Chapter 24

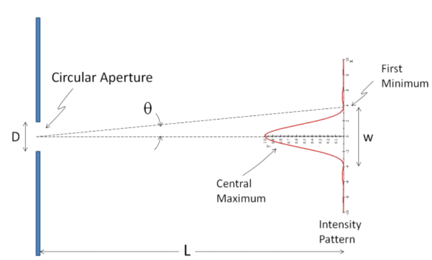
24 Resolution 3.4.5, 3.4.6, 3.4.7

Fundamental Concepts

- Circular Apertures
- Resolution is a name we give to the fundamental blurriness in geometrical optical systems due to the wave nature of light.
- Interference from circular apertures
- X-ray Diffraction
- Holography

24.1 Circular apertures

Our analysis of light going through holes has been somewhat limited by squarish holes or slits. But most optical systems, including our eyes don't have square holes. So what happens when the hole is round? The situation is as shown in the next figure.



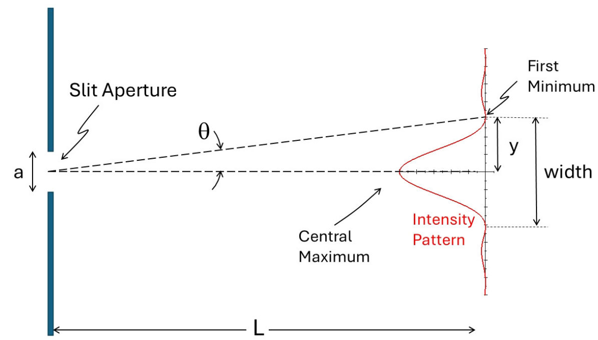
Before we discuss this situation, let's think about the width of a single slit pattern. We remember that

$$\sin(\theta) = (1)\,\frac{\lambda}{a}$$

for the first minima,or that

$$\theta \approx \frac{\lambda}{a}$$

for small angles. And from the figure



we can see that

$$\theta \approx \frac{y}{L}$$

so long as θ is small, then we find the position of the first minimum to be

$$y \approx \frac{\lambda}{a}L$$

This is the distance from the center bright spot peak intensity to the first dark spot. The width of the bright spot is twice this distance

$$w \approx 2\frac{\lambda}{a}L$$

We expect something like this for our circular aperture. The derivation is not to hard, but it involves Bessel functions, which are beyond the math requirement for this course[1]. So I will give you the answer

$$\theta \approx 1.22\frac{\lambda}{D}$$

Where we have replaced a with D, the diameter of the aperture. But they are both the size of the hole.

Let's think about this last equation. Compare it to our equation for θ for a square opening. That's right, the circular aperture (hole) only adds a factor of 1.22. And as with the slit

$$\theta \approx \frac{y}{L}$$

[1] We find the whole diffraction pattern for a circular aperture in PH375, *Optics*.

