distance that the central leg has been moved from the readings on the vertical scale (millimeter) and the circular scale (1/100 millimeter)

(l) the distance between the central leg and the other three legs

5. Plot the relation between r^2 (y-axis) and m (x-axis).
6. Calculate $slope$ from the plot
7. Calculate the wave length

$$\boxed{\lambda = \frac{slope}{R}}$$

8.explain the shape of the fringes produced from the Newton's rings experiment and describe its properties .

Michelson Interferometer There are many two-beam interferometers which allow the surfaces producing the two wavefronts to be physically separated by a large distance . These instruments allow the two wavefronts to travel along dif ferent optical paths . One of these is the Michelson interferometer diagramed in Fig . 16 a . The two interfering wavefronts are produced by the reflections from the two mirrors . A plate beamsplitter with one face partially silvered is used , and an identical block of glass is placed in one of the arms of the interferometer to provide the same amount of glass path in each arm . This cancels the ef fects of the dispersion of the glass beamsplitter and allows the system to be used with white light since the optical path dif ference is the same for all wavelengths . Figure 16 b provides a folded view of this interferometer and shows the relative optical

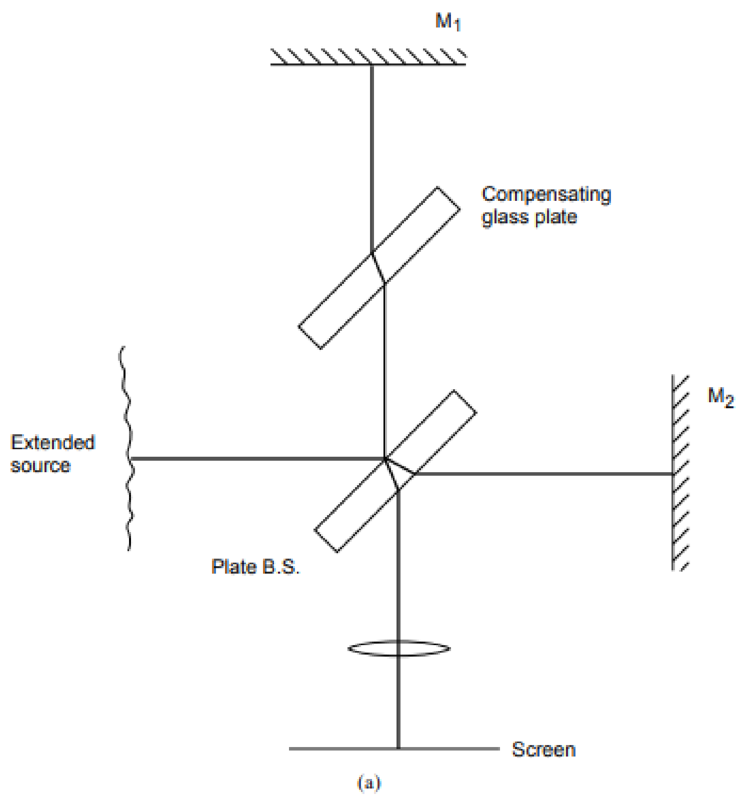
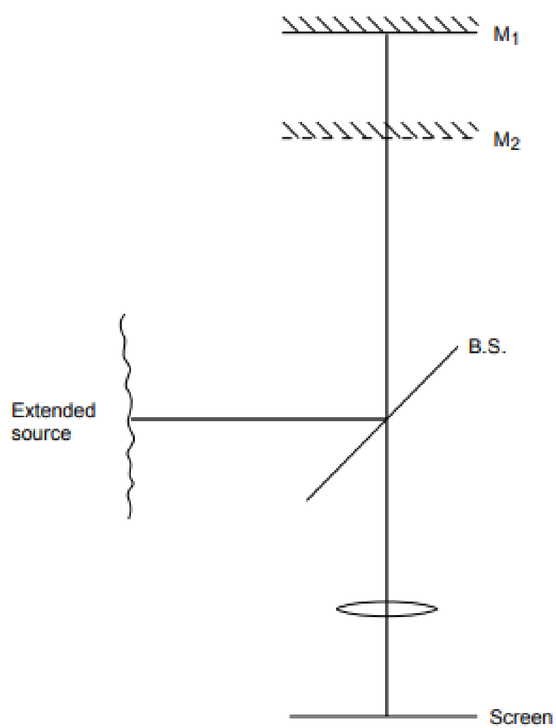


FIGURE 16 Michelson interferometer: (*a*) schematic view; and (*b*) folded view showing the relative optical position of the two mirrors.

position of the two mirrors as seen by the viewing screen . It should be obvious that the two mirrors can be thought of as the two surfaces of a ‘‘glass’’ plate that is illuminated by the source . In this case , the index of the fictitious plate is one , and the reflectivity at the two surfaces is that of the mirrors . Depending on the mirror orientations and shapes , the interferometer either mimics a plane-parallel plate of adjustable thickness , a wedge of arbitrary angle and thickness , or the comparison of a reference surface with an irregular or curved surface . The type of fringes that are produced will depend on this configuration , as well as on the source used for illumination . When a monochromatic point source is used , nonlocalized fringes are produced , and the imaging lens is not needed .

Two virtual-source images are produced , and the resulting fringes can be described by the interference of two spherical waves (discussed earlier) . If the mirrors are parallel , circular fringes centered on the line normal to the mirrors result as with a plane-parallel plate . The source separation is given by twice the apparent mirror separation . If the mirrors have a relative tilt , the two source images appear to be laterally displaced , and hyperbolic fringes result . Along a plane bisecting the source images , straight equispaced fringes are observed . When an extended monochromatic source is used , the interference fringes are localized . If the mirrors are parallel , fringes of equal inclination or Haidinger fringes (as described earlier) are produced . The fringes are localized at infinity and are observed in the rear



(b)

FIGURE 16 (*Continued*.)

focal plane of the imaging lens . Fringes of equal thickness localized at the mirrors are generated when the mirrors are tilted . The apparent mirror separation should be kept small , and the imaging lens should focus on the mirror surface .

If the extended source is polychromatic , colored fringes localized at the mirrors result . They are straight for tilted mirrors . The fringes will have high visibility only if the apparent mirror separation or OPD is smaller than the coherence length of the source . Another way of stating this is that the order of interference m must be small to view the colored fringes . As m increases , the fringes will wash out . The direct analogy here is a thin film . As the mirror separation is varied , the fringe visibility will vary . The fringe visibility as a function of mirror separation is related to the source frequency spectrum (see under ‘‘Source Spectrum’’ and ‘‘Coherence and Interference’’) , and this interferometer forms the basis of a number of spectrometers . This topic is further discussed in , ‘‘Metrology . ’’ When the source spectrum is broad , chromatic fringes cannot be viewed with the mirrors parallel . This is because the order of interference for fringes of equal inclination is a maximum at the center of the pattern .

An important variation of the Michelson interferometer occurs when monochromatic collimated light is used . This is the Twyman - Green interferometer , and is a special case of point-source illumination with the source at infinity . Plane waves fall on both mirrors , and if the mirrors are flat , nonlocalized equispaced fringes are produced . Fringes of equal thickness can be viewed by imaging the mirrors onto the observation screen . If one of the mirrors is not flat , the fringes represent changes in the surface height . The two surfaces are compared as in the Fizeau interferometer . This interferometer is an invaluable tool for optical testing .

MULTIPLE BEAM INTERFERENCE Throughout the preceding discussions , we have assumed that only two waves were being interfered . There are many situations where multiple beams are involved . Two examples are the dif fraction grating and a plane-parallel plate . We have been ignoring multiple reflections , and in some instances these extra beams are very important . The net electric field is the sum of all of the component fields . The two examples noted above present dif ferent physical situations : all of the interfering beams have a constant intensity with a dif fraction grating , and the intensity of the beams from a plane-parallel plate decreases with multiple reflections

Diffraction Grating

A *diffraction grating* can be modeled as a series of equispaced slits, and the analysis bears a strong similarity to the Young's double slit (discussed earlier). It operates by division of wavefront, and the geometry is shown in Fig. 17. The slit separation is d, the OPD between successive beams for a given observation angle θ is $d \sin(\theta)$, and the corresponding phase difference $\Delta\phi = 2\pi d \sin(\theta)/\lambda$. The field due to the nth slit at a distant observation point is

$$E_j(\theta) = Ae^{i(j-1)\Delta\phi} \qquad j = 1, 2, \ldots, N \tag{55}$$

where all of the beams have been referenced to the first slit, and there are N total slits. The net field is

$$E(\theta) = \sum_{j=1}^{N} E_j(\theta) = A\sum_{j=1}^{N} (e^{i\Delta\phi})^{j-1} \tag{56}$$

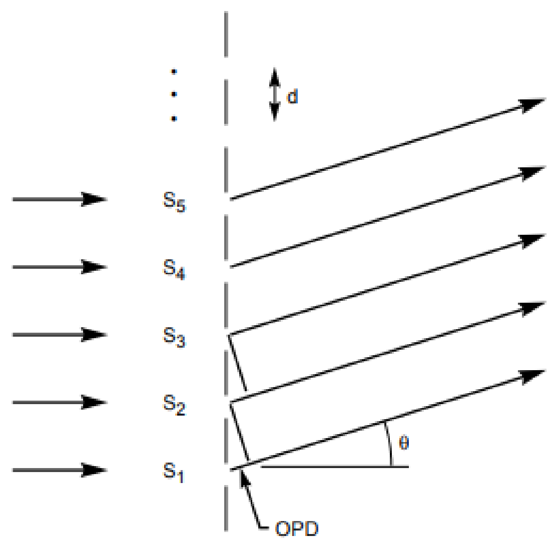


FIGURE 17 Diffraction grating: multiple-beam interference by division of wavefront.

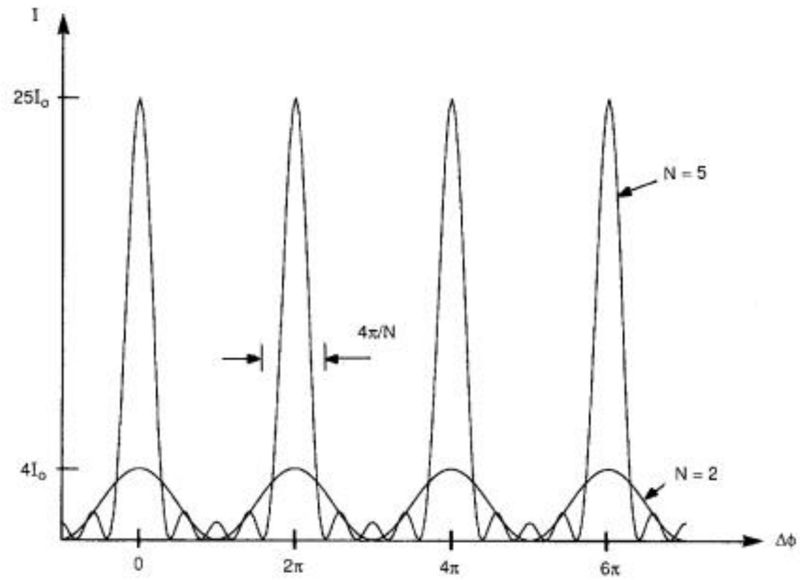


FIGURE 18 The interference patterns produced by gratings with 2 and 5 slits.

which simplifies to

$$E(\theta) = A\left(\frac{1 - e^{iN\Delta\phi}}{1 - e^{i\Delta\phi}}\right) \tag{57}$$

The resulting intensity is

$$I(\theta) = I_0\left[\frac{\sin^2\left(\frac{N\Delta\phi}{2}\right)}{\sin^2\left(\frac{\Delta\phi}{2}\right)}\right] = I_0\left[\frac{\sin^2\left(\frac{N\pi d\sin(\theta)}{\lambda}\right)}{\sin^2\left(\frac{\pi d\sin(\theta)}{\lambda}\right)}\right] \tag{58}$$

where I_0 is the intensity due to an individual slit.

This intensity pattern is plotted in Fig. 18 for $N = 5$. The result for $N = 2$, which is the double-slit experiment, is also shown. The first thing to notice is that the locations of the maxima are the same, independent of the number of slits. A maximum of intensity is obtained whenever the phase difference between adjacent slits is a multiple of 2π. These maxima occur at the diffraction angles given by

$$\sin(\theta) = \frac{m\lambda}{d} \tag{59}$$

where m is an integer. The primary difference between the two patterns is that with multiple slits, the intensity at the maximum increases to N^2 times that due to a single slit, and this energy is concentrated into a much narrower range of angles. The full width of a diffraction peak between intensity zero corresponds to a phase difference $\Delta\phi$ of $4\pi/N$.

The number of intensity zeros between peaks is N 2 1 . As the number of slits increases , the angular resolution or resolving power of the grating greatly increases . The ef fects of a finite slit width can be added by replacing I0 in Eq .

(58) by the single-slit dif fraction pattern . This intensity variation forms an envelope for the curve in Fig .

****

UNIT-3

Diffraction and the Wave Theory of Light In Chapter 35 we defined diffraction rather loosely as the flaring of light as it emerges from a narrow slit. More than just flaring occurs, however, because the light produces an interference pattern called a diffraction pattern. For example, when monochromatic light from a distant source (or a laser) passes through a narrow slit and is then intercepted by a viewing screen, the light produces on the screen a diffraction pattern like that in Fig. 36-1. This pattern consists of a broad and intense (very bright) central maximum plus a number of narrower and less intense maxima (called secondary or side maxima) to both sides. In between the maxima are minima. Light flares into those dark regions, but the light waves cancel out one another. Such a pattern would be totally unexpected in geometrical optics: If light traveled in straight lines as rays, then the slit would allow some of those rays through to form a sharp rendition of the slit on the viewing screen instead of a pattern of bright and dark bands as we see in Fig. 36-1. As in Chapter 35, we must conclude that geometrical optics is only an approximation. Diffraction is not limited to situations when light passes through a narrow opening (such as a slit or pinhole). It also occurs when light passes an edge, such as the edges of the razor blade whose diffraction pattern is shown in Fig. 36-2. Note the lines of maxima and minima that run approximately parallel to the edges, at both the inside edges of the blade and the outside edges. As the light passes, say, the vertical edge at the left, it flares left and right and undergoes interference, producing the pattern along the left edge. The rightmost portion of that pattern actually lies behind the blade, within what would be the blade's shadow if geometrical optics prevailed.

You encounter a common example of diffraction when you look at a clear blue sky and see tiny specks and hairlike structures floating in your view. These *floaters,* as they are called, are produced when light passes the edges of tiny deposits in the vitreous humor, the transparent material filling most of the eyeball. What you are seeing when a floater is in your field of vision is the diffraction pattern produced on the retina by one of these deposits. If you sight through a pinhole in a piece of cardboard so as to make the light entering your eye approximately a plane wave, you can distinguish individual maxima and minima in the patterns.

Diffraction is a wave effect. That is, it occurs because light is a wave and it occurs with other types of waves as well. For example, you have probably seen diffraction in action at football games. When a cheerleader near the playing field yells up at several thousand noisy fans, the yell can hardly be heard because the sound waves diffract when they pass through the narrow opening of the cheerleader's mouth. This flaring leaves little of the waves traveling toward the fans in front of the cheerleader. To offset the diffraction, the cheerleader can yell through a megaphone. The sound waves then emerge from the much wider opening at the end of the megaphone. The flaring is thus reduced, and much more of the sound reaches the fans in front of the cheerleader.

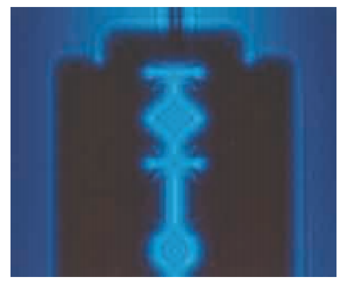


Fig. 36-2 The diffraction pattern produced by a razor blade in monochromatic light. Note the lines of alternating maximum and minimum intensity. *(Ken Kay/Fundamental Photographs)*

The Fresnel Bright Spot Diffraction finds a ready explanation in the wave theory of light. However, this theory, originally advanced in the late 1600s by Huygens and used 123 years later by Young to explain double-slit interference, was very slow in being adopted, largely because it ran counter to Newton's theory that light was a stream of particles. Newton's view was the prevailing view in French scientific circles of the early 19th century, when Augustin Fresnel was a young military engineer. Fresnel, who believed in the wave theory of light, submitted a paper to the French Academy of Sciences describing his experiments with light and his wave-theory explanations of them. In 1819, the Academy, dominated by supporters of Newton and thinking to challenge the wave point of view, organized a prize competition for an essay on the subject of diffraction. Fresnel won. The Newtonians, however, were not swayed. One of them, S. D. Poisson, pointed out the "strange result" that if Fresnel's theories were correct, then light waves should flare into the shadow region of a sphere as they pass the edge of the sphere, producing a bright spot at the center of the shadow. The prize committee arranged a test of Poisson's prediction and discovered that the predicted Fresnel bright spot, as we call it today, was indeed there (Fig. 36-3). Nothing builds confidence in a theory so much as having one of its unexpected and counterintuitive predictions verified by experiment.

