

2.13.5: Wave Optics

Electron microscopes can make images of individual atoms, but why will a visible-light microscope never be able to? Stereo speakers create the illusion of music that comes from a band arranged in your living room, but why doesn't the stereo illusion work with bass notes? Why are computer chip manufacturers investing billions of dollars in equipment to etch chips with x-rays instead of visible light?

The answers to all of these questions have to do with the subject of wave optics. So far this book has discussed the interaction of light waves with matter, and its practical applications to optical devices like mirrors, but we have used the ray model of light almost exclusively. Hardly ever have we explicitly made use of the fact that light is an electromagnetic wave. We were able to get away with the simple ray model because the chunks of matter we were discussing, such as lenses and mirrors, were thousands of times larger than a wavelength of light. We now turn to phenomena and devices that can only be understood using the wave model of light.

12.5.1 Diffraction

Figure 2.13.5.1a shows a typical problem in wave optics, enacted with water waves. It may seem surprising that we don't get a simple pattern like Figure 2.13.5.1b but the pattern would only be that simple if the wavelength was hundreds of times shorter than the distance between the gaps in the barrier and the widths of the gaps.

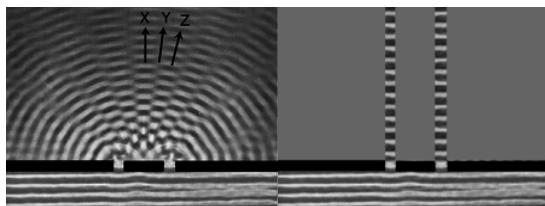


Figure 2.13.5.1: a. In this view from overhead, a straight, sinusoidal water wave encounters a barrier with two gaps in it. Strong wave vibration occurs at angles X and Z , but there is none at all at angle Y . (The figure has been retouched from a real photo of water waves. In reality, the waves beyond the barrier would be much weaker than the ones before it, and they would therefore be difficult to see.) b / This doesn't happen.

Wave optics is a broad subject, but this example will help us to pick out a reasonable set of restrictions to make things more manageable:

1. We restrict ourselves to cases in which a wave travels through a uniform medium, encounters a certain area in which the medium has different properties, and then emerges on the other side into a second uniform region.
2. We assume that the incoming wave is a nice tidy sine-wave pattern with wavefronts that are lines (or, in three dimensions, planes).
3. In Figure 2.13.5.1a we can see that the wave pattern immediately beyond the barrier is rather complex, but farther on it sorts itself out into a set of wedges separated by gaps in which the water is still. We will restrict ourselves to studying the simpler wave patterns that occur farther away, so that the main question of interest is how intense the outgoing wave is at a given angle.

The kind of phenomenon described by restriction (1) is called *diffraction*. Diffraction can be defined as the behavior of a wave when it encounters an obstacle or a nonuniformity in its medium. In general, diffraction causes a wave to bend around obstacles and make patterns of strong and weak waves radiating out beyond the obstacle. Understanding diffraction is the central problem of wave optics. If you understand diffraction, even the subset of diffraction problems that fall within restrictions (2) and (3), the rest of wave optics is icing on the cake.

Diffraction can be used to find the structure of an unknown diffracting object: even if the object is too small to study with ordinary imaging, it may be possible to work backward from the diffraction pattern to learn about the object. The structure of a crystal, for example, can be determined from its x-ray diffraction pattern.

Diffraction can also be a bad thing. In a telescope, for example, light waves are diffracted by all the parts of the instrument. This will cause the image of a star to appear fuzzy even when the focus has been adjusted correctly. By understanding diffraction, one can learn how a telescope must be designed in order to reduce this problem --- essentially, it should have the biggest possible diameter.

There are two ways in which restriction (2) might commonly be violated. First, the light might be a mixture of wavelengths. If we simply want to observe a diffraction pattern or to use diffraction as a technique for studying the object doing the diffracting (e.g., if the object is too small to see with a microscope), then we can pass the light through a colored filter before diffracting it.

