Object A is a brick. Object B is half of a similar brick. If A is heated, we have $\Delta S = \frac{Q}{T}$. Show that if this equation is valid for A, then it is also valid for B.  

Typically the atmosphere gets colder with increasing altitude. However, sometimes there is an inversion layer, in which this trend is reversed, e.g., because a less dense mass of warm air moves into a certain area, and rises above the denser colder air that was already present. Suppose that this causes the pressure $P$ as a function of height $y$ to be given by a function of the form $P = P_o e^{-ky}(1 + by)$, where constant temperature would give $b = 0$ and an inversion layer would give $b > 0$. (a) Infer the units of the constants $P_o$, $k$, and $b$. (b) Find the density of the air as a function of $y$, of the constants, and of the acceleration of gravity $g$. (c) Check that the units of your answer to part b make sense.

(a) Consider a one-dimensional ideal gas consisting of $n$ material particles, at temperature $T$. Trace back through the logic of the equipartition theorem on p. 333 to determine the total energy. (b) Explain why it should matter how many dimensions there are. (c) Gases that we encounter in everyday life are made of atoms, but there are gases made out of other things. For example, soon after the big bang, there was a period when the universe was very hot and dominated by light rather than matter. A particle of light is called a photon, so the early universe was a “photon gas.” For simplicity, consider a photon gas in one dimension. Photons are massless, and we will see in ch. 7 on relativity that for a massless particle, the energy is related to the momentum by $E = pc$, where $c$ is the speed of light. (Note that $p = mv$ does not hold for a photon.) Again, trace back through the logic of equipartition on p. 333. Does the photon gas have the same heat capacity as the one you found in part a?

You use a spoon at room temperature, $22^\circ C$, to mix your coffee, which is at $80^\circ C$. During this brief period of thermal contact, $1.3$ J of heat is transferred from the coffee to the spoon. Find the total change in the entropy of the universe.

The sun is mainly a mixture of hydrogen and helium, some of which is ionized. As a simplified model, let’s pretend that it’s made purely out of neutral, monatomic hydrogen, and that the whole mass of the sun is in thermal equilibrium. Given its mass, it would then contain $1.2 \times 10^{57}$ atoms. It generates energy from nuclear reactions at a rate of $3.8 \times 10^{26}$ W, and it is in a state of equilibrium in which this amount of energy is radiated off into space as light. Suppose that its ability to radiate light were somehow blocked. Find the rate at which its temperature would increase.
In metals, some electrons, called conduction electrons, are free to move around, rather than being bound to one atom. Classical physics gives an adequate description of many of their properties. Consider a metal at temperature $T$, and let $m$ be the mass of the electron. Find expressions for (a) the average kinetic energy of a conduction electron, and (b) the average square of its velocity, $v^2$. (It would not be of much interest to find $\bar{v}$, which is just zero.)

Numerically, $\sqrt{\langle v^2 \rangle}$, called the root-mean-square velocity, comes out to be surprisingly large — about two orders of magnitude greater than the normal thermal velocities we find for atoms in a gas. Why?

Remark: From this analysis, one would think that the conduction electrons would contribute greatly to the heat capacities of metals. In fact they do not contribute very much in most cases; if they did, Dulong and Petit’s observations would not have come out as described in the text. The resolution of this contradiction was only eventually worked out by Sommerfeld in 1933, and involves the fact that electrons obey the Pauli exclusion principle.

Key to symbols:

- easy
- typical
- challenging
- difficult
- very difficult
- $\sqrt{\langle v^2 \rangle}$

An answer check is available at www.lightandmatter.com.
The vibrations of this electric bass string are converted to electrical vibrations, then to sound vibrations, and finally to vibrations of our eardrums.

Chapter 6
Waves

Dandelion. Cello. Read those two words, and your brain instantly conjures a stream of associations, the most prominent of which have to do with vibrations. Our mental category of “dandelion-ness” is strongly linked to the color of light waves that vibrate about half a million billion times a second: yellow. The velvety throb of a cello has as its most obvious characteristic a relatively low musical pitch — the note you’re spontaneously imagining right now might be one whose sound vibrations repeat at a rate of a hundred times a second.

Evolution seems to have designed our two most important senses around the assumption that our environment is made of waves, whereas up until now, we’ve mostly taken the view that Nature can be understood by breaking her down into smaller and smaller parts, ending up with particles as her most fundamental building blocks. Does that work for light and sound? Sound waves are disturbances in air, which is made of atoms, but light, on the other hand, isn’t a vibration of atoms. Light, unlike sound, can travel through a vacuum: if you’re reading this by sunlight, you’re taking advantage of light that had to make it through millions of miles of vacuum to get to you. Waves, then, are not just a trick that vibrating atoms can do. Waves are one of the basic phenomena of the universe. At the
end of this book, we’ll even see that the things we’ve been calling particles, such as electrons, are really waves!\textsuperscript{1}

### 6.1 Free waves

#### 6.1.1 Wave motion

Let’s start with an intuition-building exercise that deals with waves in matter, since they’re easier than light waves to get your hands on. Put your fingertip in the middle of a cup of water and then remove it suddenly. You’ll have noticed two results that are surprising to most people. First, the flat surface of the water does not simply sink uniformly to fill in the volume vacated by your finger. Instead, ripples spread out, and the process of flattening out occurs over a long period of time, during which the water at the center vibrates above and below the normal water level. This type of wave motion is the topic of the present section. Second, you’ve found that the ripples bounce off of the walls of the cup, in much the same way that a ball would bounce off of a wall. In the next section we discuss what happens to waves that have a boundary around them. Until then, we confine ourselves to wave phenomena that can be analyzed as if the medium (e.g., the water) was infinite and the same everywhere.

It isn’t hard to understand why removing your fingertip creates ripples rather than simply allowing the water to sink back down uniformly. The initial crater, a/1, left behind by your finger has sloping sides, and the water next to the crater flows downhill to fill in the hole. The water far away, on the other hand, initially has no way of knowing what has happened, because there is no slope for it to flow down. As the hole fills up, the rising water at the center gains upward momentum, and overshoots, creating a little hill where there had been a hole originally. The area just outside of this region has been robbed of some of its water in order to build the hill, so a depressed “moat” is formed, a/2. This effect cascades outward, producing ripples.

There are three main ways in which wave motion differs from the motion of objects made of matter.

1. **Superposition**

   If you watched the water in the cup carefully, you noticed the ghostlike behavior of the reflected ripples coming back toward the center of the cup and the outgoing ripples that hadn’t yet been reflected: they passed right through each other. This is the first, and the most profound, difference between wave motion and the mo-

\textsuperscript{1}Speaking more carefully, I should say that every basic building block of light and matter has both wave and particle properties.
The two circular patterns of ripples pass through each other. Unlike material objects, wave patterns can overlap in space, and when this happens they combine by addition.

Superposition can occur not just with sinusoidal waves like the ones in the figure above but with waves of any shape. The figures on the following page show superposition of wave pulses. A pulse is simply a wave of very short duration. These pulses consist only of a single hump or trough. If you hit a clothesline sharply, you will observe pulses heading off in both directions. This is analogous to the way ripples spread out in all directions when you make a disturbance at one point on water. The same occurs when the hammer on a piano comes up and hits a string.

Experiments to date have not shown any deviation from the principle of superposition in the case of light waves. For other types of waves, it is typically a very good approximation for low-energy waves.
As the wave pulse goes by, the ribbon tied to the spring is not carried along. The motion of the wave pattern is to the right, but the medium (spring) is moving from side to side, not to the right. (PSSC Physics)

c / As the wave pattern passes the rubber duck, the duck stays put. The water isn’t moving with the wave.

2. The medium is not transported with the wave.

The sequence of three photos in figure c shows a series of water waves before it has reached a rubber duck (left), having just passed the duck (middle) and having progressed about a meter beyond the duck (right). The duck bobs around its initial position, but is not carried along with the wave. This shows that the water itself does not flow outward with the wave. If it did, we could empty one end of a swimming pool simply by kicking up waves! We must distinguish between the motion of the medium (water in this case) and the motion of the wave pattern through the medium. The medium vibrates; the wave progresses through space.

self-check A

In figure d, you can detect the side-to-side motion of the spring because the spring appears blurry. At a certain instant, represented by a single photo, how would you describe the motion of the different parts of the spring? Other than the flat parts, do any parts of the spring have zero velocity?  

⊿ Answer, p. 1061

A worm example 1

The worm in the figure is moving to the right. The wave pattern, a pulse consisting of a compressed area of its body, moves to the left. In other words, the motion of the wave pattern is in the opposite direction compared to the motion of the medium.
The incorrect belief that the medium moves with the wave is often reinforced by garbled secondhand knowledge of surfing. Anyone who has actually surfed knows that the front of the board pushes the water to the sides, creating a wake — the surfer can even drag his hand through the water, as in figure e. If the water was moving along with the wave and the surfer, this wouldn’t happen. The surfer is carried forward because forward is downhill, not because of any forward flow of the water. If the water was flowing forward, then a person floating in the water up to her neck would be carried along just as quickly as someone on a surfboard. In fact, it is even possible to surf down the back side of a wave, although the ride wouldn’t last very long because the surfer and the wave would quickly part company.

