CHAPTER 11

Waves: Mechanical and Electromagnetic

When we think of transporting mechanical energy, we might first think of a ball or bullet. The kinetic energy that we give one of these projectiles at one end of its path is available at the other end. There is another way. Think of a group of people standing close to each other. Someone pushes the first one, let's call him Robert. That makes him lean over and push the person next to him, who, in turn, pushes the next one, and so on. The last one, call him Richard, falls over when his neighbor pushes him. Robert never touches Richard, but the initial push starts the sequence of forces from one to the next, through the line of people, from Robert to Richard.

We have described a single disturbance, or pulse. If it is a continuing disturbance, back and forth, or up and down that is transmitted, we call it a wave.

We first explore mechanical waves, characterized by forces and displacements. We then come to electromagnetic waves, where no material object moves and only the electric and magnetic fields change as they chase each other through empty space.

11.1 What is a wave?

Different kinds of waves

Think of a quiet lake. The water surface is smooth and horizontal. The gravitational force acts on each part of it. If some of the water is higher than the rest, its weight pulls it to a lower level. If some of it is lower, water from higher levels will tend to flow there. Only when all of the surface is at the same level is the water quiet.

Now look at a stone falling in the lake. Where it hits the water, it pushes it down. The stone sinks out of sight, but the water is no longer in equilibrium. The depression is still there. The gravitational force now acts to restore the equilibrium configuration. Water fills the depression. At the moment when the low part is filled, the water level is still changing. The water moves further and overshoots, and there is now a raised portion of the surface. It experiences a downward force, again back to the horizontal equilibrium. Once more it overshoots. The water surface at the point where the stone fell oscillates up and down, and would continue to do so if frictional forces did not cause the oscillations gradually to diminish and eventually to disappear.

As the water moves up and down in the same place, there is another motion that develops. Each time the water moves down, it pushes some of the rest of the water out of the way sideways, so that the neighboring region moves up. The water there, in turn, pushes on the water further away, so that the oscillation is propagated outward, away from the original point (the source) where the stone started it.

When we look at the water we see an up-and-down oscillation at the source and also at every other point that has been affected. At each point there is a motion of the water, up and down, as a function of time.

The second variation that we see is as a function of distance, away from the source in every direction along the surface. We can take a snapshot and see the oscillation at a particular moment in time.

If you think about what you actually look at when you see a wave, it is likely to be something different still. Your eye follows a point on the wave, a high point or a low point (a crest or a trough) as it moves away from the source along the direction of propagation of the wave.

Each of these variations is characteristic also for other waves. There are waves along a guitar string, sound waves in air, and elastic waves through a solid or liquid material. All of these are *mechanical* waves. In each case an equilibrium situation is disturbed. A restoring force acts to reestablish the original equilibrium, but the motion continues: there is an overshoot. The restoring force continues to act, changing direction so that it is always toward the equilibrium configuration.

EXAMPLE 1

Go to the PhET website (http://phet.colorado.edu) and open the simulation *Wave on a String*. Set "damping" to zero. (This eliminates friction and other dissipative forces.) Select "oscillate" and "no end." This will allow the string on the screen to move up and down as if the string continued on to infinity. Change the frequency and the amplitude. Push "pause" and then "step."

(a) Describe the motion of one of the green particles. Use the "step" feature.

(b) What is the relation between the direction of motion of the particles and the direction of propagation of the wave?

Ans.:

(a) The particles oscillate along the vertical direction in simple harmonic motion.

(b) In this case the motion of the particles is in the direction perpendicular to the direction of propagation of the wave. This kind of wave is called a *transverse wave.*

(In other kinds of waves, such as sound waves in air, the particle motion is along the same direction as the wave propagation. These waves are called *longitudinal* waves.)

Some features are common to all waves, but each kind has its special characteristics that depend on the forces that act to restore the equilibrium and on the medium (the water, the string, the material) in which the waves propagate.

A sound wave in air, for example, is quite different from a wave on the water surface, but both share properties that are common to all waves.

Let's see how a sound wave in air gets started. It can be by the back-and-forth motion of your vocal cords or of the cone of a loudspeaker, or by the vibration of a string, as in a guitar or violin. In each case the air is pushed and disturbed from its equilibrium. In turn, it pushes on the neighboring region, which pushes further, and so on along the propagating wave.

If the sound wave reaches an ear, it pushes on the eardrum, and makes it move back and forth. Knowledge of this vibration is transmitted to the brain, and is perceived by it as sound.

Along the wave there is the motion of air molecules, back and forth, at each point, superimposed on their random thermal motion. There is a second quantity that varies in space. As the air is pushed at the source, the air pressure increases in the region around it. When the oscillation continues at the source, and the vocal cord or string draws back, the opposite occurs, and the pressure decreases. Where the motion of the molecules causes them to be closer together, the air pressure increases; where they move farther apart, the pressure decreases. We can describe the wave either as a variation in space and time of the motion of the molecules or as a variation of the air pressure.

EXAMPLE 2

Go to the PhET website and open the simulation *Wave Interference*. Select "Sound" at the top, "one speaker," "no barrier," speaker "on," and "grayscale." Change the amplitude and the frequency with the sliders, and observe the speaker and the wave. You can use the "pause-step" function with the buttons at the bottom of the screen. Click on "show graph."

What varies as the wave propagates? What do the light and dark bands represent?

Click on "add detector." Place the detector in front of the speaker by clicking on it and moving it with the mouse. The graph can be moved out of the way of the waves. Now you can see what varies and how it varies.

Click on "particles." The markers allow you to follow the motion of one particle at a time. How do the particles move? What does the eye follow more naturally?

Ans.:
The back-and-forth motion of the speaker pushes the air particles. It causes regions of high pressure and low pressure to move outward (to radiate) from the speaker. The light regions are seen to represent high pressure and the dark regions low pressure.

The pressure varies sinusoidally with time.

Each particle oscillates along the direction of propagation of the wave. Where the particles bunch together the pressure is high. The eye more naturally follows the path taken by the points of maximum pressure.

11.2 What can waves do? Describing waves and their properties

Interference

If I throw a baseball at you, there's no doubt that it carries energy. When it hits you, some of its energy is transferred to you. If two baseballs come at you, you feel the energies transferred from each one. They add up.

The way waves add is very different. When two waves arrive at a point their effects sometimes add, but they can also cancel.

We can see that with the water waves that we looked at when we started. As the wave passes, it alternately raises and lowers the water level. Another similar wave, passing at the same place, coming from another direction, will, by itself, have the same effect. But what happens when both get to the same place at the same time? One possibility is that each one, separately, would cause the water level to rise. When both arrive together, the effects add, and the water level rises twice as high as it does with just one of the waves.

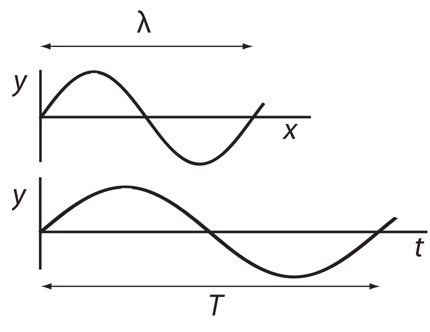
But it is also possible that when both waves get to the same place, one of them tends to make the water level rise, but the other one acts to lower it. The effects again combine, but if one of the waves, by itself, would cause a crest, and the other a trough, the two will cancel. The water level neither rises nor falls.

This possibility of cancellation is peculiar to waves. It is an essential feature that distinguishes waves from baseballs or other thrown objects. The way that waves superimpose, sometimes resulting in reinforcement and sometimes in cancellation, is called *interference*.

Pictorial and mathematical description

Let's describe a wave in detail with some mathematics. We can think of a water wave as it travels along a quiet lake, but it could also be some other kind of wave, such as a sound wave in air or through a solid material.

Here are two graphs of a wave.



They look similar, but each tells something quite different. The first is a snapshot. It shows the height of the water as a function of the displacement along a line in the lake. It could go on, along the x-axis, repeating after a distance λ (Greek *lambda*), called the *wavelength*.

The second graph shows what happens at one point on this line as the water level rises and falls when the wave passes. It shows the wave at this point as a function of time. It repeats after a time T, called the *period*.

In the time the wave travels the distance λ, the point at $x = 0$ goes through one full cycle, from $y = 0$, up, down, and back to zero. This variation is shown by the second of the two graphs. We see that the wave travels a distance λ in the time T. The wave speed is therefore $\frac{\lambda}{T}$.

One more quantity characteristic of the wave is the *frequency*. If it takes $\frac{1}{4}$ s for the wave to repeat ($T = \frac{1}{4}$ s), then four full cycles of the wave pass in every second. The frequency, f, is then $4\,\text{s}^{-1}$ or 4 hertz (4 Hz). In general, $f = \frac{1}{T}$. We see also that the wave speed is given by $v = f\lambda$.

EXAMPLE 3

Go to the PhET webvsite and open the simulation *Wave on a String*. Set "damping" to zero, "oscillate," and "no end." Select "rulers" and "timer."

Use the timer to measure how long it takes the wave to travel one wavelength. This is done most easily by going to "pause," start timer, then "step" until the wave has moved one wavelength.

How does this time (the *period* of the motion) change as the frequency is changed? Note that the number above "frequency" is proportional to the frequency, but is not the actual frequency.

Ans.:
The period, T, is the time it takes for the wave to travel one wavelength. It is the reciprocal of the frequency, f. The speed is $v = \frac{\lambda}{T} = f\lambda$.

EXAMPLE 4

Go to the PhET website and open the simulation *Wave Interference*. Select "Water," "One Drip," "No Barrier." Check "measuring tape" and "stopwatch." Set the f-slider near the middle. Pause (bottom button) and measure the wavelength with the measuring tape. Use the stopwatch to measure the period. (Pause, start, reset, and step.)

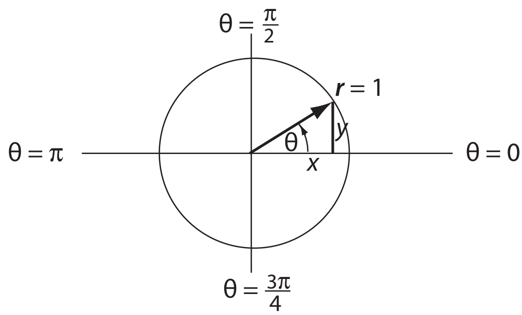
Repeat for two smaller and two larger settings of the frequency.

Make a graph of your data of λ vs. T. What quantity on your graph represents the speed of the water wave? What is your result for the speed?

What is the advantage of using the graph compared to using the individual results?

Ans.:
Since $\lambda = vT$, we expect a straight line through the origin, whose slope is the wave speed. The advantage of the graph is that drawing the straight line averages the results. Here the speed is about 2.3 cm/s.

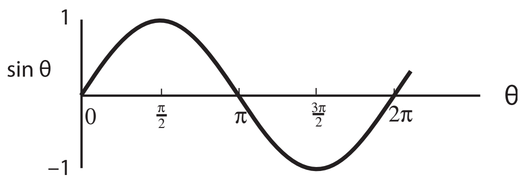
We still haven't written down a mathematical description of the two graphs, y as a function of x and y as a function of t. We need a mathematical function that repeats. Here is one: the sine function $y = \sin\theta$. We can show how this function varies with the help of a circle whose radius is one unit.



The radial vector (**r**) starts horizontally, to the right. Let the angle that it makes with the horizontal be θ, measured in radians. As this angle increases, we can look at the quantity $y = \sin\theta$. It is the vertical component of the vector **r**. ($\sin\theta = \frac{y}{r}$, and since the magnitude r is 1, $\sin\theta = y$.)

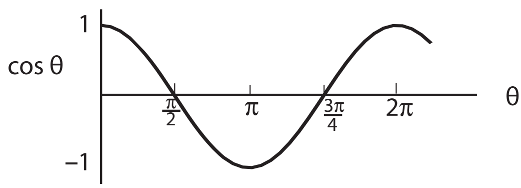
As θ goes from 0 to 90°, i.e., from 0 to $\frac{\pi}{2}$ radians, the distance y ($= \sin\theta$) goes from 0 to 1. In the next quadrant, as θ goes from 90° to 180°, or $\frac{\pi}{2}$ to π radians, y goes from 1 to zero. In the third quadrant, θ goes from π to $\frac{3\pi}{2}$ and y from 0 to -1, and finally, as θ goes from $\frac{3\pi}{2}$ to 2π, y returns from -1 to zero. The full cycle is 2π *radians*, and as θ continues beyond this value, the cycle repeats. We see that $\sin 0 = 0$, $\sin\frac{\pi}{2} = 1$, $\sin\pi = 0$, $\sin\frac{3\pi}{2} = -1$, and $\sin 2\pi = 0$, after which the sine function repeats.

That's not quite what we need. The function $\sin\theta$ repeats each time θ increases by 2π. We



want a function that repeats when x is increased by λ. This will happen if we use the function $y = \sin \frac{2\pi}{\lambda}x$. When x in this function reaches λ, $y = \sin 2\pi$, which is equal to zero, and then the function repeats.

What about the variation with time? It is similar. The function $y = \sin \frac{2\pi}{T}t$ $(= \sin 2\pi ft)$ repeats after a time T.



The cosine function $y = \cos\theta$ looks exactly like the function $y = \sin\theta$, except that it starts with the value $\cos 0$, which is equal to 1. It is shifted by $\frac{\pi}{2}$ radians from the sine function. We can also say that there is a *phase difference* of $\frac{\pi}{2}$ *radians* between the two curves, or that they are *out of phase* by $\frac{\pi}{2}$ *radians*.

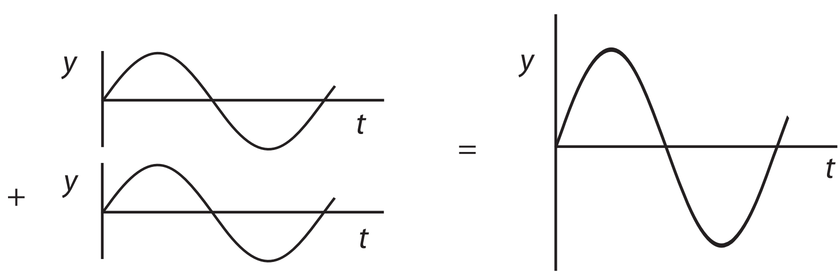
We can add one more feature to our description. The sine and the cosine vary between the values $+1$ and -1. The functions $A \sin\theta$ and $A \cos\theta$ vary between $+A$ and $-A$, where the maximum value, A, is called the *amplitude*.

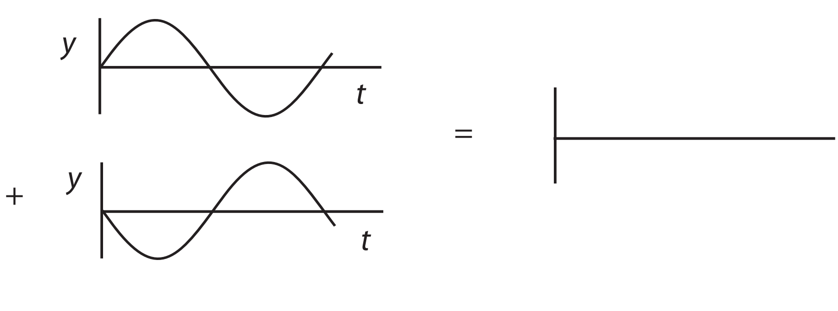
The amplitude is particularly important because it is related to the energy of the wave. Let's see what that relation is. Look again at the time variation $y = A \sin \frac{2\pi}{T}t$. As the value of y changes from zero to $+A$, to zero, to $-A$ and back, it moves with simple harmonic motion, as if it were on a spring with spring constant k. The energy of such a spring is $\frac{1}{2}kA^2$. This is the energy of the wave, proportional, as we see, to the square of the amplitude.

We can also define the *intensity* of the wave. It is the amount of energy that is transported by the wave per second, divided by the area through which it passes. Its SI unit is therefore the watt per square meter, $\frac{\mathrm{W}}{\mathrm{m}^2}$. Since it is proportional to the energy, it is also proportional to A^2.