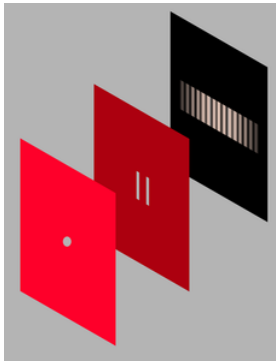
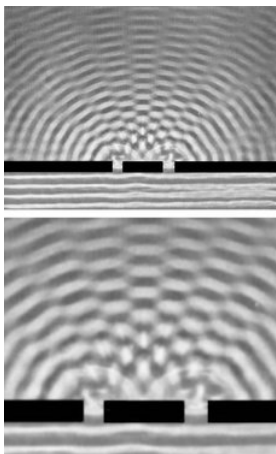


Figure 2.13.5.2: A practical, low-tech setup for observing diffraction of light.

A second issue is that light from sources such as the sun or a lightbulb does not consist of a nice neat plane wave, except over very small regions of space. Different parts of the wave are out of step with each other, and the wave is referred to as *incoherent*. One way of dealing with this is shown in Figure 2.13.5.2 After filtering to select a certain wavelength of red light, we pass the light through a small pinhole. The region of the light that is intercepted by the pinhole is so small that one part of it is not out of step with another. Beyond the pinhole, light spreads out in a spherical wave; this is analogous to what happens when you speak into one end of a paper towel roll and the sound waves spread out in all directions from the other end. By the time the spherical wave gets to the double slit it has spread out and reduced its curvature, so that we can now think of it as a simple plane wave.



d / The bottom figure is simply a copy of the middle portion of the top one, scaled up by a factor of two. All the angles are the same. Physically, the angular pattern of the diffraction fringes can't be any different if we scale both λ and d by the same factor, leaving λ/d unchanged.

If this seems laborious, you may be relieved to know that modern technology gives us an easier way to produce a single-wavelength, coherent beam of light: the laser.

The parts of the final image on the screen in Figure 2.13.5.2 are called *diffraction fringes*. The center of each fringe is a point of maximum brightness, and halfway between two fringes is a minimum.

Exercise 2.13.5.1: Discussion Question

Why would x-rays rather than visible light be used to find the structure of a crystal? Sound waves are used to make images of fetuses in the womb. What would influence the choice of wavelength

12.5.2 Scaling of diffraction

This chapter has “optics” in its title, so it is nominally about light, but we started out with an example involving water waves. Water waves are certainly easier to visualize, but is this a legitimate comparison? In fact the analogy works quite well, despite the fact that a light wave has a wavelength about a million times shorter. This is because diffraction effects scale uniformly. That is, if we enlarge or reduce the whole diffraction situation by the same factor, including both the wavelengths and the sizes of the obstacles the wave encounters, the result is still a valid solution.

This is unusually simple behavior! In subsection 0.2.2 we saw many examples of more complex scaling, such as the impossibility of bacteria the size of dogs, or the need for an elephant to eliminate heat through its ears because of its small surface-to-volume ratio, whereas a tiny shrew's life-style centers around conserving its body heat.

Of course water waves and light waves differ in many ways, not just in scale, but the general facts you will learn about diffraction are applicable to all waves. In some ways it might have been more appropriate to insert this chapter after section 6.2 on bounded waves, but many of the important applications are to light waves, and you would probably have found these much more difficult without any background in optics.

Another way of stating the simple scaling behavior of diffraction is that the diffraction angles we get depend only on the unitless ratio λ/d , where λ is the wavelength of the wave and d is some dimension of the diffracting objects, e.g., the center-to-center spacing between the slits in figure a. If, for instance, we scale up both λ and d by a factor of 37, the ratio λ/d will be unchanged.

12.5.3 The Correspondence Principle

The only reason we don't usually notice diffraction of light in everyday life is that we don't normally deal with objects that are comparable in size to a wavelength of visible light, which is about a millionth of a meter. Does this mean that wave optics contradicts ray optics, or that wave optics sometimes gives wrong results? No. If you hold three fingers out in the sunlight and cast a shadow with them, *either* wave optics or ray optics can be used to predict the straightforward result: a shadow pattern with two bright lines where the light has gone through the gaps between your fingers. Wave optics is a more general theory than ray optics, so in any case where ray optics is valid, the two theories will agree. This is an example of a general idea enunciated by the physicist Niels Bohr, called the *correspondence principle*: when flaws in a physical theory lead to the creation of a new and more general theory, the new theory must still agree with the old theory within its more restricted area of applicability. After all, a theory is only created as a way of describing experimental observations. If the original theory had not worked in any cases at all, it would never have become accepted.