Fig. 36-3 A photograph of the diffraction pattern of a disk. Note the concentric diffraction rings and the Fresnel bright spot at the center of the pattern. This experiment is essentially identical to that arranged by the committee testing Fresnel's theories, because both the sphere they used and the disk used here have a cross section with a circular edge. *(Jearl Walker)*

Diffraction by a Single Slit: Locating the Minima Let us now examine the diffraction pattern of plane waves of light of wavelength l that are diffracted by a single long, narrow slit of width a in an otherwise opaque screen B, as shown in cross section in Fig. 36-4. (In that figure, the slit's length extends into and out of the page, and the incoming wavefronts are parallel to screen B.) When the diffracted light reaches viewing screen C, waves from different points within the slit undergo interference and produce a diffraction pattern of bright and dark fringes (interference maxima and minima) on the screen. To locate the fringes, we shall use a procedure somewhat similar to the one we used to locate the fringes in a two-slit interference pattern. However, diffraction is more mathematically challenging, and here we shall be able to find equations for only the dark fringes. Before we do that, however, we can justify the central bright fringe seen in Fig. 36-1 by noting that the Huygens wavelets from all points in the slit travel about the same distance to reach the center of the pattern and thus are in phase there.As for the other bright fringes, we can say only that they are approximately halfway between adjacent dark fringes. To find the dark fringes, we shall use a clever (and simplifying) strategy that involves pairing up all the rays coming through the slit and then finding what conditions cause the wavelets of the rays in each pair to cancel each other. We apply this strategy in Fig. 36-4 to locate the first dark fringe, at point P1. First, we mentally divide the slit into two zones of equal widths a/2. Then we extend to P1 a light ray r1 from the top point of the top zone and a light ray r2 from the top point of the bottom zone. We want the wavelets along these two rays to cancel each other when they arrive at P1. Then any similar pairing of rays from the two zones will give cancellation.A central axis is drawn from the center of the slit to screen C, and P1 is located at an angle u to that axis. The wavelets of the pair of rays r1 and r2 are in phase within the slit because they originate from the same wavefront passing through the slit, along the width of the slit. However, to produce the first dark fringe they must be out of phase by l/2 when they reach P1; this phase difference is due to their path length difference, with the path traveled by the wavelet of r2 to reach P1 being longer than the path traveled by the wavelet of r1.To display this path length difference, we find a point b on ray r2 such that the path length from b to P1 matches the path length of ray r1. Then the path length difference between the two rays is the distance from the center of the slit to b.

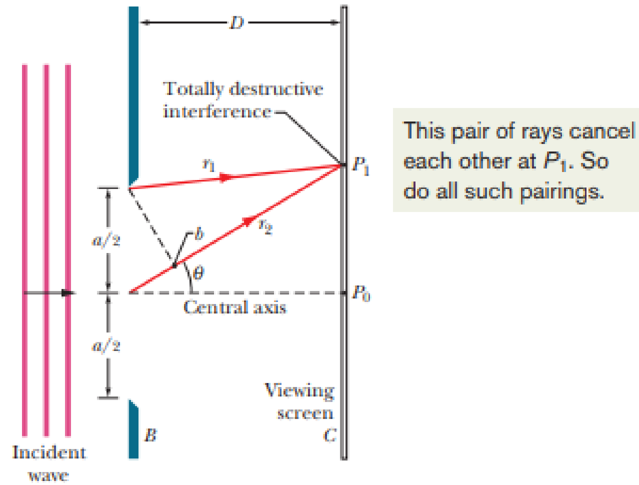


Fig. 36-4 Waves from the top points of two zones of width $a/2$ undergo fully destructive interference at point P_1 on viewing screen C.

When viewing screen C is near screen B, as in Fig. 36-4, the diffraction pattern on C is difficult to describe mathematically. However, we can simplify the mathematics considerably if we arrange for the screen separation D to be much larger than the slit width a. Then we can approximate rays r_1 and r_2 as being parallel, at angle θ to the central axis (Fig. 36-5). We can also approximate the triangle formed by point b, the top point of the slit, and the center point of the slit as being a right triangle, and one of the angles inside that triangle as being θ. The path length difference between rays r_1 and r_2 (which is still the distance from the center of the slit to point b) is then equal to $(a/2)\sin\theta$.

We can repeat this analysis for any other pair of rays originating at corresponding points in the two zones (say, at the midpoints of the zones) and extending to point P_1. Each such pair of rays has the same path length difference $(a/2)\sin\theta$. Setting this common path length difference equal to $\lambda/2$ (our condition for the first dark fringe), we have

$$\frac{a}{2}\sin\theta = \frac{\lambda}{2},$$

which gives us

$$a\sin\theta = \lambda \qquad \text{(first minimum).} \tag{36-1}$$

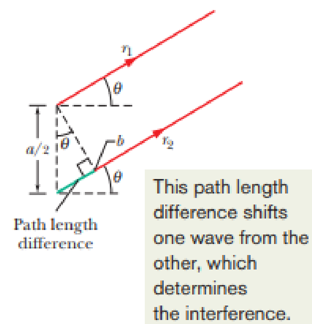


Fig. 36-5 For $D \gg a$, we can approximate rays r_1 and r_2 as being parallel, at angle θ to the central axis.

Given slit width a and wavelength λ, Eq. 36-1 tells us the angle θ of the first dark fringe above and (by symmetry) below the central axis.

Note that if we begin with $a > \lambda$ and then narrow the slit while holding the wavelength constant, we increase the angle at which the first dark fringes appear; that is, the extent of the diffraction (the extent of the flaring and the width of the pattern) is *greater* for a *narrower* slit. When we have reduced the slit width to the wavelength (that is, $a = \lambda$), the angle of the first dark fringes is 90°. Since the first dark fringes mark the two edges of the central bright fringe, that bright fringe must then cover the entire viewing screen.

We find the second dark fringes above and below the central axis as we found the first dark fringes, except that we now divide the slit into *four* zones of equal widths $a/4$, as shown in Fig. 36-6*a*. We then extend rays r_1, r_2, r_3, and r_4 from the top points of the zones to point P_2, the location of the second dark fringe above the central axis. To produce that fringe, the path length difference

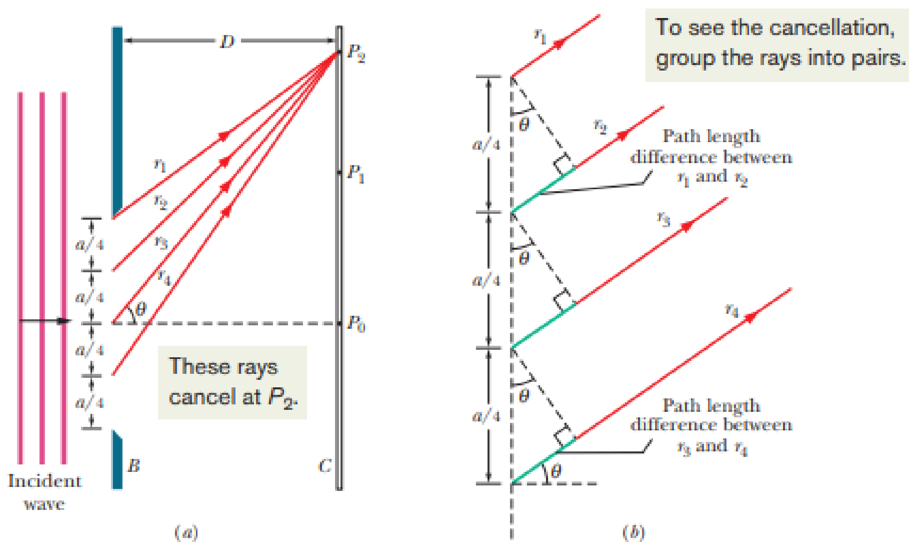


Fig. 36-6 (a) Waves from the top points of four zones of width a/4 undergo fully destructive interference at point P2. (b) For D a, we can approximate rays r1,r2,r3, and r4 as being parallel, at angle u to the central axis.

between r_1 and r_2, that between r_2 and r_3, and that between r_3 and r_4 must all be equal to $\lambda/2$.

For $D \gg a$, we can approximate these four rays as being parallel, at angle θ to the central axis. To display their path length differences, we extend a perpendicular line through each adjacent pair of rays, as shown in Fig. 36-6*b*, to form a series of right triangles, each of which has a path length difference as one side. We see from the top triangle that the path length difference between r_1 and r_2 is $(a/4) \sin \theta$. Similarly, from the bottom triangle, the path length difference between r_3 and r_4 is also $(a/4) \sin \theta$. In fact, the path length difference for any two rays that originate at corresponding points in two adjacent zones is $(a/4) \sin \theta$. Since in each such case the path length difference is equal to $\lambda/2$, we have

$$\frac{a}{4} \sin \theta = \frac{\lambda}{2},$$

which gives us

$$a \sin \theta = 2\lambda \qquad \text{(second minimum)}. \tag{36-2}$$

We could now continue to locate dark fringes in the diffraction pattern by splitting up the slit into more zones of equal width. We would always choose an even number of zones so that the zones (and their waves) could be paired as we have been doing. We would find that the dark fringes above and below the central axis can be located with the general equation

$$a \sin \theta = m\lambda, \qquad \text{for } m = 1, 2, 3, \ldots \quad \text{(minima—dark fringes)}. \tag{36-3}$$

You can remember this result in the following way. Draw a triangle like the one in Fig. 36-5, but for the full slit width a, and note that the path length difference between the top and bottom rays equals $a \sin \theta$. Thus, Eq. 36-3 says:

This may seem to be wrong because the waves of those two particular rays will be exactly in phase with each other when their path length difference is an integer number of wavelengths. However, they each will still be part of a pair of waves that are exactly out of phase with each other; thus, *each* wave will be canceled by some other wave, resulting in darkness. (Two light waves that are exactly out of phase will always cancel each other, giving a net wave of zero, even if they happen to be exactly in phase with other light waves.)

Equations 36-1, 36-2, and 36-3 are derived for the case of $D \gg a$. However, they also apply if we place a converging lens between the slit and the viewing screen and then move the screen in so that it coincides with the focal plane of the lens. The lens ensures that rays which now reach any point on the screen are *exactly* parallel (rather than approximately) back at the slit. They are like the initially parallel rays of Fig. 34-14*a* that are directed to the focal point by a converging lens.

Sample Problem

Single-slit diffraction pattern with white light

A slit of width a is illuminated by white light.

(a) For what value of a will the first minimum for red light of wavelength $\lambda = 650$ nm appear at $\theta = 15°$?

KEY IDEA

Diffraction occurs separately for each wavelength in the range of wavelengths passing through the slit, with the locations of the minima for each wavelength given by Eq. 36-3 ($a \sin \theta = m\lambda$).

Calculation: When we set $m = 1$ (for the first minimum) and substitute the given values of θ and λ, Eq. 36-3 yields

$$a = \frac{m\lambda}{\sin \theta} = \frac{(1)(650 \text{ nm})}{\sin 15°}$$
$$= 2511 \text{ nm} \approx 2.5\ \mu\text{m}. \qquad \text{(Answer)}$$

For the incident light to flare out that much ($\pm 15°$ to the first minima) the slit has to be very fine indeed—in this case, a mere four times the wavelength. For comparison, note that a fine human hair may be about 100 μm in diameter.

(b) What is the wavelength λ' of the light whose first side diffraction maximum is at 15°, thus coinciding with the first minimum for the red light?

KEY IDEA

The first side maximum for any wavelength is about halfway between the first and second minima for that wavelength.

Calculations: Those first and second minima can be located with Eq. 36-3 by setting $m = 1$ and $m = 2$, respectively. Thus, the first side maximum can be located *approximately* by setting $m = 1.5$. Then Eq. 36-3 becomes

$$a \sin \theta = 1.5\lambda'.$$

Solving for λ' and substituting known data yield

$$\lambda' = \frac{a \sin \theta}{1.5} = \frac{(2511 \text{ nm})(\sin 15°)}{1.5}$$
$$= 430 \text{ nm}. \qquad \text{(Answer)}$$

Light of this wavelength is violet (far blue, near the short-wavelength limit of the human range of visible light). From the two equations we used, can you see that the first side maximum for light of wavelength 430 nm will always coincide with the first minimum for light of wavelength 650 nm, no matter what the slit width is? However, the angle θ at which this overlap occurs does depend on slit width. If the slit is relatively narrow, the angle will be relatively large, and conversely.

36-4 Intensity in Single-Slit Diffraction, Qualitatively

In Section 36-3 we saw how to find the positions of the minima and the maxima in a single-slit diffraction pattern. Now we turn to a more general problem: find an expression for the intensity I of the pattern as a function of θ, the angular position of a point on a viewing screen.

To do this, we divide the slit of Fig. 36-4 into N zones of equal widths Δx small enough that we can assume each zone acts as a source of Huygens wavelets. We wish to superimpose the wavelets arriving at an arbitrary point P on the viewing screen, at angle θ to the central axis, so that we can determine the amplitude E_θ of the electric component of the resultant wave at P. The intensity of the light at P is then proportional to the square of that amplitude.

To find E_θ, we need the phase relationships among the arriving wavelets. The phase difference between wavelets from adjacent zones is given by

$$\begin{pmatrix} \text{phase} \\ \text{difference} \end{pmatrix} = \left(\frac{2\pi}{\lambda}\right) \begin{pmatrix} \text{path length} \\ \text{difference} \end{pmatrix}.$$

For point P at angle θ, the path length difference between wavelets from adjacent zones is $\Delta x \sin \theta$; so the phase difference $\Delta\phi$ between wavelets from adjacent zones is

$$\Delta\phi = \left(\frac{2\pi}{\lambda}\right)(\Delta x \sin \theta). \qquad (36\text{-}4)$$

36-5 Intensity in Single-Slit Diffraction, Quantitatively

Equation 36-3 tells us how to locate the minima of the single-slit diffraction pattern on screen C of Fig. 36-4 as a function of the angle θ in that figure. Here we wish to derive an expression for the intensity $I(\theta)$ of the pattern as a function of θ. We state, and shall prove below, that the intensity is given by

$$I(\theta) = I_m\left(\frac{\sin\alpha}{\alpha}\right)^2, \tag{36-5}$$

where

$$\alpha = \tfrac{1}{2}\phi = \frac{\pi a}{\lambda}\sin\theta. \tag{36-6}$$

The symbol α is just a convenient connection between the angle θ that locates a point on the viewing screen and the light intensity $I(\theta)$ at that point. The intensity I_m is the greatest value of the intensities $I(\theta)$ in the pattern and occurs at the central maximum (where $\theta = 0$), and ϕ is the phase difference (in radians) between the top and bottom rays from the slit of width a.

Study of Eq. 36-5 shows that intensity minima will occur where

$$\alpha = m\pi, \qquad \text{for } m = 1, 2, 3, \ldots. \tag{36-7}$$

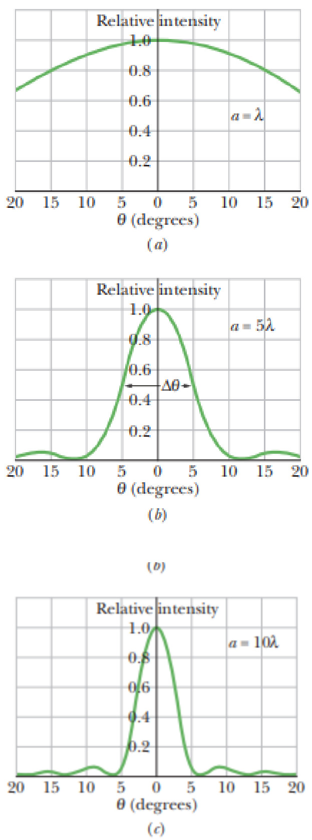


Fig. 36-8 The relative intensity in single-slit diffraction for three values of the ratio a/λ. The wider the slit is, the narrower is the central diffraction maximum.

If we put this result into Eq. 36-6, we find

$$m\pi = \frac{\pi a}{\lambda}\sin\theta, \qquad \text{for } m = 1, 2, 3, \ldots,$$

or

$$a\sin\theta = m\lambda, \qquad \text{for } m = 1, 2, 3, \ldots \quad \text{(minima—dark fringes)}, \tag{36-8}$$

which is exactly Eq. 36-3, the expression that we derived earlier for the location of the minima.

Figure 36-8 shows plots of the intensity of a single-slit diffraction pattern, calculated with Eqs. 36-5 and 36-6 for three slit widths: $a = \lambda$, $a = 5\lambda$, and $a = 10\lambda$. Note that as the slit width increases (relative to the wavelength), the width of the *central diffraction maximum* (the central hill-like region of the graphs) decreases; that is, the light undergoes less flaring by the slit. The secondary maxima also decrease in width (and become weaker). In the limit of slit width a being much greater than wavelength λ, the secondary maxima due to the slit disappear; we then no longer have single-slit diffraction (but we still have diffraction due to the edges of the wide slit, like that produced by the edges of the razor blade in Fig. 36-2).