3. A wave’s velocity depends on the medium.

A material object can move with any velocity, and can be sped up or slowed down by a force that increases or decreases its kinetic energy. Not so with waves. The speed of a wave, depends on the properties of the medium (and perhaps also on the shape of the wave, for certain types of waves). Sound waves travel at about 340 m/s in air, 1000 m/s in helium. If you kick up water waves in a pool, you will find that kicking harder makes waves that are taller (and therefore carry more energy), not faster. The sound waves from an exploding stick of dynamite carry a lot of energy, but are no faster than any other waves. In the following section we will give an example of the physical relationship between the wave speed and the properties of the medium.

Breaking waves example 3

The velocity of water waves increases with depth. The crest of a wave travels faster than the trough, and this can cause the wave to break.

Once a wave is created, the only reason its speed will change is if it enters a different medium or if the properties of the medium change. It is not so surprising that a change in medium can slow down a wave, but the reverse can also happen. A sound wave traveling through a helium balloon will slow down when it emerges into the air, but if it enters another balloon it will speed back up again! Similarly, water waves travel more quickly over deeper water, so a wave will slow down as it passes over an underwater ridge, but speed up again as it emerges into deeper water.

Hull speed example 4

The speeds of most boats, and of some surface-swimming animals, are limited by the fact that they make a wave due to their motion through the water. The boat in figure g is going at the same speed as its own waves, and can’t go any faster. No matter how hard the boat pushes against the water, it can’t make
the wave move ahead faster and get out of the way. The wave’s speed depends only on the medium. Adding energy to the wave doesn’t speed it up, it just increases its amplitude.

A water wave, unlike many other types of wave, has a speed that depends on its shape: a broader wave moves faster. The shape of the wave made by a boat tends to mold itself to the shape of the boat’s hull, so a boat with a longer hull makes a broader wave that moves faster. The maximum speed of a boat whose speed is limited by this effect is therefore closely related to the length of its hull, and the maximum speed is called the hull speed. Sailboats designed for racing are not just long and skinny to make them more streamlined — they are also long so that their hull speeds will be high.

Wave patterns

If the magnitude of a wave’s velocity vector is preordained, what about its direction? Waves spread out in all directions from every point on the disturbance that created them. If the disturbance is small, we may consider it as a single point, and in the case of water waves the resulting wave pattern is the familiar circular ripple, h/1. If, on the other hand, we lay a pole on the surface of the water and wiggle it up and down, we create a linear wave pattern, h/2. For a three-dimensional wave such as a sound wave, the analogous patterns would be spherical waves and plane waves, i.

Infinitely many patterns are possible, but linear or plane waves are often the simplest to analyze, because the velocity vector is in the same direction no matter what part of the wave we look at. Since all the velocity vectors are parallel to one another, the problem is effectively one-dimensional. Throughout this chapter and the next, we will restrict ourselves mainly to wave motion in one dimension, while not hesitating to broaden our horizons when it can be done without too much complication.
Discussion Questions

A The left panel of the figure shows a sequence of snapshots of two positive pulses on a coil spring as they move through each other. In the right panel, which shows a positive pulse and a negative one, the fifth frame has the spring just about perfectly flat. If the two pulses have essentially canceled each other out perfectly, then why does the motion pick up again? Why doesn’t the spring just stay flat?

B Sketch two positive wave pulses on a string that are overlapping but not right on top of each other, and draw their superposition. Do the same for a positive pulse running into a negative pulse.

C A traveling wave pulse is moving to the right on a string. Sketch the velocity vectors of the various parts of the string. Now do the same for a pulse moving to the left.

D In a spherical sound wave spreading out from a point, how would the energy of the wave fall off with distance?
6.1.2 Waves on a string

So far you’ve learned some counterintuitive things about the behavior of waves, but intuition can be trained. The first half of this subsection aims to build your intuition by investigating a simple, one-dimensional type of wave: a wave on a string. If you have ever stretched a string between the bottoms of two open-mouthed cans to talk to a friend, you were putting this type of wave to work. Stringed instruments are another good example. Although we usually think of a piano wire simply as vibrating, the hammer actually strikes it quickly and makes a dent in it, which then ripples out in both directions. Since this chapter is about free waves, not bounded ones, we pretend that our string is infinitely long.

After the qualitative discussion, we will use simple approximations to investigate the speed of a wave pulse on a string. This quick and dirty treatment is then followed by a rigorous attack using the methods of calculus, which turns out to be both simpler and more general.

**Intuitive ideas**

Consider a string that has been struck, l/1, resulting in the creation of two wave pulses, l/2, one traveling to the left and one to the right. This is analogous to the way ripples spread out in all directions from a splash in water, but on a one-dimensional string, “all directions” becomes “both directions.”

We can gain insight by modeling the string as a series of masses connected by springs, m. (In the actual string the mass and the springiness are both contributed by the molecules themselves.) If we look at various microscopic portions of the string, there will be some areas that are flat, 1, some that are sloping but not curved, 2, and some that are curved, 3 and 4. In example 1 it is clear that both the forces on the central mass cancel out, so it will not accelerate. The same is true of 2, however. Only in curved regions such as 3 and 4 is an acceleration produced. In these examples, the vector sum of the two forces acting on the central mass cancel out, so it will not accelerate. The same is true of 2, however. Only in curved regions such as 3 and 4 is an acceleration produced. In these examples, the vector sum of the two forces acting on the central mass is not zero. The important concept is that curvature makes force: the curved areas of a wave tend to experience forces resulting in an acceleration toward the mouth of the curve. Note, however, that an uncurved portion of the string need not remain motionless. It may move at constant velocity to either side.

**Approximate treatment**

We now carry out an approximate treatment of the speed at which two pulses will spread out from an initial indentation on a string. For simplicity, we imagine a hammer blow that creates a triangular dent, n/1. We will estimate the amount of time, t, required until each of the pulses has traveled a distance equal to the width of the pulse itself. The velocity of the pulses is then $\pm v/t$. 

---

$k$ / Hitting a key on a piano causes a hammer to come up from underneath and hit a string (actually a set of three). The result is a pair of pulses moving away from the point of impact.

$l$ / A pulse on a string splits in two and heads off in both directions.

$m$ / Modeling a string as a series of masses connected by springs.
As always, the velocity of a wave depends on the properties of the medium, in this case the string. The properties of the string can be summarized by two variables: the tension, $T$, and the mass per unit length, $\mu$ (Greek letter mu).

If we consider the part of the string encompassed by the initial dent as a single object, then this object has a mass of approximately $\mu w$ (mass/length $\times$ length = mass). (Here, and throughout the derivation, we assume that $h$ is much less than $w$, so that we can ignore the fact that this segment of the string has a length slightly greater than $w$.) Although the downward acceleration of this segment of the string will be neither constant over time nor uniform across the pulse, we will pretend that it is constant for the sake of our simple estimate. Roughly speaking, the time interval between $n/1$ and $n/2$ is the amount of time required for the initial dent to accelerate from rest and reach its normal, flattened position. Of course the tip of the triangle has a longer distance to travel than the edges, but again we ignore the complications and simply assume that the segment as a whole must travel a distance $h$. Indeed, it might seem surprising that the triangle would so neatly spring back to a perfectly flat shape. It is an experimental fact that it does, but our analysis is too crude to address such details.

The string is kinked, i.e., tightly curved, at the edges of the triangle, so it is here that there will be large forces that do not cancel out to zero. There are two forces acting on the triangular hump, one of magnitude $T$ acting down and to the right, and one of the same magnitude acting down and to the left. If the angle of the sloping sides is $\theta$, then the total force on the segment equals $2T \sin \theta$. Dividing the triangle into two right triangles, we see that $\sin \theta$ equals $h$ divided by the length of one of the sloping sides. Since $h$ is much less than $w$, the length of the sloping side is essentially the same as $w/2$, so we have $\sin \theta = 2h/w$, and $F = 4T h/w$. The acceleration of the segment (actually the acceleration of its center of mass) is

\[
\frac{F}{m} = \frac{4T h}{\mu w^2}.
\]

The time required to move a distance $h$ under constant acceleration $a$ is found by solving $h = (1/2)at^2$ to yield

\[
t = \sqrt{\frac{2h}{a}} = w \sqrt{\frac{\mu}{2T}}.
\]
Our final result for the speed of the pulses is

\[ v = \frac{w}{t} \]

\[ = \sqrt{\frac{2T}{\mu}}. \]

The remarkable feature of this result is that the velocity of the pulses does not depend at all on \( w \) or \( h \), i.e., any triangular pulse has the same speed. It is an experimental fact (and we will also prove rigorously below) that any pulse of any kind, triangular or otherwise, travels along the string at the same speed. Of course, after so many approximations we cannot expect to have gotten all the numerical factors right. The correct result for the speed of the pulses is

\[ v = \sqrt{\frac{T}{\mu}}. \]

The importance of the above derivation lies in the insight it brings — that all pulses move with the same speed — rather than in the details of the numerical result. The reason for our too-high value for the velocity is not hard to guess. It comes from the assumption that the acceleration was constant, when actually the total force on the segment would diminish as it flattened out.