In the case of optics, the correspondence principle tells us that when λ/d is small, both the ray and the wave model of light must give approximately the same result. Suppose you spread your fingers and cast a shadow with them using a coherent light source. The quantity λ/d is about 10^{-4} , so the two models will agree very closely. (To be specific, the shadows of your fingers will be outlined by a series of light and dark fringes, but the angle subtended by a fringe will be on the order of 10^{-4} radians, so they will be invisible and washed out by the natural fuzziness of the edges of sun-shadows, caused by the finite size of the sun.)

Exercise 2.13.5.1

What kind of wavelength would an electromagnetic wave have to have in order to diffract dramatically around your body? Does this contradict the correspondence principle?

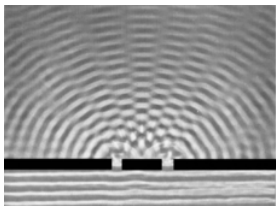
Answer

(answer in the back of the PDF version of the book)

12.5.4 Huygens' Principle

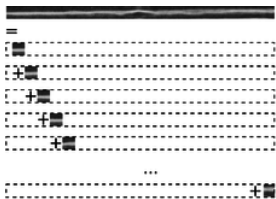


e / Christiaan Huygens (1629-1695).



f / Double-slit diffraction.

Returning to the example of double-slit diffraction, f, note the strong visual impression of two overlapping sets of concentric semicircles. This is an example of *Huygens' principle*, named after a Dutch physicist and astronomer. (The first syllable rhymes with “boy.”) Huygens' principle states that any wavefront can be broken down into many small side-by-side wave peaks, g, which then spread out as circular ripples, h, and by the principle of superposition, the result of adding up these sets of ripples must give the same result as allowing the wave to propagate forward, i.



g / A wavefront can be analyzed by the principle of superposition, breaking it down into many small parts.



h / If it was by itself, each of the parts would spread out as a circular ripple.



i / Adding up the ripples produces a new wavefront.

In the case of sound or light waves, which propagate in three dimensions, the “ripples” are actually spherical rather than circular, but we can often imagine things in two dimensions for simplicity.

In double-slit diffraction the application of Huygens' principle is visually convincing: it is as though all the sets of ripples have been blocked except for two. It is a rather surprising mathematical fact, however, that Huygens' principle gives the right result in the case of an unobstructed linear wave, h and i. A theoretically infinite number of circular wave patterns somehow conspire to add together and produce the simple linear wave motion with which we are familiar.

Since Huygens' principle is equivalent to the principle of superposition, and superposition is a property of waves, what Huygens had created was essentially the first wave theory of light. However, he imagined light as a series of pulses, like hand claps, rather

than as a sinusoidal wave.

The history is interesting. Isaac Newton loved the atomic theory of matter so much that he searched enthusiastically for evidence that light was also made of tiny particles. The paths of his light particles would correspond to rays in our description; the only significant difference between a ray model and a particle model of light would occur if one could isolate individual particles and show that light had a “graininess” to it. Newton never did this, so although he thought of his model as a particle model, it is more accurate to say he was one of the builders of the ray model.

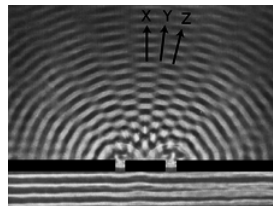
Almost all that was known about reflection and refraction of light could be interpreted equally well in terms of a particle model or a wave model, but Newton had one reason for strongly opposing Huygens' wave theory. Newton knew that waves exhibited diffraction, but diffraction of light is difficult to observe, so Newton believed that light did not exhibit diffraction, and therefore must not be a wave. Although Newton's criticisms were fair enough, the debate also took on the overtones of a nationalistic dispute between England and continental Europe, fueled by English resentment over Leibniz's supposed plagiarism of Newton's calculus. Newton wrote a book on optics, and his prestige and political prominence tended to discourage questioning of his model.



j / Thomas Young

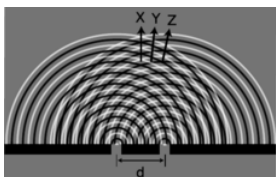
Thomas Young (1773-1829) was the person who finally, a hundred years later, did a careful search for wave interference effects with light and analyzed the results correctly. He observed double-slit diffraction of light as well as a variety of other diffraction effects, all of which showed that light exhibited wave interference effects, and that the wavelengths of visible light waves were extremely short. The crowning achievement was the demonstration by the experimentalist Heinrich Hertz and the theorist James Clerk Maxwell that light was an *electromagnetic* wave. Maxwell is said to have related his discovery to his wife one starry evening and told her that she was the only other person in the world who knew what starlight was.