Proof of Eqs. 36-5 and 36-6

To find an expression for the intensity at a point in the diffraction pattern, we need to divide the slit into many zones and then add the phasors corresponding to those zones, as we did in Fig. 36-7. The arc of phasors in Fig. 36-9 represents the wavelets that reach an arbitrary point P on the viewing screen of Fig. 36-4, corresponding to a particular small angle θ. The amplitude E_θ of the resultant wave at P is the vector sum of these phasors. If we divide the slit of Fig. 36-4 into infinitesimal zones of width Δx, the arc of phasors in Fig. 36-9 approaches the arc of a circle; we call its radius R as indicated in that figure. The length of the arc must be E_m, the amplitude at the center of the diffraction pattern, because if we straightened out the arc we would have the phasor arrangement of Fig. 36-7*a* (shown lightly in Fig. 36-9).

The angle ϕ in the lower part of Fig. 36-9 is the difference in phase between the infinitesimal vectors at the left and right ends of arc E_m. From the geometry, ϕ is also the angle between the two radii marked R in Fig. 36-9. The dashed line in that figure, which bisects ϕ, then forms two congruent right triangles. From either triangle we can write

$$\sin\tfrac{1}{2}\phi = \frac{E_\theta}{2R}. \tag{36-9}$$

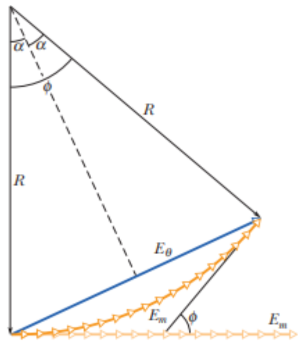


Fig. 36-9 A construction used to calculate the intensity in single-slit diffraction. The situation shown corresponds to that of Fig. 36-7*b*.

In radian measure, ϕ is (with E_m considered to be a circular arc)

$$\phi = \frac{E_m}{R}.$$

Solving this equation for R and substituting in Eq. 36-9 lead to

$$E_\theta = \frac{E_m}{\frac{1}{2}\phi}\sin\tfrac{1}{2}\phi. \tag{36-10}$$

In Section 33-5 we saw that the intensity of an electromagnetic wave is proportional to the square of the amplitude of its electric field. Here, this means that the maximum intensity I_m (which occurs at the center of the diffraction pattern) is proportional to E_m^2 and the intensity $I(\theta)$ at angle θ is proportional to E_θ^2. Thus, we may write

$$\frac{I(\theta)}{I_m} = \frac{E_\theta^2}{E_m^2}. \tag{36-11}$$

Substituting for E_θ with Eq. 36-10 and then substituting $\alpha = \frac{1}{2}\phi$, we are led to the following expression for the intensity as a function of θ:

$$I(\theta) = I_m\left(\frac{\sin\alpha}{\alpha}\right)^2.$$

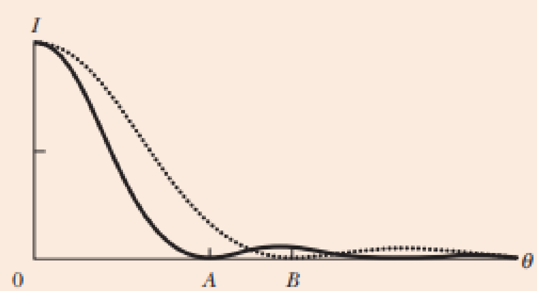
This is exactly Eq. 36-5, one of the two equations we set out to prove.

The second equation we wish to prove relates α to θ. The phase difference ϕ between the rays from the top and bottom of the entire slit may be related to a path length difference with Eq. 36-4; it tells us that

$$\phi = \left(\frac{2\pi}{\lambda}\right)(a\sin\theta),$$

where a is the sum of the widths Δx of the infinitesimal zones. However, $\phi = 2\alpha$, so this equation reduces to Eq. 36-6.

Two wavelengths, 650 and 430 nm, are used separately in a single-slit diffraction experiment. The figure shows the results as graphs of intensity I versus angle θ for the two diffraction patterns. If both wavelengths are then used simultaneously, what color will be seen in the combined diffraction pattern at (a) angle A and (b) angle B?



Sample Problem

Intensities of the maxima in a single-slit interference pattern

Find the intensities of the first three secondary maxima (side maxima) in the single-slit diffraction pattern of Fig. 36-1, measured as a percentage of the intensity of the central maximum.

KEY IDEAS

The secondary maxima lie approximately halfway between the minima, whose angular locations are given by Eq. 36-7 ($\alpha = m\pi$). The locations of the secondary maxima are then given (approximately) by

$$a = (m + \tfrac{1}{2})\pi, \qquad \text{for } m = 1, 2, 3, \ldots,$$

with α in radian measure. We can relate the intensity I at any point in the diffraction pattern to the intensity I_m of the central maximum via Eq. 36-5.

Calculations: Substituting the approximate values of α for the secondary maxima into Eq. 36-5 to obtain the relative intensities at those maxima, we get

$$\frac{I}{I_m} = \left(\frac{\sin\alpha}{\alpha}\right)^2 = \left(\frac{\sin(m + \frac{1}{2})\pi}{(m + \frac{1}{2})\pi}\right)^2, \quad \text{for } m = 1, 2, 3, \ldots.$$

The first of the secondary maxima occurs for $m = 1$, and its relative intensity is

$$\frac{I_1}{I_m} = \left(\frac{\sin(1 + \frac{1}{2})\pi}{(1 + \frac{1}{2})\pi}\right)^2 = \left(\frac{\sin 1.5\pi}{1.5\pi}\right)^2$$

$$= 4.50 \times 10^{-2} \approx 4.5\%. \qquad \text{(Answer)}$$

For $m = 2$ and $m = 3$ we find that

$$\frac{I_2}{I_m} = 1.6\% \quad \text{and} \quad \frac{I_3}{I_m} = 0.83\%. \qquad \text{(Answer)}$$

As you can see from these results, successive secondary maxima decrease rapidly in intensity. Figure 36-1 was deliberately overexposed to reveal them.

36-6 Diffraction by a Circular Aperture

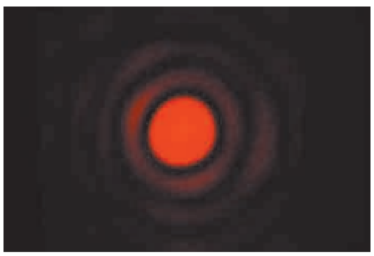


Fig. 36-10 The diffraction pattern of a circular aperture. Note the central maximum and the circular secondary maxima. The figure has been overexposed to bring out these secondary maxima, which are much less intense than the central maximum. *(Jearl Walker)*

Here we consider diffraction by a circular aperture—that is, a circular opening, such as a circular lens, through which light can pass. Figure 36-10 shows the image formed by light from a laser that was directed onto a circular aperture with a very small diameter. This image is not a point, as geometrical optics would suggest, but a circular disk surrounded by several progressively fainter secondary rings. Comparison with Fig. 36-1 leaves little doubt that we are dealing with a diffraction phenomenon. Here, however, the aperture is a circle of diameter d rather than a rectangular slit.

The (complex) analysis of such patterns shows that the first minimum for the diffraction pattern of a circular aperture of diameter d is located by

$$\sin\theta = 1.22\frac{\lambda}{d} \qquad \text{(first minimum—circular aperture).} \qquad (36\text{-}12)$$

The angle θ here is the angle from the central axis to any point on that (circular) minimum. Compare this with Eq. 36-1,

$$\sin\theta = \frac{\lambda}{a} \qquad \text{(first minimum—single slit),} \qquad (36\text{-}13)$$

which locates the first minimum for a long narrow slit of width a. The main difference is the factor 1.22, which enters because of the circular shape of the aperture.

Resolvability

The fact that lens images are diffraction patterns is important when we wish to *resolve* (distinguish) two distant point objects whose angular separation is small. Figure 36-11 shows, in three different cases, the visual appearance and corresponding intensity pattern for two distant point objects (stars, say) with small angular separation. In Figure 36-11*a*, the objects are not resolved because of diffraction; that is, their diffraction patterns (mainly their central maxima) overlap so much that the two objects cannot be distinguished from a single point object. In Fig. 36-11*b* the objects are barely resolved, and in Fig. 36-11*c* they are fully resolved.

In Fig. 36-11*b* the angular separation of the two point sources is such that the central maximum of the diffraction pattern of one source is centered on the first minimum of the diffraction pattern of the other, a condition called **Rayleigh's criterion** for resolvability. From Eq. 36-12, two objects that are barely resolvable

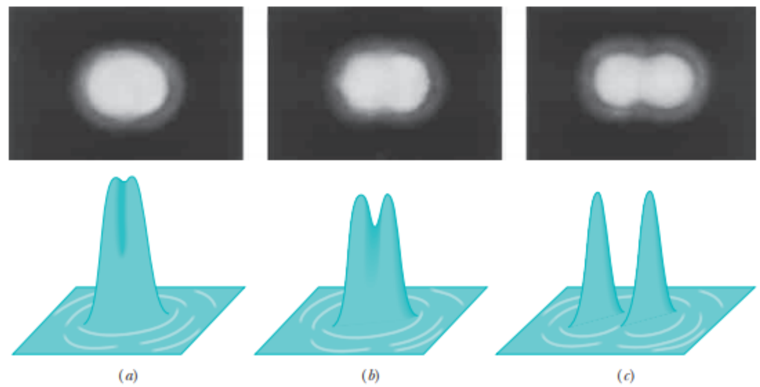


Fig. 36-11 At the top, the images of two point sources (stars) formed by a converging lens. At the bottom, representations of the image intensities. In (*a*) the angular separation of the sources is too small for them to be distinguished, in (*b*) they can be marginally distinguished, and in (*c*) they are clearly distinguished. Rayleigh's criterion is satisfied in (*b*), with the central maximum of one diffraction pattern coinciding with the first minimum of the other.

by this criterion must have an angular separation θ_R of

$$\theta_R = \sin^{-1}\frac{1.22\lambda}{d}.$$

Since the angles are small, we can replace $\sin\theta_R$ with θ_R expressed in radians:

$$\theta_R = 1.22\frac{\lambda}{d} \qquad \text{(Rayleigh's criterion).} \qquad (36\text{-}14)$$

Applying Rayleigh's criterion for resolvability to human vision is only an approximation because visual resolvability depends on many factors, such as the relative brightness of the sources and their surroundings, turbulence in the air between the sources and the observer, and the functioning of the observer's visual system. Experimental results show that the least angular separation that can actually be resolved by a person is generally somewhat greater than the value given by Eq. 36-14. However, for calculations here, we shall take Eq. 36-14 as being a precise criterion: If the angular separation θ between the sources is greater than θ_R, we can visually resolve the sources; if it is less, we cannot.

Rayleigh's criterion can explain the arresting illusions of color in the style of painting known as pointillism (Fig. 36-12). In this style, a painting is made not with brush strokes in the usual sense but rather with a myriad of small colored dots. One fascinating aspect of a pointillistic painting is that when you change your distance from it, the colors shift in subtle, almost subconscious ways. This color shifting has to do with whether you can resolve the colored dots. When you stand close enough to the painting, the angular separations θ of adjacent dots are greater than θ_R and thus the dots can be seen individually. Their colors are the true colors of the paints used. However, when you stand far enough from the painting, the angular separations θ are less than θ_R and the dots cannot be seen individually. The resulting blend of colors coming into your eye from any group of dots can then cause your brain to "make up" a color for that group—a color that may not actually exist in the group. In this way, a pointillistic painter uses your visual system to create the colors of the art.

Fig. 36-12 The pointillistic painting *The Seine at Herblay* by Maximilien Luce consists of thousands of colored dots. With the viewer very close to the canvas, the dots and their true colors are visible. At normal viewing distances, the dots are irresolvable and thus blend. *(Maximilien Luce, The Seine at Herblay, 1890. Musee d'Orsay, Paris, France. Photo by Erich Lessing/Art Resource)*

When we wish to use a lens instead of our visual system to resolve objects of small angular separation, it is desirable to make the diffraction pattern as small as possible. According to Eq. 36-14, this can be done either by increasing the lens diameter or by using light of a shorter wavelength. For this reason ultraviolet light is often used with microscopes because its wavelength is shorter than a visible light wavelength.

Sample Problem

Pointillistic paintings use the diffraction of your eye

Figure 36-13*a* is a representation of the colored dots on a pointillistic painting. Assume that the average center-to-center separation of the dots is $D = 2.0$ mm. Also assume that the diameter of the pupil of your eye is $d = 1.5$ mm and that the least angular separation between dots you can resolve is set only by Rayleigh's criterion. What is the least viewing distance from which you cannot distinguish any dots on the painting?

KEY IDEA

Consider any two adjacent dots that you can distinguish when you are close to the painting. As you move away, you continue to distinguish the dots until their angular separation θ (in your view) has decreased to the angle given by Rayleigh's criterion:

$$\theta_R = 1.22\frac{\lambda}{d}. \tag{36-15}$$

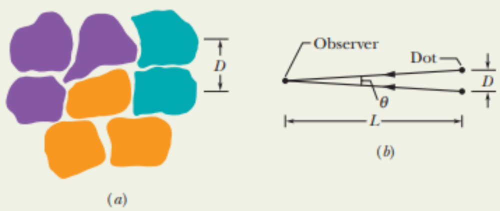


Fig. 36-13 (*a*) Representation of some dots on a pointillistic painting, showing an average center-to-center separation D. (*b*) The arrangement of separation D between two dots, their angular separation θ, and the viewing distance L.

Calculations: Figure 36-13*b* shows, from the side, the angular separation θ of the dots, their center-to-center separation D, and your distance L from them. Because D/L is small, angle θ is also small and we can make the approximation

$$\theta = \frac{D}{L}. \tag{36-16}$$

Setting θ of Eq. 36-16 equal to θ_R of Eq. 36-15 and solving for L, we then have

$$L = \frac{Dd}{1.22\lambda}. \tag{36-17}$$

Equation 36-17 tells us that L is larger for smaller λ. Thus, as you move away from the painting, adjacent red dots (long wavelengths) become indistinguishable before adjacent blue dots do. To find the least distance L at which *no* colored dots are distinguishable, we substitute $\lambda = 400$ nm (blue or violet light) into Eq. 36-17:

$$L = \frac{(2.0 \times 10^{-3}\text{ m})(1.5 \times 10^{-3}\text{ m})}{(1.22)(400 \times 10^{-9}\text{ m})} = 6.1\text{ m}. \quad \text{(Answer)}$$

At this or a greater distance, the color you perceive at any given spot on the painting is a blended color that may not actually exist there.

Sample Problem

Rayleigh's criterion for resolving two distant objects

A circular converging lens, with diameter $d = 32$ mm and focal length $f = 24$ cm, forms images of distant point objects in the focal plane of the lens. The wavelength is $\lambda = 550$ nm.

(a) Considering diffraction by the lens, what angular separation must two distant point objects have to satisfy Rayleigh's criterion?

KEY IDEA

Figure 36-14 shows two distant point objects P_1 and P_2, the lens, and a viewing screen in the focal plane of the lens. It also shows, on the right, plots of light intensity I versus position on the screen for the central maxima of the images formed by the lens. Note that the angular separation θ_o of the objects equals the angular separation θ_i of the images. Thus, if the images are to satisfy Rayleigh's criterion

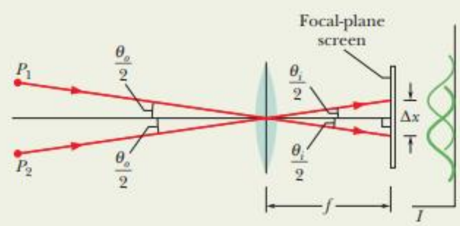


Fig. 36-14 Light from two distant point objects P_1 and P_2 passes through a converging lens and forms images on a viewing screen in the focal plane of the lens. Only one representative ray from each object is shown. The images are not points but diffraction patterns, with intensities approximately as plotted at the right. The angular separation of the objects is θ_o and that of the images is θ_i; the central maxima of the images have a separation Δx.

for resolvability, the angular separations on both sides of the lens must be given by Eq. 36-14 (assuming small angles).

Calculations: From Eq. 36-14, we obtain

$$\theta_o = \theta_i = \theta_R = 1.22\frac{\lambda}{d}$$
$$= \frac{(1.22)(550 \times 10^{-9}\text{ m})}{32 \times 10^{-3}\text{ m}} = 2.1 \times 10^{-5}\text{ rad. (Answer)}$$

At this angular separation, each central maximum in the two intensity curves of Fig. 36-14 is centered on the first minimum of the other curve.

(b) What is the separation Δx of the centers of the *images* in the focal plane? (That is, what is the separation of the *central* peaks in the two intensity-versus-position curves?)