*Treatment using calculus*

After expending considerable effort for an approximate solution, we now display the power of calculus with a rigorous and completely general treatment that is nevertheless much shorter and easier. Let the flat position of the string define the \( x \) axis, so that \( y \) measures how far a point on the string is from equilibrium. The motion of the string is characterized by \( y(x,t) \), a function of two variables. Knowing that the force on any small segment of string depends on the curvature of the string in that area, and that the second derivative is a measure of curvature, it is not surprising to find that the infinitesimal force \( dF \) acting on an infinitesimal segment \( dx \) is given by

\[ dF = T \frac{\partial^2 y}{\partial x^2} \, dx. \]

(This can be proved by vector addition of the two infinitesimal forces acting on either side.) The symbol \( \partial \) stands for a partial derivative, e.g., \( \partial/\partial x \) means a derivative with respect to \( x \) that is evaluated while treating \( t \) as a constant. The acceleration is then \( a = dF/\, dm \), or, substituting \( dm = \mu \, dx \),

\[ \frac{\partial^2 y}{\partial t^2} = \frac{T}{\mu} \frac{\partial^2 y}{\partial x^2}. \]

The second derivative with respect to time is related to the second derivative with respect to position. This is no more than a fancy
mathematical statement of the intuitive fact developed above, that the string accelerates so as to flatten out its curves.

Before even bothering to look for solutions to this equation, we note that it already proves the principle of superposition, because the derivative of a sum is the sum of the derivatives. Therefore the sum of any two solutions will also be a solution.

Based on experiment, we expect that this equation will be satisfied by any function \( y(x, t) \) that describes a pulse or wave pattern moving to the left or right at the correct speed \( v \). In general, such a function will be of the form \( y = f(x - vt) \) or \( y = f(x + vt) \), where \( f \) is any function of one variable. Because of the chain rule, each derivative with respect to time brings out a factor of \( v \). Evaluating the second derivatives on both sides of the equation gives

\[
(\pm v)^2 f'' = \frac{T}{\mu} f''.
\]

Squaring gets rid of the sign, and we find that we have a valid solution for any function \( f \), provided that \( v \) is given by

\[
v = \sqrt{\frac{T}{\mu}}.
\]

6.1.3 Sound and light waves

Sound waves

The phenomenon of sound is easily found to have all the characteristics we expect from a wave phenomenon:

- Sound waves obey superposition. Sounds do not knock other sounds out of the way when they collide, and we can hear more than one sound at once if they both reach our ear simultaneously.

- The medium does not move with the sound. Even standing in front of a titanic speaker playing earsplitting music, we do not feel the slightest breeze.

- The velocity of sound depends on the medium. Sound travels faster in helium than in air, and faster in water than in helium. Putting more energy into the wave makes it more intense, not faster. For example, you can easily detect an echo when you clap your hands a short distance from a large, flat wall, and the delay of the echo is no shorter for a louder clap.

Although not all waves have a speed that is independent of the shape of the wave, and this property therefore is irrelevant to our collection of evidence that sound is a wave phenomenon, sound does nevertheless have this property. For instance, the music in a large concert hall or stadium may take on the order of a second to reach someone seated in the nosebleed section, but we do not notice or
care, because the delay is the same for every sound. Bass, drums, and vocals all head outward from the stage at 340 m/s, regardless of their differing wave shapes. (The speed of sound in a gas is related to the gas’s physical properties in example 13 on p. 389.)

If sound has all the properties we expect from a wave, then what type of wave is it? It is a series of compressions and expansions of the air. Even for a very loud sound, the increase or decrease compared to normal atmospheric pressure is no more than a part per million, so our ears are apparently very sensitive instruments. In a vacuum, there is no medium for the sound waves, and so they cannot exist. The roars and whooshes of space ships in Hollywood movies are fun, but scientifically wrong.

**Light waves**

Entirely similar observations lead us to believe that light is a wave, although the concept of light as a wave had a long and tortuous history. It is interesting to note that Isaac Newton very influentially advocated a contrary idea about light. The belief that matter was made of atoms was stylish at the time among radical thinkers (although there was no experimental evidence for their existence), and it seemed logical to Newton that light as well should be made of tiny particles, which he called corpuscles (Latin for “small objects”). Newton’s triumphs in the science of mechanics, i.e., the study of matter, brought him such great prestige that nobody bothered to question his incorrect theory of light for 150 years. One persuasive proof that light is a wave is that according to Newton’s theory, two intersecting beams of light should experience at least some disruption because of collisions between their corpuscles. Even if the corpuscles were extremely small, and collisions therefore very infrequent, at least some dimming should have been measurable. In fact, very delicate experiments have shown that there is no dimming.

The wave theory of light was entirely successful up until the 20th century, when it was discovered that not all the phenomena of light could be explained with a pure wave theory. It is now believed that both light and matter are made out of tiny chunks which have both wave and particle properties. For now, we will content ourselves with the wave theory of light, which is capable of explaining a great many things, from cameras to rainbows.

If light is a wave, what is waving? What is the medium that wiggles when a light wave goes by? It isn’t air. A vacuum is impenetrable to sound, but light from the stars travels happily through zillions of miles of empty space. Light bulbs have no air inside them, but that doesn’t prevent the light waves from leaving the filament. For a long time, physicists assumed that there must be a mysterious medium for light waves, and they called it the ether (not to be confused with the chemical). Supposedly the ether existed everywhere in space, and was immune to vacuum pumps. The details of
the story are more fittingly reserved for later in this course, but the end result was that a long series of experiments failed to detect any evidence for the ether, and it is no longer believed to exist. Instead, light can be explained as a wave pattern made up of electrical and magnetic fields.

6.1.4 Periodic waves

Period and frequency of a periodic wave

You choose a radio station by selecting a certain frequency. We have already defined period and frequency for vibrations,

\[ T = \text{period} = \text{seconds per cycle} \]
\[ f = \text{frequency} = 1/T = \text{cycles per second} \]
\[ \omega = \text{angular frequency} = 2\pi f = \text{radians per second} \]

but what do they signify in the case of a wave? We can recycle our previous definition simply by stating it in terms of the vibrations that the wave causes as it passes a receiving instrument at a certain point in space. For a sound wave, this receiver could be an eardrum or a microphone. If the vibrations of the eardrum repeat themselves over and over, i.e., are periodic, then we describe the sound wave that caused them as periodic. Likewise we can define the period and frequency of a wave in terms of the period and frequency of the vibrations it causes. As another example, a periodic water wave would be one that caused a rubber duck to bob in a periodic manner as they passed by it.

The period of a sound wave correlates with our sensory impression of musical pitch. A high frequency (short period) is a high note. The sounds that really define the musical notes of a song are only the ones that are periodic. It is not possible to sing a nonperiodic sound like “sh” with a definite pitch.

The frequency of a light wave corresponds to color. Violet is the high-frequency end of the rainbow, red the low-frequency end. A color like brown that does not occur in a rainbow is not a periodic light wave. Many phenomena that we do not normally think of as light are actually just forms of light that are invisible because they fall outside the range of frequencies our eyes can detect. Beyond the red end of the visible rainbow, there are infrared and radio waves. Past the violet end, we have ultraviolet, x-rays, and gamma rays.

Graphs of waves as a function of position

Some waves, like sound waves, are easy to study by placing a detector at a certain location in space and studying the motion as a function of time. The result is a graph whose horizontal axis is time. With a water wave, on the other hand, it is simpler just to look at the wave directly. This visual snapshot amounts to a graph of the height of the water wave as a function of position. Any wave can be represented in either way.
An easy way to visualize this is in terms of a strip chart recorder, an obsolescing device consisting of a pen that wiggles back and forth as a roll of paper is fed under it. It can be used to record a person’s electrocardiogram, or seismic waves too small to be felt as a noticeable earthquake but detectable by a seismometer. Taking the seismometer as an example, the chart is essentially a record of the ground’s wave motion as a function of time, but if the paper was set to feed at the same velocity as the motion of an earthquake wave, it would also be a full-scale representation of the profile of the actual wave pattern itself. Assuming, as is usually the case, that the wave velocity is a constant number regardless of the wave’s shape, knowing the wave motion as a function of time is equivalent to knowing it as a function of position.

**Wavelength**

Any wave that is periodic will also display a repeating pattern when graphed as a function of position. The distance spanned by one repetition is referred to as one wavelength. The usual notation for wavelength is \( \lambda \), the Greek letter lambda. Wavelength is to space as period is to time.