12.5.5 Double-slit diffraction



k / Double-slit diffraction.

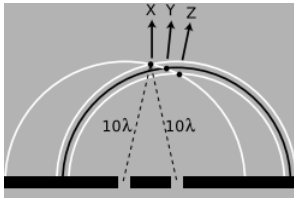
Let's now analyze double-slit diffraction, k, using Huygens' principle. The most interesting question is how to compute the angles such as X and Z where the wave intensity is at a maximum, and the in-between angles like Y where it is minimized. Let's measure all our angles with respect to the vertical center line of the figure, which was the original direction of propagation of the wave.



l / Use of Huygens' principle.

If we assume that the width of the slits is small (on the order of the wavelength of the wave or less), then we can imagine only a single set of Huygens ripples spreading out from each one, l. White lines represent peaks, black ones troughs. The only dimension

of the diffracting slits that has any effect on the geometric pattern of the overlapping ripples is then the center-to-center distance, d , between the slits.



m / Constructive interference along the center-line.

We know from our discussion of the scaling of diffraction that there must be some equation that relates an angle like θ_Z to the ratio λ/d ,

$$\frac{\lambda}{d} \leftrightarrow \theta_Z.$$

If the equation for θ_Z depended on some other expression such as $\lambda + d$ or λ^2/d , then it would change when we scaled λ and d by the same factor, which would violate what we know about the scaling of diffraction.

Along the central maximum line, X, we always have positive waves coinciding with positive ones and negative waves coinciding with negative ones. (I have arbitrarily chosen to take a snapshot of the pattern at a moment when the waves emerging from the slit are experiencing a positive peak.) The superposition of the two sets of ripples therefore results in a doubling of the wave amplitude along this line. There is constructive interference. This is easy to explain, because by symmetry, each wave has had to travel an equal number of wavelengths to get from its slit to the center line, m : Because both sets of ripples have ten wavelengths to cover in order to reach the point along direction X, they will be in step when they get there.

At the point along direction Y shown in the same figure, one wave has traveled ten wavelengths, and is therefore at a positive extreme, but the other has traveled only nine and a half wavelengths, so it is at a negative extreme. There is perfect cancellation, so points along this line experience no wave motion.

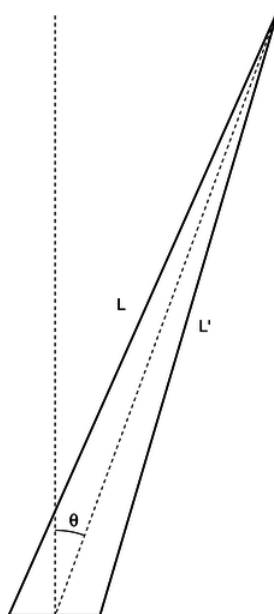
But the distance traveled does not have to be equal in order to get constructive interference. At the point along direction Z, one wave has gone nine wavelengths and the other ten. They are both at a positive extreme.

self-check:

At a point half a wavelength below the point marked along direction X, carry out a similar analysis.

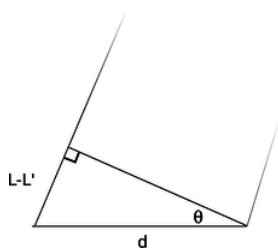
(answer in the back of the PDF version of the book)

To summarize, we will have perfect constructive interference at any point where the distance to one slit differs from the distance to the other slit by an integer number of wavelengths. Perfect destructive interference will occur when the number of wavelengths of path length difference equals an integer plus a half.



n / The waves travel distances L and L' from the two slits to get to the same point in space, at an angle θ from the center line.