Finally, we can combine the space and time variations in a single relation, $y = A \sin(\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t)$. At the time $t = 0$ this relation becomes the space variation $y = A \sin(\frac{2\pi}{\lambda}x)$. At this time (and at any other time) y varies *sinusoidally* with x. Similarly, at the point $x = 0$ (or at any other point) y varies sinusoidally with T.

Think of the quantity in brackets as an angle in radians. As it goes through its cycle, from zero





to 2π, y goes from zero to A, back to zero, to $-A$, and to zero again. Each crest ($y = A$), each trough ($y = -A$), and each of the points between represents a point with a certain *phase*. For a constant phase the quantity in brackets is constant. For a phase of zero, for instance, $\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t = 0$, or $\frac{x}{\lambda} = \frac{t}{T}$, or $\frac{x}{t} = \frac{\lambda}{T}$. We see that a point with constant phase moves with the speed $\frac{\lambda}{T}$, as we saw before. This also shows again that as we watch a crest, a trough, or any point with a particular phase, it moves with the wave speed $\frac{\lambda}{T}$ or $f\lambda$.

When two similar waves (same A, λ, T) arrive at a point, they might be *in phase*, i.e., with their crests arriving at the same time. In that case they reinforce each other, resulting in a variation at that point between $+2A$ and $-2A$. This is called *constructive interference*. If they are water waves, they will combine to have a crest twice that of each wave alone. If they are sound waves, the pressure will be twice that of a single one.

But if they arrive so that the crest of one ($y = A$) occurs at the same time as the trough ($y = -A$) of the other, they will cancel. This is called *destructive interference*. At the point where two water waves interfere destructively, the water will not be displaced. If two sound waves interfere destructively, there will be no sound.

EXAMPLE 5

What is the relation that describes a wave with a wavelength of 3 m, a period of 4 s, and an amplitude of 20 cm?

Ans.:
$y = 0.2\sin(\frac{2\pi}{3}x - \frac{2\pi}{4}t)$ or $y = 0.2\sin(2.09x - 1.57t)$

(We are not explicitly putting the units into this relation. But note that the "3" in the first fraction is a distance, 3 m, and the "4" in the second fraction is a time, 4 s, so that the quantity in the brackets has no units at all.)

It could also be the cosine function $y = 0.2\cos(2.09x - 1.57t)$.

This function has the value $A = 0.2$ m when $x = 0$ and $t = 0$, while the sine function is equal to zero when $x = 0$ and $t = 0$. The way in which y varies with x and t is the same in the two cases. One starts at the origin with the value zero, the other with its maximum value. Except for the phase difference between them the two graphs look the same.

EXAMPLE 6

What is the frequency of a wave described by the relation $y = 4\sin(2x - 3t)$, where all quantities are in SI units?

Ans.:
Here the coefficient of t is equal to 3 s. In the general equation it is $\frac{2\pi}{T}$ or $2\pi f$. Comparing the two we see that $2\pi f = 3$ and $f = \frac{3}{2\pi} = 0.48\,\text{s}^{-1} = 0.48\,\text{Hz}$.

11.3 Sound and musical scales

From the source and through the medium

How is a sound wave generated? If we sing a note, we cause our vocal cords to vibrate. If the sound comes from a string instrument it is the string that vibrates. The frequency is determined by the length of the string, by its mass, and by the tension. With a wind instrument, such as a flute or clarinet, it is the length of the vibrating air column that determines the frequency, and with it the "pitch" of the note (from squeaky high to booming low) that is produced. Quite generally, when the source of a wave is a vibrating object, the frequency of the wave is equal to the frequency of the vibration of the source.

The wave then propagates through a *medium*. For a wave along a string the string is the medium, for a sound wave in air it is the air. As the wave travels through the medium, it is no longer connected to the source, and its properties no longer depend on the source (except for the frequency given to it by the source). In the medium there is a displacement from equilibrium and a restoring force back to the equilibrium configuration. The restoring force depends on the medium. For a wave along a string, for example, it is determined by how tight the string is, i.e., by the tension of the string.

The velocity of the wave depends on the restoring force and also on the mass of the material that is displaced. We see that the velocity is determined by the properties of the medium in which the wave travels. For a string under tension with a force F, and mass per unit length, $\frac{M}{L}$, the wave velocity is $\sqrt{\frac{F}{M/L}}$.

Once the frequency and the velocity are fixed, the value of the wavelength follows from $v = f\lambda$. In the cases that we have described it is a dependent variable that takes its value from the frequency of the source and from the velocity as determined by the medium. The two together determine the magnitude of the wavelength in the medium.

Standing waves

The waves that we have been talking about are waves that transport energy away from the source. They are *traveling waves*. If they encounter a barrier they are (completely or partially) reflected. This happens, for instance, when a water wave or a sound wave hits a wall, or when a wave along a rope arrives at the end of the rope where it is tied to a wall or tree.

Such a reflection can give rise to a quite different type of wave. The most direct demonstration is to take a string or rope, fixed at one end, and to shake the other end up and down. If the frequency at which you shake the string is just right, there will be a *standing wave* on the string: each point on the string goes up and down in simple harmonic motion, while the amplitude of the variation varies from point to point.

We can see in more detail what happens mathematically. We'll need the trigonometric relation that says that $\sin(A+B) = \sin A \cos B + \cos A \sin B$, and the one that says that $\sin(A-B) = \sin A \cos B - \cos A \sin B$, so that $\sin(A+B) + \sin(A-B) = 2 \sin A \cos B$.

Let's talk about a wave on a string, traveling to the right, described by $y = A \sin\left(\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t\right)$. The reflected wave goes to the left, but is otherwise identical, with the same values of A, λ, and T, so that it is described by $y = A \sin\left(\frac{2\pi}{\lambda}x + \frac{2\pi}{T}t\right)$. Their sum represents the standing wave. We can write this sum, using our trigonometric relation, as $y = 2A \cos \frac{2\pi}{T}t \sin \frac{2\pi}{\lambda}x$.

What does this standing wave look like? The variation in space is described by $\sin \frac{2\pi}{\lambda}x$. The rest is the amplitude: $2A \cos \frac{2\pi}{T}t$. We have a sine wave in space, whose amplitude goes up and down sinusoidally with time.

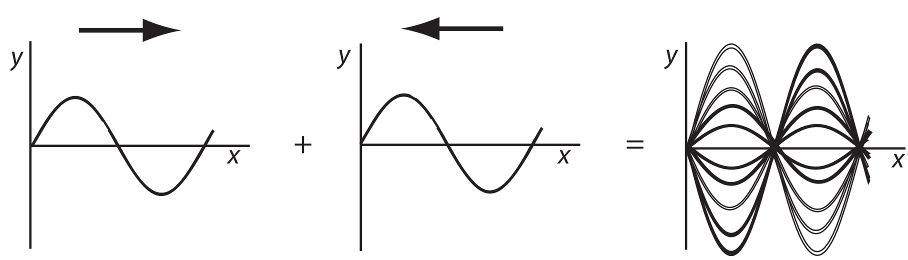
You can see why the frequency with which you move the free end of the string has to be just right. The fixed end is a point with no vibration at all. (Such a point is called a *node*.) The end that you are holding is a point of maximum vibration. (It is called an *antinode*.) The distance between two adjacent nodes is $\frac{1}{2}\lambda$. The distance between a node and an adjacent antinode is $\frac{1}{4}\lambda$. The length of the string between the antinode at your hand and the node at the fixed end can therefore be $\frac{1}{4}\lambda$, or $\frac{1}{4}\lambda$ plus some number of half wavelengths. Standing waves are possible only at these wavelengths.

The velocity is, as usual, determined by the mass of the string and its tension. Since v is fixed, and only certain wavelengths (λ) are possible, it follows from $v = f\lambda$ that only certain discrete frequencies are possible.

The relation between the length of the string and the wavelength is even simpler when both ends of the string are fixed. The length of the string can then only be a whole number of half wavelengths.

Look at the unexpected and significant result that we have come to. We started with quantities, the frequency and the wavelength, that seemed to be able to take on any value. We described a situation where the ends are constrained, either by keeping one or both fixed or by holding one end and causing it to oscillate up and down. The conclusion that we came to is that then only certain values of the wavelength and frequency are possible. We have gone from continuous quantities to discrete or *quantized* values. What is it that brought about this change?

In an infinitely long string there would be no reflection and no limit to the possible values of the wavelength or frequency. The discreteness of



the values, i.e., the quantization of the spectrum of allowed values of the frequency, comes about as a result of what we do at the ends of the string. In other words, it comes about as a result of what are called the *boundary conditions*. We will see in the next chapter that there are strong similarities to the quantization of the energy spectrum of atoms, i.e., to the way in which the discrete allowed energies of atoms arise.

EXAMPLE 7

Go to the PhET website and open the simulation *Wave on a String*. Set the tension to its maximum and "damping" to zero. Select "oscillate" and "fixed end." Set the frequency to 50 and the amplitude to 1 or 2. Observe the standing wave. Why is it called a "standing wave"?

Try to find another combination of frequency and tension that also leads to a standing wave.

Ans.:
The points where there are maxima and minima (the antinodes) and the points where there is no motion (the nodes) remain in place. There is no wave propagation.

It is not easy to find other combinations of frequency and tension that produce a standing wave. This is mainly because the simulation does not allow these two quantities (or at least one of them) to be varied continuously. The boundary conditions determine what happens at the ends. They can be nodes, antinodes, or something between the two. The distance between two nodes is one-half wavelength. The pattern of nodes and antinodes has to "fit" between the two ends. (Here is another combination that you can try. It comes close: tension at 0.8 of the maximum value and frequency setting at 25.)

We have talked about waves on strings because they are so visible. Vibrating air columns have similar characteristics, and are particularly interesting because they form the basis of musical wind instruments. In a flute or recorder (the wooden, now often plastic flute that was common in the renaissance and baroque eras and that had a strong revival in the twentieth century) the air columns are set into vibration by the breath of the player. In the clarinet and the oboe the mouthpiece has reeds that vibrate and transmit their motion to the air column.

Standing waves can be set up in air columns, similar to those on oscillating strings. Here also the boundary conditions determine the relation between the length of the column and the wavelength of the standing wave. There are two quantities that vary sinusoidally. One is the pressure, the other is the displacement, i.e., the distance by which the air moves back and forth. We will use the pressure. There is a pressure node at the open end of the tube where the pressure is (approximately) equal to that of the air outside the tube. There is a pressure antinode at the closed end.

In a tube either open at both ends or closed at both ends, i.e., with the same boundary conditions at the two ends, the length of the vibrating air column is a whole number of half wavelengths. In a tube closed at one end and open at the other it is $\frac{1}{4}\lambda$ plus some whole number (including zero) of half wavelengths.

EXAMPLE 8

A string is 90 cm long and is fixed at both ends. The speed of waves along the string is 135 m/s. What are the three longest possible wavelengths and the corresponding frequencies of standing waves on this string?

Ans.:
There are nodes at each end of the string. The distance between nodes is $\frac{1}{2}\lambda$. Hence the three values are given by $L = \frac{1}{2}\lambda$, $L = (2)(\frac{1}{2}\lambda)$, and $L = (3)(\frac{1}{2}\lambda)$, or $\lambda = 2L$, L, and $\frac{2}{3}L$, i.e., 1.80 m, 0.90 m, and 0.60 m.

The corresponding frequencies are 75 Hz, 150 Hz, and 225 Hz.

EXAMPLE 9

An air column, 60 cm long, is open at one end and closed at the other. What are the three longest wavelengths and the corresponding frequencies of standing waves on this air column?

Ans.:
There is a node at one end and an antinode at the other. The distance between one antinode and the nearest node is $\frac{1}{4}\lambda$. The next distance is $\frac{1}{4}\lambda + \frac{1}{2}\lambda$, or $\frac{3}{4}\lambda$. The one after that is $\frac{1}{4}\lambda + \lambda$ or $\frac{5}{4}\lambda$, so that $\lambda = 4L$, $\frac{4}{3}L$, and $\frac{4}{5}L$, i.e., 2.40 m, 0.80 m, and 0.48 m.

To find the corresponding frequencies we have to know the wave velocity. The velocity of sound in air at 20°C (68°F) is 344 m/s. With this value the frequencies are $\frac{344}{2.40} = 143\,\text{Hz}$, $\frac{344}{0.80} = 430\,\text{Hz}$, and $\frac{344}{0.48} = 717\,\text{Hz}$.

Resonance

Frequencies at which a system such as a string or an air column can vibrate easily are called its *natural* or *resonant* frequencies. It requires a relatively small amount of energy to produce sustained oscillations at these frequencies.

Here is a demonstration of a natural frequency: sit on a table or high chair so that your leg dangles. Let it swing freely from the knee. Now make it swing, first at half the frequency and then at twice the frequency. This illustrates that there is a natural frequency that is easy to bring about. Other frequencies are quite unnatural, and require more effort.

Another example is that of two similar musical strings near each other, with the same length and under the same tension, so that they have the same natural frequency. When one is made to sound, the other will "resonate," i.e., it will vibrate also. The first string is the source of the waves. The air between them transmits the waves. It serves as the coupling between the two. The second string receives the wave and is pushed back and forth so as to vibrate also. The amplitude of its vibration will be largest at its own natural or resonant frequency. Some instruments, such as the sitar, have a set of strings that vibrate only through this mechanism.

Scales

What makes a pleasing sound? We are now asking a question about how sounds are perceived, i.e., how our brain interprets the vibrations received by the ears. We are therefore leaving the objective description provided by physics for the subjective question of how we *feel* when we hear certain sounds. This is different for different people and depends on what we are famliar with.There are nevertheless some guidelines. It turns out that two notes together sound "special" to us when the ratio of their frequencies is small. If the ratio is two, for example, we say that the two notes are an *octave* apart. We can subdivide the octave interval, and this is done differently in different cultures. The western eight-note scale has two versions, each of seven intervals. In the *major scale* the intervals are in the ratios to the first note of $1, \frac{9}{8}, \frac{5}{4}, \frac{4}{3}, \frac{3}{2}, \frac{5}{3}, \frac{15}{8}, 2$. In the *minor scale* the ratios are $1, \frac{9}{8}, \frac{6}{5}, \frac{4}{3}, \frac{3}{2}, \frac{8}{5}, \frac{9}{5}, 2$.

These notes are sometimes called the members of the *natural* scales. A problem arises if we want to start the scale with notes of different frequencies. In the *C-major* scale we start with C. The next note, D, has a frequency $\frac{9}{8}$ as large. The other notes follow with their frequencies, D, E, F, G, A, B, and again C, now an octave higher than at the starting point. If we want to start with G, to get the *G-major* scale, we get a whole new set of frequencies. This is, in fact, what happens when people sing together, or play instruments that produce notes whose frequencies can be varied continuously, such as those in the violin family. But what do we do if we want to use instruments, such as the piano, where each key produces a sound with a fixed frequency?

The scale that is used for pianos is a compromise. We divide the octave into 12 equal intervals, so that the frequency of each note is larger than that of the previous one by $\sqrt[12]{2}$ or 1.0595. The resulting notes of what is called the *equal tempered scale* do not sound as "pure" as those of the natural scales, but with these fixed intervals we can start our scales with any note. *A capella* singers (without instruments) and string quartets are not constrained by the compromise, and can play the notes of the natural scales. In the equal tempered scale the intervals can be *half notes* with ratios of 1.0595 or *whole notes* with ratios of 1.0595^2 or 1.225. The natural scale is approximated by using five whole-note intervals and two half-note intervals. In the C-major scale, for instance, the half-note intervals are between E and F and between B and C.

The Doppler effect

When a police car comes toward you with its sirens on and then passes you, you hear a change in the pitch (the frequency) of the siren's sound. That's the Doppler effect, first suggested for lightwaves from stars by Christian Doppler in 1842, and studied for sound waves soon after.

Let's see how it comes about. Think of a bus stopped near a store. Passengers leave it, and immediately start walking toward the store. The bus is our source, "emitting" passengers. The store is the receiver.

Suppose a passenger leaves the bus every 5 s and walks at a velocity of 1.2 m/s toward the store. One passenger arrives at the store every 5 s. The frequency of passengers leaving as well as arriving is $\frac{1}{5}$ or 0.2 passengers per second. In 5 s a passenger goes $(1.2)(5) = 6$ m, so that the passengers are 6 m apart as they walk.