**Wave velocity related to frequency and wavelength**

Suppose that we create a repetitive disturbance by kicking the surface of a swimming pool. We are essentially making a series of wave pulses. The wavelength is simply the distance a pulse is able to travel before we make the next pulse. The distance between pulses is \( \lambda \), and the time between pulses is the period, \( T \), so the speed of the wave is the distance divided by the time,

\[
v = \frac{\lambda}{T}.
\]

This important and useful relationship is more commonly written in terms of the frequency,

\[
v = f\lambda.
\]

**Example 5**

The speed of light is \( 3.0 \times 10^8 \) m/s. What is the wavelength of the radio waves emitted by KMHD, a station whose frequency is 89.1 MHz?

Solving for wavelength, we have

\[
\lambda = \frac{v}{f}
\]

\[
= \frac{(3.0 \times 10^8 \text{ m/s})}{(89.1 \times 10^6 \text{ s}^{-1})}
\]

\[
= 3.4 \text{ m}
\]

The size of a radio antenna is closely related to the wavelength of the waves it is intended to receive. The match need not be exact.
Ultrasound, i.e., sound with frequencies higher than the range of human hearing, was used to make this image of a fetus. The resolution of the image is related to the wavelength, since details smaller than about one wavelength cannot be resolved. High resolution therefore requires a short wavelength, corresponding to a high frequency.

A water wave traveling into a region with different depth will change its wavelength. (since after all one antenna can receive more than one wavelength!), but the ordinary “whip” antenna such as a car’s is 1/4 of a wavelength. An antenna optimized to receive KMHD’s signal would have a length of \((3.4 \, \text{m})/4 = 0.85 \, \text{m}\).

The equation \(v = f\lambda\) defines a fixed relationship between any two of the variables if the other is held fixed. The speed of radio waves in air is almost exactly the same for all wavelengths and frequencies (it is exactly the same if they are in a vacuum), so there is a fixed relationship between their frequency and wavelength. Thus we can say either “Are we on the same wavelength?” or “Are we on the same frequency?”

A different example is the behavior of a wave that travels from a region where the medium has one set of properties to an area where the medium behaves differently. The frequency is now fixed, because otherwise the two portions of the wave would otherwise get out of step, causing a kink or discontinuity at the boundary, which would be unphysical. (A more careful argument is that a kink or discontinuity would have infinite curvature, and waves tend to flatten out their curvature. An infinite curvature would flatten out infinitely fast, i.e., it could never occur in the first place.) Since the frequency must stay the same, any change in the velocity that results from the new medium must cause a change in wavelength.

The velocity of water waves depends on the depth of the water, so based on \(\lambda = v/f\), we see that water waves that move into a region of different depth must change their wavelength, as shown in figure U. This effect can be observed when ocean waves come up to the shore. If the deceleration of the wave pattern is sudden enough, the tip of the wave can curl over, resulting in a breaking wave.

A note on dispersive waves

The discussion of wave velocity given here is actually a little bit of an oversimplification for a wave whose velocity depends on its frequency and wavelength. Such a wave is called a dispersive wave. Nearly all the waves we deal with in this course are nondispersive, but the issue becomes important in chapter 13, where it is discussed in detail.

Sinusoidal waves

Sinusoidal waves are the most important special case of periodic waves. In fact, many scientists and engineers would be uncomfortable with defining a waveform like the “ah” vowel sound as having a definite frequency and wavelength, because they consider only sine waves to be pure examples of a certain frequency and wavelengths. Their bias is not unreasonable, since the French mathematician Fourier showed that any periodic wave with frequency \(f\) can be constructed as a superposition of sine waves with frequencies \(f, 2f, 3f, \ldots\) In this sense, sine waves are the basic, pure building blocks of any waveform.
blocks of all waves. (Fourier’s result so surprised the mathematical community of France that he was ridiculed the first time he publicly presented his theorem.)

However, what definition to use is really a matter of convenience. Our sense of hearing perceives any two sounds having the same period as possessing the same pitch, regardless of whether they are sine waves or not. This is undoubtedly because our ear-brain system evolved to be able to interpret human speech and animal noises, which are periodic but not sinusoidal. Our eyes, on the other hand, judge a color as pure (belonging to the rainbow set of colors) only if it is a sine wave.

**Discussion Questions**

A Suppose we superimpose two sine waves with equal amplitudes but slightly different frequencies, as shown in the figure. What will the superposition look like? What would this sound like if they were sound waves?

[Image of two sine waves superimposed]

Discussion question A.

### 6.1.5 The Doppler effect

Figure w shows the wave pattern made by the tip of a vibrating rod which is moving across the water. If the rod had been vibrating in one place, we would have seen the familiar pattern of concentric circles, all centered on the same point. But since the source of the waves is moving, the wavelength is shortened on one side and lengthened on the other. This is known as the Doppler effect.

Note that the velocity of the waves is a fixed property of the medium, so for example the forward-going waves do not get an extra boost in speed as would a material object like a bullet being shot forward from an airplane.

We can also infer a change in frequency. Since the velocity is constant, the equation \( v = f \lambda \) tells us that the change in wavelength must be matched by an opposite change in frequency: higher frequency for the waves emitted forward, and lower for the ones emitted backward. The frequency Doppler effect is the reason for the familiar dropping-pitch sound of a race car going by. As the car approaches us, we hear a higher pitch, but after it passes us we hear
a frequency that is lower than normal.

The Doppler effect will also occur if the observer is moving but the source is stationary. For instance, an observer moving toward a stationary source will perceive one crest of the wave, and will then be surrounded by the next crest sooner than she otherwise would have, because she has moved toward it and hastened her encounter with it. Roughly speaking, the Doppler effect depends only on the relative motion of the source and the observer, not on their absolute state of motion (which is not a well-defined notion in physics) or on their velocity relative to the medium.

Restricting ourselves to the case of a moving source, and to waves emitted either directly along or directly against the direction of motion, we can easily calculate the wavelength, or equivalently the frequency, of the Doppler-shifted waves. Let $u$ be the velocity of the source. The wavelength of the forward-emitted waves is shortened by an amount $uT$ equal to the distance traveled by the source over the course of one period. Using the definition $f = 1/T$ and the equation $v = f\lambda$, we find for the wavelength $\lambda'$ of the Doppler-shifted wave the equation

$$\lambda' = \left(1 - \frac{u}{v}\right) \lambda.$$

A similar equation can be used for the backward-emitted waves, but with a plus sign rather than a minus sign.

### Example 6

If a race car moves at a velocity of 50 m/s, and the velocity of sound is 340 m/s, by what percentage are the wavelength and frequency of its sound waves shifted for an observer lying along its line of motion?

For an observer whom the car is approaching, we find

$$1 - \frac{u}{v} = 0.85,$$

so the shift in wavelength is 15%. Since the frequency is inversely proportional to the wavelength for a fixed value of the speed of sound, the frequency is shifted upward by

$$1/0.85 = 1.18,$$

i.e., a change of 18%. (For velocities that are small compared to the wave velocities, the Doppler shifts of the wavelength and frequency are about the same.)

### Example 7

What is the percent shift in the wavelength of the light waves emitted by a race car’s headlights?

Looking up the speed of light in the back of the book, $v = 3.0 \times 10^8$ m/s, we find

$$1 - \frac{u}{v} = 0.99999983,$$
The galaxy M100 in the constellation Coma Berenices. Under higher magnification, the milky clouds reveal themselves to be composed of trillions of stars.

The telescope at Mount Wilson used by Hubble.

i.e., the percentage shift is only 0.000017%.

The second example shows that under ordinary earthbound circumstances, Doppler shifts of light are negligible because ordinary things go so much slower than the speed of light. It’s a different story, however, when it comes to stars and galaxies, and this leads us to a story that has profound implications for our understanding of the origin of the universe.

The Big Bang

As soon as astronomers began looking at the sky through telescopes, they began noticing certain objects that looked like clouds in deep space. The fact that they looked the same night after night meant that they were beyond the earth’s atmosphere. Not knowing what they really were, but wanting to sound official, they called them “nebulae,” a Latin word meaning “clouds” but sounding more impressive. In the early 20th century, astronomers realized that although some really were clouds of gas (e.g., the middle “star” of Orion’s sword, which is visibly fuzzy even to the naked eye when conditions are good), others were what we now call galaxies: virtual island universes consisting of trillions of stars (for example the Andromeda Galaxy, which is visible as a fuzzy patch through binoculars). Three hundred years after Galileo had resolved the Milky Way into individual stars through his telescope, astronomers realized that the universe is made of galaxies of stars, and the Milky Way is simply the visible part of the flat disk of our own galaxy, seen from inside.

This opened up the scientific study of cosmology, the structure and history of the universe as a whole, a field that had not been seriously attacked since the days of Newton. Newton had realized that if gravity was always attractive, never repulsive, the universe would have a tendency to collapse. His solution to the problem was to posit a universe that was infinite and uniformly populated with matter, so that it would have no geometrical center. The gravitational forces in such a universe would always tend to cancel out by symmetry, so there would be no collapse. By the 20th century, the belief in an unchanging and infinite universe had become conventional wisdom in science, partly as a reaction against the time that had been wasted trying to find explanations of ancient geological phenomena based on catastrophes suggested by biblical events like Noah’s flood.