Now we are ready to find the equation that predicts the angles of the maxima and minima. The waves travel different distances to get to the same point in space, n. We need to find whether the waves are in phase (in step) or out of phase at this point in order to predict whether there will be constructive interference, destructive interference, or something in between.



o / A close-up view of figure n, showing how the path length difference $L - L'$ is related to d and to the angle θ .

One of our basic assumptions in this chapter is that we will only be dealing with the diffracted wave in regions very far away from the object that diffracts it, so the triangle is long and skinny. Most real-world examples with diffraction of light, in fact, would have triangles with even skinner proportions than this one. The two long sides are therefore very nearly parallel, and we are justified in drawing the right triangle shown in figure o, labeling one leg of the right triangle as the difference in path length, $L - L'$, and labeling the acute angle as θ . (In reality this angle is a tiny bit greater than the one labeled θ in figure n.)

The difference in path length is related to d and θ by the equation

$$\frac{L - L'}{d} = \sin \theta.$$

Constructive interference will result in a maximum at angles for which $L - L'$ is an integer number of wavelengths,

$$L - L' = m\lambda.$$

[condition for a maximum; m is an integer]

Here m equals 0 for the central maximum, -1 for the first maximum to its left, $+2$ for the second maximum on the right, etc. Putting all the ingredients together, we find $m\lambda/d = \sin \theta$, or

$$\frac{\lambda}{d} = \frac{\sin \theta}{m}.$$

[condition for a maximum; m is an integer]

Similarly, the condition for a minimum is

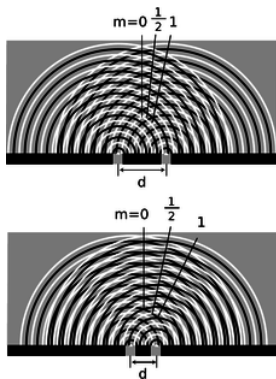
$$\frac{\lambda}{d} = \frac{\sin \theta}{m}.$$

[condition for a minimum; m is an integer plus $1/2$]

That is, the minima are about halfway between the maxima.

As expected based on scaling, this equation relates angles to the unitless ratio λ/d . Alternatively, we could say that we have proven the scaling property in the special case of double-slit diffraction. It was inevitable that the result would have these scaling properties, since the whole proof was geometric, and would have been equally valid when enlarged or reduced on a photocopying machine!

Counterintuitively, this means that a diffracting object with smaller dimensions produces a bigger diffraction pattern, p.



p / Cutting d in half doubles the angles of the diffraction fringes.

Example 12: Double-slit diffraction of blue and red light

Blue light has a shorter wavelength than red. For a given double-slit spacing d , the smaller value of λ/d for leads to smaller values of $\sin \theta$, and therefore to a more closely spaced set of diffraction fringes, q



q / Double-slit diffraction patterns of long-wavelength red light (top) and short-wavelength blue light (bottom).

Example 13: The correspondence principle

Let's also consider how the equations for double-slit diffraction relate to the correspondence principle. When the ratio λ/d is very small, we should recover the case of simple ray optics. Now if λ/d is small, $\sin \theta$ must be small as well, and the spacing between the diffraction fringes will be small as well. Although we have not proven it, the central fringe is always the brightest, and the fringes get dimmer and dimmer as we go farther from it. For small values of λ/d , the part of the diffraction pattern that is bright enough to be detectable covers only a small range of angles. This is exactly what we would expect from ray optics: the rays passing through the two slits would remain parallel, and would continue moving in the $\theta = 0$ direction. (In fact there would be images of the two separate slits on the screen, but our analysis was all in terms of angles, so we should not expect it to address the issue of whether there is structure within a set of rays that are all traveling in the $\theta = 0$ direction.)