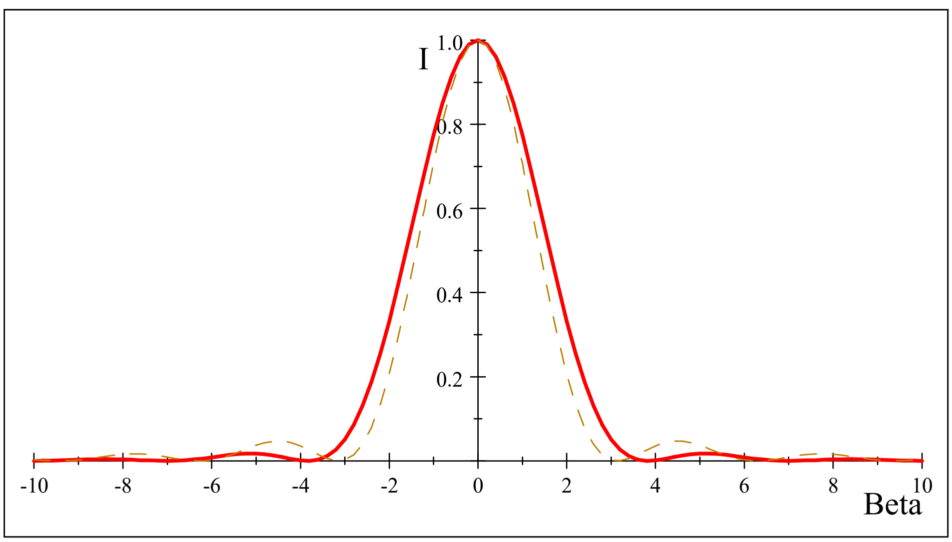
so

$$y \approx 1.22 \frac{\lambda}{D} L \tag{24.1}$$

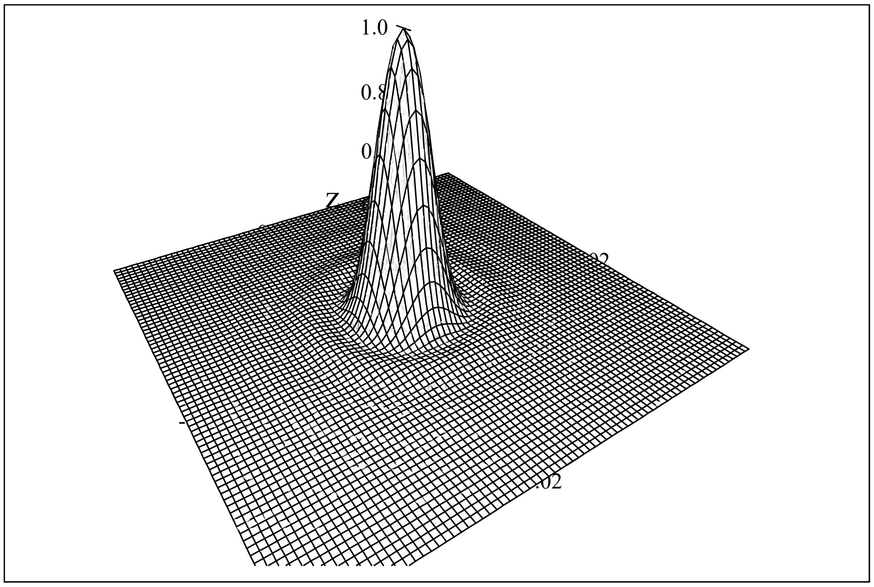
and

$$w \approx 2.44 \frac{\lambda}{D} L \tag{24.2}$$

The pattern looks a little like the slit pattern. But the secondary maxima are actually smaller for the circular aperture case. Here is a larger version in red (solid curve) with the single slit pattern in brown (dashed line).

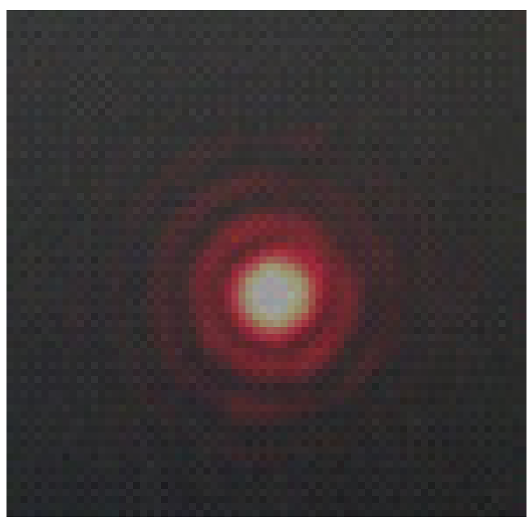


A three dimensional version of the intensity pattern from the circular aperture.



Graph of the Airy disk for the case $a = 0.05\,\mathrm{mm}$, $\lambda = 500\,\mathrm{nm}$, and $R = 1\,\mathrm{m}$.

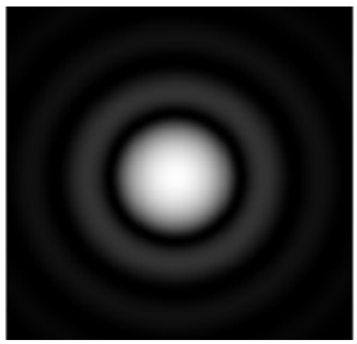
With a bright enough laser, the pattern becomes visible.



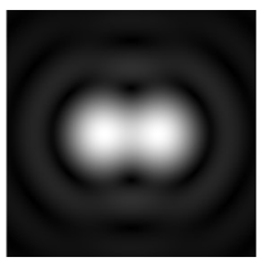
24.2 Resolution

We have emphasized that an extended object can be viewed as a collection of point objects. Then the image is formed from the collection if images of those point objects. We thought about the design of cameras and telescopes and other optical systems using ray optics. It would be great if optical systems could form images with infinite precision, but it turns out they can't. And it is the fact that light acts as a wave prevents this from being true! Our wave nature of light comes back to complicate our simple optical designs!

Because light is really a wave, the images of the point objects won't be points, themselves. And because our eyes and cameras have circular apertures, the image if the point objects will be little circular central maxima with dim concentric ring patterns.



If we have more than one point of light, we will get two such patterns.



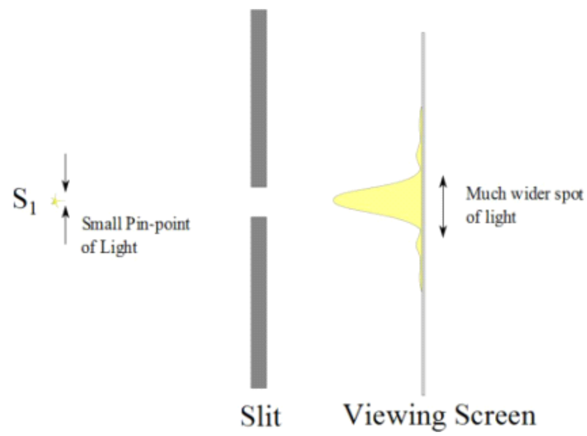
and our image of an extended object (like Aunt Sally) is made up of many, many points all reflecting light into our camera lens. We want each of those points of light to create corresponding points on our image. But instead they are creating small circles of light, and form the above figures we can see that those circles will overlap. All this makes our images fuzzy.