In the 1920’s astronomer Edwin Hubble began studying the Doppler shifts of the light emitted by galaxies. A former college football player with a serious nicotine addiction, Hubble did not set out to change our image of the beginning of the universe. His autobiography seldom even mentions the cosmological discovery for which he is now remembered. When astronomers began to study the
Doppler shifts of galaxies, they expected that each galaxy’s direction and velocity of motion would be essentially random. Some would be approaching us, and their light would therefore be Doppler-shifted to the blue end of the spectrum, while an equal number would be expected to have red shifts. What Hubble discovered instead was that except for a few very nearby ones, all the galaxies had red shifts, indicating that they were receding from us at a hefty fraction of the speed of light. Not only that, but the ones farther away were receding more quickly. The speeds were directly proportional to their distance from us.

Did this mean that the earth (or at least our galaxy) was the center of the universe? No, because Doppler shifts of light only depend on the relative motion of the source and the observer. If we see a distant galaxy moving away from us at 10% of the speed of light, we can be assured that the astronomers who live in that galaxy will see ours receding from them at the same speed in the opposite direction. The whole universe can be envisioned as a rising loaf of raisin bread. As the bread expands, there is more and more space between the raisins. The farther apart two raisins are, the greater the speed with which they move apart.

The universe’s expansion is presumably decelerating because of gravitational attraction among the galaxies. We do not presently know whether there is enough mass in the universe to cause enough attraction to halt the expansion eventually. But perhaps more interesting than the distant future of the universe is what its present expansion implies about its past. Extrapolating backward in time using the known laws of physics, the universe must have been denser and denser at earlier and earlier times. At some point, it must have been extremely dense and hot, and we can even detect the radiation from this early fireball, in the form of microwave radiation that permeates space. The phrase Big Bang was originally coined by the doubters of the theory to make it sound ridiculous, but it stuck, and today essentially all astronomers accept the Big Bang theory based on the very direct evidence of the red shifts and the cosmic microwave background radiation.

Finally it should be noted what the Big Bang theory is not. It is not an explanation of why the universe exists. Such questions belong to the realm of religion, not science. Science can find ever simpler and ever more fundamental explanations for a variety of phenomena, but ultimately science takes the universe as it is according to observations.

Furthermore, there is an unfortunate tendency, even among many scientists, to speak of the Big Bang theory as a description of the very first event in the universe, which caused everything after it. Although it is true that time may have had a beginning (Einstein’s theory of general relativity admits such a possibility), the methods
Shock waves are created by the X-15 rocket plane, flying at 3.5 times the speed of sound. This fighter jet has just accelerated past the speed of sound. The sudden decompression of the air causes water droplets to condense, forming a cloud.

A note on Doppler shifts of light

If Doppler shifts depend only on the relative motion of the source and receiver, then there is no way for a person moving with the source and another person moving with the receiver to determine who is moving and who isn’t. Either can blame the Doppler shift entirely on the other’s motion and claim to be at rest herself. This is entirely in agreement with the principle stated originally by Galileo that all motion is relative.

On the other hand, a careful analysis of the Doppler shifts of water or sound waves shows that it is only approximately true, at low speeds, that the shifts just depend on the relative motion of the source and observer. For instance, it is possible for a jet plane to keep up with its own sound waves, so that the sound waves appear to stand still to the pilot of the plane. The pilot then knows she is moving at exactly the speed of sound. The reason this doesn’t disprove the relativity of motion is that the pilot is not really determining her absolute motion but rather her motion relative to the air, which is the medium of the sound waves.

Einstein realized that this solved the problem for sound or water waves, but would not salvage the principle of relative motion in the case of light waves, since light is not a vibration of any physical medium such as water or air. Beginning by imagining what a beam of light would look like to a person riding a motorcycle alongside it, Einstein eventually came up with a radical new way of describing the universe, in which space and time are distorted as measured by observers in different states of motion. As a consequence of this
Theory of Relativity, he showed that light waves would have Doppler shifts that would exactly, not just approximately, depend only on the relative motion of the source and receiver.

**Discussion Questions**

**A** If an airplane travels at exactly the speed of sound, what would be the wavelength of the forward-emitted part of the sound waves it emitted? How should this be interpreted, and what would actually happen? What happens if it's going faster than the speed of sound? Can you use this to explain what you see in figures aa and ab?

**B** If bullets go slower than the speed of sound, why can a supersonic fighter plane catch up to its own sound, but not to its own bullets?

**C** If someone inside a plane is talking to you, should their speech be Doppler shifted?
6.2 Bounded waves

Speech is what separates humans most decisively from animals. No other species can master syntax, and even though chimpanzees can learn a vocabulary of hand signs, there is an unmistakable difference between a human infant and a baby chimp: starting from birth, the human experiments with the production of complex speech sounds.

Since speech sounds are instinctive for us, we seldom think about them consciously. How do we control sound waves so skillfully? Mostly we do it by changing the shape of a connected set of hollow cavities in our chest, throat, and head. Somehow by moving the boundaries of this space in and out, we can produce all the vowel sounds. Up until now, we have been studying only those properties of waves that can be understood as if they existed in an infinite, open space with no boundaries. In this chapter we address what happens when a wave is confined within a certain space, or when a wave pattern encounters the boundary between two different media, such as when a light wave moving through air encounters a glass windowpane.

6.2.1 Reflection, transmission, and absorption

Reflection and transmission

Sound waves can echo back from a cliff, and light waves are reflected from the surface of a pond. We use the word reflection, normally applied only to light waves in ordinary speech, to describe any such case of a wave rebounding from a barrier. Figure (a) shows a circular water wave being reflected from a straight wall. In this chapter, we will concentrate mainly on reflection of waves that move in one dimension, as in figure c/1.

Wave reflection does not surprise us. After all, a material object such as a rubber ball would bounce back in the same way. But waves are not objects, and there are some surprises in store.

First, only part of the wave is usually reflected. Looking out through a window, we see light waves that passed through it, but a person standing outside would also be able to see her reflection in the glass. A light wave that strikes the glass is partly reflected and partly transmitted (passed) by the glass. The energy of the original wave is split between the two. This is different from the behavior of the rubber ball, which must go one way or the other, not both.

Second, consider what you see if you are swimming underwater and you look up at the surface. You see your own reflection. This is utterly counterintuitive, since we would expect the light waves to burst forth to freedom in the wide-open air. A material projectile shot up toward the surface would never rebound from the water-air boundary!

What is it about the difference between two media that causes
waves to be partly reflected at the boundary between them? Is it their density? Their chemical composition? Typically all that matters is the speed of the wave in the two media.\(^2\) A wave is partially reflected and partially transmitted at the boundary between media in which it has different speeds. For example, the speed of light waves in window glass is about 30% less than in air, which explains why windows always make reflections. Figure c shows examples of wave pulses being reflected at the boundary between two coil springs of different weights, in which the wave speed is different.

![Image](image_url)

Reflections such as b and c/1, where a wave encounters a massive fixed object, can usually be understood on the same basis as cases like c/2 and c/3 where two media meet. Example c/1, for instance, is like a more extreme version of example c/2. If the heavy coil spring in c/2 was made heavier and heavier, it would end up acting like the fixed wall to which the light spring in c/1 has been attached.

**self-check B**

In figure c/1, the reflected pulse is upside-down, but its depth is just as big as the original pulse’s height. How does the energy of the reflected pulse compare with that of the original?

\(\text{Answer, p. 1061}\)

\(^1\) *Fish have internal ears.*

Why don’t fish have ear-holes? The speed of sound waves in a fish’s body is not much different from their speed in water, so sound waves are not strongly reflected from a fish’s skin. They pass right through its body, so fish can have internal ears.

\(^2\) Some exceptions are described in sec. 6.2.5, p. 389.
Sound waves travel at drastically different speeds through rock, water, and air. Whale songs are thus strongly reflected both at both the bottom and the surface. The sound waves can travel hundreds of miles, bouncing repeatedly between the bottom and the surface, and still be detectable. Sadly, noise pollution from ships has nearly shut down this cetacean version of the internet.

Radio communication can occur between stations on opposite sides of the planet. The mechanism is entirely similar to the one explained in the previous example, but the three media involved are the earth, the atmosphere, and the ionosphere.

Sonar is a method for ships and submarines to detect each other by producing sound waves and listening for echoes. What properties would an underwater object have to have in order to be invisible to sonar? (Answer, p. 1061)

The use of the word “reflection” naturally brings to mind the creation of an image by a mirror, but this might be confusing, because we do not normally refer to “reflection” when we look at surfaces that are not shiny. Nevertheless, reflection is how we see the surfaces of all objects, not just polished ones. When we look at a sidewalk, for example, we are actually seeing the reflecting of the sun from the concrete. The reason we don’t see an image of the sun at our feet is simply that the rough surface blurs the image so drastically.