Example 14: Spacing of the fringes at small angles

At small angles, we can use the approximation $\sin \theta \approx \theta$, which is valid if θ is measured in radians. The equation for double-slit diffraction becomes simply

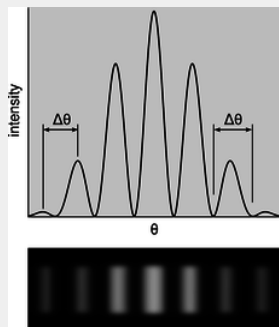
$$\frac{\lambda}{d} = \frac{\theta}{m},$$

which can be solved for θ to give

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

The difference in angle between successive fringes is the change in θ that results from changing m by plus or minus one,

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



Interpretation of the angular spacing $\Delta\theta$ in example 14. It can be defined either from maximum to maximum or from minimum to minimum. Either way, the result is the same. It does not make sense to try to interpret $\Delta\theta$ as the width of a fringe; one can see from the graph and from the image below that it is not obvious either that such a thing is well defined or that it would be the same for all fringes. For example, if we write θ_7 for the angle of the seventh bright fringe on one side of the central maximum and θ_8 for the neighboring one, we have

$$\begin{aligned}\theta_8 - \theta_7 &= \frac{8\lambda}{d} - \frac{7\lambda}{d} \\ &= \frac{\lambda}{d},\end{aligned}$$

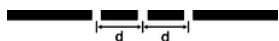
and similarly for any other neighboring pair of fringes.

Although the equation $\lambda/d = \sin \theta/m$ is only valid for a double slit, it can still be a guide to our thinking even if we are observing diffraction of light by a virus or a flea's leg: it is always true that

- (1) large values of λ/d lead to a broad diffraction pattern, and
- (2) diffraction patterns are repetitive.

In many cases the equation looks just like $\lambda/d = \sin \theta/m$ but with an extra numerical factor thrown in, and with d interpreted as some other dimension of the object, e.g., the diameter of a piece of wire.

12.5.6 Repetition

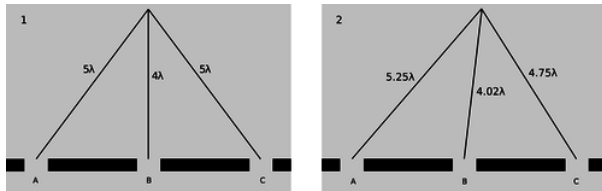


A triple slit.

Suppose we replace a double slit with a triple slit, s . We can think of this as a third repetition of the structures that were present in the double slit. Will this device be an improvement over the double slit for any practical reasons?

The answer is yes, as can be shown using figure u. For ease of visualization, I have violated our usual rule of only considering points very far from the diffracting object. The scale of the drawing is such that a wavelength is one cm. In u/1, all three waves travel an integer number of wavelengths to reach the same point, so there is a bright central spot, as we would expect from our experience with the double slit. In figure u/2, we show the path lengths to a new point. This point is farther from slit A by a quarter

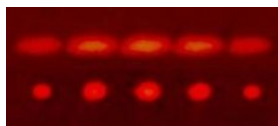
of a wavelength, and correspondingly closer to slit C. The distance from slit B has hardly changed at all. Because the paths lengths traveled from slits A and C differ by half a wavelength, there will be perfect destructive interference between these two waves. There is still some uncanceled wave intensity because of slit B, but the amplitude will be three times less than in figure u/1, resulting in a factor of 9 decrease in brightness. Thus, by moving off to the right a little, we have gone from the bright central maximum to a point that is quite dark.



u / 1. There is a bright central maximum. 2. At this point just off the central maximum, the path lengths traveled by the three waves have changed.

Now let's compare with what would have happened if slit C had been covered, creating a plain old double slit. The waves coming from slits A and B would have been out of phase by 0.23 wavelengths, but this would not have caused very severe interference. The point in figure u/2 would have been quite brightly lit up.

To summarize, we have found that adding a third slit narrows down the central fringe dramatically. The same is true for all the other fringes as well, and since the same amount of energy is concentrated in narrower diffraction fringes, each fringe is brighter and easier to see, t.



t / A double-slit diffraction pattern (top), and a pattern made by five slits (bottom).

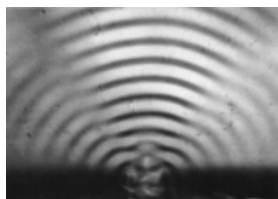
This is an example of a more general fact about diffraction: if some feature of the diffracting object is repeated, the locations of the maxima and minima are unchanged, but they become narrower.

Taking this reasoning to its logical conclusion, a diffracting object with thousands of slits would produce extremely narrow fringes. Such an object is called a diffraction grating.