24.3 Resolution

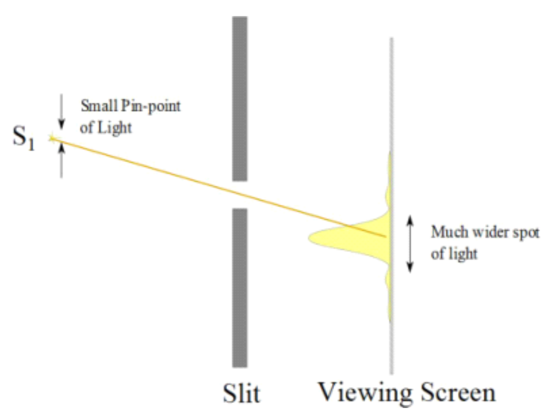
The quality of our image depends on how poorly a point object is imaged. If each point object makes a large circle of light on the screen or detector array, we get a very confusing image (it will look blurry to us). Let's review why this will happen so we can know how to minimize the effect.

Since we are using eyes or cameras, we expect circular apertures. Let's review what we know but include some ray diagrams.

We already know that if we take light and pass it through a single slit, we get an intensity pattern that has a central bright region.

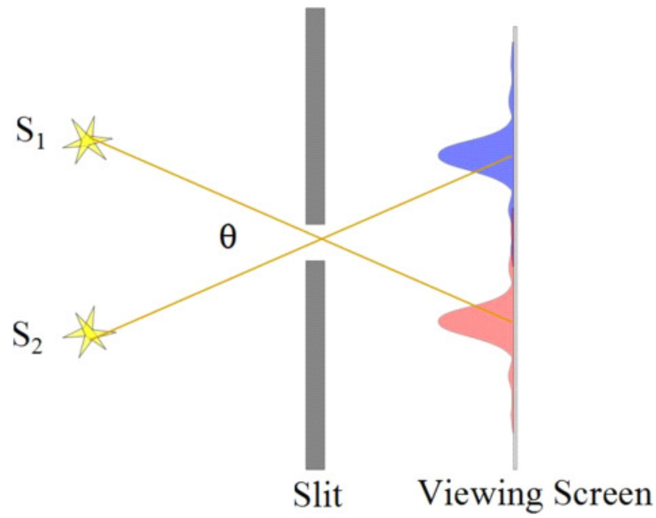


Remember that normal objects will be made up of many small points of light (either due to reflection or glowing) and each of these will form such an intensity pattern on a screen. Here is a bright point source that is not on the axis, and we see that it too makes a bright spot on the screen (and smaller bright spots or rings, depending on the shape of the aperture)

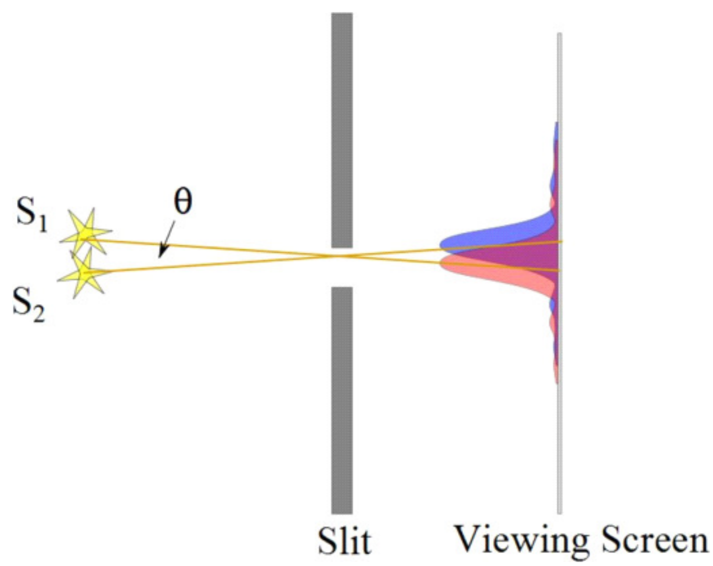


So our images will be made up of many central bright spots, each of which represents a point of light from the object. These central bright spots may overlap, (and their secondary maxima certainly will overlap).

Let's take a simple case of two points of light, S_1 and S_2. If we take a single slit and pass light from two distant point sources through the slit, we do not get two sharp images of the point sources. Instead, we get two diffraction patterns.



If these patterns are formed sufficiently far from each other, it is easy to tell they were formed from two distinct objects. Each point became a small blur, but that is really not so bad. We can still tell that the two blurs came from different sources. If our pixel size is about the same size of the blur, we may not even notice the blurriness in the digital imagery.