Calculations: From either triangle between the lens and the screen in Fig. 36-14, we see that $\tan\,\theta_i/2 = \Delta x/2f$. Rearranging this equation and making the approximation $\tan\theta \approx \theta$, we find

$$\Delta x = f\theta_i, \qquad (36\text{-}18)$$

where θ_i is in radian measure. Substituting known data then yields

$$\Delta x = (0.24\text{ m})(2.1 \times 10^{-5}\text{ rad}) = 5.0\ \mu\text{m}. \qquad \text{(Answer)}$$

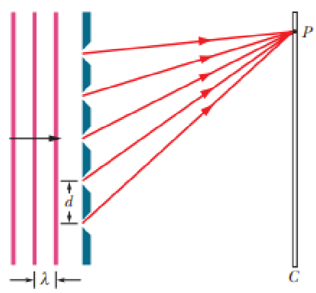


Fig. 36-18 An idealized diffraction grating, consisting of only five rulings, that produces an interference pattern on a distant viewing screen C.

36-8 Diffraction Gratings

One of the most useful tools in the study of light and of objects that emit and absorb light is the **diffraction grating.** This device is somewhat like the double-slit arrangement of Fig. 35-10 but has a much greater number N of slits, often called *rulings,* perhaps as many as several thousand per millimeter. An idealized grating consisting of only five slits is represented in Fig. 36-18. When monochromatic light is sent through the slits, it forms narrow interference fringes that can be analyzed to determine the wavelength of the light. (Diffraction gratings can also be opaque surfaces with narrow parallel grooves arranged like the slits in Fig. 36-18. Light then scatters back from the grooves to form interference fringes rather than being transmitted through open slits.)

With monochromatic light incident on a diffraction grating, if we gradually increase the number of slits from two to a large number N, the intensity plot changes from the typical double-slit plot of Fig. 36-15*c* to a much more complicated one and then eventually to a simple graph like that shown in Fig. 36-19*a*. The pattern you would see on a viewing screen using monochromatic red light from, say, a helium–neon laser is shown in Fig. 36-19*b*. The maxima are now very narrow (and so are called *lines*); they are separated by relatively wide dark regions.

We use a familiar procedure to find the locations of the bright lines on the viewing screen. We first assume that the screen is far enough from the grating so that the rays reaching a particular point P on the screen are approximately parallel when they leave the grating (Fig. 36-20). Then we apply to each pair of adjacent rulings the same reasoning we used for double-slit interference. The separation d between rulings is called the *grating spacing.* (If N rulings occupy a total width w, then $d = w/N$.) The path length difference between adjacent rays is again $d \sin\theta$ (Fig. 36-20), where θ is the angle from the central axis of the grating (and of the diffraction pattern) to point P. A line will be located at P if the path length difference between adjacent rays is an integer number of wavelengths—that is, if

$$d \sin \theta = m\lambda, \quad \text{for } m = 0, 1, 2, \ldots \quad \text{(maxima—lines)}, \qquad (36\text{-}25)$$

where λ is the wavelength of the light. Each integer m represents a different line; hence these integers can be used to label the lines, as in Fig. 36-19. The integers are then called the *order numbers,* and the lines are called the zeroth-order line (the central line, with $m = 0$), the first-order line ($m = 1$), the second-order line ($m = 2$), and so on.

If we rewrite Eq. 36-25 as $\theta = \sin^{-1}(m\lambda/d)$, we see that, for a given diffraction grating, the angle from the central axis to any line (say, the third-order line) depends on the wavelength of the light being used. Thus, when light of an unknown wavelength is sent through a diffraction grating, measurements of the angles to the higher-order lines can be used in Eq. 36-25 to determine the wavelength. Even light of several unknown wavelengths can be distinguished and identified in this way. We cannot do that with the double-slit arrangement of Section 35-4, even though the same equation and wavelength dependence apply there. In double-slit interference, the bright fringes due to different wavelengths overlap too much to be distinguished.

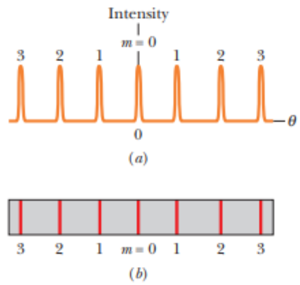


Fig. 36-19 (*a*) The intensity plot produced by a diffraction grating with a great many rulings consists of narrow peaks, here labeled with their order numbers m. (*b*) The corresponding bright fringes seen on the screen are called lines and are here also labeled with order numbers m.

Width of the Lines

A grating's ability to resolve (separate) lines of different wavelengths depends on the width of the lines. We shall here derive an expression for the *half-width* of the central line (the line for which $m = 0$) and then state an expression for the half-widths of the higher-order lines. We define the **half-width** of the central line as being the angle $\Delta\theta_{hw}$ from the center of the line at $\theta = 0$ outward to where the line effectively ends and darkness effectively begins with the first minimum (Fig. 36-21). At such a minimum, the N rays from the N slits of the grating cancel one another. (The actual width of the central line is, of course, $2(\Delta\theta_{hw})$, but line widths are usually compared via half-widths.)

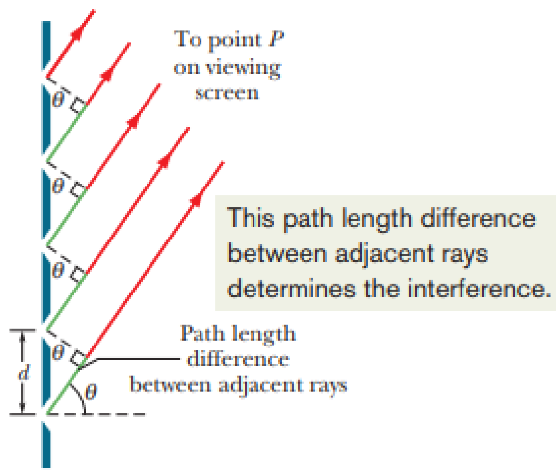


Fig. 36-20 The rays from the rulings in a diffraction grating to a distant point P are approximately parallel. The path length difference between each two adjacent rays is $d \sin \theta$, where θ is measured as shown. (The rulings extend into and out of the page.)

In Section 36-3 we were also concerned with the cancellation of a great many rays, there due to diffraction through a single slit. We obtained Eq. 36-3, which, because of the similarity of the two situations, we can use to find the first minimum here. It tells us that the first minimum occurs where the path length difference between the top and bottom rays equals λ. For single-slit diffraction, this difference is $a \sin \theta$. For a grating of N rulings, each separated from the next by distance d, the distance between the top and bottom rulings is Nd (Fig. 36-22), and so the path length difference between the top and bottom rays here is $Nd \sin \Delta\theta_{hw}$. Thus, the first minimum occurs where

$$Nd \sin \Delta\theta_{hw} = \lambda. \qquad (36\text{-}26)$$

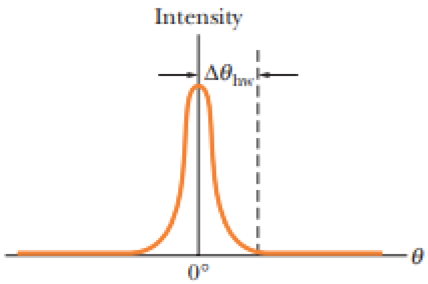


Fig. 36-21 The half-width $\Delta\theta_{hw}$ of the central line is measured from the center of that line to the adjacent minimum on a plot of I versus θ like Fig. 36-19*a*.

Because $\Delta\theta_{hw}$ is small, $\sin \Delta\theta_{hw} = \Delta\theta_{hw}$ (in radian measure). Substituting this in Eq. 36-26 gives the half-width of the central line as

$$\Delta\theta_{hw} = \frac{\lambda}{Nd} \qquad \text{(half-width of central line).} \tag{36-27}$$

We state without proof that the half-width of any other line depends on its location relative to the central axis and is

$$\Delta\theta_{hw} = \frac{\lambda}{Nd\cos\theta} \qquad \text{(half-width of line at } \theta\text{).} \tag{36-28}$$

Note that for light of a given wavelength λ and a given ruling separation d, the widths of the lines decrease with an increase in the number N of rulings. Thus, of two diffraction gratings, the grating with the larger value of N is better able to distinguish between wavelengths because its diffraction lines are narrower and so produce less overlap.

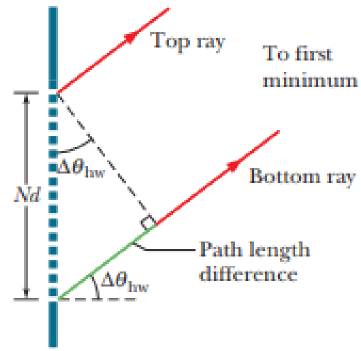


Fig. 36-22 The top and bottom rulings of a diffraction grating of N rulings are separated by Nd. The top and bottom rays passing through these rulings have a path length difference of $Nd \sin \Delta\theta_{hw}$, where $\Delta\theta_{hw}$ is the angle to the first minimum. (The angle is here greatly exaggerated for clarity.)

Gratings: Dispersion and Resolving Power

The fine rulings, each 0.5 μm wide, on a compact disc function as a diffraction grating. When a small source of white light illuminates a disc, the diffracted light forms colored "lanes" that are the composite of the diffraction patterns from the rulings.

(Kristen Brochmann/Fundamental Photographs)

To be useful in distinguishing wavelengths that are close to each other (as in a grating spectroscope), a grating must spread apart the diffraction lines associated with the various wavelengths. This spreading, called **dispersion,** is defined as

$$D = \frac{\Delta\theta}{\Delta\lambda} \quad \text{(dispersion defined).} \tag{36-29}$$

Here $\Delta\theta$ is the angular separation of two lines whose wavelengths differ by $\Delta\lambda$. The greater D is, the greater is the distance between two emission lines whose wavelengths differ by $\Delta\lambda$. We show below that the dispersion of a grating at angle θ is given by

$$D = \frac{m}{d\cos\theta} \quad \text{(dispersion of a grating).} \tag{36-30}$$

Thus, to achieve higher dispersion we must use a grating of smaller grating spacing d and work in a higher-order m. Note that the dispersion does not depend on the number of rulings N in the grating. The SI unit for D is the degree per meter or the radian per meter.

Resolving Power

To *resolve* lines whose wavelengths are close together (that is, to make the lines distinguishable), the line should also be as narrow as possible. Expressed otherwise, the grating should have a high **resolving power** R, defined as

$$R = \frac{\lambda_{avg}}{\Delta\lambda} \quad \text{(resolving power defined).} \tag{36-31}$$

Here λ_{avg} is the mean wavelength of two emission lines that can barely be recognized as separate, and $\Delta\lambda$ is the wavelength difference between them. The greater R is, the closer two emission lines can be and still be resolved. We shall show below that the resolving power of a grating is given by the simple expression

$$R = Nm \quad \text{(resolving power of a grating).} \tag{36-32}$$

To achieve high resolving power, we must use many rulings (large N).

Proof of Eq. 36-30

Let us start with Eq. 36-25, the expression for the locations of the lines in the diffraction pattern of a grating:

$$d \sin\theta = m\lambda.$$

Let us regard θ and λ as variables and take differentials of this equation. We find

$$d(\cos\theta)\, d\theta = m\, d\lambda.$$

For small enough angles, we can write these differentials as small differences, obtaining

$$d(\cos\theta)\,\Delta\theta = m\,\Delta\lambda \tag{36-33}$$

or

$$\frac{\Delta\theta}{\Delta\lambda} = \frac{m}{d\cos\theta}.$$

The ratio on the left is simply D (see Eq. 36-29), and so we have indeed derived Eq. 36-30.

Proof of Eq. 36-32

We start with Eq. 36-33, which was derived from Eq. 36-25, the expression for the locations of the lines in the diffraction pattern formed by a grating. Here $\Delta\lambda$ is the small wavelength difference between two waves that are diffracted by the grating, and $\Delta\theta$ is the angular separation between them in the diffraction pattern. If $\Delta\theta$ is to be the smallest angle that will permit the two lines to be resolved, it must (by Rayleigh's criterion) be equal to the half-width of each line, which is given by Eq. 36-28:

$$\Delta\theta_{hw} = \frac{\lambda}{Nd\cos\theta}.$$

If we substitute $\Delta\theta_{hw}$ as given here for $\Delta\theta$ in Eq. 36-33, we find that

$$\frac{\lambda}{N} = m\,\Delta\lambda,$$

from which it readily follows that

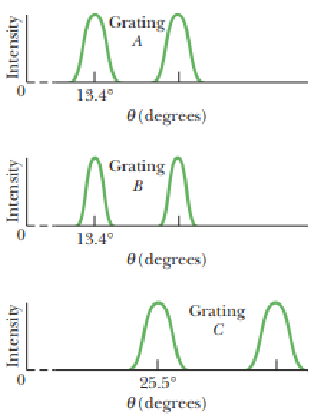


Fig. 36-26 The intensity patterns for light of two wavelengths sent through the gratings of Table 36-1. Grating B has the highest resolving power, and grating C the highest dispersion.

from which it readily follows that

$$R = \frac{\lambda}{\Delta\lambda} = Nm.$$

This is Eq. 36-32, which we set out to derive.

Dispersion and Resolving Power Compared

The resolving power of a grating must not be confused with its dispersion. Table 36-1 shows the characteristics of three gratings, all illuminated with light of wavelength $\lambda = 589$ nm, whose diffracted light is viewed in the first order ($m = 1$ in Eq. 36-25). You should verify that the values of D and R as given in the table can be calculated with Eqs. 36-30 and 36-32, respectively. (In the calculations for D, you will need to convert radians per meter to degrees per micrometer.)

For the conditions noted in Table 36-1, gratings A and B have the same *dispersion* D and A and C have the same *resolving power* R.

Figure 36-26 shows the intensity patterns (also called *line shapes*) that would be produced by these gratings for two lines of wavelengths λ_1 and λ_2, in the vicinity of $\lambda = 589$ nm. Grating B, with the higher resolving power, produces narrower lines and thus is capable of distinguishing lines that are much closer together in wavelength than those in the figure. Grating C, with the higher dispersion, produces the greater angular separation between the lines.

Table 36-1

Three Gratings[a]

Grating	N	d (nm)	θ	D (°/μm)	R
A	10 000	2540	13.4°	23.2	10 000
B	20 000	2540	13.4°	23.2	20 000
C	10 000	1360	25.5°	46.3	10 000

[a]Data are for $\lambda = 589$ nm and $m = 1$.

Sample Problem

Dispersion and resolving power of a diffraction grating

A diffraction grating has 1.26×10^4 rulings uniformly spaced over width $w = 25.4$ mm. It is illuminated at normal incidence by yellow light from a sodium vapor lamp. This light contains two closely spaced emission lines (known as the sodium doublet) of wavelengths 589.00 nm and 589.59 nm.

(a) At what angle does the first-order maximum occur (on either side of the center of the diffraction pattern) for the wavelength of 589.00 nm?

KEY IDEA

The maxima produced by the diffraction grating can be determined with Eq. 36-25 ($d \sin\theta = m\lambda$).

Calculations: The grating spacing d is

$$d = \frac{w}{N} = \frac{25.4 \times 10^{-3}\ \text{m}}{1.26 \times 10^4} = 2.016 \times 10^{-6}\ \text{m} = 2016\ \text{nm}.$$

The first-order maximum corresponds to $m = 1$. Substituting these values for d and m into Eq. 36-25 leads to

$$\theta = \sin^{-1}\frac{m\lambda}{d} = \sin^{-1}\frac{(1)(589.00\ \text{nm})}{2016\ \text{nm}} = 16.99^\circ \approx 17.0^\circ. \qquad \text{(Answer)}$$

(b) Using the dispersion of the grating, calculate the angular separation between the two lines in the first order.

KEY IDEAS

(1) The angular separation $\Delta\theta$ between the two lines in the first order depends on their wavelength difference $\Delta\lambda$ and the dispersion D of the grating, according to Eq. 36-29 ($D = \Delta\theta/\Delta\lambda$). (2) The dispersion D depends on the angle θ at which it is to be evaluated.

Calculations: We can assume that, in the first order, the two sodium lines occur close enough to each other for us to evaluate D at the angle $\theta = 16.99^\circ$ we found in part (a) for one of those lines. Then Eq. 36-30 gives the dispersion as

$$D = \frac{m}{d\cos\theta} = \frac{1}{(2016\ \text{nm})(\cos 16.99^\circ)} = 5.187 \times 10^{-4}\ \text{rad/nm}.$$

From Eq. 36-29 and with $\Delta\lambda$ in nanometers, we then have

$$\Delta\theta = D\,\Delta\lambda = (5.187 \times 10^{-4}\ \text{rad/nm})(589.59 - 589.00) = 3.06 \times 10^{-4}\ \text{rad} = 0.0175^\circ. \qquad \text{(Answer)}$$

You can show that this result depends on the grating spacing d but not on the number of rulings there are in the grating.

(c) What is the least number of rulings a grating can have and still be able to resolve the sodium doublet in the first order?

KEY IDEAS

(1) The resolving power of a grating in any order m is physically set by the number of rulings N in the grating according to Eq. 36-32 ($R = Nm$). (2) The smallest wavelength difference $\Delta\lambda$ that can be resolved depends on the average wavelength involved and on the resolving power R of the grating, according to Eq. 36-31 ($R = \lambda_{avg}/\Delta\lambda$).