Now let's get to the interesting part. Let the bus move toward the store at a speed of 0.4 m/s. The passengers leave the bus, one every 5 s, as before, and walk toward the store with the same speed as earlier. The frequency with which they leave the bus is still the same, but the frequency with which they arrive at the store is different. Each successive passenger has a smaller distance to go from the bus to the store. In the 5 s between passengers leaving the bus, the bus travels $(5)(0.4) = 2$ m. Instead of being 6 m apart as they walk they are only 4 m apart. Still moving at 1.2 m/s, they now arrive at intervals of $\frac{x}{v} = \frac{4}{1.2}$ or 3.33 s, i.e., with a frequency of $\frac{1}{3.33} =$ 0.3 passengers/second.

EXAMPLE 10

Go to the Java Applet on the Doppler effect, e.g., at http://lectureonline.cl.msu.edu/~mmp/applist/doppler/d.htm.

(a) Click on the gray rectangle to make a blue dot appear. This is the source of the waves. You can stop the wave pattern by clicking on "s." It will resume when you click on "s" again. (Think of the dot as representing the siren of a police car, and the pattern representing the spread of the sound waves.) Stop the pattern and estimate the wavelength on the screen.

(b) Use the mouse to drag an arrow toward the right-hand bottom corner. Use a velocity less than $\frac{v}{v_s} = 1$. You are the observer of the waves at that corner. What has happened to the wavelength, frequency, and velocity that you observe?

Ans.:
While the source moves toward you the waves arrive with greater frequency. The speed of sound depends on the medium and does not change. The wavelength (the distance between successive lines on the wave pattern) decreases.

You can see that if the bus moves in the opposite direction, the frequency at the store (the "receiver") is smaller. The same considerations can be used to describe the frequency changes that occur when the receiver moves toward or away from the source.

The situation is analogous when a wave is emitted by a moving source. When it arrives at a stationary receiver, the frequency is larger when the source moves toward the receiver and smaller when it moves away. Motion by the receiver has a similar effect.

The applications of the Doppler effect are of great importance. The original paper of 1842 described the possibility of measuring the speed of stars with respect to the earth. However, the Doppler effect for light was not observed until 1901. Later (in 1912) it was shown that light from other galaxies was shifted to lower frequencies ("redshifted"). The observations led to *Hubble's law*, formulated by Edwin Hubble in 1929, which says that the universe is expanding at a rate that is now called *Hubble's constant*.

The reflection of microwaves ("Radar") is widely used for the location of weather patterns. Rain, snow, sleet, and hail, even insects and dust reflect the electromagnetic waves, and the Doppler effect allows the speed of the reflecting objects to be determined. Other applications are the measurement of the speed of moving cars with microwaves, using a *Radar detector*, and of the rate of blood flow.

11.4 Maxwell's great contribution: electromagnetic waves

Maxwell's equations

They are known as *Maxwell's equations* and are the fundamental relations of electromagnetism, but Maxwell didn't invent them. All four of them were there before him. Maxwell saw that one of them was incomplete, added one more term, and so opened up a whole new world. Let's see what they are.

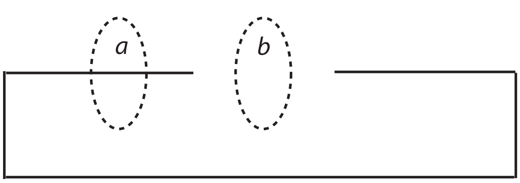
The first is Gauss's law, equivalent, as we know, to Coulomb's law: the flux of the electric field out of any closed surface is proportional to the net charge inside it.

The second is Gauss's law for magnetic fields, easy to write down, since there seem to be no isolated poles, no *monopoles*: the magnetic flux out of any closed surface is zero.

The third is Faraday's law, $\mathcal{E} = -\frac{d\Phi}{dt}$. The induced emf around any loop is equal to the rate of change of the magnetic flux through the loop. We can also describe it differently: the emf drives the current in a circuit; it is another way to talk about the electric field that makes the charges go around. The magnetic flux depends on the magnetic field. In other words, the electric field that is created is proportional to the rate of change of the magnetic field.

The fourth is the one we have to look at in more detail. In its incomplete form it is Ampere's law. For any path around a current, I, the length of the path multiplied by the component of the magnetic field along the path is equal to $\mu_0 I$.

Now what did Maxwell see? He said suppose the wire is not continuous. There can still be a current, either for a short time, or one going back and forth (an alternating current, or "AC").

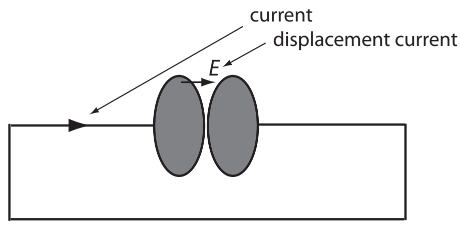


What happens as we take loops at *a* or *b*? Experimentally, there is a magnetic field at *b* as well as at *a*. If Ampere's law is to hold, the result must be the same regardless of whether the path encircles the wire or the empty space. There could also be a more strangely shaped path that is neither clearly around the wire nor around the part where there is no current. There must be something else, something like a current, also in the broken part!

He showed what it was and gave it the name *displacement current*. (The name is not very helpful, since it doesn't tell you anything about the concept.)

Let's connect the ends of the wires to large plates. (Assume that they are infinitely large.) The electric field between them, E, is equal to $\frac{\sigma}{\epsilon_0}$. If there is (temporarily) a current I, it is equal to $\frac{dQ}{dt}$ or to $\epsilon_0 A\frac{dE}{dt}$.

While there is a current in the wire, charge is moving to and from the plates. In the space



between the plates there is then a changing electric field. The quantity $\epsilon_0 A\frac{dE}{dt}$ is equal to the current in the wire leading to and from the plates. By adding this term to I there are now two terms that together remove the earlier discontinuity. In each part of the circuit there is either a current or a displacement current. The field times the path, $2\pi rB$, is now equal to $\mu_0(I + \epsilon_0 A\frac{dE}{dt})$.

The adding of this one term to the fourth of the equations was so significant that the whole set is now known as *Maxwell's equations*.

What is it that is so important about the term with $\frac{dE}{dt}$? It is like Faraday's law, but with the role of the electric field and the magnetic field reversed. The two together lead directly to the existence of electromagnetic waves.

Let's see how this comes about. Faraday's law shows that an emf, and therefore an electric field, can be produced by a changing magnetic field. The magnetic field is produced in the whole region, not just at a point. The term in $\frac{dE}{dt}$ (the displacement current) shows that, in turn, a changing electric field produces a magnetic field, again in the surrounding region.

If the field (B or E) changes with time like a wave (i.e., sinusoidally), the slope, or derivative ($\frac{dB}{dt}$ or $\frac{dE}{dt}$), will also change with time in a similar way. Faraday's law shows that a changing magnetic field gives rise to a changing electric field. The fourth equation now shows that a changing electric field produces a changing magnetic field. The two equations together show that a change in either one of the fields gives rise to a changing field of the other kind. The changing magnetic field produces a changing electric field, leading, in turn, to a changing magnetic field, which leads to a changing electric field, and so on. The two kinds of fields chase each other through space, each giving rise to the other, propagating on their own, cast loose from their source, continuing their journey as an electromagnetic wave.

Of course this can be shown rigorously, mathematically. The velocity of the waves can be calculated from the equations, and is $\frac{1}{\sqrt{\mu_0\epsilon_0}}$. Since $\frac{\mu_0}{2\pi} = k'$ and $\frac{1}{4\pi\epsilon_0} = k$, $\mu_0\epsilon_0 = \frac{k'}{2k} = \frac{10^{-7}}{9\times10^9}$ or $\frac{1}{9\times10^{16}}$ SI units, and $c = \frac{1}{\sqrt{\mu_0\epsilon_0}}$ is 3×10^8 m/s.

These are not the units or quantities that were used in Maxwell's days, but the result is the same. Maxwell knew (from experiments by others) that the speed of light was about 3×10^8 m/s. He not only showed that the equations predict the existence of electromagnetic waves, but it also became clear that lightwaves must be just this kind of wave.

Look at what a remarkable result this is: μ_0 and ϵ_0 are the constants that are needed to describe the forces between charges at rest and between currents. Now they are seen to describe also the velocity of electromagnetic waves, which evidently include lightwaves.

Does this mean that electromagnetic waves can be generated with only charges and currents? Yes, Hertz showed that, but not until much later, in 1888, more than two decades after Maxwell's work. Today every radio and TV transmittor, every microwave oven and cell phone system do just that.

The electromagnetic spectrum

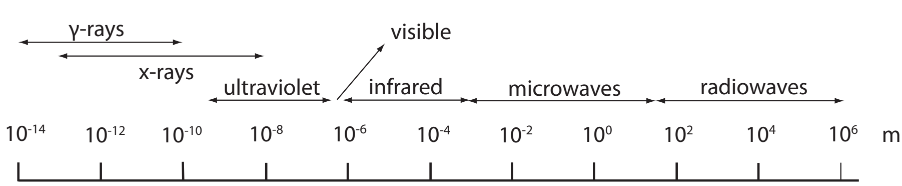
Electromagnetic waves can have wavelengths larger than hundreds of meters and smaller than the diameter of a proton. They are united in their speed through empty space, c $(= f\lambda)$, equal to 3×10^8 m/s. They can be classified by their frequency, or equivalently by their wavelength, and all form part of the *electromagnetic spectrum*. We give names to the different regions of the spectrum, but the way we do it is not very consistent. Some, like visible light, are characterized by their receiver, and some, like gamma rays, by their emitter. The regions overlap, and their boundaries are not precise.

Light is what we see. The wavelengths to which the eye is sensitive range roughly from 400 nm at the small-wavelength, high-frequency (violet) limit to 700 nm at the large-wavelength, low-frequency (red) end of the visible spectrum.

All matter consists of charges, and all matter is in motion. In gases the atoms move and collide with other atoms. In solids they oscillate about their equilibrium positions. This is their *thermal* motion. The accelerating charges lead to changing electric fields, hence to changing magnetic fields, and so to electromagnetic waves. Every object radiates electromagnetic waves with a range of frequencies that depends on its temperature. The sun and the other stars are so hot that they emit all parts of the visible spectrum and appear white. The part of the spectrum emitted by hot bodies with frequencies larger than those that are visible is called ultraviolet, and the part with lower frequencies is called infrared. A hot stove may emit radiation with frequencies into the visible part of the spectrum and be "red hot," but all objects, regardless of their temperature, emit infrared radiation. This is so for everything around us, including our own bodies.

Gamma rays are electromagnetic waves emitted by nuclei. The name "x-rays" is given to the radiation emitted by atoms. The frequencies of x-rays are generally smaller than those of gamma rays, but there is some overlap. (We will see later that the frequency of the radiation is related to its energy. The greater frequency of gamma rays is related to the much greater spacing of the energy levels of nuclei compared to those of atoms. This is because nuclei are held together by the strong nuclear force, while the energy that holds the electrons to the nucleus in atoms is the much weaker electric force.)

The term *x-rays* is also used for the electromagnetic waves that are emitted by electrons that are accelerated when they are not bound in atoms. The reason is that both kinds are emitted by an *x-ray tube*, where a beam of electrons is stopped when it hits a target made of tungsten or another heavy metal. (There is more about this in the next chapter.) A particularly strong x-ray source is the *synchrotron*, a device in which electrons are accelerated by a magnetic field so that they move in a circle.



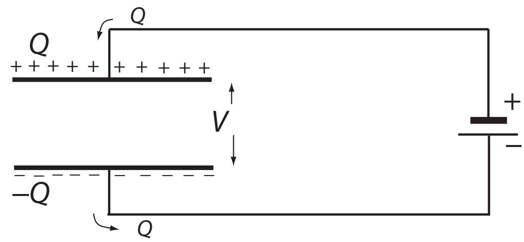
Electrons moving back and forth in a circuit or in an antenna give rise to *radio waves*. They usually have wavelengths from hundreds of meters down to a few meters. The alternating current that we normally use has a frequency of 60 Hz, and generates electromagnetic waves with a correspondingly long wavelength. Microwaves are generated by electrons oscillating in vacuum tubes and have smaller wavelengths, of the order of centimeters.

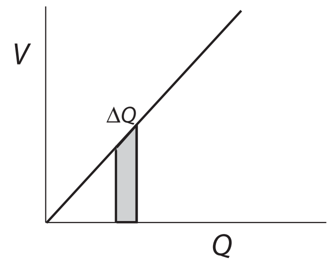
The capacitor and the energy of the electric field

We introduced the capacitor in the example at the end of Chapter 8. It consists of two plates, one carrying a charge Q and the other a charge $-Q$, with a potential difference V between them. If we use the approximation that the area of each plate, A, is large, and the distance, d, between them is small, then the electric field, E, in the volume between the plates is uniform, and $V = Ed$. The surface charge density, σ, is equal to $\frac{Q}{A}$. The *capacitance* is defined as $\frac{Q}{V}$. It is the amount of charge that can be stored on one of the plates for each volt of potential difference between them. It is equal to $\frac{\sigma A}{Ed}$, and since the field between two large (assumed infinite) plates is $\frac{\sigma}{\epsilon_0}$, it is equal to $\frac{\epsilon_0 A}{d}$. We see that the capacitance depends only on the geometric configuration, i.e., on the size of the plates and the distance between them, and not on the amount of charge or the potential difference.

Consider what happens as we charge a capacitor by connecting it to a battery. Charge transfers from one plate to the other. At first there is no potential difference, and it takes little work to transfer the first bit of charge, ΔQ. As more charge accumulates it takes more and more work to transfer an amount ΔQ, as the work, $V\Delta Q$, becomes larger. To transfer the amount Q takes an amount equal to the area under the curve of V against Q. This is equal to $\frac{1}{2}QV$.

This is the energy stored in the capacitor. It can also be written as $\frac{1}{2}CV^2$. If we substitute the expression for the capacitance, $C = \frac{\epsilon_0 A}{d}$, and $V = Ed$, we see that the energy, U, is $\frac{1}{2}\epsilon_0 AdE^2$, and the energy per unit volume, $\frac{U}{Ad}$, is $\frac{1}{2}\epsilon_0 E^2$. This expression for the energy per unit volume, or the *energy density*, holds also for other electric fields.

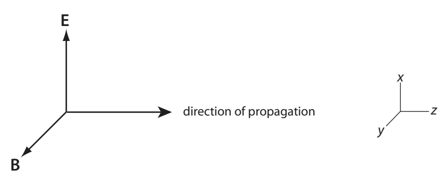




It is possible to show also that the energy density of the magnetic field in a long solenoid is $\frac{B^2}{2\mu_0}$. This is also the energy density in any other magnetic field.

The propagating fields

Look again at a loop of wire in which a current is induced by a changing magnetic field. The induced current is there as a result of the induced emf. There is an induced electric field along the wire of the loop in accord with Faraday's law. The electric field is in the plane of the loop, while the magnetic field is at right angles to the plane of the loop: **E** and **B** are perpendicular to each other.



In an electromagnetic wave **E** and **B** are at right angles to each other and perpendicular to the direction in which the wave travels. (The vector **E** × **B** is at right angles to both fields and is in the direction of propagation.)

As the wave travels, both **E** and **B** oscillate, in phase, both in space and in time. In other words, as the wave passes a particular point, the magnitude E at the point varies from its maximum value E_{max} to 0 to $-E_{max}$ and back with the frequency of the wave, as $E = E_{max} \sin \omega t$. At the same time the magnitude B of the magnetic field varies from B_{max} to 0 to $-B_{max}$ and back, in phase with E. Similarly, a snapshot of the fields at a particular moment shows the sinusoidal variation as a function of z, the direction of propagation, i.e., both $E_{max} \sin kz$ and $B_{max} \sin kz$, along the direction of propagation of the wave.

In an electromagnetic wave both kinds of fields travel together and both carry energy. This is one of the most important features of the waves: the fields have energy, and the electromagnetic waves transport that energy.

Every electromagnetic wave, whether it is a radio wave, microwave, lightwave, x-ray, or gamma ray, consists of electric and magnetic fields moving through space together. Each wave carries energy. This is how we get the energy from the sun that makes our life on earth possible. This is how we get the signals to our radios, TV sets, and cell phones, the radiation that warms our skin, and that which we see with our eyes.