But if the patterns are formed close to each other, it gets hard to tell whether they were formed from two objects or one bright object. We now have a problem. Suppose you are trying to look at a star and see if it has a planet. But all you can see is a blur. You can't tell if there is one source of light or two.

Long ago an early researcher titled Lord Rayleigh developed a test to determine if you can distinguish between two diffraction patterns. When the central maximum of one point's image falls on the first minimum of anther point's image, the images are said to be just resolved. This test is known as *Rayleigh's criterion*.

We can find the required separation for a slit. Remember that

$$\sin(\theta) = m\frac{\lambda}{a} \quad m = \pm 1, \pm 2, \pm 3 \ldots \tag{24.3}$$

gives the minima (dark spots) for a single slit. We want the first minimum, so $m = 1$

$$\sin(\theta) = \frac{\lambda}{a} \tag{24.4}$$

Remember that

$$\tan\theta = \frac{y}{L}$$

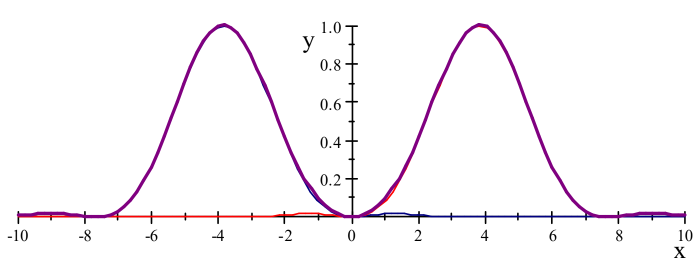
where L is the distance from the slit to the viewing screen, and y is the vertical location of the dark spot. If we place the second image maximum so it is just at this location, the two images will be just barely resolvable. In the small angle approximation, $\sin(\theta) \approx \theta$ so

$$\theta_{\min} = \frac{\lambda}{a} \tag{24.5}$$

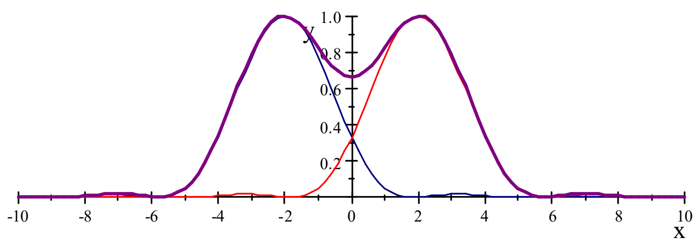
Now you may be saying to yourself that you don't often take pictures through single illuminated slits, so this is nice, but not really very interesting. But suppose, instead, that we image a circular aperture. We just pick up a factor of 1.22 and change a to D for diameter.

$$\theta_{\min} = 1.22\frac{\lambda}{D} \tag{24.6}$$

The Rayleigh criteria tells you, based on your camera aperture size, how a point source will be imaged on the film or sensor array. If we consider extended sources (like your favorite car or Aunt Sally) to be collections of many point sources, then we have a way to tell what features will be clearly resolved on the image and what features will not (like you may not be able to see the lettering on the car to tell what model it is, or you may not be able to distinguish between the gem stones in Aunt Sally's necklace because the image is too blurry to see these features clearly).

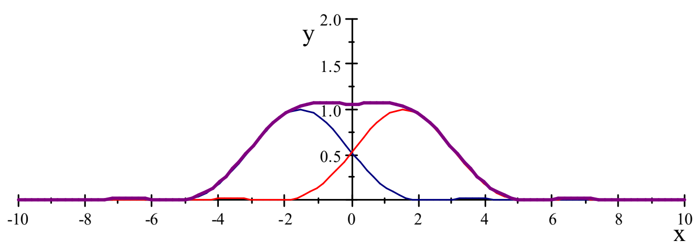


Pattern from two resolved circular slits.



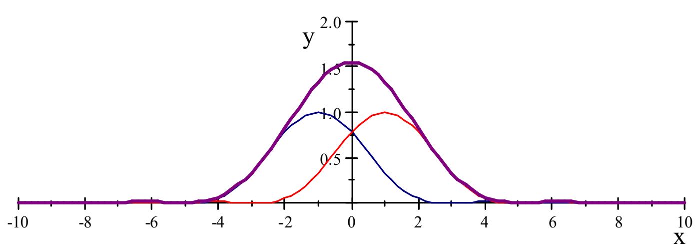
Rayleigh Criteria: Pattern from two circular soruces where the sources are close enough that the maximum from one pattern is placed on the minimum of the other. Lord Rayleigh gave this as the criteria for just barily being resolved.

Astronomers sometimes use Sparrow's criteria for two sources being resolved. It is shown below.