Calculation: For the sodium doublet to be barely resolved, $\Delta\lambda$ must be their wavelength separation of 0.59 nm, and λ_{avg} must be their average wavelength of 589.30 nm. Thus, we find that the smallest number of rulings for a grating to resolve the sodium doublet is

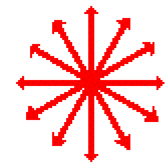
$$N = \frac{R}{m} = \frac{\lambda_{avg}}{m\,\Delta\lambda} = \frac{589.30\ \text{nm}}{(1)(0.59\ \text{nm})} = 999\ \text{rulings}. \qquad \text{(Answer)}$$

UNIT-4

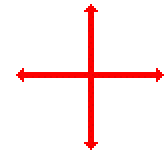
Polarization

A light wave is an electromagnetic wave that travels through the vacuum of outer space. Light waves are produced by vibrating electric charges. The nature of such electromagnetic waves is beyond the scope of The Physics Classroom Tutorial. For our purposes, it is sufficient to merely say that an electromagnetic wave is a transverse wave that has both an electric and a magnetic component.

A light wave is known to vibrate in a multitude of directions ...



... In general, a light wave can be thought of as vibrating in a vertical and in a horizontal plane.



The transverse nature of an electromagnetic wave is quite different from any other type of wave that has been discussed in The Physics Classroom Tutorial. Let's suppose that we use the customary slinky to model the behavior of an electromagnetic wave. As an electromagnetic wave traveled towards you, then you would observe the vibrations of the slinky occurring in more than one plane of vibration. This is quite different than what you might notice if you were to look along a slinky and observe a slinky wave traveling towards you. Indeed, the coils of the slinky would be vibrating back and forth as the slinky approached; yet these vibrations would occur in a single plane of space. That is, the coils of the slinky might vibrate up and down or left and right. Yet regardless of their direction of vibration, they would be moving along the same linear direction as you sighted along the slinky. If a slinky wave were an electromagnetic wave, then the vibrations of the slinky would occur in multiple planes. Unlike a usual slinky wave, the electric and magnetic vibrations of an electromagnetic wave occur in numerous planes. A light wave that is vibrating in more than one plane is referred to as **unpolarized light**. Light emitted by the sun, by a lamp in the classroom, or by a candle flame is unpolarized light. Such light waves are created by electric charges that vibrate in a variety of directions, thus creating an electromagnetic wave that vibrates in a variety of directions. This concept of unpolarized light is rather difficult to visualize. In general, it is helpful to picture unpolarized light as a wave that has an average of half its vibrations in a horizontal plane and half of its vibrations in a vertical plane.

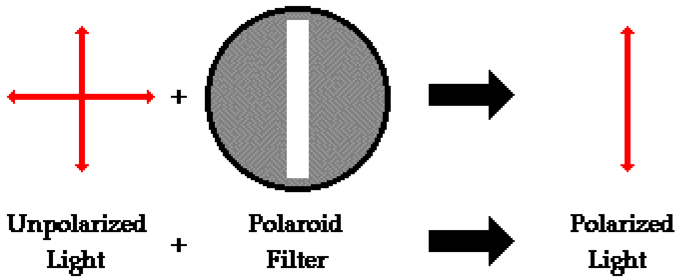
It is possible to transform unpolarized light into **polarized light**. Polarized light waves are light waves in which the vibrations occur in a single plane. The process of transforming unpolarized light into polarized light is known as **polarization**. There are a variety of methods of polarizing light. The four methods discussed on this page are:

- Polarization by Transmission
- Polarization by Reflection
- Polarization by Refraction
- Polarization by Scattering

Polarization by Use of a Polaroid Filter

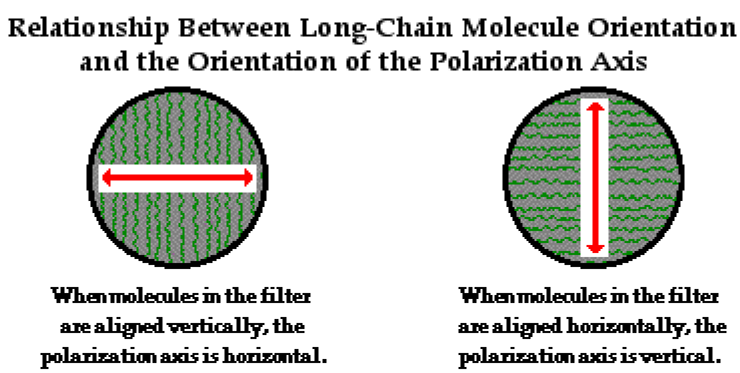
The most common method of polarization involves the use of a **Polaroid filter**. Polaroid filters are made of a special material that is capable of blocking one of the two planes of vibration of an electromagnetic wave. (Remember, the notion of two planes or directions of vibration is merely a simplification that helps us to visualize the wavelike nature of the electromagnetic wave.) In this sense, a Polaroid serves as a device that filters out one-half of the vibrations upon transmission of the light through the filter.

When unpolarized light is transmitted through a Polaroid filter, it emerges with one-half the intensity and with vibrations in a single plane; it emerges as polarized light.

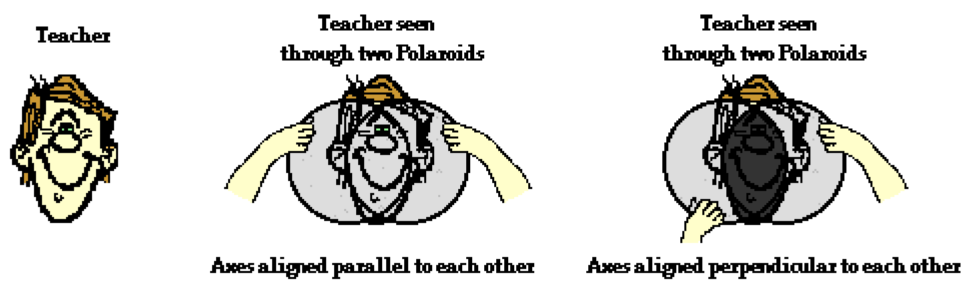


A Polaroid filter is able to polarize light because of the chemical composition of the filter material. The filter can be thought of as having long-chain molecules that are aligned within the filter in the same direction. During the fabrication of the filter, the long-chain molecules are stretched across the filter so that each molecule is (as much as possible) aligned in say the vertical direction. As unpolarized light strikes the filter, the portion of the waves vibrating in the vertical direction are absorbed by the filter. The general rule is that the electromagnetic vibrations that are in a direction parallel to the alignment of the molecules are absorbed.

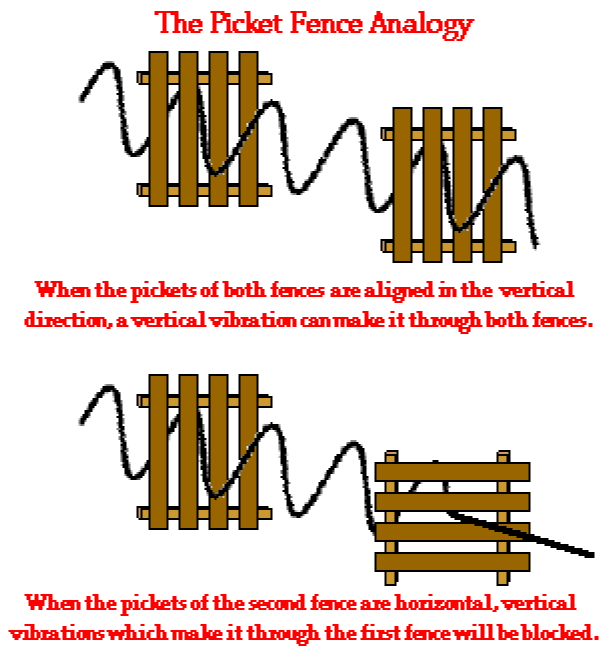
The alignment of these molecules gives the filter a **polarization axis**. This polarization axis extends across the length of the filter and only allows vibrations of the electromagnetic wave that are parallel to the axis to pass through. Any vibrations that are perpendicular to the polarization axis are blocked by the filter. Thus, a Polaroid filter with its long-chain molecules aligned horizontally will have a polarization axis aligned vertically. Such a filter will block all horizontal vibrations and allow the vertical vibrations to be transmitted (see diagram above). On the other hand, a Polaroid filter with its long-chain molecules aligned vertically will have a polarization axis aligned horizontally; this filter will block all vertical vibrations and allow the horizontal vibrations to be transmitted.



Polarization of light by use of a Polaroid filter is often demonstrated in a Physics class through a variety of demonstrations. Filters are used to look through and view objects. The filter does not distort the shape or dimensions of the object; it merely serves to produce a dimmer image of the object since one-half of the light is blocked as it passed through the filter. A pair of filters is often placed back to back in order to view objects looking through two filters. By slowly rotating the second filter, an orientation can be found in which all the light from an object is blocked and the object can no longer be seen when viewed through two filters. What happened? In this demonstration, the light was polarized upon passage through the first filter; perhaps only vertical vibrations were able to pass through. These vertical vibrations were then blocked by the second filter since its polarization filter is aligned in a horizontal direction. While you are unable to see the axes on the filter, you will know when the axes are aligned perpendicular to each other because with this orientation, all light is blocked. So by use of two filters, one can completely block all of the light that is incident upon the set; this will only occur if the polarization axes are rotated such that they are perpendicular to each other.



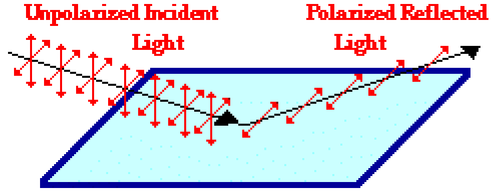
A picket-fence analogy is often used to explain how this dual-filter demonstration works. A picket fence can act as a polarizer by transforming an unpolarized wave in a rope into a wave that vibrates in a single plane. The spaces between the pickets of the fence will allow vibrations that are parallel to the spacings to pass through while blocking any vibrations that are perpendicular to the spacings. Obviously, a vertical vibration would not have the room to make it through a horizontal spacing. If two picket fences are oriented such that the pickets are both aligned vertically, then vertical vibrations will pass through both fences. On the other hand, if the pickets of the second fence are aligned horizontally, then the vertical vibrations that pass through the first fence will be blocked by the second fence. This is depicted in the diagram below.



In the same manner, two Polaroid filters oriented with their polarization axes perpendicular to each other will block all the light. Now that's a pretty cool observation that could never be explained by a particle view of light.

Polarization by Reflection

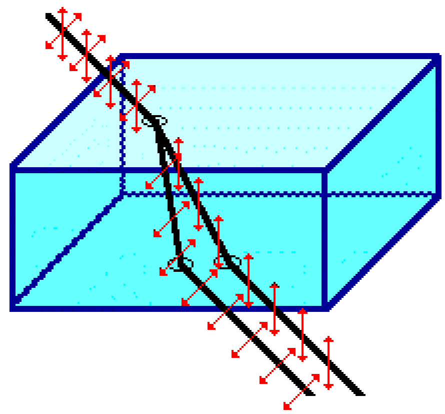
Unpolarized light can also undergo polarization by reflection off of nonmetallic surfaces. The extent to which polarization occurs is dependent upon the angle at which the light approaches the surface and upon the material that the surface is made of. Metallic surfaces reflect light with a variety of vibrational directions; such reflected light is unpolarized. However, nonmetallic surfaces such as asphalt roadways, snowfields and water reflect light such that there is a large concentration of vibrations in a plane parallel to the reflecting surface. A person viewing objects by means of light reflected off of nonmetallic surfaces will often perceive a glare if the extent of polarization is large. Fishermen are familiar with this glare since it prevents them from seeing fish that lie below the water. Light reflected off a lake is partially polarized in a direction parallel to the water's surface. Fishermen know that the use of glare-reducing sunglasses with the proper polarization axis allows for the blocking of this partially polarized light. By blocking the plane-polarized light, the glare is reduced and the fisherman can more easily see fish located under the water.



Reflection of light off of non-metallic surfaces results in some degree of polarization parallel to the surface.

Polarization by Refraction

Polarization can also occur by the refraction of light. Refraction occurs when a beam of light passes from one material into another material. At the surface of the two materials, the path of the beam changes its direction. The refracted beam acquires some degree of polarization. Most often, the polarization occurs in a plane perpendicular to the surface. The polarization of refracted light is often demonstrated in a Physics class using a unique crystal that serves as a double-refracting crystal. Iceland Spar, a rather rare form of the mineral calcite, refracts incident light into two different paths. The light is *split* into two beams upon entering the crystal. Subsequently, if an object is viewed by looking through an Iceland Spar crystal, two images will be seen. The two images are the result of the double refraction of light. Both refracted light beams are polarized - one in a direction parallel to the surface and the other in a direction perpendicular to the surface. Since these two refracted rays are polarized with a perpendicular orientation, a polarizing filter can be used to completely block one of the images. If the polarization axis of the filter is aligned perpendicular to the plane of polarized light, the light is completely blocked by the filter; meanwhile the second image is as bright as can be. And if the filter is then turned 90-degrees in either direction, the second image reappears and the first image disappears. Now that's pretty neat observation that could never be observed if light did not exhibit any wavelike behavior.



The two refracted rays passing through the Iceland Spar crystal are polarized with perpendicular orientations.

Polarization by Scattering

Polarization also occurs when light is scattered while traveling through a medium. When light strikes the atoms of a material, it will often set the electrons of those atoms into vibration. The vibrating electrons then produce their own electromagnetic wave that is radiated outward in all directions. This newly generated wave strikes neighboring atoms, forcing their electrons into vibrations at the same original frequency. These vibrating electrons produce another electromagnetic wave that is once more radiated outward in all directions. This absorption and reemission of light waves causes the light to be scattered about the medium. (This process of scattering contributes to the blueness of our skies, a topic to be discussed later.) This scattered light is partially polarized. Polarization by scattering is observed as light passes through our atmosphere. The scattered light often produces a glare in the skies. Photographers know that this partial polarization of scattered light leads to photographs characterized by a *washed-out* sky. The problem can easily be corrected by the use of a Polaroid filter. As the filter is rotated, the partially polarized light is blocked and the glare is reduced. The photographic secret of capturing a vivid blue sky as the backdrop of a beautiful foreground lies in the physics of polarization and Polaroid filters.

Applications of Polarization

Polarization has a wealth of other applications besides their use in glare-reducing sunglasses. In industry, Polaroid filters are used to perform stress analysis tests on transparent plastics. As light passes through a plastic, each color of visible light is polarized with its own orientation. If such a plastic is placed between two polarizing plates, a colorful pattern is revealed. As the top plate is turned, the color pattern changes as new colors become blocked and the formerly blocked colors are transmitted. A common Physics demonstration involves placing a plastic protractor between two Polaroid plates and placing them on top of an overhead projector. It is known that structural stress in plastic is signified at locations where there is a large concentration of colored bands. This location of stress is usually the location where structural failure will most likely occur. Perhaps you wish that a more careful stress analysis were performed on the plastic case of the CD that you recently purchased.

Polarization is also used in the entertainment industry to produce and show 3-D movies. Three-dimensional movies are actually two movies being shown at the same time through two projectors. The two movies are filmed from two slightly different camera locations. Each individual movie is then projected from different sides of the audience onto a metal screen. The movies are projected through a polarizing filter. The polarizing filter used for the projector on the left may have its polarization axis aligned horizontally while the polarizing filter used for the projector on the right would have its polarization axis aligned vertically. Consequently, there are two slightly different movies being projected onto a screen. Each movie is cast by light that is polarized with an orientation perpendicular to the other movie. The audience then wears glasses that have two Polaroid filters. Each filter has a different polarization axis - one is horizontal and the other is vertical. The result of this arrangement of projectors and filters is that the left eye sees the movie that is projected from the right projector while the right eye sees the movie that is projected from the left projector. This gives the viewer a perception of depth.

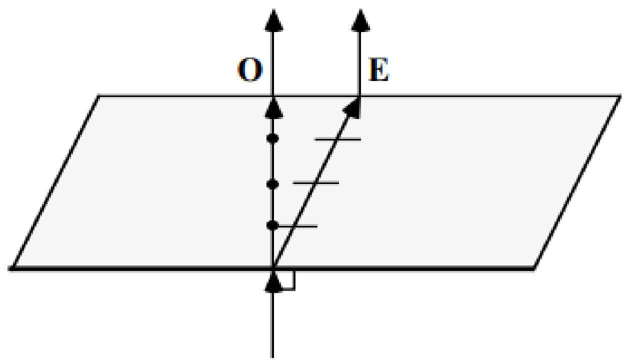
Our model of the polarization of light provides some substantial support for the wavelike nature of light. It would be extremely difficult to explain polarization phenomenon using a particle view of light. Polarization would only occur with a transverse wave. For this reason, polarization is one more reason why scientists believe that light exhibits wavelike behavior.