How do we know the waves are there? We don't see the fields or feel them directly. We know about them only when they exert forces on electric charges. We have to return to the beginning: in an electric field there is a force on a charge. In an electromagnetic wave there is an oscillating electric field. A charge in the path of such a wave is set in motion, oscillating under the action of the force provided by the wave's electric field. This is what happens when an electromagnetic wave is detected, whether it is in an antenna, on our skin, or in the eye.

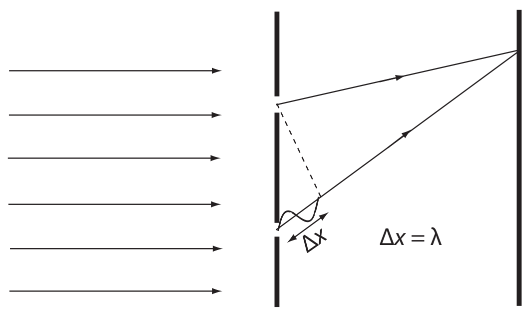
When we discussed what we mean by the word "real" (in Chapter 5) we shied away from a definition of reality. Regardless of the definition, there are few who would doubt that electric fields, magnetic fields, and their joint dance through space are real.

11.5 Observing interference of light

Young's double slit experiment

That light waves are electromagnetic waves became clear from Maxwell's work in about 1862. But the wave properties of light had been established by 1801, principally when Thomas Young showed that light could exhibit the properties of interference.

Wave properties become apparent primarily when the waves encounter an opening or an obstacle that is not too much larger than their wavelength. For visible light this is between about 400 nm for violet light and 700 nm at the red end of the spectrum. These distances are so small that we are normally not aware of the wave properties of light, and it takes special efforts to observe them. Young's double slit experiment provides the most famous and direct demonstration of the interference of light, and hence of the wave nature of light. Here is the experiment.



Two narrow openings, or slits, are illuminated by a source of light of a single color. Each then acts as a new source from which light spreads out. When the waves from the two slits arrive at a point on a surface or screen, they have traveled different distances. One of them has gone further than the other by an amount Δx. If Δx is just one wavelength long, the crests of the two waves will arrive at the same time.

There is then constructive interference, and there will be a bright line on the screen with the addition of the light from the two slits. The same is true if Δx is equal to 2, 3, or some other whole number of wavelengths. But if, for some place on the screen, Δx is $\frac{1}{2}\lambda$, the crest of one wave arrives there at the same time as the trough of the other, and there is destructive interference. The two waves cancel, and there is no light on the screen.

The result is that there are alternating bright and dark lines on the screen. They are the result of the interference of the lightwaves from the two slits. This pattern could not be there if light did not have the properties of waves. If you imagine shooting "bullets" of light rather than waves, there would be no interference, and no such pattern of bright and dark lines.

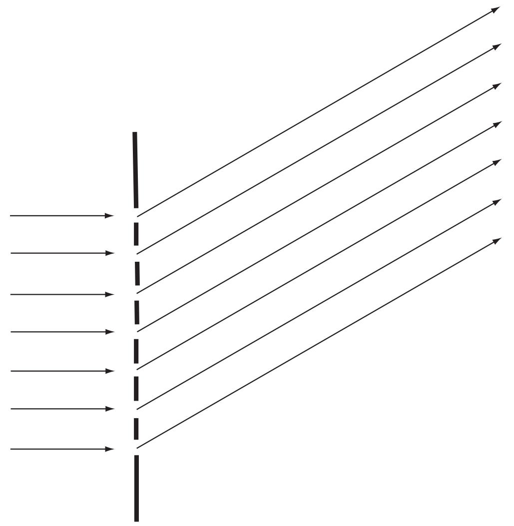
The spacing between the lines gives us a measure of the wavelength of light. The experiment shows that it ranges from about 400 nm to about 700 nm, and that each value of the wavelength in that range corresponds to a particular color. For the interference pattern to be clearly seen, the source of light needs to be of a single wavelength and color, i.e., to be *monochromatic*.

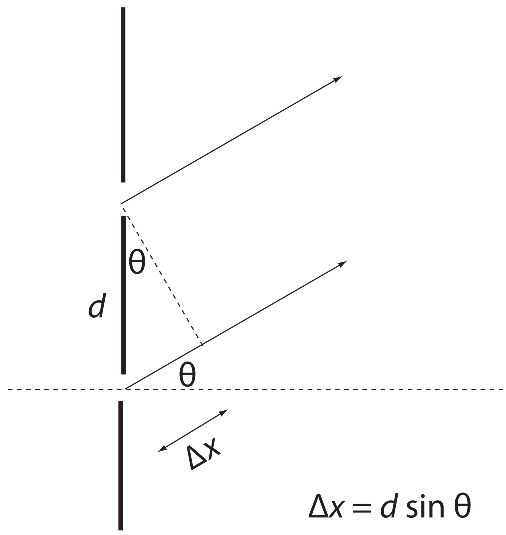
EXAMPLE 11

Go to the PhET website and open the simulation *Wave Interference*. Select "Light." Click on "show screen" and observe the waves.

Select "two slits" and move the barrier close to the source. (You can drag it with the mouse or move the pointer in the set of tools on the right. You may also want to change the slit width.) Again observe what you see. Click on "intensity graph" and look at the result. How does the pattern change when you change the wavelength?

Ans.:
The pattern on the screen shows the alternating bands of light and dark that result from the interference of the light from the two slits. The intensity graph shows the variation of the intensity along the screen. The distance between successive minima becomes smaller when the wavelength decreases.





The diffraction grating

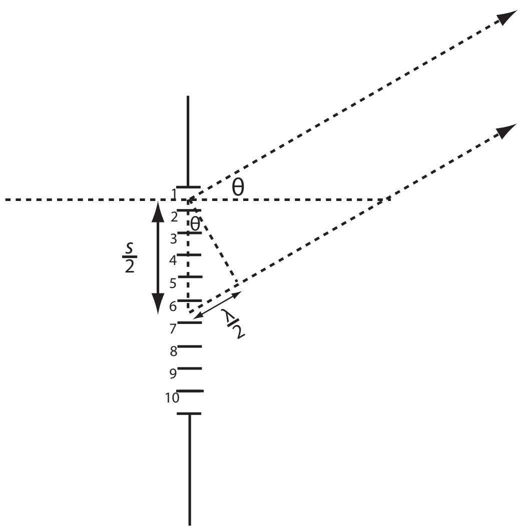
There are other experiments that show the interference of light, and we will explore some of them. The first is the diffraction grating. The idea is the same as in Young's double slit, but there are many slits.

Again each slit acts as a source of waves. As in the double slit experiment, the interference pattern is determined by the path difference between the waves from the different slits. Look at the direction at the angle θ to the original direction of the light where the path difference between the light from the first slit and that from the next one is one wavelength. Then for the next one it is 2λ, then $3\lambda, 4\lambda$, and so on. The light from all of the slits will interfere constructively. If the distance between adjacent slits is d, we see that the angle θ is given by $\sin\theta = \frac{\lambda}{d}$. There is also constructive interference from all the slits when $\sin\theta = \frac{2\lambda}{d}$, and, in general, for $\sin\theta = \frac{n\lambda}{d}$, where n is an integer. (For all other angles there is destructive interference. Since there are many slits, there is always another one from which the light is out of phase by half a wavelength.)

For light with a single wavelength there is constructive interference at the values of θ for which $\sin\theta = \frac{n\lambda}{d}$. If the beam consists of light with different wavelengths the angle for constructive interference is different for each. The grating then separates the light, and allows us to see the range of wavelengths that are present. The distribution of colors is called a *spectrum*. More precisely, a spectrum is a graph of the intensity of radiation as a function of wavelength (or another quantity related to the wavelength).

Single slit diffraction

A diffraction grating could also be called an "interference grating." The word "diffraction" is commonly used instead for some interference phenomena, such as the pattern that occurs when light passes through a single slit. In that case alternating dark and bright bands occur because of the interference of light from different parts of the slit.



To see this we divide the slit into parts. The figure shows a slit of width s divided into 10 strips. We're going to compare the light from the five slits of the upper half to the light from the five strips of the lower half. Let's look at the first strip in each of the halves, i.e., the first and the sixth of these strips, at an angle such that there is destructive interference of the light from the two. The two strips are separated by a distance $\frac{s}{2}$. For destructive interference the path difference Δx is $\frac{1}{2}\lambda$, so that $\sin\theta = \frac{\frac{1}{2}\lambda}{\frac{1}{2}s}$, or $\sin\theta = \frac{\lambda}{s}$.

At this angle there is also destructive interference from the second and seventh strips, as well as for the third and eighth, the fourth and ninth, and the fifth and tenth. We see that at this angle there is destructive interference for every part of one-half of the slit and a corresponding part of the other half. At this angle, therefore, there is a dark band on the screen.

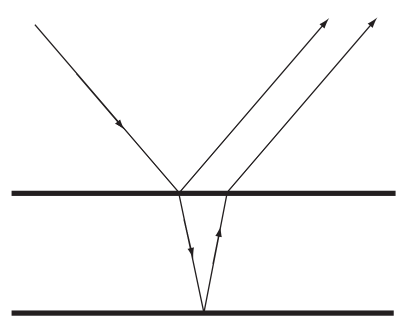
A similar analysis can be made for other angles and other path differences. The result is that there are again bands of bright and dark fringes, as in the double-slit experiment, but their origin and the angles at which they occur are different.

We can extend the same considerations to another situation. We have looked at light stopped by a barrier, except for the single slit in it. Suppose we reverse the arrangement, so that light continues everywhere *except* for the space that was occupied by the slit. We might guess that the same kind of pattern will occur, and it does! Other obstacles in the path of a light beam can also give rise to *diffraction patterns*.

The lines or bands that we see in Young's double slit experiment, with a diffraction grating, and in single-slit diffraction, have the shapes of the slits. A similar effect occurs when sunlight passes through the spaces between the leaves of a tree. The pattern of light and shadows is to a large extent the result of diffraction, and shows the image of the sun. We are not usually aware of that, and it is startling to see the effect at the time of a solar eclipse, when the source of the light, the sun, is no longer round. When the sun is half obscured, for example, the light between the shadows, still consisting of images of the sun, is seen to be made up of half circles.

Thin films

A common interference effect occurs when light is reflected from both the bottom and the top of a thin film.

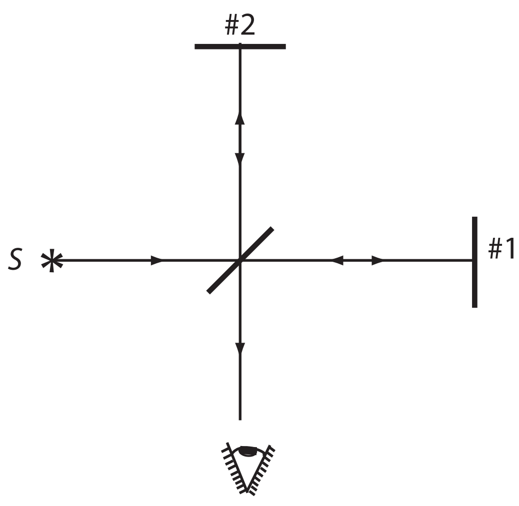


For a beam incident at right angles to the film, the light reflected from the bottom goes farther by a distance $2d$, where d is the film thickness. The nature of the interference depends on the relation between this path difference and the wavelength of the light. Here we have to know that the wavelength in the film is different from that in air: it is smaller by the factor n, the index of refraction, i.e., $\lambda_{\text{film}} = \frac{\lambda_{\text{air}}}{n}$. (On the figure the light rays are shown at angles different from 90°, so that the rays reflected from the top and from the bottom are separated.)

Since light of different colors has different wavelengths, the conditions for constructive and destructive interference will be different for different colors. The colors seen in soap films, and in oil films on water, are interference patterns. They come about because of the differences in wavelength, because the path difference is different at different angles, and as a result of thickness variations in the films.

The Michelson interferometer

A fascinating application of interference that had a profound influence on the history of physics is the Michelson interferometer. It allows measurements of distance with great precision, because they are made in terms of units equal to the wavelength of light. It can also be used to compare the velocities of light in two directions at right angles to each other. This was done in 1887 in the *Michelson–Morley experiment*. The experiment showed that there was a fundamental difference between the behavior of lightwaves and mechanical waves, indicating that Newtonian mechanics could not adequately describe electromagnetic phenomena. It paved the way for the development, 18 years later, of the special theory of relativity.



A half-silvered mirror lets some of the light from a monochromatic source through to mirror #1. Some of it is reflected to mirror #2. The light returns from mirror #1, and some of it is reflected toward the eye of the observer. Some of the light that is reflected from mirror #2 goes through the half-silvered mirror to the eye. The two rays combine and give rise to interference, the nature of which depends, as before, on the path difference. If the path difference is equal to a whole number of wavelengths, there is constructive interference. If it is a whole number of wavelengths plus a half wavelength, there is destructive interference.

The interferometer can be used to measure distances in terms of the wavelength of light. If one of the mirrors, #1 or #2, is moved through $\frac{1}{4}\lambda$, the path difference changes by $\frac{1}{2}\lambda$. Constructive interference changes to destructive interference, and what the eye sees as it looks along the line of the diagram changes from bright to dark.

At an angle to the lines of the light path in the diagram the distances are larger. What is observed by the eye is an interference pattern consisting of a series of concentric rings, alternately bright and dark. As one of the mirrors is moved, the rings move inward or outward. A bright ring is replaced by a dark ring when one of the mirrors is moved through a distance of $\frac{1}{4}\lambda$.

Coherence

In all our examples of ways to exhibit the interference of light there was a single source, with a split into different paths from the source to a place where they recombined. You might ask why we cannot observe interference with beams from two separate light bulbs, instead of going through the more or less elaborate splitting procedure. The answer lies in the nature of the radiation emitted from a hot object, such as the filament in a light bulb. The atoms of the filament emit flashes of light, first one, then another, and another, randomly and independently. The flashes come from different places and at different times.

This is very different from the beams that we described earlier, which come from the same source, split into two or more parts along different paths, and then recombine. The two different parts in Young's double-slit experiment, for example, start out together (in phase) and then are out of phase because one goes further than the other by the path difference Δx.

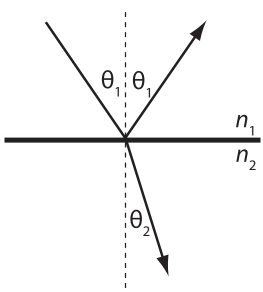
The flashes from the hot filaments arrive at different times, and so the light from two separate light bulbs cannot lead to interference. Is it ever possible to have interference effects from two separate light sources? They have to have the same wavelength and there has to be a definite phase relationship between the two sources. This is possible with *lasers*. A laser emits a beam of light with a single wavelength, rather than the random flashes with a range of wavelengths emitted by a hot object. The light from two similar lasers can therefore lead to interference patterns.

The light from lasers is referred to as *coherent*, while the light from a hot body, such as a light bulb filament, is *incoherent*.

11.6 Reflection and refraction

The laws of reflection and refraction

Normally, we are used to thinking of light as traveling in straight lines. In the absence of the openings and obstacles that lead to interference this is, in fact, what happens. The straight-line rays can still change direction. One way is by *reflection*, as from a mirror. Another is by *refraction*, the change of direction that occurs when a ray goes from one medium to another.



The figure shows an incident ray, a reflected ray, and a refracted ray. The incident ray makes an angle θ_1 with the normal to the surface (a line perpendicular to the surface), which is called the angle of incidence. The angle of the reflected ray with the normal, also θ_1, is called the angle of reflection. The fact that the two angles are equal is called *the law of reflection*.

The angle of a ray changes when it goes from one medium to another in which the speed of light is different. In empty space, i.e., in a vacuum, the speed of light is $c = 3 \times 10^8$ m/s. In any other medium it is $\frac{c}{n}$, where n is called the *index of refraction*.

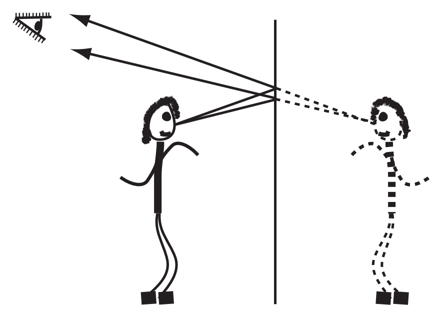
If the angle with the normal in one medium, with index of refraction n_1, is θ_1, and in the second medium the values are n_2 and θ_2, then $n_1 \sin \theta_1 = n_2 \sin \theta_2$. This is the *law of refraction*, or *Snell's law*.