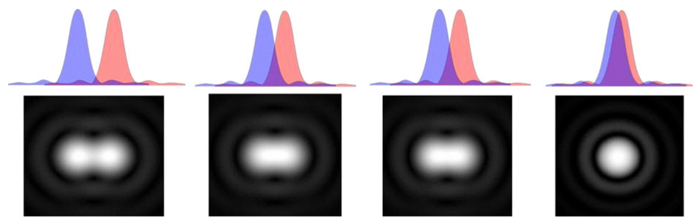
Sparrow Criteria: This is a less concervative resolution criteria than Rayleigh. When the intenisty pattern is flat on the top, there must be two sources. This criterial is used in astronomy.

Since astronomers just want to know if there are two stars or one, it is enough to see that the intensity pattern went flat at the center. That must mean that there are two stars. But if the stars are any closer, the flat center becomes a peak and the stars are unresolved.



Two circular sources unresolved

Here is what the easily resolved, Rayleigh resolved, Sparrow resolved, and unresolved cases look like.

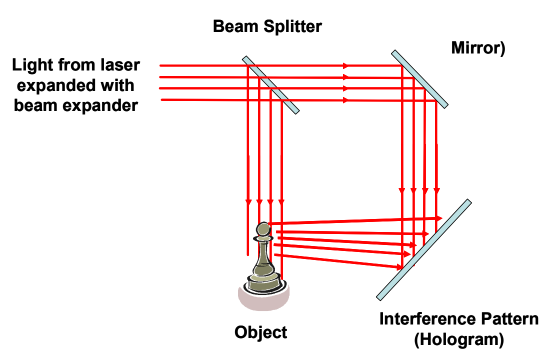


Of course, astronomers have it easy. They really image points of light. But for photos of more normal things we don't want to not be able to tell part of Aunt Sally's eye from part of her nose. Our aperture limits how good the image will look regardless of the lens that we use in the camera. Because this limit is due to diffraction patterns we call this resolution limit a *diffraction limit*. If you are designing a camera system you generally want your sensor pixel size to match the size of the central diffraction spot, because smaller pixels won't get you a less blurry image once you reach the size of the spot created by imaging a point source.

This really concludes our study of waves and optics! But let's look at two more applications of what we have learned.

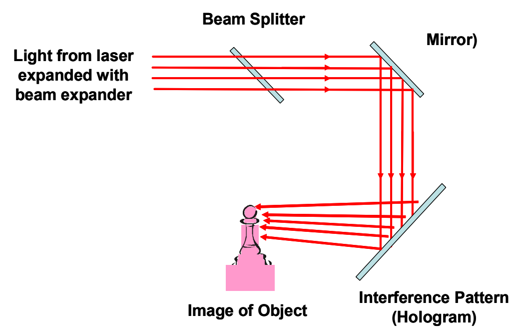
24.4 Holography

You may have seen holograms in the past. We have enough understanding of light to understand how they are generated now.



A device for generating a hologram is shown in the figure above. Light from a laser or other coherent source is expanded and split into two beams. One travels to a photographic plate, the other is directed to an object. At the object, light is scattered and the scattered light also reaches the photographic plate. The combination of the direct and scattered beams generates a complicated interference pattern.

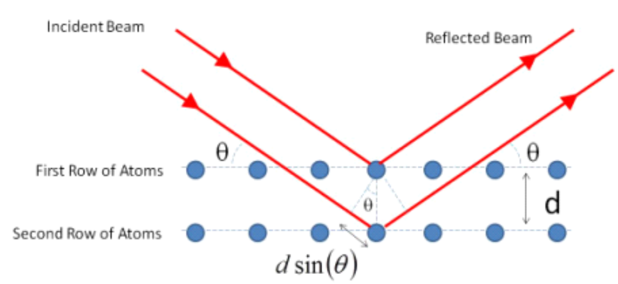
The pattern can be developed (like you develop photographic film). Once developed, it can be re-illuminated with a direct beam. The emulsion on the plate creates complicated patterns of light transmission, which combine to create interference. It is like a very complicated slit pattern or grating pattern. The result is a three-dimensional image generated by the interference. The interference pattern generates an image that looks like the original object.



24.5 Diffraction of X-rays by Crystals

If we make the wavelength of light very small, then we can deal with very small diffraction gratings. This concept is used to investigate the structure of crystals with x-rays. The crystal latus of molecules or atoms creates the regular pattern we need for a grating. The pattern is three dimensional, so the patterns are complex.

Let's start with a simple crystal with a square regular latus. $NaCl$ has such a structure.