Inverted and uninverted reflections

Notice how the pulse reflected back to the right in example c/2 comes back upside-down, whereas the one reflected back to the left in c/3 returns in its original upright form. This is true for other waves as well. In general, there are two possible types of reflections, a reflection back into a faster medium and a reflection back into a slower medium. One type will always be an inverting reflection and one noninverting.

It’s important to realize that when we discuss inverted and uninverted reflections on a string, we are talking about whether the wave is flipped across the direction of motion (i.e., upside-down in these drawings). The reflected pulse will always be reversed front to back, as shown in figures d and e. This is because it is traveling in the other direction. The leading edge of the pulse is what gets reflected first, so it is still ahead when it starts back to the left — it’s just that “ahead” is now in the opposite direction.
Absorption

So far we have tacitly assumed that wave energy remains as wave energy, and is not converted to any other form. If this was true, then the world would become more and more full of sound waves, which could never escape into the vacuum of outer space. In reality, any mechanical wave consists of a traveling pattern of vibrations of some physical medium, and vibrations of matter always produce heat, as when you bend a coathanger back and forth and it becomes hot. We can thus expect that in mechanical waves such as water waves, sound waves, or waves on a string, the wave energy will gradually be converted into heat. This is referred to as absorption. The reduction in the wave’s energy can also be described as a reduction in amplitude, the relationship between them being, as with a vibrating object, \( E \propto A^2 \).

The wave suffers a decrease in amplitude, as shown in figure f. The decrease in amplitude amounts to the same fractional change for each unit of distance covered. For example, if a wave decreases from amplitude 2 to amplitude 1 over a distance of 1 meter, then after traveling another meter it will have an amplitude of 1/2. That is, the reduction in amplitude is exponential. This can be proved as follows. By the principle of superposition, we know that a wave of amplitude 2 must behave like the superposition of two identical waves of amplitude 1. If a single amplitude-1 wave would die down to amplitude 1/2 over a certain distance, then two amplitude-1 waves superposed on top of one another to make amplitude 1+1=2 must die down to amplitude 1/2+1/2=1 over the same distance.

**self-check D**

As a wave undergoes absorption, it loses energy. Does this mean that it slows down?  
> Answer, p. 1061

In many cases, this frictional heating effect is quite weak. Sound waves in air, for instance, dissipate into heat extremely slowly, and the sound of church music in a cathedral may reverberate for as much as 3 or 4 seconds before it becomes inaudible. During this time it has traveled over a kilometer! Even this very gradual dissipation of energy occurs mostly as heating of the church’s walls and by the leaking of sound to the outside (where it will eventually end up as heat). Under the right conditions (humid air and low frequency), a sound wave in a straight pipe could theoretically travel hundreds of kilometers before being noticeable attenuated.

In general, the absorption of mechanical waves depends a great deal on the chemical composition and microscopic structure of the medium. Ripples on the surface of antifreeze, for instance, die out extremely rapidly compared to ripples on water. For sound waves and surface waves in liquids and gases, what matters is the viscosity of the substance, i.e., whether it flows easily like water or mercury or more sluggishly like molasses or antifreeze. This explains why
our intuitive expectation of strong absorption of sound in water is incorrect. Water is a very weak absorber of sound (viz. whale songs and sonar), and our incorrect intuition arises from focusing on the wrong property of the substance: water’s high density, which is irrelevant, rather than its low viscosity, which is what matters.

Light is an interesting case, since although it can travel through matter, it is not itself a vibration of any material substance. Thus we can look at the star Sirius, $10^{14}$ km away from us, and be assured that none of its light was absorbed in the vacuum of outer space during its 9-year journey to us. The Hubble Space Telescope routinely observes light that has been on its way to us since the early history of the universe, billions of years ago. Of course the energy of light can be dissipated if it does pass through matter (and the light from distant galaxies is often absorbed if there happen to be clouds of gas or dust in between).

**Soundproofing**

Typical amateur musicians setting out to soundproof their garages tend to think that they should simply cover the walls with the densest possible substance. In fact, sound is not absorbed very strongly even by passing through several inches of wood. A better strategy for soundproofing is to create a sandwich of alternating layers of materials in which the speed of sound is very different, to encourage reflection.

The classic design is alternating layers of fiberglass and plywood. The speed of sound in plywood is very high, due to its stiffness, while its speed in fiberglass is essentially the same as its speed in air. Both materials are fairly good sound absorbers, but sound waves passing through a few inches of them are still not going to be absorbed sufficiently. The point of combining them is that a sound wave that tries to get out will be strongly reflected at each of the fiberglass-plywood boundaries, and will bounce back and forth many times like a ping pong ball. Due to all the back-and-forth motion, the sound may end up traveling a total distance equal to ten times the actual thickness of the soundproofing before it escapes. This is the equivalent of having ten times the thickness of sound-absorbing material.

**Radio transmission**

A radio transmitting station must have a length of wire or cable connecting the amplifier to the antenna. The cable and the antenna act as two different media for radio waves, and there will therefore be partial reflection of the waves as they come from the cable to the antenna. If the waves bounce back and forth many times between the amplifier and the antenna, a great deal of their energy will be absorbed. There are two ways to attack the problem. One possibility is to design the antenna so that the speed of the waves in it is as close as possible to the speed of the waves
in the cable; this minimizes the amount of reflection. The other method is to connect the amplifier to the antenna using a type of wire or cable that does not strongly absorb the waves. Partial reflection then becomes irrelevant, since all the wave energy will eventually exit through the antenna.

**Discussion Questions**

A. A sound wave that underwent a pressure-inverting reflection would have its compressions converted to expansions and vice versa. How would its energy and frequency compare with those of the original sound? Would it sound any different? What happens if you swap the two wires where they connect to a stereo speaker, resulting in waves that vibrate in the opposite way?

6.2.2 **Quantitative treatment of reflection**

In this section we use the example of waves on a string to analyze the reasons why a reflection occurs at the boundary between media, predict quantitatively the intensities of reflection and transmission, and discuss how to tell which reflections are inverting and which are noninverting. Some technical details are relegated to sec. 6.2.5, p. 389.

**Why reflection occurs**

To understand the fundamental reasons for what does occur at the boundary between media, let’s first discuss what doesn’t happen. For the sake of concreteness, consider a sinusoidal wave on a string. If the wave progresses from a heavier portion of the string, in which its velocity is low, to a lighter-weight part, in which it is high, then the equation \( v = f\lambda \) tells us that it must change its frequency, or its wavelength, or both. If only the frequency changed, then the parts of the wave in the two different portions of the string would quickly get out of step with each other, producing a discontinuity in the wave, g/1. This is unphysical, so we know that the wavelength must change while the frequency remains constant, g/2.

But there is still something unphysical about figure g/2. The sudden change in the shape of the wave has resulted in a sharp kink at the boundary. This can’t really happen, because the medium tends to accelerate in such a way as to eliminate curvature. A sharp kink corresponds to an infinite curvature at one point, which would produce an infinite acceleration, which would not be consistent with the smooth pattern of wave motion envisioned in fig. g/2. Waves can have kinks, but not stationary kinks.

We conclude that without positing partial reflection of the wave, we cannot simultaneously satisfy the requirements of (1) continuity of the wave, and (2) no sudden changes in the slope of the wave. In other words, we assume that both the wave and its derivative are continuous functions.)
Does this amount to a proof that reflection occurs? Not quite. We have only proved that certain types of wave motion are not valid solutions. In the following subsection, we prove that a valid solution can always be found in which a reflection occurs. Now in physics, we normally assume (but seldom prove formally) that the equations of motion have a unique solution, since otherwise a given set of initial conditions could lead to different behavior later on, but the Newtonian universe is supposed to be deterministic. Since the solution must be unique, and we derive below a valid solution involving a reflected pulse, we will have ended up with what amounts to a proof of reflection.

Intensity of reflection

I will now show, in the case of waves on a string, that it is possible to satisfy the physical requirements given above by constructing a reflected wave, and as a bonus this will produce an equation for the proportions of reflection and transmission and a prediction as to which conditions will lead to inverted and which to uninverted reflection. We assume only that the principle of superposition holds, which is a good approximation for waves on a string of sufficiently small amplitude.

Let the unknown amplitudes of the reflected and transmitted waves be \( R \) and \( T \), respectively. An inverted reflection would be represented by a negative value of \( R \). We can without loss of generality take the incident (original) wave to have unit amplitude. Superposition tells us that if, for instance, the incident wave had double this amplitude, we could immediately find a corresponding solution simply by doubling \( R \) and \( T \).