These are all of the fundamental principles of the subject called *geometric optics*. There are, however, many important and interesting applications. We can use the law of reflection to study different kinds of mirrors. The law of refraction leads to an understanding of lenses and optical instruments, such as the microscope and the telescope.

Mirrors

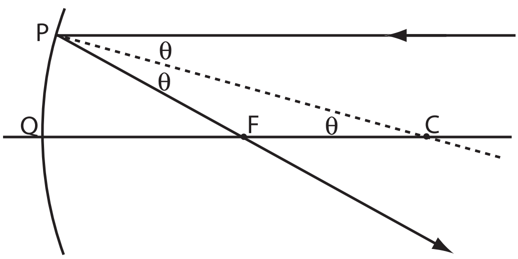
How do we see ourselves in a plane mirror?

The figure shows a mirror, and a person, Lucy, in front of it. It also shows two rays, coming from her nose, and reflected by the mirror.



The two rays are looked at by another person. They come from Lucy's nose, but they *seem* to come from a point behind the mirror. That point is called the *image* of the point on her nose where the rays originate. The rays only seem to come from there. There are no actual rays from her at the image. This kind of image is called a *virtual* image.

With a parabolic mirror rays coming to the mirror parallel to the axis are reflected so that they cross at a point on the axis called the *focus*. The fact that they do that is a geometric property of a parabola. It is much easier to make spherical mirrors than parabolic mirrors. Rays parallel to the axis of a spherical mirror do not come to a focus *exactly*. But for rays near the axis they do so approximately. Let's see how this comes about.



Look at a ray that comes to the mirror parallel to the axis. It comes to the mirror at a point *P*, and is reflected toward the axis, which it crosses at the point *F*, the *focus*. The line from *P* to the center of curvature of the mirror (*C*) is the radius of the spherical surface. It is perpendicular to the mirror, i.e., it is the normal to the surface, so that both the incoming and the reflected rays make the same angle, θ, with it. The angle of the line *PC* with the axis is also θ. The triangle *PCF* has two angles equal to θ so that it is isosceles and its sides *PF* and *FC* have the same length.

The axis crosses the mirror at the point *Q*. The lengths *PC* and *QC* are the same, since they both go from the mirror to its center of curvature. The lengths *PF* and *QF* are not the same, but we will make the approximation that they are. The smaller the diameter of the disc of the mirror is compared to its radius of curvature, the more closely this will be true. In this approximation it doesn't matter how far *P* is from the axis—all the rays coming to the mirror parallel to the axis cross the axis at the point *F*. Moreover the *focal length* $QF = f$ is seen to be equal to half the radius of curvature. (In fact, *QF* changes as θ changes, so that the point where the reflected ray crosses the axis also changes.)

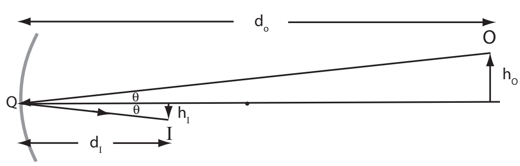
A spherical mirror can produce an image that is different from the virtual image of a plane mirror. In the diagram the rays originate at the *object*, and after reflection they converge at the *image*. This time the rays actually go to the image. It is a *real* image. It can be caught on a piece of paper or on a screen. This is not so for a virtual image, a point on the other side of the mirror from which the rays seem to come.

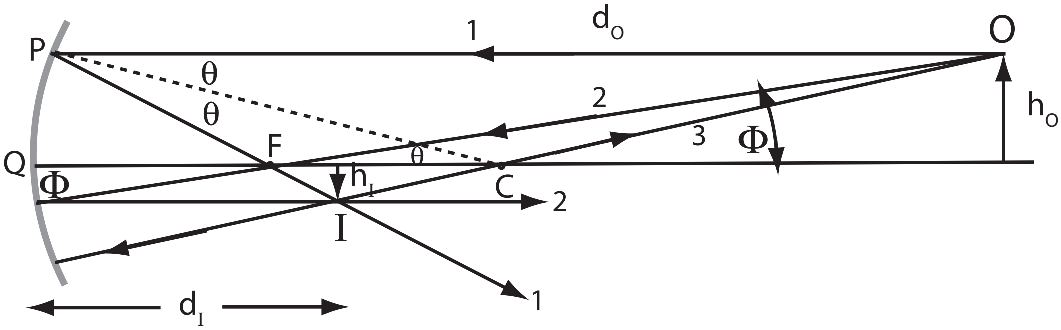
We can see where the image is by following some representative rays. A ray coming to the mirror parallel to the axis (1) is reflected so as to go through the focus. Conversely, a ray through the focus (2) is reflected so that it is then parallel to the axis. A ray along the line through the center, *C*, of the spherical surface (3) is reflected so that it returns along the same line.

The figure shows an object at a distance further from the mirror than *C*, and a real image, smaller and inverted, between the focus and the center.

Since the rays follow the same path in either direction, we can interchange the image (I) and the object (O), and see that for an object between the focus and the center there is a real, inverted, and magnified image.

There is a simple relation between the object distance d_O, the image distance d_I, and the focal length f. The figure shows that the height of the object (h_O) and the height of the image (h_I) are related in two ways. First, we see that because the two angles marked ϕ (Greek *phi*) are the same, $\frac{h_O}{d_O - f} = \frac{h_I}{f}$, or $\frac{h_O}{h_I} = \frac{d_O - f}{f}$.





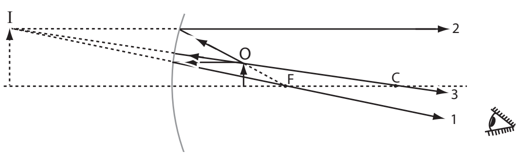
A ray from the object to Q goes to the image, forming two similar triangles that show that $\frac{h_O}{d_O} = \frac{h_I}{d_I}$ or $\frac{h_O}{h_I} = \frac{d_o}{d_I}$.

Equating the two expressions that are each equal to $\frac{h_O}{h_I}$, we get $\frac{d_O - f}{f} = \frac{d_O}{d_I}$, which can be written as $\frac{d_O}{f} - 1 = \frac{d_O}{d_I}$. Dividing each term by d_O we get $\frac{1}{f} - \frac{1}{d_O} = \frac{1}{d_I}$ or $\frac{1}{d_I} + \frac{1}{d_O} = \frac{1}{f}$.

This relation also holds for mirrors that curve outward, or *convex* mirrors, if f is taken to be negative, and for virtual images if d_I is then taken to be negative. (Mirrors that curve inward are called *concave*.)

We also see that the *magnification* $\frac{h_I}{h_O}$ is equal to $\frac{d_I}{d_O}$. We will take h_O to be positive above the axis. To avoid an extra negative sign we will take h_I to be positive when it is below the axis, as it is when the image is real and inverted.

What happens when the object is between the mirror and the focus? The figure shows the representative rays. A virtual and magnified image is produced, right-side up. This is what happens in a magnifying mirror like those used in bathrooms for shaving and makeup.



EXAMPLE 12

A concave mirror has a radius of curvature of 30 cm.

(a) An object is 90 cm from the mirror. What is the position of the image? What kind of image is it? What is the magnification?

(b) Repeat part (a) for $d_O = 20$ cm.

(c) Repeat part (a) for $d_O = 10$ cm.

Ans.:

(a) $f = \frac{1}{2}r = 15$ cm.

$\frac{1}{d_I} = \frac{1}{f} - \frac{1}{d_O} = \frac{1}{15} - \frac{1}{90}$ so that $d_I = 18$ cm.

[We can use $\frac{1}{f} - \frac{1}{d_O} = \frac{d_O - f}{f d_O}$ so that $d_I = \frac{f d_O}{d_O - f} = \frac{(15)(90)}{90-15} = 18$ cm.]

$\frac{h_I}{h_O} = \frac{d_I}{d_O} = \frac{18}{90} = 0.20$.

The image is real and inverted, 18 cm in front of the mirror. The magnification is 0.20, i.e., the size of the image is 0.20 times that of the object.

(b) $\frac{1}{f} - \frac{1}{d_O} = \frac{1}{15} - \frac{1}{20} = \frac{4-3}{60} = \frac{1}{60}$. $d_i = 60$ cm.

Magnification = $\frac{d_I}{d_O} = 3$.

The image is real and inverted, 60 cm in front of the mirror. The magnification is 3, so that the image is three times as large as the object.

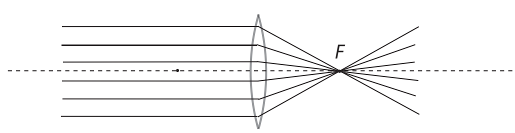
(c) $\frac{1}{d_I} = \frac{1}{15} - \frac{1}{10} = \frac{2-3}{30} = \frac{-1}{30}$. $d_i = -30$ cm.

The image is virtual, right-side up, 30 cm behind the mirror. The magnification is -3. The image is again three times as large as the object, but not inverted.

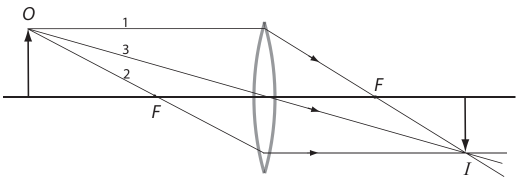
Lenses

Lenses are among the most familiar optical devices. They are used as pocket magnifiers, camera lenses, and eyeglasses. In combination they are the constituents of telescopes and microscopes. Lenses usually have spherical surfaces. Each ray's path can be followed and analyzed by using Snell's law. Computers now allow this to be done much more efficiently than earlier, and this has led to the development of complex lenses, consisting of many parts, even in inexpensive cameras.

It is also possible to use an approximate analysis similar to that which we have used for mirrors. Just as in that case, it is most accurate for rays close to the lens axis. Lenses, like mirrors, have focus points. There are two of them, one on each side of the lens. We will start with lenses that are thicker in the middle, called *convex* or converging lenses. (Lenses that are thinner in the middle are called *concave* or diverging lenses.)



The figure illustrates that rays arriving parallel to the axis are refracted so as to cross at the focus. This leads us to two representative rays: one that comes to the lens parallel to the axis (1), and is refracted so as to go through the focus, and a second that comes through the focus (2), and is refracted so as to leave the lens parallel to the axis. A third representative ray goes through the center of the lens (3). For a sufficiently thin lens, and for rays close to the axis, this ray goes through the lens, to a good approximation, without changing direction.



The figure shows the three representative rays coming from an object and converging to an image. In this case the image is real, i.e., the rays actually go to the image. It is inverted and reduced in size. Since the rays can follow the same path in both directions, the image and object can be interchanged. If this is done the image is still real and inverted, but it is magnified.

The situation is quite different for an object between the lens and one of the focal points. In that case the rays, after being refracted by the lens, diverge. There is no real image. After they go through the lens, however, the rays seem to come from a point on the same side of the lens as the object. This is the virtual image. It is right-side up and magnified.

Concave lenses, by themselves, produce only virtual images, right-side up and reduced in size, regardless of the position of the object.

EXAMPLE 13

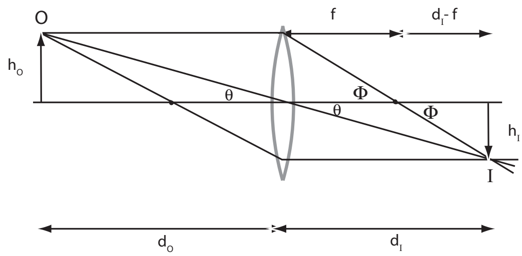
Go to the PhET website and open the simulation *Geometric Optics*. Select "principal rays," "curvature 0.8," "refractive index 1.87," and "virtual image."

Move the object back and forth and observe what happens to the image and the magnification. What happens as the object moves into the region between the focus and the lens?

Ans.:
When the object is outside the focus the image becomes larger as the object moves closer to the focus. When the object moves past the focus the image becomes virtual and magnified. It is then on the same side of the lens as the object.

The thin-lens relation

Let's look once more at the representative rays that show the formation of a real image by a thin convex lens. Call the heights of the object and the image h_O and h_I. Call the distances from the object and the image to the lens d_O and d_I. Look at the two similar triangles whose sides are h_O and d_O for the one on the left and h_I and d_I for the one on the right. They both contain the angle marked θ, so that $\frac{h_O}{d_O} = \frac{h_I}{d_I}$ or $\frac{h_I}{h_O} = \frac{d_I}{d_O}$.



Let f be the *focal length*, i.e., the distance from either focus to the lens. A second pair of similar triangles, both with the angles marked ϕ, has sides h_O and f between the lens and the focus on the right and h_I and $d_I - f$ between the same focus and the image. They show that $\frac{h_O}{f} = \frac{h_I}{d_I - f}$ or $\frac{h_I}{h_O} = \frac{d_I - f}{f}$.

Combining the two relations we get $\frac{d_I - f}{f} = \frac{d_I}{d_O}$ or $\frac{d_I}{f} - 1 = \frac{d_I}{d_O}$. Divide by d_I to get $\frac{1}{f} - \frac{1}{d_I} = \frac{1}{d_O}$, or $\frac{1}{d_O} + \frac{1}{d_I} = \frac{1}{f}$.

This *thin-lens relation* describes not only the situation that we started with, for a convex lens with a real image. It can also be used for a concave lens if the focal length is taken to be negative. For a virtual image d_I needs to be taken to be negative.

Just as for the spherical mirror, the magnification $\frac{h_I}{h_O}$ is equal to $\frac{d_I}{d_O}$.

EXAMPLE 14

Thin lens calculation

A thin convex lens has a focal length of 15 cm.

(a) An object is 90 cm to the left of the lens. What are the position and nature of the image? What is the magnification?

(b) Repeat part (a) for $d_O = 20$ cm.

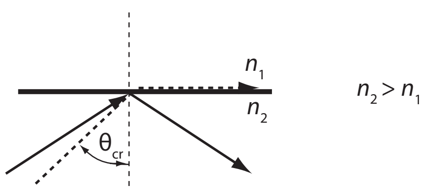
(c) Repeat part (a) for $d_O = 10$ cm.

Ans.:
The calculations are the same as for the previous example with the spherical mirror. Since the focal

length is the same in the two cases, and the object distances are the same, the answers are the same. The only differences are that the real images (parts a and b) are on the right-hand side of the lens and the virtual image (part c) is on the left-hand side of the lens.

Total internal reflection

Look at what happens when a ray in one medium comes to a boundary with a medium whose index of refraction is smaller. The angle of reflection is then larger than the angle of incidence. If the angle of incidence, θ_1, is now gradually increased, the angle of refraction, θ_2, becomes larger, and eventually reaches 90°. The angle of incidence at that point is called the critical angle θ_{cr}. What happens if the angle of incidence is increased still further? There is then no refracted ray at all! The ray is reflected back into the first medium. This is called *total internal reflection.*

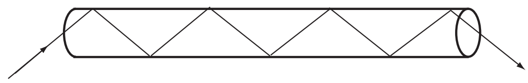


EXAMPLE 15

What is the critical angle at the interface of glass with an index of refraction of 1.5 and air?

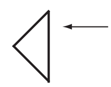
Ans.:
Snell's law is $n_1 \sin\theta_1 = n_2 \sin\theta_2$. Use the subscripts 2 for air and 1 for glass. In air n_2 is close to 1. The critical angle is $\theta_{cr} = \theta_1$ when $\theta_2 = 90°$ so that $1.5 \sin\theta_{cr} = 1$. Hence $\sin\theta_{cr} = \frac{1}{1.5} = 0.67$ and $\theta_{cr} = 42°$.

Light entering a glass or plastic cylinder at one end is totally reflected each time it hits the wall of the cylinder at an angle greater than the critical angle, until it emerges at the other end. This is the principle of *fiber optics.* The attenuation of the light as it travels along a thin fiber can be very small, and light signals can be carried by optical fibers with much less loss than for electric signals along copper wires.