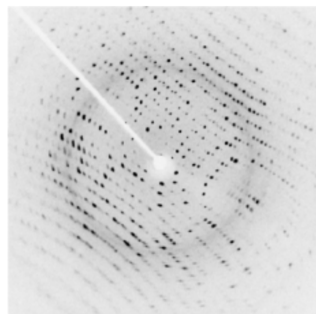
If we illuminate the crystal with x-rays, the x-rays can reflect off the top layer of atoms, or off the second layer of atoms (or off any other layer, but for now let's just consider two layers). If the spacing between the layers is d, then the path difference will be

$$\delta = 2\left(d\sin\left(\theta\right)\right) \tag{24.7}$$

then for constructive interference

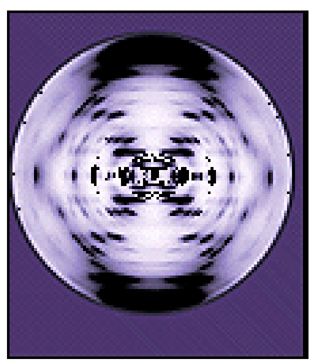
$$2d \sin(\theta) = m\lambda \qquad m = 1, 2, 3, \ldots \tag{24.8}$$

This is known as *Bragg's law.* This relationship can be used to measure the distance between the crystal planes. As the crystal structure gets more complicated, the interference pattern gets more complicated as well. Here is an example



Diffraction image of protein crystal. Hen egg lysozyme, X-ray souce Bruker IμS, $\lambda = 0.154188\,\mathrm{nm}$, $45\,\mathrm{kV}$, Exposure $10\,\mathrm{s}$. (image in the public domain)

DNA makes in interesting diffraction pattern.

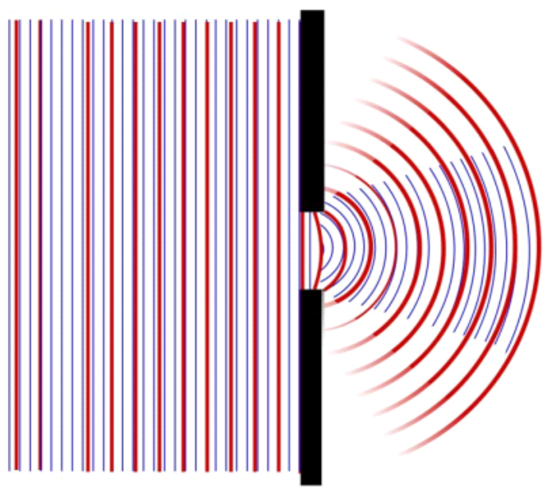


X-ray diffraction pattern of DNA (image courtesy of the National Institute of Health, image in the public domain)

24.6 Ray Model vs. Wave Model

We have studied all the new information that we will get on waves and optics. But notice that we started our study of light using rays, but knowing that light is a wave. We did most of our work with mirrors and lenses without using wave properties of light. But then we looked at apertures and found the wave nature of light was very important. When do we need to use the equations for light as a wave, and when can we get away with ignoring the wave nature of light?

In the next figure, two waves of different wavelengths go through a single opening. The wave representing the central maxima is shown in each case, but not the secondary maxima.



Notice that the smaller wavelength has a narrower central maxima as we would expect from

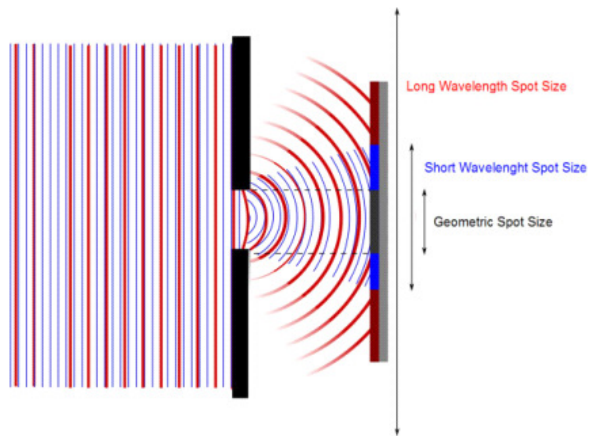
$$\sin(\theta_{dark}) = 1.22\frac{\lambda}{D}$$

or

$$\theta_{dark} \approx 1.22\frac{\lambda}{D}$$

we see that the ratio of the wavelength to the hole size determines the angular extent of the central maxima. The smaller the ratio, the smaller the central region. We can use this to explain why the wave nature of light was so hard to find.

The patch of light on a screen that is created by light passing through the aperture is created by the central maximum.



For the long wavelength (red) the central maximum is larger than the screen. The short wavelength spot will be wholly on the screen as shown. The geometric spot is what we would see if the light traveled straight through the opening. Notice that the short wavelength spot is closer to the size of the geometric spot. In the limit that

$$\lambda \ll a$$

or for circular openings

$$\lambda \ll D$$

then

$$\theta \approx \frac{\lambda}{a} \approx 0$$

or

$$\theta \approx \frac{\lambda}{D} \approx 0$$

and the spot size would be very nearly equal to the geometric spot size.

This is the limit we will called the *ray approximation*.

For most of mankind's time on the Earth, it was very hard to build holes that were comparable to the size of a wavelength of visible light (around 500 nm). So it is no wonder that the waviness of light was missed for so many years.

But this ray limit is very useful for apertures the size of camera lenses. So it was find for lenses and mirrors to use this small λ, large aperture approximation.