Double Refraction, Polarized Light

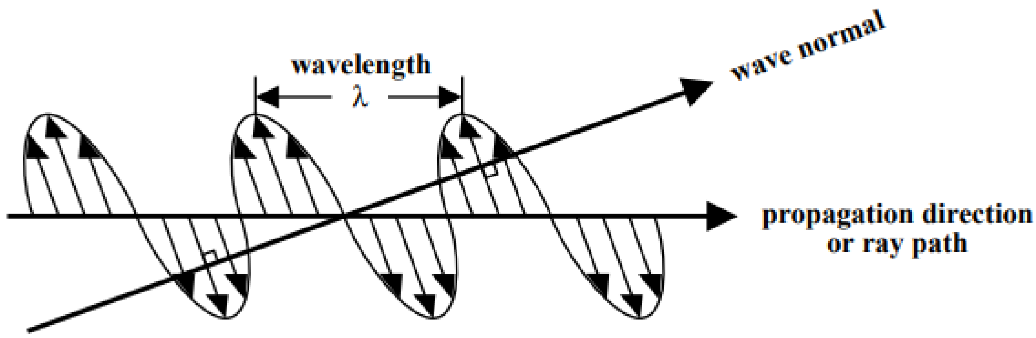
- Experiment: observations with optical calcite.

- Light passing through a calcite crystal is split into two rays. This process, first reported by Erasmus Bartholinus in 1669, is called **double refraction**. The two rays of light are each **plane polarized** by the calcite such that the planes of polarization are mutually perpendicular. For **normal incidence** (a Snell's law angle of 0°), the two planes of polarization are also perpendicular to the plane of incidence.

- For **normal incidence** (a 0° angle of incidence), Snell's law predicts that the angle of refraction will be 0°. In the case of double refraction of a normally incident ray of light, at least one of the two rays must violate Snell's Law as we know it. For calcite, one of the two rays does indeed obey Snell's Law; this ray is called the **ordinary ray** (or **O**-ray). The other ray (and any ray that does not obey Snell's Law) is an **extraordinary ray** (or **E**-ray).

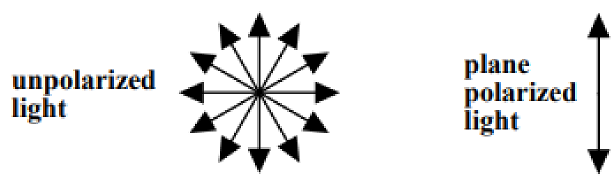


- For ordinary rays the vibration direction, indicated by the electric vectors in our illustrations, is perpendicular to the ray path. For extraordinary rays, the vibration direction is **not** perpendicular to the ray path. The direction perpendicular to the vibration direction is called the **wave normal**. Although Snell's Law is not satisfied by the ray path for extraordinary rays, it is satisfied by the wave normals of extraordinary rays. In other words, the wave normal direction for the refracted ray is related to the wave normal direction for the incident ray by Snell's Law.

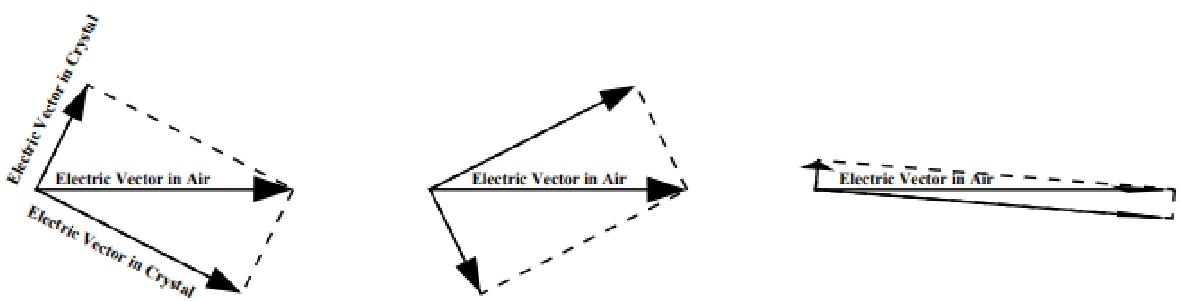


- All transparent crystals except those in the cubic system have the property of double refraction. For most crystals the image separation is not large enough to be visible. However, we will observe other optical properties that result from the double refraction. For hexagonal and tetragonal crystals, there will be one **O**-ray and one **E**-ray. For orthorhombic, monoclinic, and triclinic crystals, there will be two **E**-rays. In general, the refractive indices for non-cubic crystals depend on vibration direction. Non-cubic crystals are, therefore, said to be optically **anisotropic**. In most cases the refractive indices for the two rays produced by double refraction are not the same. One of the two rays will have a higher refractive index (and a lower velocity); this ray is called the **slow ray**. The other ray is then the **fast ray**.

- Ordinary light is not polarized. Looking along a ray of light, the electric vectors make all angles with the vertical. Light that is plane polarized in the vertical plane has only vertical electric vectors. The plane of polarization is the plane that includes both the vibration direction and the ray path. Light may be polarized by crystals, by polarizing filters, and by reflection. Reflected light is partially polarized, favoring the vibration direction perpendicular to the plane of the ray path (including both the incident and reflected rays).

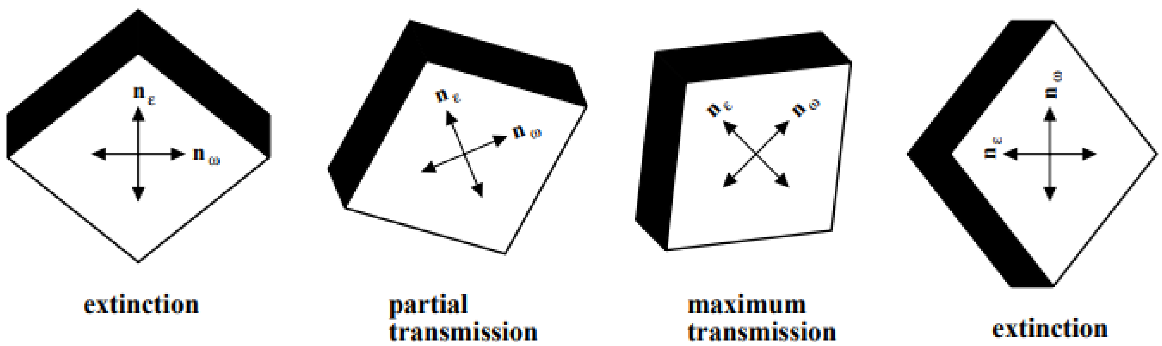


- Polarizing filters **exclude** all light not vibrating in the preferred direction of the filter. Polarizing sunglasses, by orienting their polarizing material vertically, selectively exclude the polarized portion of light reflected by horizontal surfaces. Transparent crystals do not exclude light, whatever its plane of polarization. Transparent anisotropic crystals **resolve** the electric vectors of incident light into two perpendicular electric vectors by the process of double refraction. Upon emergence from the crystal, the two rays add together again according to the rules of **vector addition**. However, because the two rays have not traveled through the crystal with the same velocity, the combined emerging ray will not be identical to the incident ray.

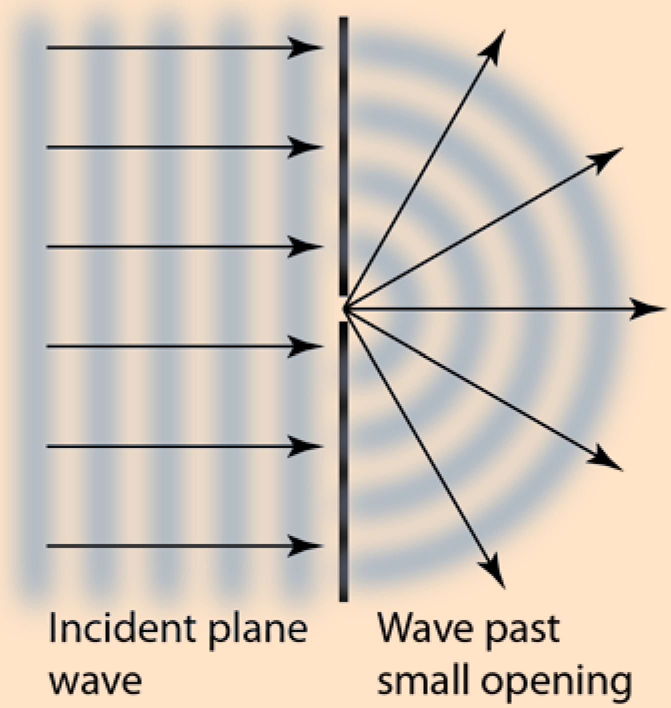


- Polarizing microscopes are equipped with polarizing filters both below and above the stage of the microscope. The lower filter (fixed, but rotatable) is called the **polarizer** and on our microscopes has its direction of polarization oriented E-W when viewed from above. (Beware, some older microscopes have their polarizing filters oriented N-S!) The upper filter (removable, but not rotatable) is called the **analyzer** and has its direction of polarization oriented N-S when viewed from above. Because the polariztion directions of these two filters are perpendicular, all of the light passing through the polarizer will be blocked by the analyzer, unless the analyzer is removed or an anisotropic crystal is placed between the two filters.

- Note that if the incident light is already polarized, special orientations exist for which **all** of the incident light is resolved along **one** of the two preferred vibration directions in the crystal. In this special case, there is no double refraction and the emerging ray **will** be indentical to the incident ray. If a crystal is oriented so that one of its two vibration directions is "parallel to the polarizer" (that is parallel to the plane of polarization of the polarizing filter), then all light emerging from the crystal would be polaized E-W and would be blocked by the analyzer. When this happens, the crystal appears dark and is said to be "at **extinction**."

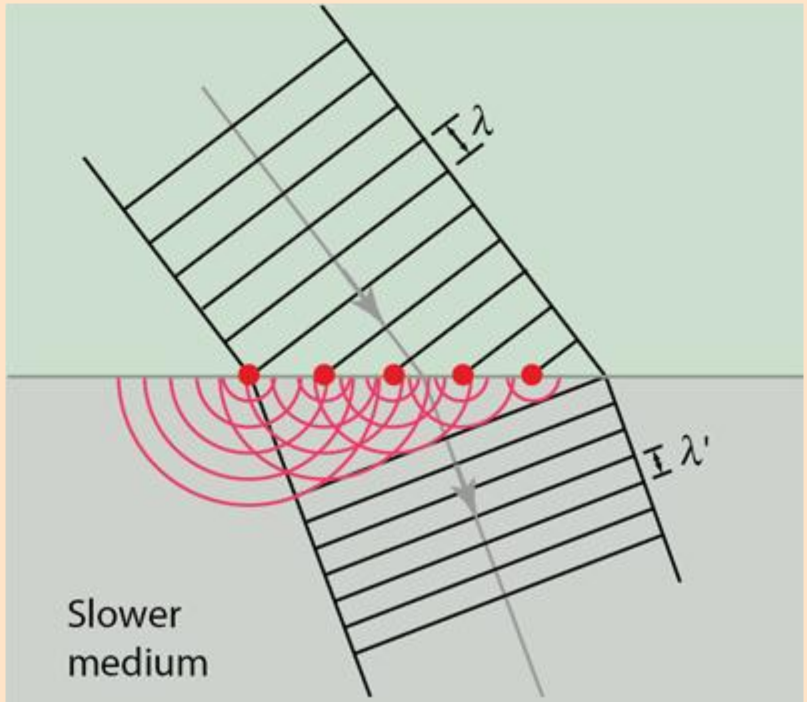


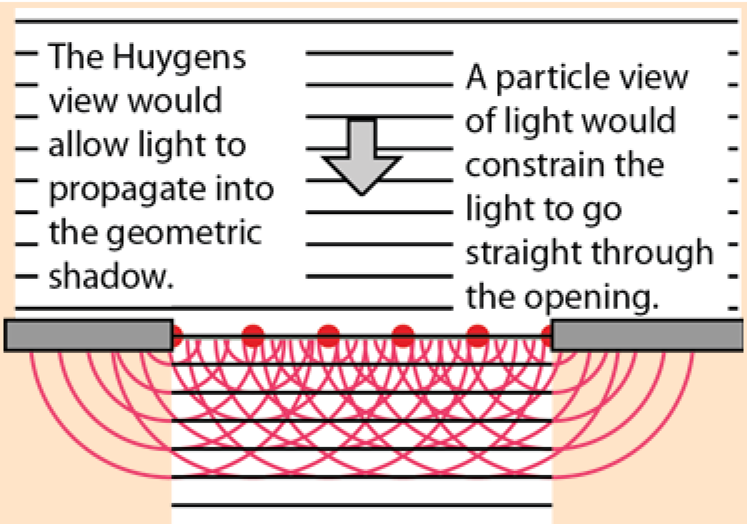
Huygens' Principle



In 1678 Huygens proposed a model where each point on a wavefront may be regarded as a source of waves expanding from that point. The expanding waves may be demonstrated in a ripple tank by sending plane waves toward a barrier with a small opening. If waves approaching a beach strike a barrier with a small opening, the waves may be seen to expand from the opening.

Huygens' principle provides a convenient way to visualize refraction. If points on the wavefront at the boundary of a different medium serve as sources for the propagating light, one can see why the direction of the light propagation changes.





The Huygens' principle view permitted a visualization of how light could penetrate into the geometric shadow in a way that the particle view could not.

Though helpful in establishing a wave view rather than a particle view of light for ordinary optics, Huygens' principle left a number of unanswered questions. For example, with its view of each point on a wavefront as a source, it gave no explanation of why it didn't propagate backward as well as forward. Miller and Fresnel further developed the theory of light propagation including diffraction. The theory of light propagation was made more rigorous by Kirchhoff.

Crystal Optics

Outline

1 Homogeneous, Anisotropic Media

2 Crystals

3 Plane Waves in Anisotropic Media

4 Wave Propagation in Uniaxial Media 5 Reflection and Transmission at Interfaces

Homogeneous, Anisotropic Media

Introduction

- material equations for homogeneous anisotropic media

$$\vec{D} = \epsilon \vec{E}$$
$$\vec{B} = \mu \vec{H}$$

- tensors of rank 2, written as 3 by 3 matrices
  - ϵ: *dielectric tensor*
  - μ: *magnetic permeability tensor*
- examples:
  - crystals, liquid crystals
  - external electric, magnetic fields acting on isotropic materials (glass, fluids, gas)
  - anisotropic mechanical forces acting on isotropic materials

Properties of Dielectric Tensor

- Maxwell equations imply symmetric dielectric tensor

$$\epsilon = \epsilon^T = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{12} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{13} & \epsilon_{23} & \epsilon_{33} \end{pmatrix}$$

- symmetric tensor of rank 2 $\Rightarrow$ coordinate system exists where tensor is diagonal
- orthogonal axes of this coordinate system: *principal axes*
- elements of diagonal tensor: *principal dielectric constants*
- 3 *principal indices of refraction* in coordinate system spanned by principal axes

$$\vec{D} = \begin{pmatrix} n_x^2 & 0 & 0 \\ 0 & n_y^2 & 0 \\ 0 & 0 & n_z^2 \end{pmatrix} \vec{E}$$

- x, y, z because principal axes form Cartesian coordinate system

Uniaxial Materials

- isotropic materials: $n_x = n_y = n_z$ for any coordinate system
- anisotropic materials: $n_x \neq n_y \neq n_z$
- *uniaxial materials*: $n_x = n_y \neq n_z$
- *ordinary index of refraction*: $n_o = n_x = n_y$
- *extraordinary index of refraction*: $n_e = n_z$
- rotation of coordinate system around z does not change anything
- most materials used in polarimetry are (almost) uniaxial

Crystals

Crystal Axes Terminology

- *optic axis* is the axis that has a different index of refraction
- also called *c* or *crystallographic axis*
- *fast axis*: axis with smallest index of refraction
- ray of light going through uniaxial crystal is (generally) split into two rays
- *ordinary ray* (*o-ray*) passes the crystal without any deviation
- *extraordinary ray* (*e-ray*) is deviated at air-crystal interface
- two emerging rays have orthogonal polarization states
- common to use indices of refraction for ordinary ray (n_o) and extraordinary ray (n_e) instead of indices of refraction in crystal coordinate system
- $n_e < n_o$: *negative* uniaxial crystal
- $n_e > n_o$: *positive* uniaxial crystal

Plane Waves in Anisotropic Media

Displacement and Electric Field Vectors

- plane-wave ansatz for $\vec{D}$, $\vec{E}$, $\vec{H}$

$$\vec{E} = \vec{E}_0 e^{i\left(\vec{k}\cdot\vec{x}-\omega t\right)}$$
$$\vec{D} = \vec{D}_0 e^{i\left(\vec{k}\cdot\vec{x}-\omega t\right)}$$
$$\vec{H} = \vec{H}_0 e^{i\left(\vec{k}\cdot\vec{x}-\omega t\right)}$$

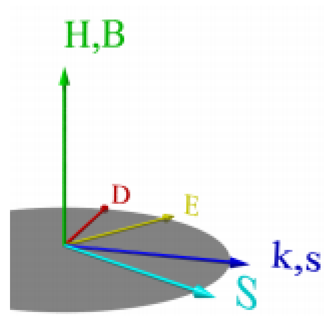
- no net charges in medium ($\nabla \cdot \vec{D} = 0$)

$$\vec{D} \cdot \vec{k} = 0$$

  $\vec{D}$ perpendicular to $\vec{k}$
- $\vec{D}$ and $\vec{E}$ not parallel $\Rightarrow$ $\vec{E}$ not perpendicular to $\vec{k}$
- wave normal, energy flow not in same direction, same speed