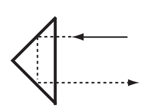
Bundles of thin fibers, a few μm in diameter, can transmit images from one end to the other, even when they are bent. This has led to a number of important medical applications. One is the exploration of the inside of the human body, as for the stomach with an *endoscope* and the colon with a *colonoscope.* The images provided by fiber optics can be combined with miniaturized instrumentation for *arthroscopic surgery*, which requires only very small incisions.

EXAMPLE 16



The figure shows a beam of light entering a 45° prism made of glass ($n = 1.5$). What is the subsequent path of the beam?

Ans.:
Since the angle of incidence is 45°, and therefore larger than the critical angle of 42°, the beam is totally reflected. It is reflected a second time at the lower surface and emerges in the direction opposite to its original direction.



Resolution

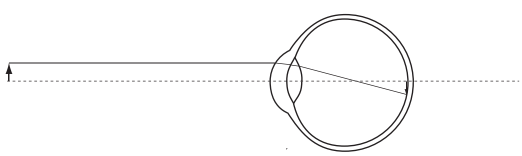
When light passes through a hole, or *aperture*, or through a lens, what we see is a diffraction pattern. When the aperture is large compared to the wavelength, we are not usually aware of that. The diffraction effects become larger and more apparent when the aperture is small or when we are trying to resolve very small or distant objects. They provide a limit to the validity of geometric optics. When they become observable it is no longer appropriate to consider light to be propagated as rays that travel through the aperture in straight lines.

Camera and eye

In a camera a lens produces a real image, which is then recorded by a film or other "sensor."

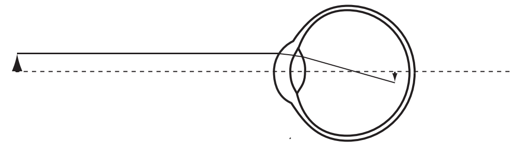
Even in simple cameras the lens is made of several components, designed to reduce the deficiencies of single lenses. "Chromatic aberration," for instance, refers to the fact that for each color (wavelength) there is a different index of refraction and focal length, so that the image of a blue object is formed at a different place than the image of a red object. This effect can be partially compensated for with a lens made of two or more components.

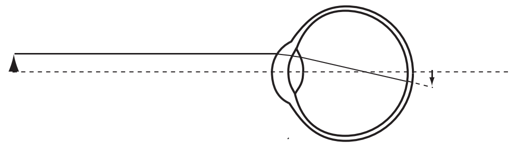
Lenses generally have spherical surfaces because other shapes are much more difficult to make. Rays that come from the axis to such a lens at a point far from the axis do not come to an image at the same distance as rays close to the axis. This is called "spherical aberration." In addition distortions can occur from objects that extend far from the axis.



The same considerations apply to the eye. The lens of the eye forms an image on the *retina*. The retina contains the sensors that send signals to the brain via the *optic nerve*. In contrast to the lens of a camera the eye's lens can change its curvature and hence its focal length under the action of the muscles that surround it. In other words, it can *adapt* so as to create images of objects that are at different distances from it.

In cases where the eye's lens cannot adapt correctly, it is usually possible to to use external lenses ("eyeglasses") to correct for the deficiencies of the lens of the eye. In a *nearsighted* eye the rays converge more than is required to form an image on the retina so that the image is in front of the retina. A diverging lens in front of the eye can correct for that. In a *farsighted* eye the rays from an object do not converge sufficiently, and are refracted so that they head toward an image behind the retina. A converging lens is used to correct this problem. The strength of eyeglasses is usually measured in *diopters*, equal to $\frac{100}{f}$, where f is the focal length in cm.





The magnifying glass

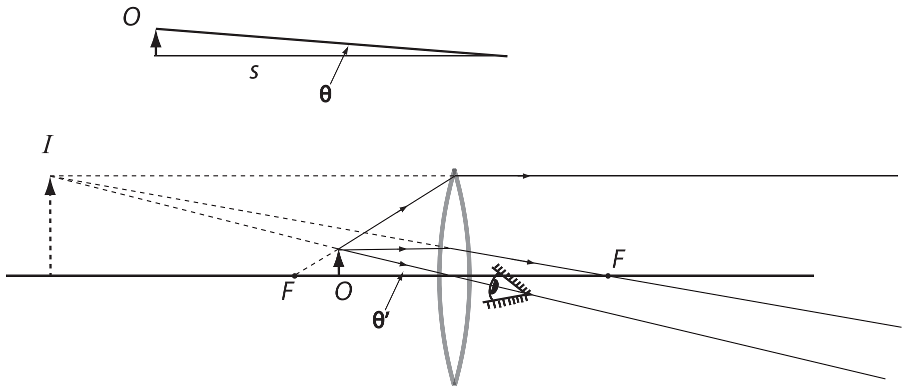
When a lens is used as a magnifying glass, the object is between the lens and one of the focal points. The virtual image is on the same side of the lens as the object, right-side up and magnified. When the object is moved closer to the focus, the image becomes larger.

You can make the image huge by letting the object get really close to the focus. The magnification ($\frac{h_I}{h_o}$) can be as large as you like. So why does that not give you all the magnification that you would ever want? The downside is that as the image becomes larger it moves further away. Since you have to stay on the other side of the lens from it, the fact that it gets so much larger doesn't do you any good. The image doesn't *appear* to become larger.

We need a new criterion to describe what happens. The magnification $\frac{h_I}{h_o}$ isn't very useful here. What matters is the *angle* that the image subtends at the eye. That's $\frac{h_I}{\ell_I}$, where ℓ_I is the distance from the image to the eye. You want this angle to be as large as you can make it. What the lens does for you is to make this angle larger than the angle that the object subtends at your eye without the lens.

How large is that angle without the lens? Can't you make it larger just by moving your eye close to the object? Yes you can, but there's a limit. At some distance the eye can't focus any more. Your eye has a *near point*, and when you move an object closer than that it looks blurred.

That gives us our criterion: we need to use the *angular magnification*. The best we can do without the lens is to put the object at the near point. If this point is a distance s from the eye, the angle subtended by the object at the eye is $\frac{h_o}{s}$. The lens makes it possible for both the object



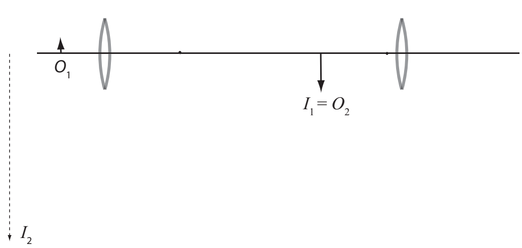
and the image to subtend the angle $\theta' = \frac{h_o}{d_o}$. The angular magnification is then $\frac{\theta'}{\theta} = \frac{h_o}{d_o}\frac{s}{h_o} = \frac{s}{d_o}$.

We can express this relation in terms of d_I by using $\frac{1}{d_o} + \frac{1}{d_I} = \frac{1}{f}$, or $\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_I}$ so that $\frac{\theta'}{\theta} = \frac{s}{d_o} = s(\frac{1}{f} - \frac{1}{d_I})$. When the object is at the focus ($d_o = f$), $\frac{1}{d_I} = 0$, and the angular magnification is $\frac{s}{f}$.

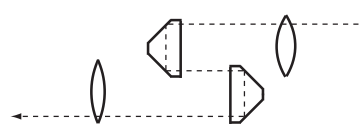
It can be made a little larger by moving the image to the near point, so that $d_I = -s$. (The distance to a virtual image is negative!). In this case the angular magnification is $\frac{\theta'}{\theta} = s(\frac{1}{f} - \frac{1}{-s}) = \frac{s}{f} + 1$.

Microscope and telescope

Some of the most interesting applications of lenses are those that use them in combinations where the image formed by one lens is the object for a second lens. Two examples are the microscope and the telescope. In a microscope one lens is near the object and is called the *objective*. It forms a real enlarged image. This image is the object for a second lens, the *eyepiece*, which forms a virtual enlarged image of it.



A telescope can also be constructed of an objective that produces a real image, which is then observed with a magnifying glass. There are, however, some different requirements. The object is far away and you want the first image to be as large as possible. This is best done with a lens that has a long focal length. The image is inverted, which may not be a problem for astronomical observations but needs to be corrected for use on earth. A third lens can be put between the objective and the eyepiece to change the orientation of the image. The resulting *spyglass* has the disadvantage that it is very long. The modern equivalent is the *prism binocular*. Two right-angled prisms are inserted between the objective and the eyepiece. This shortens the distance betwen the two lenses and also inverts the image.



11.7 Where Einstein started: electromagnetism and relativity

The ether and the speed of light

Sound waves are propagated by the vibration of atoms or molecules in air or other materials, water waves by the motion of the water. Both are examples of mechanical waves in a medium that is disturbed as the wave passes. Sound cannot travel through empty space. Electromagnetic waves are different. The generation of electric fields from changing magnetic fields, described by Faraday's law, and the generation of magnetic fields by changing electric fields do not require a material, a *medium*. This is confirmed by the fact that electromagnetic radiation reaches us through empty space from the sun and other stars, while the sound of the violent explosions that take place there cannot reach us. This seems perfectly natural to us.

It didn't seem so natural in the nineteenth century, when every other known kind of wave required a medium whose constituents could move back and forth. So a new one was imagined, the *ether*, as a medium for electromagnetic waves. It was not a very comfortable idea right from the beginning. It would have to be there even where nothing else seemed to be, and it would have to be able to vibrate with the almost unimaginably high frequencies required by the high speed of the waves. For example, at one end of the visible spectrum the frequency is about 10^{15} Hz, and the ether would have to be able to vibrate that many times per second.

There was a more subtle and more interesting difficulty. Was the ether attached to the earth or at rest with respect to the sun or some other part of the universe? Just as the speed of sound waves is measured with respect to the medium in which they propagate, and is different in systems moving with respect to it, so it was expected for the electromagnetic waves as they propagated in the ether. Experiments of great sophistication and precision were made to detect motion with respect to the ether, but no evidence for such a motion was ever found.

The most famous and decisive of these experiments was done by A. A. Michelson, who had already made the most precise measurement of the speed of light, and who now recruited his colleague Edward Morley to collaborate with him. It compared the speed of light along two directions at right angles to each other. We would expect a difference depending on the angle between the earth's motion and the motion of the ether. However, no evidence for any such "ether wind" was detected. The experiment led to the startling conclusion that irrespective of the direction and the time of day or year, the speed of light was the same.

This result is entirely in accord with Maxwell's equations. They lead to the same speed, c, with no indication of any dependence on a particular coordinate system. This is fundamentally different from the way mechanical waves, like sound and water waves, behave. A sound wave on a moving train has different speeds with respect to the train and with respect to the ground.

That the different kinds of waves should behave so differently was not at all understood at the time, and all attempts at reconciliation over the next 17 years failed.

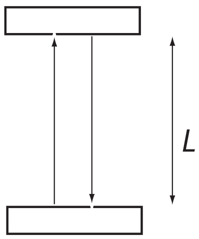
Only Einstein, in 1905, followed Maxwell's equations seriously when he said that the speed of electromagnetic waves is c $(= \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3 \times 10^8$ m/s) regardless of the frame of reference. He was the only one not to try to make lightwaves conform to the behavior of waves in materials, i.e., to mechanical waves such as sound waves. He took as his point of departure the simple fact that Maxwell was right, that the speed of electromagnetic waves was c, and that it remained so, regardless of any velocity of the frame of reference with respect to which the waves might be observed.

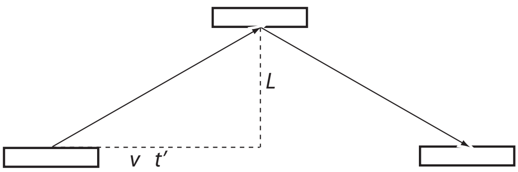
The results were startling. If c is constant, the dimensions, i.e., the observed lengths of objects, are no longer constant and independent of the motion of the object with respect to the observer. They are subject to the *Lorentz contraction* by the factor $\sqrt{1 - \frac{v^2}{c^2}}$, when determined by an observer moving with a velocity v with respect to the object. Time intervals also depend on the relative motion, and are larger by the reciprocal of the same factor, $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$.

Kinematics of the special theory of relativity: time dilation and length contraction

Let's see how these results come about. We will do two "thought experiments." For the first we start with two mirrors a distance L apart. A beam of light with its velocity c goes back and forth between them. This is a primitive clock, marking time intervals Δt between reflections, so that $L = c\Delta t$ and $\Delta t = \frac{L}{c}$.

Look at this clock while it flies by at a speed v. The light beam has to go further between reflections. If its speed continues to be c, the time interval between reflections must be different. Call it $\Delta t'$.





The beam has to go along the hypotenuse of the triangle whose sides are L and $v\Delta t'$, i.e., a distance $\sqrt{L^2+(v\Delta t')^2}$. It takes a time $\Delta t'$ to traverse this distance, so that $\sqrt{L^2+(v\Delta t')^2} = c\Delta t'$, or $L^2+v^2(\Delta t')^2 = (c\Delta t')^2$.

Dividing by c^2 we get $\frac{L^2}{c^2}+\frac{v^2}{c^2}(\Delta t')^2 = (\Delta t')^2$, or $\frac{L^2}{c^2} = (\Delta t')^2(1-\frac{v^2}{c^2})$.

But $\frac{L}{c} = \Delta t$, so that $(\Delta t)^2 = (\Delta t')^2(1-\frac{v^2}{c^2})$, or $\Delta t' = \frac{\Delta t}{\sqrt{1-(v^2/c^2)}}$.

The denominator is less than one, so that $\Delta t'$ is greater than Δt. The time interval in the moving system is greater, and we refer to the phenomenon as *time dilation*.

Δt is the time interval in the system in which the observer is at rest with respect to the clock. It is called the "proper time interval." In all other systems the observer is moving with respect to the clock, and the time interval ($\Delta t'$) is longer.

For the second thought experiment we look at a spaceship, with its astronaut, A, traveling between points in two cities on earth, where you are the observer. As the spaceship travels between the two cities, A, at rest with respect to his clock, records the time interval Δt. You, on earth, see the longer time interval, $\Delta t' = \frac{\Delta t}{\sqrt{1-(v^2/c^2)}}$. Both you and A measure the same relative velocity v.

What about the distance between the two cities? We'll call that distance, as you observe it, L. You can take your time measuring it because the cities are at rest in your coordinate system. That's why we call this the "proper distance." As you watch the spaceship fly by at the speed v, you see that it takes a time $\Delta t'$ to traverse the distance L, so that $v = \frac{L}{\Delta t'}$.

A, in the spaceship, sees the two cities pass her in the shorter ("proper") time interval Δt. The relative velocity is still v. The distance between the cities must therefore, as she sees it (L'), be smaller than the proper distance, L, which is the one observed by you on earth: $v = \frac{L'}{\Delta t}$.

We see that the proper distance (L) goes with the longer time interval ($\Delta t'$), while the proper time interval (Δt) goes with the shorter distance (L'), and $L' = v\Delta t = L\frac{\Delta t}{\Delta t'}$, or $L' = L\sqrt{1-\frac{v^2}{c^2}}$. The flying astronaut sees the cities closer to each other, i.e., the distance between them is *contracted*. We speak of *length contraction* and *time dilation*. This goes both ways: the length of the spaceship is seen by you as smaller than by A.

Dynamics of the special theory: $E = mc^2$

These are the kinematic results. In addition there are the even more interesting and important dynamic aspects, the ones that deal with mass, energy, and momentum, and that lead to $E = mc^2$.

We will quote some of the results for the dynamic quantities, momentum, and energy, that are also part of Einstein's special theory of relativity. The "relativistic momentum," p, is no longer mv, but is $\frac{mv}{\sqrt{1-\frac{v^2}{c^2}}}$.

The most startling result is that the *total energy* is $E = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}$, of which $\frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} - mc^2$ is the relativistic kinetic energy. The rest, mc^2, is called the *rest energy*.

We can derive a relation between energy and momentum from these relations: from the energy relation we have $\frac{E}{c^2} = \frac{m}{\sqrt{1-\frac{v^2}{c^2}}}$, which, from the momentum relation, is equal to $\frac{p}{v}$, so that $\frac{v}{c} = \frac{pc}{E}$.