Just to the left of the boundary, the height of the wave is given by the height 1 of the incident wave, plus the height \( R \) of the part of the reflected wave that has just been created and begun heading back, for a total height of \( 1 + R \). On the right side immediately next to the boundary, the transmitted wave has a height \( T \). To avoid a discontinuity, we must have

\[
1 + R = T.
\]

Next we turn to the requirement of equal slopes on both sides of the boundary. Let the slope of the incoming wave be \( s \) immediately to the left of the junction. If the wave was 100% reflected, and without inversion, then the slope of the reflected wave would be \( -s \), since the wave has been reversed in direction. In general, the slope of the reflected wave equals \( -sR \), and the slopes of the superposed waves on the left side add up to \( s - sR \). On the right, the slope depends on the amplitude, \( T \), but is also changed by the stretching or compression of the wave due to the change in speed. If, for example, the wave speed is twice as great on the right side, then the slope is cut in half by this effect. The slope on the right is
Therefore \( s(v_1/v_2)T \), where \( v_1 \) is the velocity in the original medium and \( v_2 \) the velocity in the new medium. Equality of slopes gives \( s - sR = s(v_1/v_2)T \), or

\[
1 - R = \frac{v_1}{v_2} T.
\]

Solving the two equations for the unknowns \( R \) and \( T \) gives

\[
R = \frac{v_2 - v_1}{v_2 + v_1}
\]

and

\[
T = \frac{2v_2}{v_2 + v_1}.
\]

The first equation shows that there is no reflection unless the two wave speeds are different, and that the reflection is inverted in reflection back into a fast medium.

The energies of the transmitted and reflected waves always add up to the same as the energy of the original wave. There is never any abrupt loss (or gain) in energy when a wave crosses a boundary; conversion of wave energy to heat occurs for many types of waves, but it occurs throughout the medium. The equation for \( T \), surprisingly, allows the amplitude of the transmitted wave to be greater than 1, i.e., greater than that of the incident wave. This does not violate conservation of energy, because this occurs when the second string is less massive, reducing its kinetic energy, and the transmitted pulse is broader and less strongly curved, which lessens its potential energy. In other words, the constant of proportionality in \( E \propto A^2 \) is different in the two different media.

We have attempted to develop some general facts about wave reflection by using the specific example of a wave on a string, which raises the question of whether these facts really are general. These issues are discussed in more detail in optional section 6.2.5, p. 389, but here is a brief summary.

The following facts are more generally true for wave reflection in one dimension.

- The wave is partially reflected and partially transmitted, with the reflected and transmitted parts sharing the energy.

- For an interface between media 1 and 2, there are two possible reflections: back into 1, and back into 2. One of these is inverting \((R < 0)\) and the other is noninverting \((R > 0)\).

The following aspects of our analysis may need to be modified for different types of waves.
- In some cases, the expressions for the reflected and transmitted amplitudes depend not on the ratio \( v_1/v_2 \) but on some more complicated ratio \( v_1\ldots/v_2\ldots \), where \( \ldots \) stands for some additional property of the medium.

- The sign of \( R \), depends not just on this ratio but also on the type of the wave and on what we choose as a measure of amplitude.

### 6.2.3 Interference effects

If you look at the front of a pair of high-quality binoculars, you will notice a greenish-blue coating on the lenses. This is advertised as a coating to prevent reflection. Now reflection is clearly undesirable — we want the light to go in the binoculars — but so far I’ve described reflection as an unalterable fact of nature, depending only on the properties of the two wave media. The coating can’t change the speed of light in air or in glass, so how can it work? The key is that the coating itself is a wave medium. In other words, we have a three-layer sandwich of materials: air, coating, and glass. We will analyze the way the coating works, not because optical coatings are an important part of your education but because it provides a good example of the general phenomenon of wave interference effects.

There are two different interfaces between media: an air-coating boundary and a coating-glass boundary. Partial reflection and partial transmission will occur at each boundary. For ease of visualization let’s start by considering an equivalent system consisting of three dissimilar pieces of string tied together, and a wave pattern consisting initially of a single pulse. Figure i/1 shows the incident pulse moving through the heavy rope, in which its velocity is low. When it encounters the lighter-weight rope in the middle, a faster medium, it is partially reflected and partially transmitted. (The transmitted pulse is bigger, but nevertheless has only part of the original energy.) The pulse transmitted by the first interface is then partially reflected and partially transmitted by the second boundary, i/3. In figure i/4, two pulses are on the way back out to the left, and a single pulse is heading off to the right. (There is still a weak pulse caught between the two boundaries, and this will rattle back and forth, rapidly getting too weak to detect as it leaks energy to the outside with each partial reflection.)

Note how, of the two reflected pulses in i/4, one is inverted and one uninverted. One underwent reflection at the first boundary (a reflection back into a slower medium is uninverted), but the other was reflected at the second boundary (reflection back into a faster medium is inverted).

Now let’s imagine what would have happened if the incoming wave pattern had been a long sinusoidal wave train instead of a single pulse. The first two waves to reemerge on the left could be
A pulse bounces back and forth. We model a guitar string attached to the guitar’s body at both ends as a light-weight string attached to extremely heavy strings at its ends.

This whole analysis applies directly to our original case of optical coatings. Visible light from most sources does consist of a stream of short sinusoidal wave-trains such as the ones drawn above. The only real difference between the waves-on-a-rope example and the case of an optical coating is that the first and third media are air and glass, in which light does not have the same speed. However, the general result is the same as long as the air and the glass have light-wave speeds that are either both greater than the coating’s or both less than the coating’s.

The business of optical coatings turns out to be a very arcane one, with a plethora of trade secrets and “black magic” techniques handed down from master to apprentice. Nevertheless, the ideas you have learned about waves in general are sufficient to allow you to come to some definite conclusions without any further technical knowledge. The self-check and discussion questions will direct you along these lines of thought.

**self-check E**

Color corresponds to wavelength of light waves. Is it possible to choose a thickness for an optical coating that will produce destructive interference for all colors of light?  

> Answer, p. 1061

This example was typical of a wide variety of wave interference effects. With a little guidance, you are now ready to figure out for yourself other examples such as the rainbow pattern made by a compact disc or by a layer of oil on a puddle.

**Discussion Questions**

A Is it possible to get complete destructive interference in an optical coating, at least for light of one specific wavelength?

B Sunlight consists of sinusoidal wave-trains containing on the order of a hundred cycles back-to-back, for a length of something like a tenth of a millimeter. What happens if you try to make an optical coating thicker than this?

C Suppose you take two microscope slides and lay one on top of the other so that one of its edges is resting on the corresponding edge of the bottom one. If you insert a sliver of paper or a hair at the opposite end, a wedge-shaped layer of air will exist in the middle, with a thickness that changes gradually from one end to the other. What would you expect to see if the slides were illuminated from above by light of a single color? How would this change if you gradually lifted the lower edge of the top slide until the two slides were finally parallel?
An observation like the one described in discussion question C was used by Newton as evidence against the wave theory of light! If Newton didn’t know about inverting and noninverting reflections, what would have seemed inexplicable to him about the region where the air layer had zero or nearly zero thickness?

6.2.4 Waves bounded on both sides

In the example of the previous section, it was theoretically true that a pulse would be trapped permanently in the middle medium, but that pulse was not central to our discussion, and in any case it was weakening severely with each partial reflection. Now consider a guitar string. At its ends it is tied to the body of the instrument itself, and since the body is very massive, the behavior of the waves when they reach the end of the string can be understood in the same way as if the actual guitar string was attached on the ends to strings that were extremely massive. Reflections are most intense when the two media are very dissimilar. Because the wave speed in the body is so radically different from the speed in the string, we should expect nearly 100% reflection.

Although this may seem like a rather bizarre physical model of the actual guitar string, it already tells us something interesting about the behavior of a guitar that we would not otherwise have understood. The body, far from being a passive frame for attaching the strings to, is actually the exit path for the wave energy in the strings. With every reflection, the wave pattern on the string loses a tiny fraction of its energy, which is then conducted through the body and out into the air. (The string has too little cross-section to make sound waves efficiently by itself.) By changing the properties of the body, moreover, we should expect to have an effect on the manner in which sound escapes from the instrument. This is clearly demonstrated by the electric guitar, which has an extremely massive, solid wooden body. Here the dissimilarity between the two wave media is even more pronounced, with the result that wave energy leaks out of the string even more slowly. This is why an electric guitar with no electric pickup can hardly be heard at all, and it is also the reason why notes on an electric guitar can be sustained for longer than notes on an acoustic guitar.

If we initially create a disturbance on a guitar string, how will the reflections behave? In reality, the finger or pick will give the string a triangular shape before letting it go, and we may think of this triangular shape as a very broad “dent” in the string which will spread out in both directions. For simplicity, however, let’s just imagine a wave pattern that initially consists of a single, narrow pulse traveling up the neck, k/1. After reflection from the top end, it is inverted, k/3. Now something interesting happens: figure k/5 is identical to figure k/1. After two reflections, the pulse has been inverted twice and has changed direction twice. It is now back where it started. The motion is periodic. This is why a guitar produces
The period of this double-pulse pattern is half of what we'd otherwise expect.

**Self-check F**

Notice that from k/1 to k/5, the pulse has passed by every point on the string exactly twice. This means that the total distance it has traveled equals 2L, where L is the length of the string. Given this fact, what are the period and frequency of the sound it produces, expressed in terms of L and v, the velocity of the wave?  