12.5.7 Single-slit diffraction

If we use only a single slit, is there diffraction? If the slit is not wide compared to a wavelength of light, then we can approximate its behavior by using only a single set of Huygens ripples. There are no other sets of ripples to add to it, so there are no constructive or destructive interference effects, and no maxima or minima. The result will be a uniform spherical wave of light spreading out in all directions, like what we would expect from a tiny lightbulb. We could call this a diffraction pattern, but it is a completely featureless one, and it could not be used, for instance, to determine the wavelength of the light, as other diffraction patterns could.

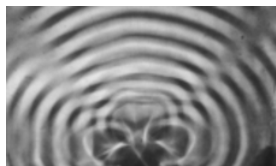
All of this, however, assumes that the slit is narrow compared to a wavelength of light. If, on the other hand, the slit is broader, there will indeed be interference among the sets of ripples spreading out from various points along the opening. Figure v shows an example with water waves, and figure w with light.



v / Single-slit diffraction of water waves.



w / Single-slit diffraction of red light. Note the double width of the central maximum.



x / A pretty good simulation of the single-slit pattern of figure v, made by using three motors to produce overlapping ripples from three neighboring points in the water.

self-check:

How does the wavelength of the waves compare with the width of the slit in figure v?

(answer in the back of the PDF version of the book)

We will not go into the details of the analysis of single-slit diffraction, but let us see how its properties can be related to the general things we've learned about diffraction. We know based on scaling arguments that the angular sizes of features in the diffraction pattern must be related to the wavelength and the width, a , of the slit by some relationship of the form

$$\frac{\lambda}{a} \leftrightarrow \theta.$$

This is indeed true, and for instance the angle between the maximum of the central fringe and the maximum of the next fringe on one side equals $1.5\lambda/a$. Scaling arguments will never produce factors such as the 1.5, but they tell us that the answer must involve λ/a , so all the familiar qualitative facts are true. For instance, shorter-wavelength light will produce a more closely spaced diffraction pattern.



y / An image of the Pleiades star cluster. The circular rings around the bright stars are due to single-slit diffraction at the mouth of the telescope's tube.

An important scientific example of single-slit diffraction is in telescopes. Images of individual stars, as in figure y, are a good way to examine diffraction effects, because all stars except the sun are so far away that no telescope, even at the highest magnification, can image their disks or surface features. Thus any features of a star's image must be due purely to optical effects such as diffraction. A prominent cross appears around the brightest star, and dimmer ones surround the dimmer stars. Something like this is seen in most telescope photos, and indicates that inside the tube of the telescope there were two perpendicular struts or supports. Light diffracted around these struts. You might think that diffraction could be eliminated entirely by getting rid of all obstructions in the tube, but the circles around the stars are diffraction effects arising from single-slit diffraction at the mouth of the telescope's tube! (Actually we have not even talked about diffraction through a circular opening, but the idea is the same.) Since the angular sizes of the diffracted images depend on λ/a , the only way to improve the resolution of the images is to increase the diameter, a , of the tube. This is one of the main reasons (in addition to light-gathering power) why the best telescopes must be very large in diameter.



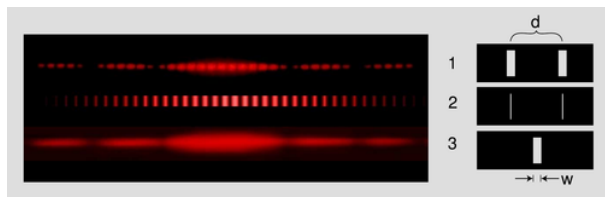
z / A radio telescope.

self-check:

What would this imply about radio telescopes as compared with visible-light telescopes?

(answer in the back of the PDF version of the book)

Double-slit diffraction is easier to understand conceptually than single-slit diffraction, but if you do a double-slit diffraction experiment in real life, you are likely to encounter a complicated pattern like figure aa/1, rather than the simpler one, 2, you were expecting. This is because the slits are fairly big compared to the wavelength of the light being used. We really have two different distances in our pair of slits: d , the distance between the slits, and w , the width of each slit. Remember that smaller distances on the object the light diffracts around correspond to larger features of the diffraction pattern. The pattern 1 thus has two spacings in it: a short spacing corresponding to the large distance d , and a long spacing that relates to the small dimension w .



aa / 1. A diffraction pattern formed by a real double slit. The width of each slit is fairly big compared to the wavelength of the light. This is a real photo. 2. This idealized pattern is not likely to occur in real life. To get it, you would need each slit to be so narrow that its width was comparable to the wavelength of the light, but that's not usually possible. This is not a real photo. 3. A real photo of a single-slit diffraction pattern caused by a slit whose width is the same as the widths of the slits used to make the top pattern.