Magnetic Field

- constant, scalar μ, vanishing current density $\Rightarrow$ $\vec{H} \parallel \vec{B}$
- $\nabla \cdot \vec{H} = 0 \Rightarrow \vec{H} \perp \vec{k}$
- $\nabla \times \vec{H} = \frac{1}{c}\frac{\partial \vec{D}}{\partial t} \Rightarrow \vec{H} \perp \vec{D}$
- $\nabla \times \vec{E} = -\frac{\mu}{c}\frac{\partial \vec{H}}{\partial t} \Rightarrow \vec{H} \perp \vec{E}$
- $\vec{D}$, $\vec{E}$, and $\vec{k}$ all in one plane
- $\vec{H}$, $\vec{B}$ perpendicular to that plane
- Poynting vector $\vec{S} = \frac{c}{4\pi}\vec{E} \times \vec{H}$ perpendicular to $\vec{E}$ and $\vec{H}$ $\Rightarrow$ $\vec{S}$ (in general) not parallel to $\vec{k}$
- energy (in general) not transported in direction of wave vector $\vec{k}$



Relation between $\vec{D}$ and $\vec{E}$

- combine Maxwell, material equations in principal coordinate system

$$\mu D_i = \mu\epsilon_i \vec{E}_i = n^2 \left(\vec{E}_i - s_i \left(\vec{E} \cdot \vec{s} \right) \right) \quad i = 1 \cdots 3$$

- $\vec{s} = \vec{k}/|\vec{k}|$: unit vector in direction of wave vector $\vec{k}$
- n: refractive index associated with direction $\vec{s}$, i.e. $n = n(\vec{s})$
- 3 equations for 3 unknowns $\vec{E}_i$
- eliminate $\vec{E}$ assuming $\vec{E} \neq \vec{0} \Rightarrow$ *Fresnel equation*

$$\frac{s_x^2}{n^2 - \mu\epsilon_x} + \frac{s_y^2}{n^2 - \mu\epsilon_y} + \frac{s_z^2}{n^2 - \mu\epsilon_z} = \frac{1}{n^2} ,$$

- with $n_i^2 = \mu\epsilon_i$

$$s_x^2 n_x^2 \left(n^2 - n_y^2\right)\left(n^2 - n_z^2\right) + s_y^2 n_y^2 \left(n^2 - n_x^2\right)\left(n^2 - n_z^2\right) + s_z^2 n_z^2 \left(n^2 - n_x^2\right)\left(n^2 - n_y^2\right) = 0$$

Electric Field in Anisotropic Material

- with arbitrary constant a, electric field vector given by

$$\vec{E} = a \begin{pmatrix} \frac{s_x}{n^2 - n_x^2} \\ \frac{s_y}{n^2 - n_y^2} \\ \frac{s_z}{n^2 - n_z^2} \end{pmatrix}$$

- quadratic equation in $n \Rightarrow$ generally two solutions for given direction $\vec{s}$
- electric field can also be written as

$$\vec{E}_k = \frac{n^2 \vec{s}_k \left(\vec{E} \cdot \vec{s} \right)}{n^2 - \mu\epsilon_k}$$

- system of 3 equations can be solved for E_k
- denominator vanishes if $\vec{k}$ parallel to a principal axis $\Rightarrow$ treat separately

Non-Absorbing, Non-Active, Anisotropic Materials

- $\vec{k}$ not parallel to a principal axis $\Rightarrow$ ratio of 2 electric field components k and l

$$\frac{E_k}{E_l} = \frac{s_k\left(n^2 - \mu\epsilon_l\right)}{s_l\left(n^2 - \mu\epsilon_k\right)}$$

- ratio is independent of electric field components
- n^2 and ϵ_i real $\Rightarrow$ ratios are real $\Rightarrow$ electric field is linearly polarized
- in non-absorbing, non-active, anisotropic material, 2 waves propagate that have different linear polarization states and different directions of energy flows
- direction of vibration of $\vec{D}$ corresponding to 2 solutions are orthogonal to each other (without proof)

$$\vec{D}_1 \cdot \vec{D}_2 = 0$$

Wave Propagation in Uniaxial Media

Introduction

- uniaxial media $\Rightarrow$ dielectric constants:

$$\epsilon_x = \epsilon_y = n_o^2$$
$$\epsilon_z = n_e^2$$

- second form of Fresnel equation reduces to

$$\left(n^2 - n_o^2\right)\left[n_o^2\left(s_x^2 + s_y^2\right)\left(n^2 - n_e^2\right) + s_z^2 n_e^2\left(n^2 - n_o^2\right)\right] = 0$$

- two solutions n_1, n_2 given by

$$n_1^2 = n_o^2$$
$$\frac{1}{n_2^2} = \frac{s_x^2 + s_y^2}{n_e^2} + \frac{s_z^2}{n_o^2}$$

Propagation in General Direction

- (unit) wave vector direction in spherical coordinates

$$\vec{s} = \begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}$$

  - θ: angle between wave vector and optic axis
  - ϕ: azimuth angle in plane perpendicular to optic axis
-

$$\begin{aligned} \frac{1}{n_2^2} &= \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2} \\ n_2(\theta) &= \frac{n_o n_e}{\sqrt{n_o^2 \sin^2\theta + n_e^2 \cos^2\theta}} \end{aligned}$$

- take positive root, negative value corresponds to waves propagating in opposite direction

Ordinary and Extraordinary Rays

- from before

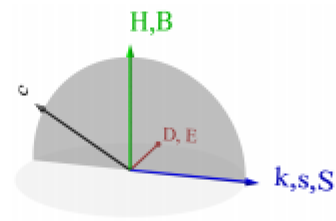
$$\begin{aligned} \frac{1}{n_2^2} &= \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2} \\ n_2(\theta) &= \frac{n_o n_e}{\sqrt{n_o^2 \sin^2\theta + n_e^2 \cos^2\theta}} \end{aligned}$$

- n_2 varies between n_o for $\theta = 0$ and n_e for $\theta = 90^\circ$
- first solution propagates according to ordinary index of refraction, independent of direction $\Rightarrow$ *ordinary* beam or ray
- second solution corresponds to *extraordinary* beam or ray
- index of refraction of extraordinary beam is (in general) *not* the extraordinary index of refraction

Ordinary Beam

- ordinary beam speed independent of wave vector direction
- for $\mu D_i = \mu\epsilon_i \vec{E}_i = n^2 \left(\vec{E}_i - s_i \left(\vec{E} \cdot \vec{s}\right)\right), i = 1 \cdots 3$ to hold for any direction $\vec{s}$, $\vec{E}_o \cdot \vec{s} = 0$ and $E_{o,z} = 0$
- electric field vector of ordinary beam (with real constant $a_o \neq 0$)

$$\vec{E}_o = a_o \begin{pmatrix} \sin\phi \\ -\cos\phi \\ 0 \end{pmatrix}$$



- ordinary beam is linearly polarized
- $\vec{E}_o$ perpendicular to plane formed by wave vector $\vec{k}$ and c-axis
- displacement vector $\vec{D}_o = n_o \vec{E}_o \parallel \vec{E}_o$
- Poynting vector $\vec{S}_o \parallel \vec{k}$

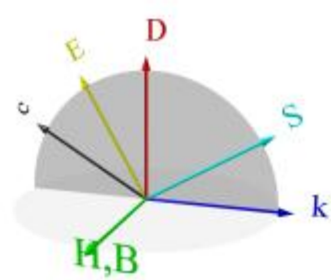
Extraordinary Ray

- since $\vec{D}_e \cdot \vec{k} = 0$ and $\vec{D}_e \cdot \vec{D}_o = 0 \Rightarrow$ unique solution (up to real constant a_e)

$$\vec{D}_e = a_e \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{pmatrix}$$

- since $E_e \cdot D_o = 0$, $D_e = \epsilon \vec{E}_e$

$$\vec{E}_e = a \begin{pmatrix} n_e^2 \cos\theta\cos\phi \\ n_e^2 \cos\theta\sin\phi \\ -n_o^2 \sin\theta \end{pmatrix}$$



- uniaxial medium $\Rightarrow \vec{E}_o \cdot \vec{E}_e = 0$
- however, $\vec{E}_e \cdot \vec{k} \neq 0$

Dispersion Angle

- angle between $\vec{k}$ and Poynting vector $\vec{S}$ = angle between $\vec{E}$ and $\vec{D}$ = *dispersion angle*

$$\tan\alpha = \frac{\left|\vec{E}_e \times \vec{D}_e\right|}{\vec{E}_e \cdot \vec{D}_e} = \frac{(n_e^2 - n_o^2)\tan\theta}{n_e^2 + n_o^2\tan^2\theta} = \frac{\sin 2\theta}{2}\frac{(n_e^2 - n_o^2)}{n_o^2\sin^2\theta + n_e^2\cos^2\theta}$$

- equivalent expression

$$\alpha = \theta - \arctan\left(\frac{n_o^2}{n_e^2}\tan\theta\right)$$

- for given $\vec{k}$ in principal axis system, α fully determines direction of energy propagation in uniaxial medium
- for θ approaching $\pi/2$, $\alpha = 0$
- for $\theta = 0$, $\alpha = 0$

Propagation Direction of Extraordinary Beam

- angle θ' between Poynting vector $\vec{S}$ and optic axis

$$\tan\theta' = \frac{n_o^2}{n_e^2}\tan\theta$$

- ordinary and extraordinary wave do (in general) not travel at the same speed
- phase difference in radians between the two waves given by

$$\frac{\omega}{c}\left(n_2(\theta)d_e - n_o d_o\right)$$

- $d_{o,e}$: geometrical distances traveled by ordinary and extraordinary rays

Propagation Along c Axis

- plane wave propagating along c-axis $\Rightarrow \theta = 0$
- ordinary and extraordinary beams propagate at same speed $\frac{c}{n_o}$
- electric field vectors are perpendicular to c-axis and only depend on azimuth ϕ
- ordinary and extraordinary rays are indistinguishable
- uniaxial medium behaves like an isotropic medium
- example: "c-cut" sapphire windows

Propagation Perpendicular to c Axis

- plane wave propagating perpendicular to c-axis $\Rightarrow \theta = \pi/2$

$$\vec{E}_o = \begin{pmatrix} \sin\phi \\ -\cos\phi \\ 0 \end{pmatrix}$$

- $\vec{E}_o$ perpendicular to plane formed by $\vec{k}$ and c-axis
- electric field vector of extraordinary wave

$$\vec{E}_e = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

- $\vec{E}_e$ parallel to c-axis
- direction of energy propagation of extraordinary wave parallel to $\vec{k}$ since $\vec{E}_e \parallel \vec{D}_e$

Phase Delay between Ordinary and Extraordinary Rays

- ordinary and extraordinary wave propagate in same direction
- ordinary ray propagates with speed $\frac{c}{n_o}$
- extraordinary beam propagates at different speed $\frac{c}{n_e}$
- $\vec{E}_o$, $\vec{E}_e$ perpendicular to each other $\Rightarrow$ plane wave with arbitrary polarization can be (coherently) decomposed into components parallel to $\vec{E}_o$ and $\vec{E}_e$
- 2 components will travel at different speeds
- (coherently) superposing 2 components after distance $d \Rightarrow$ phase difference between 2 components $\frac{\omega}{c}(n_e - n_o)d$ radians
- phase difference $\Rightarrow$ change in polarization state
- basis for constructing linear retarders

Summary: Wave Propagation in Uniaxial Media

- ordinary ray propagates like in an isotropic medium with index n_o
- extraordinary ray sees direction-dependent index of refraction

$$n_2(\theta) = \frac{n_o n_e}{\sqrt{n_o^2 \sin^2\theta + n_e^2 \cos^2\theta}}$$

n_2 direction-dependent index of refraction of the extraordinary ray
n_o ordinary index of refraction
n_e extraordinary index of refraction
θ angle between extraordinary wave vector and optic axis

- extraordinary ray is not parallel to its wave vector
- angle between the two is *dispersion angle*

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

Reflection and Transmission at Uniaxial Interfaces

General case

- from isotropic medium (n_I) into uniaxial medium (n_o, n_e)
- θ_I: angle between surface normal and $\vec{k}_I$ for incoming beam
- $\theta_{1,2}$: angles between surface normal and wave vectors of (refracted) ordinary wave $\vec{k}_1$ and extraordinary wave $\vec{k}_2$
- phase matching at interface requires

$$\vec{k}_I \cdot \vec{x} = \vec{k}_1 \cdot \vec{x} = \vec{k}_2 \cdot \vec{x}$$

- $\vec{x}$: position vector of a point on interface surface

$$n_I \sin\theta_I = n_1 \sin\theta_1 = n_2 \sin\theta_2$$

- $n_1 = n_o$: index of refraction of ordinary wave
- n_2: index of refraction of extraordinary wave

Ordinary and Extraordinary Rays

- ordinary wave $\Rightarrow$ Snell's law

$$\sin\theta_1 = \frac{n_I}{n_1} \sin\theta_I$$

- law for extraordinary ray not trivial

$$n_I \sin\theta_I = n_2\left(\theta(\theta_2)\right) \sin\theta_2$$

- (in general) θ_2 and therefore $\vec{k}_2$ will *not* determine direction of extraordinary beam since Poynting vector (in general) not parallel to wave vector
- solve for $\theta_2 \Rightarrow$ determine direction of Poynting vector
- special cases reduce complexity of equations

Extraordinary Ray Refraction for General Case

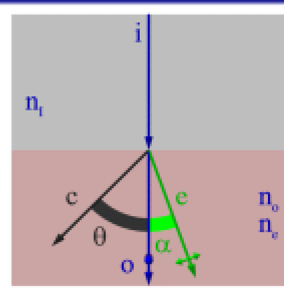
$$\cot \theta_2 = \frac{c_x c_y \left(n_o^2 - n_e^2\right) \pm n_o \sqrt{\frac{n_o^2 n_e^2 + n_e^2 c_x^2 \left(n_e^2 - n_o^2\right)}{\sin^2 \theta_I} - n_o^2 - \left(n_e^2 - n_o^2\right)\left(c_x^2 + c_y^2\right)}}{n_o^2 + c_x^2 \left(n_e^2 - n_o^2\right)}$$

propagation vector of extraordinary ray

$$\begin{aligned}
S_x &= \cos\alpha \cos\theta_2 + \frac{\sin\alpha \sin\theta_2 \left(c_x \sin\theta_2 - c_y \cos\theta_2\right)}{\sqrt{c_z^2 + \left(c_x \sin\theta_2 - c_y \cos\theta_2\right)^2}} \\
S_y &= \cos\alpha \sin\theta_2 - \frac{\sin\alpha \cos\theta_2 \left(c_x \sin\theta_2 - c_y \cos\theta_2\right)}{\sqrt{c_z^2 + \left(c_x \sin\theta_2 - c_y \cos\theta_2\right)^2}} \\
S_z &= c_z \cdot \frac{\sin\alpha}{\sqrt{c_z^2 + \left(c_x \sin\theta_2 - c_y \cos\theta_2\right)^2}}
\end{aligned}$$

- $\vec{c}$ optic axis vector $\vec{c} = (c_x, c_y, c_z)^T$
- $\vec{S}$ propagation direction of extraordinary ray $\vec{S} = (S_x, S_y, S_z)^T$
- θ_I angle between $\vec{k}_I$ and interface normal
- θ_2 angle between $\vec{k}_e$ and interface normal
- α dispersion angle

Normal Incidence



- normal incidence $\Rightarrow \theta_I = 0$, $\theta_1 = \theta_2 = 0$
- choose plane formed by surface normal and crystal axis
- both wave vectors and ordinary ray not refracted
- extraordinary ray refracted by dispersion angle α

$$\alpha = \theta - \arctan\left(\frac{n_o^2}{n_e^2} \tan\theta\right)$$

Optic Axis in Plane of Incidence and Plane of Interface

- $\theta + \theta_2 = \pi/2 \Rightarrow \cot\theta_2 = \frac{n_e}{n_o}\cot\theta_1$
- θ_1: angle between surface normal and *ordinary* ray or wave vector ($\sin\theta_I = n_o \sin\theta_1$)
- extraordinary wave sees equivalent refractive index

$$n_y = \sqrt{n_e^2 + \sin^2\theta_I\left(1 - \frac{n_e^2}{n_o^2}\right)}$$

- direction of Poynting vector

$$\begin{aligned} S_x &= \cos(\theta_2 + \alpha) \\ S_y &= \sin(\theta_2 + \alpha) \\ S_z &= 0 \end{aligned}$$

- determine dispersion angle α and *add* to θ_2 to obtain direction of extraordinary *ray*

Optic Axis Perpendicular to Plane of Incidence

- c-axis perpendicular to plane of incidence $\Rightarrow \theta = \frac{\pi}{2}$, $n_2\left(\frac{\pi}{2}\right) = n_e$

$$n_I \sin\theta_I = n_e \sin\theta_2$$

- extraordinary wave vector obeys Snell's law with index n_e
- $\theta = \frac{\pi}{2} \Rightarrow$ dispersion angle $\alpha = 0$
- Poynting vector $||$ wave vector, extraordinary beam itself obeys Snell's law with n_e
- double refraction only for non-normal incidence