24.7 The Ray Approximation in Geometric Optics

In the last section we said that when the geometric spot size was larger about the same size as the spot due to diffraction, we could ignore diffraction and use the simpler ray model. This is usually true in our personal experiences. But this may not be true in experiments or devices we design. We should see where the crossover point is.

Intuitively, if the aperture and the spot are the same size, that ought to be some sort of critical point. That is when the aperture size is equal to the spot size. We found that for a circular aperture the spot width is

$$w \approx 2.44 \frac{\lambda}{D_{aperture}} L$$

We want the case where $w = D_{aperture}$

$$D_{aperture} = 2.44 \frac{\lambda}{D_{aperture}} L$$

This gives

$$D_{aperture} = \sqrt{2.44 \lambda L}$$

Of course this is for round apertures, but for square apertures we know we remove the 2.44. This gives about a millimeter for visible wavelengths.

$$\begin{aligned} D_{aperture} &= \sqrt{2.44 \, (500 \, \mathrm{nm}) \, (1 \, \mathrm{m})} \\ &= 1.104\,5 \times 10^{-3} \, \mathrm{m} \end{aligned}$$

for apertures much larger than a millimeter, we expect interference effects due to diffraction through the aperture to be much harder to see. We expect them to be easy to see if the aperture is smaller than a millimeter. But what about when the aperture is about a millimeter in size? That is a subject for PH375, and so we will avoid this case in this class. But this is not too restrictive. Most good optical systems have apertures larger than 1 mm. Cell phone cameras may be an exception (but they play tricks to beat this by having multiple apertures). Even our eyes have an aperture that varies from about 2 mm to about 7 mm, so most common experiences in visible wavelengths will work fine with what we learn. Note that for microwave or radio wave systems this may really not be true!

How about the other extreme? Suppose $\lambda \gg D$. This is really beyond our class (requires partial differential equations), but in the extreme case, we can use reason to find out what happens. If the opening is much smaller than the wavelength, then the wave does not see the opening, and no wave is produced on the other side. This is the case of a microwave oven door. If the wavelength is much larger than the spacing of the little dots or lines that span the door, then the waves will not leave the interior of the microwave oven. Of course as the wavelength becomes closer to D this is less true, and this case is more challenging to calculate, and we will save it for a 300 level electrodynamics course.

To summarize

$\lambda \ll D$	Wave nature of light is not visible
$\lambda \approx D$	Wave nature of light is apparent
$\lambda \gg D$	Little to no penetration of aperture by the wave

24.8 The ray model and phase

There is a further complication that helps to explain why the wave nature of light was not immediately apparent. Let's consider a light source.

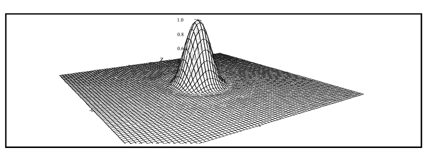
For a typical light source, the filament is larger than about a millimeter, so we should expect that diffraction should be hard to see. But the filament is made of hot metal. The atoms of the hot metal emit light because of the extra energy they have. The method of producing this light is that the atom's excited electrons are in upper shells because of the extra thermal energy provided by the electricity flowing through the filament. But the electrons eventually fall to their proper shell, and in doing so they give off the extra energy as light. It is not too hard to believe that this process of exciting electrons and having them fall back down is a random process. Each electron that moves starts a wave. The atoms have different positions, so there will be a path difference Δr between each atom's waves. There will also be a time difference Δt between when the waves start. We can model this with a $\Delta \phi_o$.

It is also true that not all of the electrons fall from the same shell. This gives us different frequencies, so we expect beating between different waves from different atoms. It is also true that we have millions of atoms, so we have millions of waves.

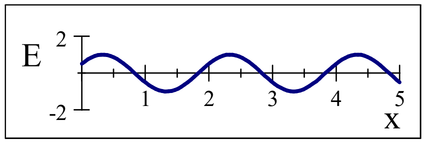
Let's look at just two of these waves

$$\begin{array}{l} \lambda = 2 \\ k = \frac{2\pi}{\lambda} \\ \omega = 1 \\ \phi_o = \frac{\pi}{6} \\ t = 0 \\ E_o = 1\,\frac{\mathrm{N}}{\mathrm{C}} \end{array}$$

$$E_1 = E_o \sin\left(kx - \omega t\right)$$



$$E_2 = E_o \sin(kx - \omega t + \phi_o)$$

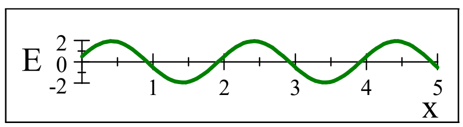


then

$$E_r = E_o \sin(kx - \omega t) + E_o \sin(kx - \omega t + \phi_o)$$

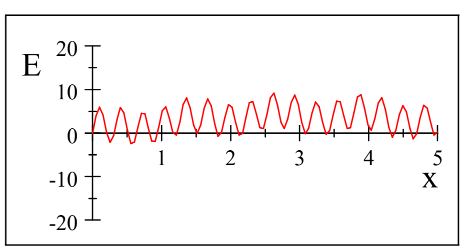
We found for these two waves

$$E_r = 2E_o \cos\left(\frac{\phi_o}{2}\right) \sin\left(kx - \omega t + \frac{\phi_o}{2}\right)$$