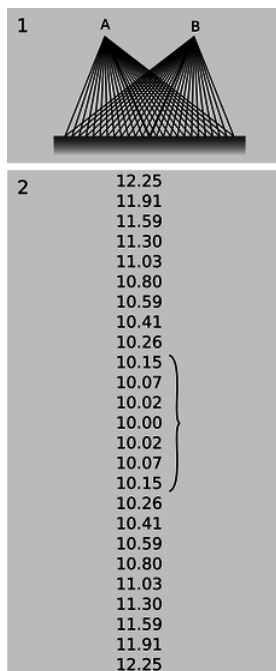
Discussion Question

◇ Why is it optically impossible for bacteria to evolve eyes that use visible light to form images?

The principle of least time

In subsection 12.1.5 and 12.4.5, we saw how in the ray model of light, both refraction and reflection can be described in an elegant and beautiful way by a single principle, the principle of least time. We can now justify the principle of least time based on the wave model of light. Consider an example involving reflection, ab. Starting at point A, Huygens' principle for waves tells us that we can think of the wave as spreading out in all directions. Suppose we imagine all the possible ways that a ray could travel from A to B. We show this by drawing 25 possible paths, of which the central one is the shortest. Since the principle of least time connects the wave model to the ray model, we should expect to get the most accurate results when the wavelength is much shorter than the distances involved --- for the sake of this numerical example, let's say that a wavelength is 1/10 of the shortest reflected path from A to B. The table, 2, shows the distances traveled by the 25 rays.

Note how similar are the distances traveled by the group of 7 rays, indicated with a bracket, that come closest to obeying the principle of least time. If we think of each one as a wave, then all 7 are again nearly in phase at point B. However, the rays that are farther from satisfying the principle of least time show more rapidly changing distances; on reuniting at point B, their phases are a random jumble, and they will very nearly cancel each other out. Thus, almost none of the wave energy delivered to point B goes by these longer paths. Physically we find, for instance, that a wave pulse emitted at A is observed at B after a time interval corresponding very nearly to the shortest possible path, and the pulse is not very "smeared out" when it gets there. The shorter the wavelength compared to the dimensions of the figure, the more accurate these approximate statements become.



ab / Light could take many different paths from A to B.

Instead of drawing a finite number of rays, such as 25, what happens if we think of the angle, θ , of emission of the ray as a continuously varying variable? Minimizing the distance L requires

$$\frac{dL}{d\theta} = 0.$$

Because L is changing slowly in the vicinity of the angle that satisfies the principle of least time, all the rays that come out close to this angle have very nearly the same L , and remain very nearly in phase when they reach B. This is the basic reason why the discrete table, ab/2, turned out to have a group of rays that all traveled nearly the same distance.

As discussed in subsection 12.1.5, the principle of least time is really a principle of least *or greatest* time. This makes perfect sense, since $dL/d\theta = 0$ can in general describe either a minimum or a maximum

The principle of least time is very general. It does not apply just to refraction and reflection --- it can even be used to prove that light rays travel in a straight line through empty space, without taking detours! This general approach to wave motion was used by Richard Feynman, one of the pioneers who in the 1950's reconciled quantum mechanics with relativity. A very readable explanation is given in a book Feynman wrote for laypeople, QED: The Strange Theory of Light and Matter.

Contributors and Attributions

- Benjamin Crowell (Fullerton College). [Conceptual Physics](#) is copyrighted with a CC-BY-SA license.

This page titled 2.13.5: Wave Optics is shared under a [CC BY-SA](#) license and was authored, remixed, and/or curated by Benjamin Crowell.

- 13.5: Wave Optics by Benjamin Crowell is licensed [CC BY-SA 4.0](#).