We can then write $E^2 = \frac{m^2c^4}{1-\frac{v^2}{c^2}}$ as $E^2 = \frac{m^2c^4}{1-\frac{p^2c^2}{E^2}}$, and finally $E^2 - p^2c^2 = m^2c^4$, or $E^2 = p^2c^2 + m^2c^4$.

Can we reconcile the expression for the relativistic kinetic energy with the Newtonian form that we are used to? Let's see how it changes when the velocity changes. We can write it as

$$K = mc^2 \left(\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right),$$

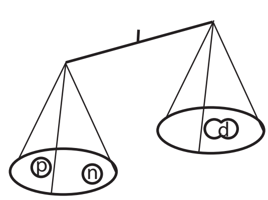
$$\text{or } mc^2[(1 - \frac{v^2}{c^2})^{-1/2} - 1].$$

We can expand the square-root part by using the *binomial theorem*, in the form $(1+x)^n = 1 + nx + \cdots$ [This is the most useful part and is sufficient here. The next terms are $\frac{n(n-1)}{2}$ and $\frac{n(n-1)(n-2)}{3!}$.] Here $x = -\frac{v^2}{c^2}$ and $n = -\frac{1}{2}$, so that $(1 - \frac{v^2}{c^2})^{-1/2}$ becomes $1 + (-\frac{1}{2})(-\frac{v^2}{c^2}) = 1 + \frac{v^2}{2c^2}$, and the kinetic energy becomes $mc^2(1 + \frac{v^2}{2c^2} - 1)$, or $\frac{1}{2}mv^2$! The next terms are successively smaller and smaller, and we see that as $\frac{v^2}{c^2}$ becomes smaller, the expression for the kinetic energy gets closer and closer to the familiar classical expression, $\frac{1}{2}mv^2$.

Let's look at $E = mc^2$. It is a real equivalence of mass and energy. If I heat an object and so increase its internal energy, and then measure its mass, I have every reason to expect that the measurement will lead to a greater mass than when it was cold. Of course the difference, equivalent to the increase in the internal energy divided by c^2, is so small that there is no hope of actually making the measurement.

It takes 13.6 eV to pull the electron away from the proton in a hydrogen atom. Adding this energy (the *binding energy*) is equivalent to increasing the mass, so that once you have separated the two particles you have that much more mass. Since the mass of the hydrogen atom (almost entirely that of the proton) is equivalent to about 937 MeV, the change corresponds to one part in 10^8, still very small.

In the nuclear realm the difference becomes significant. Pulling the proton away from the neutron in the deuteron, the simplest nucleus consisting of more than one particle, takes about 2 MeV, or about 1 part in 1000.



Is there a process in which all of the mass disappears? Yes, that can happen also. Let a positron hit a material. It slows down as it makes collisions, and eventually it comes to rest (or nearly so) near an electron. The two annihilate. Both disappear. That means they're gone, completely and forever. What happens to the energy that is equivalent to the mass that disappears? It is now that of two photons, each with an energy of about $\frac{1}{2}$ MeV. The electron and the positron are gone, but the energy equivalent to their mass is still there as pure photon energy.

Small wonder that Einstein's 1905 paper *"On the Electrodynamics of Moving Bodies"* on what came to be known as the *special theory of relativity* was met with skepticism and even with hostility. In spite of the strange and counterintuitive conclusions, however, it became clear as time went on that the theory was right, that it described experiments and observations correctly, and that it necessitated a revolutionary reappraisal of the concepts of time and space, of mass and energy. Now, more than a hundred years later, it has been so widely and completely verified that it has become a cornerstone of science without which many modern developments are unthinkable.

In nuclear and particle physics the theory plays a fundamental role. The relativistic relations are essential for the understanding, design, and operation of particle accelerators such as cyclotrons. Without a knowledge of the equivalence of mass and energy there could be no understanding of nuclear processes and transformations, including those in nuclear reactors and the fusion reactions that light the stars. Among applications that impinge directly on the experiences of nonscientists are the *Global Positioning Systems* that are able to provide remarkably precise information on location by signals from earth satellites.

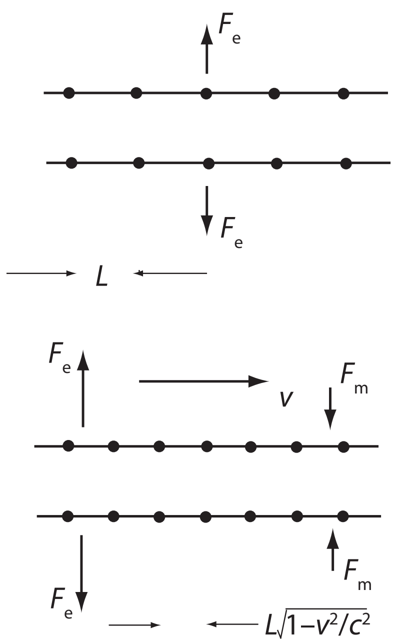
The theory is so much part of the fabric of today's science that it is often regarded as part of "classical" physics, i.e., that part of physics that is no longer questioned as to its correctness within its realm of applicability.

Magnetism and electricity: inseparable, but interchangeable

The fact that electricity and magnetism are closely related was already known from the discoveries of Oersted, Faraday, and others, but our knowledge of that synthesis was enormously expanded by Maxwell when he combined them in what came to be known as Maxwell's equations, and showed how they cooperate in electromagnetic waves. Later, in 1905, a higher level of understanding was achieved by Einstein. Even as a student, Einstein was already on the path that led to the theory of relativity when he tried to imagine what would be observed by someone riding at the crest of an electromagnetic wave.

We will look here at the much simpler ride, moving along with the charges of a current. This will allow us to see how the electric and magnetic forces change, one into the other, depending on the frame of reference in which they are observed.

Imagine two parallel lines with equally spaced charges, at rest. The two lines repel, as usual, as described by Coulomb's law. Now look at the same two lines from a frame of reference that moves to the left, so that all the charges are seen to move to the right, parallel to the two lines of charge. The two lines are now currents, and there is a magnetic force of attraction between them.



How can we reconcile this conclusion with the expectation that the actual, observed, measured force cannot depend on the frame of reference? In Newtonian, classical physics there seems to be no way out. It takes the theory of relativity to resolve the dilemma. It tells us that the spacing between the charges is now changed. All distances parallel to the direction of motion are reduced by the Lorentz contraction. There are now more charges in any given length along the lines, and the electric force per meter is therefore greater. But the greater electric repulsion between the charges, together with the magnetic attraction between them, now that they are moving, add up to the same net force as before, when there was only an electric force and the charges were at rest.

We could also take the point of view that we know the electric and magnetic forces. We could then show that to get the same result, independent of the frame of reference, there must be a Lorentz contraction, and we can calculate how large it has to be.

Alternatively, we can start from the Lorentz contraction by the factor $\sqrt{1-\frac{v^2}{c^2}}$ and calculate how much larger the electric force becomes. This gives us the magnitude of the attractive force that is required so that the total force remains unchanged. In other words if, somehow, we already know that there is such a thing as the Lorentz contraction, we can show that there must be an additional force, an attractive force between two lines of moving charges. This force is, of course, the magnetic force, and it can be seen to be a consequence of the existence of the electric force together with the theory of relativity.

There is an interesting feature of this story. It is sometimes thought that the kinematic consequences of the special theory of relativity are observable only when the relative velocities are close to the speed of light. This is not so in the present example. Think of the electrons moving in a copper wire carrying a current.They move very fast, but their average velocity, called the *drift velocity*, is of the order of a few cm/s or even smaller for ordinary currents. This is the relative velocity of the two frames of reference that we have used. The phenomena of magnetism arising from currents, including all the forces between currents in wires and coils, are seen to be the consequence of these very small relative velocities.

As we reflect on the place of the theory of relativity we can look back to the view of science before it was developed. The great achievement of Newton was to create *classical mechanics*, the theory that seemed to structure all of physics, all of science, and perhaps all of knowledge.

Understanding a phenomenon meant to know its *mechanism*, i.e., the way it comes about in accord with Newton's laws. When Newton's laws came in conflict with the more recently discovered laws of electromagnetism, as demonstrated, for example, by the Michelson–Morley experiment, it was natural to look for flaws in the newer theory. How astonishing then that it was Maxwell's synthesis of electricity and magnetism that was given new meaning by the theory of relativity, and that it was the electromagnetic equations that survived unchanged. Newton's mechanics, on the other hand, and his views of time and space, no longer hold their place as the bedrock of science after Einstein's fresh look and bold reexamination.

11.8 Summary

A wave along a string or a water wave can be seen to move fast and far, but the pieces of the string and the parts of the water are only oscillating in place. Although the string and the water do not move very far, energy and momentum are transported. One particle pushes on its neighbor, and that's how the wave propagates. Mechanical waves also include sound waves, produced by the oscillation of strings or vocal cords or columns of air.

The propagation of electromagnetic waves is more subtle. A changing electric field produces a changing magnetic field, which in turn produces a changing electric field. The disembodied fields propagate even when there are no material bodies. That's how energy reaches us from the sun through the empty space between.

Waves reach us when we hear or see, or when we feel the warmth of the sun or the stove.

In a wave there is a quantity that oscillates in space and in time. The motion repeats in space after one wavelength (λ), and in time after one period (T).

The frequency is the reciprocal of the period. $f = \frac{1}{T}$. The wave's speed is $v = \frac{\lambda}{T} = \lambda f$.

When two waves come together, their amplitudes add. (They "interfere.") They can produce constructive interference or destructive interference. Two waves are "in phase" when they move up and down together. Their space dependence has a difference in phase by an angle θ when one varies as $\sin x$ and the other as $\sin(x - \theta)$. The sine function and the cosine function are out of phase by $\frac{\pi}{2}$ radians.

The intensity of a wave (in watts/m^2) is the amount of energy transported per second divided by the cross-sectional area through which it passes.

The expression $y = A \sin(\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t)$ represents a wave.

The frequency of a wave is determined by the vibration of its source. The speed of a (mechanical) wave depends on the properties of the medium through which it travels. The two properties on which the speed depends are the displaced mass and the restoring force.

A standing wave is produced by two traveling waves in opposite directions. In a standing wave there is an oscillation at each point but no propagation. The amplitude varies (sinusoidally) with distance.

On a string fixed at both ends, standing waves are possible only with certain frequencies and wavelengths, namely those for which $L = n\frac{\lambda}{2}$. At a fixed end there is a node. At a free end there is (approximately) an antinode. The frequencies corresponding to $L = n\frac{\lambda}{2}$ are the natural or resonant frequencies. In an air column also, if the boundary conditions are the same at both ends (closed or open), $L = n\frac{\lambda}{2}$. If one end is open and the other closed, $L = \frac{\lambda}{4} + n\frac{\lambda}{2}$.

Two notes whose frequencies differ by a factor of two are said to be an octave apart. The octave interval is divided into smaller intervals to produce a *scale* of notes.

When the source of a wave and the receiver move with respect to each other, the frequency observed by the receiver is changed. This is the *Doppler effect*.

Maxwell showed that Ampere's law is incomplete when there is a break in a circuit. An additional term (the *displacement current*) then has to be added to the current. The displacement current is proportional to $\frac{dE}{dt}$. There is now a term similar to Faraday's law, but with the rate of change of the electric field instead of the magnetic field. The two laws together lead to the existence of electromagnetic waves.

All electromagnetic waves have the same speed, c, in a vacuum. They can have widely different frequencies and wavelengths.

The energy density of an electric field is $\frac{1}{2}\epsilon_0 E^2$ J/m^3. In a magnetic field it is $\frac{B^2}{2\mu_0}$.

In an electromagnetic wave electric and magnetic fields vary together (in phase with each other). The fields are at right angles to each other and to the direction of propagation.

In Young's experiment light travels to two slits, each of which acts as a source of light. The light from the two slits combines at a screen to form regions of light (constructive interference) and darkness (destructive interference). In a diffraction grating light from many slits recombines. With a single slit light from different parts of the opening leads to regions of constructive and destructive interference.

Light reflected from the top and the bottom of a thin film combines and gives rise to interference effects, leading to the colors of soap and oil films.

In the Michelson interferometer a light beam is split into two parts that travel at right angles to each other. The interference effects when they recombine can be used for precision measurements in terms of the wavelength of light. It was also used in the Michelson–Morley experiment to look for a difference in the speed of light depending on its direction.

Light from the hot filament of an incandescent light bulb is "incoherent." It consists of tiny randomly occurring flashes with different wavelengths. Light from a laser is "coherent." It consists of waves of a single frequency and wavelength, all in phase.

Law of reflection: the angle of incidence is equal to the angle of reflection.

Law of refraction, or Snell's law: $n_1 \sin \theta_1 = n_2 \sin \theta_2$.

The image formed by a plane mirror is "virtual." No rays actually go to such an image or come from it. When we look at the mirror, rays *seem* to come from the virtual image.

Rays coming to a spherical mirror parallel to the axis are reflected toward the focus. (This is approximately so for a spherical mirror and exactly so for a parabolic mirror.) Rays coming through the focus are reflected parallel to the axis. The distance from the mirror to the focus is half the distance to the center of curvature of the mirror.

A spherical mirror produces a real inverted image if the object is further from the mirror than the focus. For a closer object the image is virtual and enlarged.

Thin lens: rays coming to a lens parallel to the axis are refracted toward the focus. Rays coming through the focus are refracted parallel to the axis. Rays through the center continue in a straight line. With these three rays we can describe where an image is formed.

Thin-lens relation: $\frac{1}{d_O} + \frac{1}{d_I} = \frac{1}{f}$. For a virtual image d_I is negative, and for a diverging (concave) lens f is negative.

Total internal reflection results when a light ray is at the boundary to a medium with a smaller index of refraction, and the angle of incidence is so large that Snell's law leads to an angle of refraction greater than 90°. (This occurs in binoculars and in the filament of a fiber optic cable.)

The lens of a camera produces an image on a film or sensor. The lens of the eye produces an image on the retina. Nearsightedness and farsightedness occur when the lens of the eye cannot adapt sufficiently and the image is formed in front of the retina or behind it.

To use a converging lens as a magnifying glass, the object is between the lens and the focus, so as to produce a virtual enlarged image. This is also so for the eyepiece of a microscope or a telescope. In both there is a second lens that forms a real image of the object, which then is the object looked at with the eyepiece.

The speed of light, c (in a vacuum), is always the same. This assumption leads to length contraction, $L' = L\sqrt{1 - \frac{v^2}{c^2}}$, and time dilation, $\Delta t' = \frac{\Delta t}{\sqrt{1-\frac{v^2}{c^2}}}$.

Einstein's special theory of relativity also leads to $E = mc^2$: the mass of an object corresponds to an energy (the rest energy) equal to

mc^2. The rest energy of an object whose mass is 1 kg is 9×10^{16} J.

The relativistic momentum is $\frac{mv}{\sqrt{1-\frac{v^2}{c^2}}}$. The total energy is $\frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}$ and the kinetic energy is the total energy minus the rest energy. When v is much less than c, the kinetic energy $\frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} - mc^2$ approaches $\frac{1}{2}mv^2$.

The binding energy of a system is the energy required to separate it into its components.

11.9 Review activities and problems

Guided review

1. Go to the PhET website and open the simulation *Wave Interference*. Select "Sound" and "on." Set a low frequency and a large amplitude. Compare "Grayscale" and "Particles."

(a) Select "Particles" and "show markers." Describe the motion of one of the particles with a red x. You may want to use the "pause-step" feature at the bottom of the screen.

(b) What is the relation between the direction of motion of the particles and the direction of propagation of the wave? What is that relation for a wave on a string?

(c) Click on "Show graph." How is the pressure related to the color of the "grayscale"? Click on "Add Detector" twice. Set the two detectors a half wavelength apart and watch the pressure readings.

2. Go to the PhET website and open the simulation *Wave Interference*. Select "Sound," speaker "on," and "grayscale."

(a) Increase and decrease the frequency and observe what happens to the spacing of the rings.

(b) How does the wave speed vary?

3. Go to phet.colorado.edu and to *Wave on a String*. Set "damping" to zero, "oscillate," "no end," and check "rulers" and "timer."

(a) Set the frequency at 35. Use the timer to measure the period by measuring the time for the wave to travel one wavelength. Use the pause/play-step button. What is the frequency? What is the ratio of the frequency setting to the measured frequency? Measure the wavelength and calculate the wave speed.