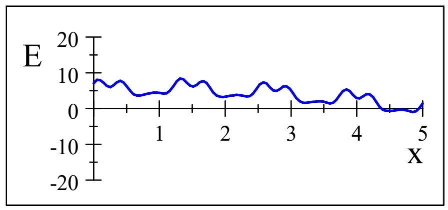


But suppose we complicate the situation by sending lots of waves at random times, each with different amplitudes and wavelengths, down the rope. If we look at a single point for a specific time, we might be experiencing interference, but it would be hard to tell. Lets try this mathematically. I will combine many waves with random phases, some coming from the right and some coming from the left.

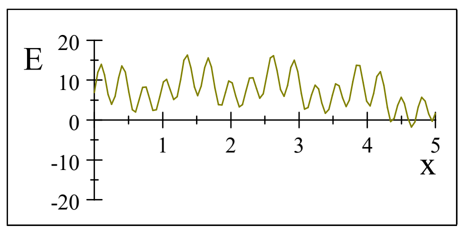
$$\begin{aligned} E_1 \quad = \quad & E_o \sin\left(5x - \omega t + \frac{\pi}{4}\right) + 0.5E_o \sin\left(0.2x - \omega t - \frac{\pi}{6}\right) \\ & +3.6E_o \sin\left(.4x - \omega t + \frac{\pi}{10}\right) + 4E_o \sin\left(20x - \omega t - \frac{\pi}{7}\right) \\ & +.2E_o \sin(15x - \omega t + 1) + 0.7E_o \sin(.7x - \omega t - .25) \end{aligned}$$



$$\begin{aligned} E_2 &= E_o \sin(0.2x + \omega t + \pi) + 2E_o \sin\left(5x + \omega t + \frac{\pi}{6}\right) \\ &\quad + 6E_o \sin\left(0.4x + \omega t + \frac{\pi}{3.5}\right) + 0.4E_o \sin(20x + \omega t - 0) \\ &\quad + E_o \sin(15x + \omega t + 1) + 0.7E_o \sin(.7x + \omega t - 4) \end{aligned}$$



Then $E_1 + E_2$ looks like



In this example, you could think about the superposition of E_1 and E_2 and predict the outcome, but if there were millions of waves, each with it's own wavelength, phase, and amplitude, the situation would be hopeless. Note that the fluctuations in these waves are much more frequent than our original waves. With all the added waves, we get a rapid change in amplitude.

Now if these waves are light waves, our eyes and most detectors are not able to react fast enough to detect the rapid fluctuations. So if there is constructive or destructive interference that might be simple enough to distinguish, we will miss it due to our detection systems' integration times. To describe this rapidly fluctuating interference pattern that we can't track with our detectors, we just say that light bulbs emit *incoherent light*. The ray approximation assumes incoherent light.

But then light bulbs and hot ovens and most things must emit incoherent light. Does any thing emit coherent light? Sure, today the easiest source of coherent light is a laser. That is why I have used lasers in the class demonstrations so far. Really though, even a laser is not perfectly coherent. One property of the laser is that it produces light with a long *coherence length*, or it produces light that can be treated under most circumstances as being monochromatic and having a single phase across the wave for much of the beam length. Radar and microwave transmitters emit coherent light (but at frequencies we can't see) and so do radio stations.

In the past, one could carefully create a monochromatic beam with filters. Then split the beam into two beams and remix the two beams. This would generate two mostly coherent sources if the distances traveled were not too large. This is what Young did. Now days we just use lasers.

24.9 Coherency

To be coherent,

1. A given part of the wave must maintain a constant phase with respect to the rest of the wave.

2. The wave must be monochromatic

These are very hard criteria to achieve. Most light, like that from our light bulb, is not coherent.

You may be left feeling that we have really just gotten started, and you would be right. If you are a physics major (or curious and have elective credit) you will study waves more in PH 279 (Modern Physics) and PH295 (Mathematical and Computational Physics) and you may study more optics in PH375. All of these are fantastic experiences. Even if you don't take a further class in these topics, you can study them on your own. There are many wonderful books on Optics.

Our goal this semester was to study the motion of many things and wave motion was a partial fulfilment of that goal. But not all motions of many objects are as uniform as wave motion. What if, say, the air molecules in our room instead of moving in simple harmonic motion like in a wave, they had random velocities? This is type of motion that we hinted about when we talked about thermal energy in PH121 and the hot atoms that emit incoherent light. We will take on this kind of motion next in this course.