Interface from Uniaxial Medium to Isotropic Medium

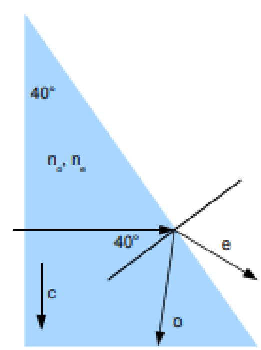
- ordinary ray follows Snell's law
- transmitted extraordinary wave vector and ray coincide
- exit of extraordinary *wave* on interface defined by extraordinary *ray*
- extraordinary wave vector follows Snell's law with index $n_2(\theta)$

$$n_I \sin\theta_E = n_2 \sin\theta_U$$

  - n_I index of isotropic medium
  - θ_E angle of wave/ray vector with surface normal in isotropic medium
  - n_2, θ_U corresponding values for extraordinary *wave vector* in uniaxial medium
- n_2 is function of θ normally already known from beam propagation in uniaxial medium
- θ_U is function of geometry of interface,
- plane-parallel slab of uniaxial medium, $\theta_E = \theta_I$, (in general) extraordinary beam displaced on exit

Total Internal Reflection (TIR)

- TIR also in anisotropic media
- $n_o \neq n_e \Rightarrow$ one beam may be totally reflected while other is transmitted
- principal of most crystal polarizers
- example: calcite prism, normal incidence, optic axis parallel to first interface, exit face inclined by 40°
- $\Rightarrow$ extraordinary ray not refracted, two rays propagate according to indices n_o,n_e
- at second interface rays (and wave vectors) at 40° to surface
- 632.8 nm: $n_o = 1.6558$, $n_e = 1.4852$
- requirement for total reflection $\frac{n_U}{n_I}\sin\theta_U > 1$
- with $n_I = 1 \Rightarrow$ extraordinary ray transmitted, ordinary ray undergoes TIR

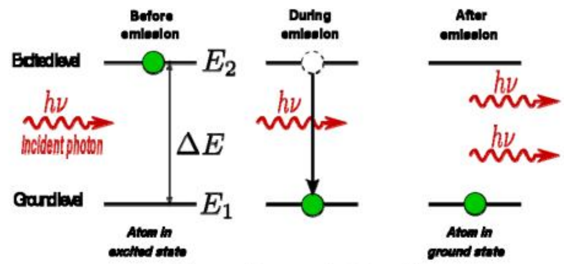


****

UNIT-5

Laser History

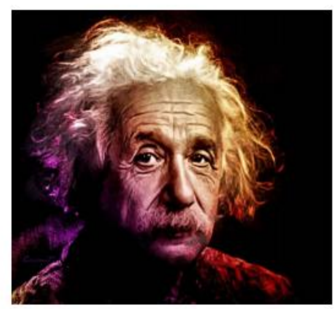
In 1916, Albert Einstein predicted the existence of stimulated emission, based on statistical physics considerations.



$$E_2 - E_1 = \Delta E = h\nu$$

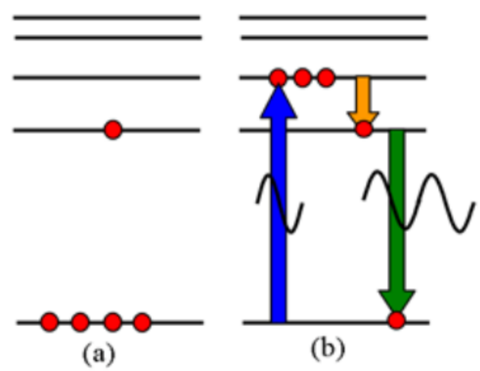
Einstein, A., "Zur Quantentheorie der Strahlung," Physikalische Gesellschaft Zürich, 18, 47-62 (1916).

He never considered the possibility of amplification.



Amplification of light through stimulated emission requires population inversion.

In equilibrium, lower energy levels are always more populated. As a result, absorption always dominates.



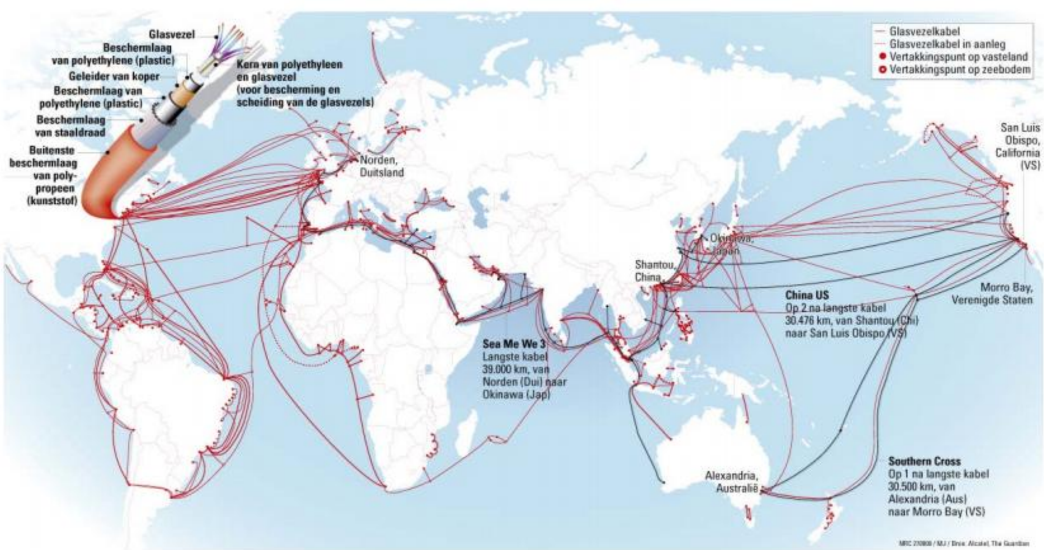
(Diagram: resourcefulphysics.org)

In order for the stimulated emission to become significant, higher energy levels should be artificially made more populated. This is called population inversion. It is achieved via «pumping».

There are many ways of pumping, including electrical and optical.

Fiber Lasers

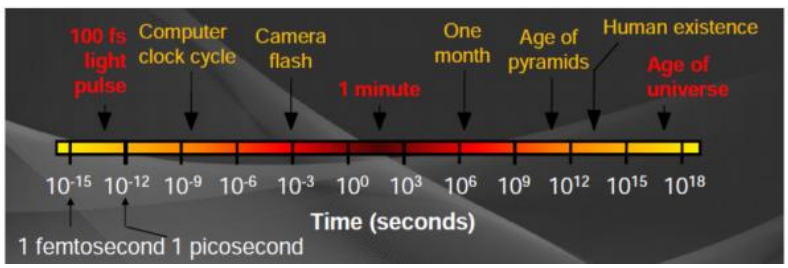
In 1964 C. J. Koester and E. Snitzer developed the first neodymium-doped fiber amplification, paving the way for fiber telecommunications. **Applied Optics, 3, 1182-1186 (1964).**



The first barcode scanner was made in 1974.

Ultrafast Laser Pulses

Today it’s routine to generate ultrashort laser pulses with durations < 100 femtosecond (10-13 s). The extreme high intensities yield many nonlinear-optical effects. The first femtosecond laser was developed in 1974 by Ippen and Shank. **Appl. Phys. Lett. 24, 373-375 (1974)**



Hologram Recording and Reconstruction Holograms are usually recorded with an optical set-up consisting of a light source (e.g. a laser), mirrors and lenses for beam guiding and a recording device (e.g. a photographic sensor). A typical set-up is shown in Fig. 2.12 [79, 121]. Light with sufficient coherence is split into two waves of reduced amplitude by a beam splitte

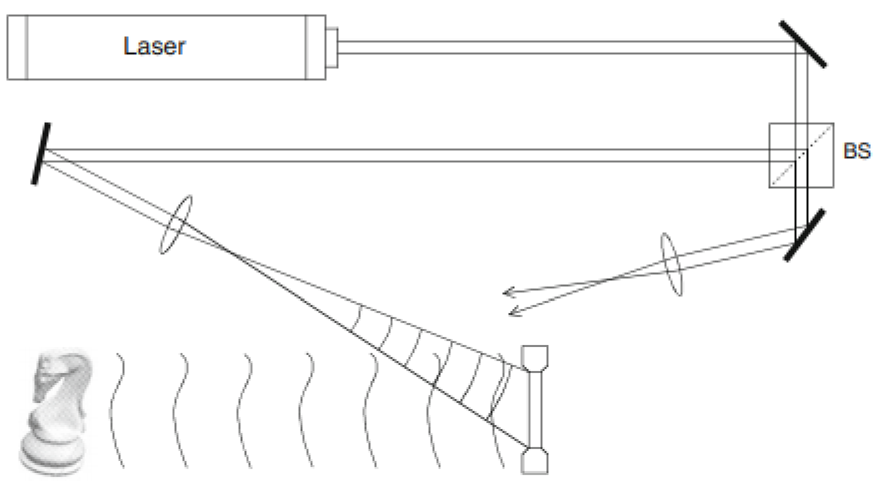


Fig. 2.12 Hologram recording

(BS). The first wave illuminates the object, is scattered at the object surface and reflected towards the recording medium. The second wave—the reference wave—directly illuminates the light sensitive medium. The waves interfere with each other to produce a characteristic interference pattern. In classical photographic holography the interference pattern is recorded on a photosensitive material such as silver halide films or plates and rendered permanent by wet chemical development of the film. In digital holography the interference pattern is recorded directly onto an electronic photosensor such as a CCD or CMOS array. The recorded interference pattern is the hologram.

The original object wave is reconstructed by illuminating the hologram with the reference wave, Fig. 2.13. An observer sees a virtual image, which is optically indistinguishable from the original object. The reconstructed image exhibits all effects of perspective, parallax and depth-of-field.

The holographic process is described mathematically using the formalism of Sect. 2.2. Across the extent of the photographic plate, the complex amplitude of the object wave is described by

$$E_O(x,y) = a_O(x,y)\exp(i\varphi_O(x,y)) \tag{2.58}$$

with real amplitude a_O and phase φ_O.

$$E_R(x,y) = a_R(x,y)\exp(i\varphi_R(x,y)) \tag{2.59}$$

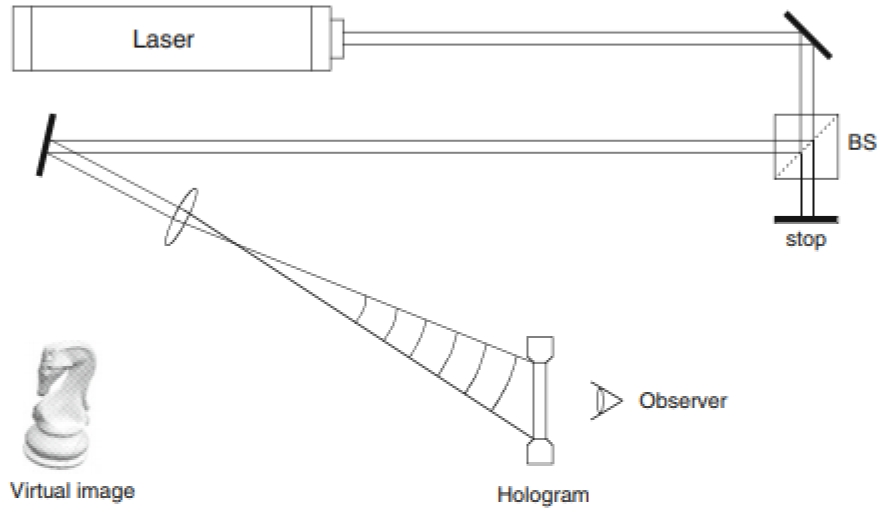


Fig. 2.13 Hologram reconstruction

is the complex amplitude of the reference wave with real amplitude a_R and phase φ_R.

Both waves interfere at the surface of the recording medium and the resultant intensity is described by

$$\begin{aligned} I(x,y) &= |E_O(x,y) + E_R(x,y)|^2 \\ &= (E_O(x,y) + E_R(x,y))(E_O(x,y) + E_R(x,y))^* \\ &= E_R(x,y)E_R^*(x,y) + E_O(x,y)E_O^*(x,y) + E_O(x,y)E_R^*(x,y) + E_R(x,y)E_O^*(x,y) \end{aligned} \tag{2.60}$$

The amplitude transmission $h(x, y)$ of the developed photographic plate (or of other recording media) is proportional to $I(x, y)$:

$$h(x,y) = h_0 + \beta\tau I(x,y) \tag{2.61}$$

The constant β is the slope of the amplitude transmittance versus exposure characteristic of the light sensitive material. For photographic emulsions β is negative. The exposure duration is denoted by τ and h_0 is the amplitude transmission of the unexposed plate; $h(x,y)$ is the hologram function. In Digital Holography using CCD or CMOS arrays as the recording medium, h_0 can be neglected.

For hologram reconstruction in classical holography, the hologram is illuminated with a replica of the original reference wave in terms of wavelength and phase. This is represented mathematically as a multiplication of the amplitude transmission of the medium with the complex amplitude of the reconstruction (reference) wave,

$$\begin{aligned} E_R(x,y)h(x,y) = {} & \left[h_0 + \beta\tau\left(a_R^2 + a_O^2\right)\right]E_R(x,y) \\ & + \beta\tau a_R^2 E_O(x,y) + \beta\tau E_R^2(x,y)E_O^*(x,y) \end{aligned} \tag{2.62}$$

The first term on the right side of this equation is the reference wave multiplied by a constant factor. It represents the non-diffracted wave passing through the hologram (zero diffraction order). The second term is the reconstructed object wave and forms the virtual image. The real factor $\beta\tau a_R^2$ only influences the brightness of the image. The third term generates a distorted real image of the object. For off-axis holography the virtual image, the real image and the non-diffracted wave are spatially separated.

The reason for the distortion of the real image is the spatially varying complex factor E_R^2, which modulates the image forming conjugate object wave E_O^*. An undistorted real image can be generated by replaying the hologram with the complex conjugate of the reference beam E_R^*. This is mathematically represented by,

$$\begin{aligned} E_R^*(x,y)h(x,y) = {} & \left[h_0 + \beta\tau\left(a_R^2 + a_O^2\right)\right]E_R^*(x,y) \\ & + \beta\tau a_R^2 E_O^*(x,y) + \beta\tau E_R^{*2}(x,y)E_O(x,y) \end{aligned} \tag{2.63}$$

What is a Phototransistor?

A **Phototransistor** is an electronic switching and current amplification component which relies on exposure to light to operate. When light falls on the junction, reverse current flows which is proportional to the luminance. Phototransistors are used extensively to detect light pulses and convert them into digital electrical signals. These are operated by light rather than electric current. Providing large amount of gain, low cost and these phototransistors might be used in numerous applications.

It is capable of converting light energy into electric energy. Phototransistors work in a similar way to photo resistors commonly known as LDR (light dependant resistor) but are able to produce both current and voltage while photo resistors are only capable of producing current due to change in resistance. Phototransistors are transistors with the base terminal exposed. Instead of sending current into the base, the photons from striking light activate the transistor. This is because a phototransistor is made of a bipolar semiconductor and focuses the energy that is passed through it. These are activated by light particles and are used in virtually all electronic devices that depend on light in some way. All silicon photo sensors (phototransistors) respond to the entire visible radiation range as well as to infrared. In fact, all diodes, transistors, Darlington's, triacs, etc. have the same basic radiation frequency response.

The **structure** of the **phototransistor** is specifically optimized for photo applications. Compared to a normal transistor, a photo transistor has a larger base and collector width and is made using diffusion or ion implantation.

Characteristics :

- Low-cost visible and near-IR photo detection.
- Available with gains from 100 to over 1500.
- Moderately fast response times.
- Available in a wide range of packages including epoxy-coated, transfer-molded and surface mounting technology.
- Electrical characteristics similar to that of signal transistors.

A **photo transistor** is nothing but an ordinary bi-poplar transistor in which the base region is exposed to the illumination. It is available in both the P-N-P and N-P-N types having different configurations like common emitter, common collector and common base. Common emitter **configuration** is generally used. It can also work while base is made open. Compared to the conventional transistor it has more base and collector areas.

Ancient photo transistors used single semiconductor materials like silicon and germanium but now a day's modern components uses materials like gallium and arsenide for high efficiency levels. The base is the lead responsible for activating the transistor. It is the gate controller device for the larger electrical supply. The collector is the positive lead and the larger electrical supply. The emitter is the negative lead and the outlet for the larger electrical supply.

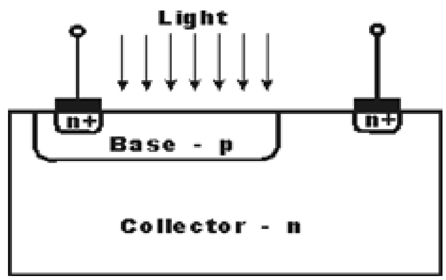


Photo Transistor Construction

****