Repeat this for three other frequencies between 25 and 65. What are your conclusions about the wave speed and the frequency setting?

4. Go to the PhET website and open the simulation *Wave Interference*. Select "Sound" and check "measuring tape" and "stopwatch."

(a) Select five different frequencies. For each of them pause and measure the wavelength. Measure the period with the timer. (Use the step feature at the bottom of the screen.)

(b) Plot your data so as to determine the wave speed. What do you need to plot so that v is the slope? Look up the speed of sound in air and in metals and compare them to your experimental value. Is the sound traveling in a gas or in a metal?

5. (a) Write an equation for a wave with a frequency of 2 Hz, a wavelength of 30 m, and an amplitude of 75 cm.

(b) Find the wave speed.

6. The equation for a certain wave is $y = 15 \cos\left(\frac{\pi}{5}x - 40\pi t\right)$. What are the wavelength, frequency, period, and wave speed?

7. Go to the PhET website and open the simulation *Wave on a String*. Set the tension to its maximum and "damping" to zero. Select "oscillate" and "fixed end." Check "rulers" and "timer." Set the frequency to 50. Since this is a resonant frequency, and so requires little energy to keep it going, the amplitude of the oscillator can be quite small. Set it at 1 or 2.

(a) Measure the wavelength by measuring the distance between points with the same phase. (Two successive points with the same phase are a wavelength apart.) (Use "pause.")

(b) Measure the frequency with the timer by measuring the time for some number of cycles. How does it compare to the frequency setting?

(c) What is the number of half-wavelength segments equal to the length of the string? What is the effect of the boundary condition at the right-hand end?

(d) Calculate the wave velocity from your measurements.

(e) Select "pulse." (Increase the amplitude.) Measure the wave speed with the timer and the ruler. Compare it to the value of part (d).

8. A string, 1.2 m long, is fixed at both ends. The wave speed along the string is 220 m/s. What are the three lowest frequencies of standing waves that can be set up?

9. An organ pipe contains an air column 80 cm long, open at one end and closed at the other. What are the three lowest frequencies of standing waves that can be set up in the air column?

10. Go to the Java Applet on the Doppler effect as in Example 10, e.g. at http://lectureonline.cl.msu.edu/~mmp/applist/doppler/d.htm.

(a) Click on the gray rectangle to make the blue dot (the source) appear. Pause the pattern, and estimate the wavelength (the distance between successive circles).

(b) Drag an arrow with the mouse, representing a velocity less than the sound velocity, v_s, toward the bottom right corner. What would you observe happening to the wavelength, frequency, and wave speed, standing at the upper left corner? What would you hear there, compared to the sound from the stationary source?

11. Go to the website phet.colorado.edu and open *Wave Interference*. Select "Light," "show screen," "no barrier," and "measuring tape."

Pause and measure the wavelength with the tape. Leave the tape in place and change the color using the wavelength slider. Measure the wavelength again with the tape. Measure the wavelength at the two ends of the spectrum. What are the approximate limits of the wavelength of visible light?

12. A concave mirror has a focal length of 20 cm.

(a) What is its radius of curvature?

(b) At what distance from the mirror must an object and an image be so that the magnification is 1?

(c) Where is the image for an object distance of 15 cm? What kind of image is it?

(d) For what range of object distances is the image larger than the object? For what part of this range is the image real and for what part is it virtual?

13. Go to the PhET website and open *Geometric Optics*. Select "principal rays," "curvature 0.8," "refractive index 1.87," "diameter 0.3," and "virtual image."

(a) Check "ruler" and use it to measure the focal distance.

(b) Select an object distance larger than the focal length and measure it. Measure the image distance.

(c) Calculate the image distance from the object distance with the thin-lens equation and compare the result with your measurement.

(d) Measure the magnification. Calculate the magnification and compare the result with your measurement.

(e) Move the object until the magnification is one. What are the object and image distances? What is their relation to the focal length?

(f) Use the thin-lens formula to find the answers to part (e) and compare them to the results of your measurement.

14. Go to phet.colorado.edu and open *Geometric Optics*, "curvature 0.8," "refractive index 1.87," "diameter 0.3," and "virtual image."

(a) Use the ruler to measure the focal distance.

(b) Select an object distance smaller than the focal length and measure it. Measure the image distance.

(c) Calculate the image distance from the object distance with the thin-lens equation and compare the result with your measurement.

(d) Measure the magnification. Calculate the magnification and compare the result with your measurement.

15. You are swimming in a quiet lake. You put your head under water and look up. What do you see?

16. What is the constraint on the index of refraction of the prism to produce the reversal of the path described in Example 16.

Problems and reasoning skill building

1. A wave travels along a long (assume infinite) rope under constant tension. The rope is marked off in 1 m intervals. At the 0 m mark the rope is observed to reach its maximum transverse displacement every 8 s. The distance between maxima at any instant is 25 m. Write a function that describes this wave, assuming that it has its maximum displacement at $x = 0, t = 0$.

2. A sinusoidal wave can be described by its frequency, speed, and wavelength. Which of these

quantities depends on the medium which on the source, and which on both?

3. You have a rope that is 4 m long. You move one end up and down to create a standing wave. What are some possible wavelengths of standing waves

(a) when the other end is fixed

(b) when the other end is free to move.

4. A plastic coating ($n = 1.4$) is used to reduce the amount of light reflected from a lens whose index of refraction is 1.5. What is the thickness required to cause destructive interference of light in the middle of the spectrum ($\lambda = 550$ nm)?

5. (a) A traveling wave is described by $y(x, t) = 0.2 \sin(\pi t - \frac{\pi}{2}x)$. What are the five physical quantities that you can determine from this relation and what are their values?

(b) Make a graph of $y(x, 0)$.

(c) Make a graph of $y(t)$ at x 2 m.

(d) Which of your graphs is a snapshot of the wave? What is the other graph?

6. John is shaving, using a concave mirror 20 cm in front of him. The virtual image is 60 cm from the mirror.

(a) What is the mirror's focal length?

(b) What is the magnification?

7. A lightbulb emitting light in all directions is 3 m below the surface of a lake. What is the shape and size of the illuminated area as seen from above?

8. Light of wavelength 550 nm passes through a single slit. Find the angle to the first intensity minimum for the following two slit widths:

(a) 2×10^{-4} m

(b) 2×10^{-6} m

(c) Explain how you knew right away that one of the angles was going to be so much larger for one of the answers.

9. Two sound sources are 3 m apart and emit waves in phase. A detector is on a line perpendicular to the line joining the sources, and directly in front of one of them, 4 m away. The speed of sound is 340 m/s.

What are the two lowest frequencies that lead to destructive interference at the detector?

10. Look at the diagram of the Michelson interferometer. The source emits yellow light with wavelength 620 nm. You look at the pattern (as shown by the eye in the figure) and see a bright yellow spot in the middle. Mirror 2 is now moved until the bright spot is replaced by a dark spot. Through what distance has the mirror moved?

11. A concave mirror projects the image of a slide on a wall. Its focal length is 30 cm. The slide is 35 cm from the mirror.

(a) How far from the wall should the mirror be?

(b) What is the magnification?

(c) For the image to be right-side up, what should the orientation of the slide be?

12. A standing wave on a 0.8 m string is described by the relation $y = 0.02 \sin 5\pi x \cos 100\pi t$ SI units.

(a) What are the amplitude, frequency, and wavelength?

(b) Draw the standing wave pattern. Which harmonic is this? (Counting the lowest frequency as 1, what is the number of this one?)

(c) The mass of the string is 0.008 kg. What is the tension in the string?

13. Three successive resonance frequencies of an air column are 75, 125, and 175 Hz.

(a) Does this column have one end open and one end closed, or both ends open?

(b) What is the fundamental (lowest) frequency?

(c) Draw the wave pattern for the 75 Hz wave.

14. You are holding a 1.6 m solid brass bar by a clamp at its middle. You strike the bar so that it resonates with its longest wavelength.

(a) Draw the standing wave pattern.

(b) The frequency is 1000 Hz. Describe what would have to be different for the frequency to be 2000 Hz.

(c) Determine the speed of sound in the brass bar.

15. For a demonstration of standing waves a string with length 2.5 m is attached to an oscillator operating at 60 Hz. The other end of the string passes over a pulley to a hanger where various masses can be attached to vary the tension in the string. Transverse standing waves are set up in the string with nodes at the pulley and at the oscillator. The mass per unit length of the string is 8 g/m.

(a) What must the tension be for the string to vibrate with its lowest frequency?

(b) What tensions are needed for the three next higher frequencies?

16. Yellow light with $\lambda = 620$ nm enters an oil film at right angles. The index of refraction of the film is 1.3.

(a) What is the wavelength in the film?

(b) What is the thickness for destructive interference?

17. A spaceship travels away from the earth for 10 days at a speed of 8000 m/s with respect to the earth as measured from earth. What is the difference between that time and the time measured by the astronaut in the spaceship?

Since the difference is small, it is not useful to calculate each time interval separately and then take the difference. Instead use the first two terms of the binomial series in the form $(1+x)^n = 1 + nx + \cdots$ (valid when x is much smaller than 1.) For example, with $x = -\frac{v^2}{c^2}$ and $n = \frac{1}{2}$, $(1 - \frac{v^2}{c^2})^{\frac{1}{2}} = 1 - \frac{1}{2}\frac{v^2}{c^2}$.

Multiple choice questions

1. Which of the following is not a kind of electromagnetic radiation?
   (a) gamma rays
   (b) x-rays
   (c) ultraviolet rays
   (d) electrons
   (e) microwaves

2. A tube is closed at one end and open at the other. It is placed in front of a loudspeaker that is playing the sound generated by a variable frequency audio oscillator. The frequency is slowly increased from 0. The first frequency at which standing waves are generated is 325 Hz and the next is 975 Hz. The speed of sound is 340 m/s. What is the length of the tube?
   (a) 20 cm
   (b) 30 cm
   (c) 40 cm
   (d) 50 cm
   (e) 10 cm

3. A 0.40 m string is clamped at both ends. The lowest frequency for a standing wave is 325 Hz. What is the wave speed?
   (a) 340 m/s
   (b) 260 m/s
   (c) 813 m/s
   (d) 130 m/s
   (e) 406 m/s

4. Ocean waves with a wavelength of 120 m are arriving at the rate of 8 per minute. What is their speed?
   (a) 8.0 m/s
   (b) 16 m/s
   (c) 24 m/s
   (d) 30 m/s
   (e) 4.0 m/s

5. A stretched string, fixed at both ends, is observed to vibrate in three equal segments when it is driven by a 480 Hz oscillator. What is the fundamental (lowest) frequency of standing waves for this string?
   (a) 480 Hz
   (b) 320 Hz
   (c) 160 Hz
   (d) 640 Hz
   (e) 240 Hz

6. A 2 m string, fixed at both ends, is observed to vibrate as a standing wave in six segments. The wave speed is 45 m/s. What is the frequency?
   (a) 270 Hz
   (b) 140 Hz
   (c) 200 Hz
   (d) 68 Hz
   (e) 34 Hz

7. An organ pipe, open at both ends, has successive resonances at 150 Hz and 200 Hz when the velocity of sound in air is 345 m/s. What is its length?
   (a) 5.25 m
   (b) 5.75 m
   (c) 2.76 m
   (d) 4.90 m
   (e) 3.45 m

8. A wave is described by the relation $y = 0.15\ \sin(\frac{\pi}{8}x - 4\pi t)$ SI units. What is the first positive value of x where y is a maximum at $t = 0$.
   (a) 16 m
   (b) 8 m
   (c) 13 m
   (d) 2 m
   (e) 4 m

9. A standing wave on a string is described by $y = 0.080 \sin 6x \cos 600t$ SI units. What is the distance between successive nodes?

(a) 0.24 m
(b) 0.08 m
(c) 0.02 m
(d) 0.52 m
(e) There is insufficient information.

10. Assume that the human vocal tract can be thought of as a tube open at one end. Assume that the length of the tube is 17 cm and that the speed of sound is 340 m/s. What are the lowest two resonant frequencies?

(a) 500 Hz, 1500 Hz
(b) 500 Hz, 1000 Hz
(c) 1000 Hz, 2000 Hz
(d) 1000 Hz, 3000 Hz
(e) 1500 Hz, 2500 Hz

11. Sound travels at 340 m/s in air and at 1500 m/s in water. A sound of 256 Hz is made in water. If the 256 Hz sound is made in air, which of the following is true:

(a) The frequency remains the same and the wavelength is shorter.

(b) The frequency remains the same and the wavelength is longer.

(c) The frequency is lower and the wavelength is longer.

(d) The frequency is higher and the wavelength stays the same.

(e) Both the frequency and the wavelength remain the same.

12. Two strings, fixed at both ends, are 1 m and 2 m long, respectively. Which of the following sets of wavelengths, in m, can represent waves on *both* strings?

(a) 0.8, 0.67, 0.5
(b) 1.33, 1.0, 0.5
(c) 2.0, 1.0, 0.5
(d) 2.0, 1.33, 1.0
(e) 4.0, 2.0, 1.0

13. A string, 4 m long, vibrates according to the relation $y = 0.04 \sin(\pi x) \cos(2\pi t)$ SI units. The number of nodes is

(a) 1
(b) 2
(c) 3
(d) 4
(e) 5

Synthesis problems and projects

1. Go to the PhET website and open the simulation *Fourier: Making Waves*.

Go to "discrete." The first row shows the amplitudes of the first 11 harmonics (f, 2f, 3f, ... 11f). Each can be dragged up or down. The decomposition of a complex wave into its components is called *Fourier analysis*, and their addition is *Fourier synthesis*. The components are called *Fourier components*.

The second row shows the separate waves coresponding to the selection you make in row 1. Start by experimenting with just one component and its amplitude at a time.

The third row shows the sum of the waves that are shown in the second row.

(a) Start with $A_1 = 1$ and $A_{10} = 0.3$ to see how two waves add when the frequencies are quite far apart. Try other combinations.

(b) See how close you can come to a "square wave" (up—straight across down—straight across and up again) with just three harmonics (A_1, A_3, A_5). What are the amplitudes? (Set A_1 to 1, and then drag A_3 up to its best point for what you want to do. Then A_5. Go back to A_3 and adjust it.) Select "Math form" and "Expand sum" to see the equations of the waves.

Select "function: square wave" to see what can be done with 11 components. Compare this with your selection: what kind of improvement can you get when there are more components?

(c) Select "function: wave packet." This shows a combination of waves that add up to a pattern confined to a narrow region in space. You can change the x and y amplitudes with the buttons to the right of the graphs. You can see a better wave packet when you choose "discrete to continuous." Start with $k_1 = \frac{\pi}{4}$, $k_0 = 12\pi$, and $\sigma_k = 2\pi$. Experiment with others values.

(d) Go to "wave game." It asks you to match a given wave by selecting the components. Start with level 1 (very easy) and go on to level 4. Continue as far as you can.

2. You have some string and you want to know how fast a pulse or wave travels along the string. You have a pulley, a meter stick, a scale, some masses, and a hanger for the masses.

What can you do with this equipment to predict the speed of a pulse or wave on the string? List any assumptions that you make.

What additional equipment do you need to test your prediction?

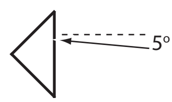
3. Parallel rays to a lens with spherical surfaces do not come together exactly at a focus. This is called "spherical aberration." The *aperture* of a camera is an opening with variable diameter in a disk to limit the part of the lens that is used. In general, a sharper image is produced when the aperture is small because spherical aberration is then smaller. However, the smaller the aperture the greater are the diffraction effects.

An aperture has a diameter of 1 mm. The film is 5 cm from the lens. What is the distance from the central maximum of the diffraction pattern to the first minimum? Use $\lambda = 500$ nm.

The distance to the first minimum produced by a point object characterizes the resolution of the lens. A second image at this distance from the center can just barely be seen separately on the film or screen. The limit imposed by diffraction effects can be overcome by using radiation with a smaller wavelength. This is done, for example, in an *electron microscope*.

4. Calculate the electric and magnetic forces for the two lines of charge near the end of Section 11.7 to show that the forces in the stationary and the moving systems are the same.

5.



Light enters a 45° prism made of glass with an index of refraction of 1.50 at an angle 5° below the horizontal. Describe the subsequent path: draw a ray diagram.