Answer, p. 1061

Note that if the waves on the string obey the principle of superposition, then the velocity must be independent of amplitude, and the guitar will produce the same pitch regardless of whether it is played loudly or softly. In reality, waves on a string obey the principle of superposition approximately, but not exactly. The guitar, like just about any acoustic instrument, is a little out of tune when played loudly. (The effect is more pronounced for wind instruments than for strings, but wind players are able to compensate for it.)

Now there is only one hole in our reasoning. Suppose we somehow arrange to have an initial setup consisting of two identical pulses heading toward each other, as in figure (g). They will pass through each other, undergo a single inverting reflection, and come back to a configuration in which their positions have been exactly interchanged. This means that the period of vibration is half as long. The frequency is twice as high.

This might seem like a purely academic possibility, since nobody actually plays the guitar with two picks at once! But in fact it is an example of a very general fact about waves that are bounded on both sides. A mathematical theorem called Fourier’s theorem states that any wave can be created by superposing sine waves. Figure n shows how even by using only four sine waves with appropriately chosen amplitudes, we can arrive at a sum which is a decent approximation to the realistic triangular shape of a guitar string being plucked. The one-hump wave, in which half a wavelength fits on the string, will behave like the single pulse we originally discussed. We call its frequency $f_0$. The two-hump wave, with one whole wavelength, is very much like the two-pulse example. For the reasons discussed above, its frequency is $2f_0$. Similarly, the three-hump and four-hump waves have frequencies of $3f_0$ and $4f_0$.

Theoretically we would need to add together infinitely many such wave patterns to describe the initial triangular shape of the string exactly, although the amplitudes required for the very high frequency parts would be very small, and an excellent approximation could be achieved with as few as ten waves.

We thus arrive at the following very general conclusion. Whenever a wave pattern exists in a medium bounded on both sides by media in which the wave speed is very different, the motion can be broken down into the motion of a (theoretically infinite) series of sine
Graphs of loudness versus frequency for the vowel “ah,” sung as three different musical notes. G is consonant with D, since every overtone of G that is close to an overtone of D (marked “*”) is at exactly the same frequency. G and C♯ are dissonant together, since some of the overtones of G (marked “x”) are close to, but not right on top of, those of C♯.

Musical applications

Many musicians claim to be able to pick out by ear several of the frequencies $2f_0$, $3f_0$, ..., called overtones or harmonics of the fundamental $f_0$, but they are kidding themselves. In reality, the overtone series has two important roles in music, neither of which depends on this fictitious ability to “hear out” the individual overtones.

First, the relative strengths of the overtones is an important part of the personality of a sound, called its timbre (rhymes with “amber”). The characteristic tone of the brass instruments, for example, is a sound that starts out with a very strong harmonic series extending up to very high frequencies, but whose higher harmonics die down drastically as the attack changes to the sustained portion of the note.

Second, although the ear cannot separate the individual harmonics of a single musical tone, it is very sensitive to clashes between the overtones of notes played simultaneously, i.e., in harmony. We tend to perceive a combination of notes as being dissonant if they have overtones that are close but not the same. Roughly speaking, strong overtones whose frequencies differ by more than 1% and less than 10% cause the notes to sound dissonant. It is important to realize that the term “dissonance” is not a negative one in music. No matter how long you search the radio dial, you will never hear more than three seconds of music without at least one dissonant combination of notes. Dissonance is a necessary ingredient in the creation of a musical cycle of tension and release. Musically knowledgeable people do not usually use the word “dissonant” as a criticism of music, and if they do, what they are really saying is that the dissonance has been used in a clumsy way, or without providing any contrast between dissonance and consonance.

Standing waves

Figure p shows sinusoidal wave patterns made by shaking a rope. I used to enjoy doing this at the bank with the pens on chains, back in the days when people actually went to the bank. You might think that I and the person in the photos had to practice for a long time in order to get such nice sine waves. In fact, a sine wave is the only shape that can create this kind of wave pattern, called a standing wave, which simply vibrates back and forth in one place without waves, with frequencies $f_o$, $2f_o$, $3f_o$, ... Except for some technical details, to be discussed below, this analysis applies to a vast range of sound-producing systems, including the air column within the human vocal tract. Because sounds composed of this kind of pattern of frequencies are so common, our ear-brain system has evolved so as to perceive them as a single, fused sensation of tone.
moving. The sine wave just creates itself automatically when you find the right frequency, because no other shape is possible.

If you think about it, it’s not even obvious that sine waves should be able to do this trick. After all, waves are supposed to travel at a set speed, aren’t they? The speed isn’t supposed to be zero! Well, we can actually think of a standing wave as a superposition of a moving sine wave with its own reflection, which is moving the opposite way. Sine waves have the unique mathematical property that the sum of sine waves of equal wavelength is simply a new sine wave with the same wavelength. As the two sine waves go back and forth, they always cancel perfectly at the ends, and their sum appears to stand still.

Standing wave patterns are rather important, since atoms are really standing-wave patterns of electron waves. You are a standing wave!

*Standing-wave patterns of air columns*

The air column inside a wind instrument behaves very much like the wave-on-a-string example we’ve been concentrating on so far, the main difference being that we may have either inverting or noninverting reflections at the ends.

Some organ pipes are closed at both ends. The speed of sound is different in metal than in air, so there is a strong reflection at the closed ends, and we can have standing waves. These reflections
**Surprisingly, sound waves undergo partial reflection at the open ends of tubes as well as closed ones.**

Graphs of excess density versus position for the lowest-frequency standing waves of three types of air columns. Points on the axis have normal air density.

Figure r shows the sound waves in and around a bamboo Japanese flute called a shakuhachi, which is open at both ends of the air column. We can only have a standing wave pattern if there are reflections at the ends, but that is very counterintuitive — why is there any reflection at all, if the sound wave is free to emerge into open space, and there is no change in medium? Recall the reason why we got reflections at a change in medium: because the wavelength changes, so the wave has to readjust itself from one pattern to another, and the only way it can do that without developing a kink is if there is a reflection. Something similar is happening here. The only difference is that the wave is adjusting from being a plane wave to being a spherical wave. The reflections at the open ends are density-inverting, so the wave pattern is pinched off at the ends. Comparing panels 1 and 2 of the figure, we see that although the wave patterns are different, in both cases the wavelength is the same: in the lowest-frequency standing wave, half a wavelength fits inside the tube. Thus, it isn’t necessary to memorize which type of reflection is inverting and which is uninverting. It’s only necessary to know that the tubes are symmetric.

Finally, we can have an asymmetric tube: closed at one end and open at the other. A common example is the pan pipes, which are closed at the bottom and open at the top. The standing wave with the lowest frequency is therefore one in which 1/4 of a wavelength fits along the length of the tube, as shown in figure s/3.

Sometimes an instrument’s physical appearance can be misleading. A concert flute, is closed at the mouth end and open at the other, so we would expect it to behave like an asymmetric air column; in reality, it behaves like a symmetric air column open at both ends, because the embouchure hole (the hole the player blows over) acts like an open end. The clarinet and the saxophone look similar, having a mouthpiece and reed at one end and an open end at the other, but they act different. In fact the clarinet’s air column has patterns of vibration that are asymmetric, the saxophone symmetric. The discrepancy comes from the difference between the conical tube of the sax and the cylindrical tube of the clarinet. The adjustment of the wave pattern from a plane wave to a spherical wave is more gradual at the flaring bell of the saxophone.

**Self-check G**

Draw a graph of pressure versus position for the first overtone of the air column in a tube open at one end and closed at the other. This will be the next-to-longest possible wavelength that allows for a point of maximum vibration at one end and a point of no vibration at the other. How many times shorter will its wavelength be compared to the frequency of the lowest-frequency standing wave, shown in the figure? Based on
A pan pipe is an asymmetric air column, open at the top and closed at the bottom.

A concert flute looks like an asymmetric air column, open at the mouth end and closed at the other. However, its patterns of vibration are symmetric, because the embouchure hole acts like an open end.

This chapter is summarized on page 1082. Notation and terminology are tabulated on pages 1070-1071.

6.2.5 * Some technical aspects of reflection

In this section we address some technical details of the treatment of reflection and transmission of waves.

Dependence of reflection on other variables besides velocity

In section 6.2.2 we derived the expressions for the transmitted and reflected amplitudes by demanding that two things match up on both sides of the boundary: the height of the wave and the slope of the wave. These requirements were stated purely in terms of kinematics (the description of how the wave moves) rather than dynamics (the explanation for the wave motion in terms of Newton’s laws). For this reason, the results depended only on the purely kinematic quantity $\alpha = v_2/v_1$, as can be seen more clearly if we rewrite the expressions in the following form:

$$R = \frac{\alpha - 1}{\alpha + 1} \quad \text{and} \quad T = \frac{2\alpha}{\alpha + 1}.$$  

But this purely kinematical treatment only worked because of a dynamical fact that we didn’t emphasize. We assumed equality of the slopes, $s_1 = s_2$, because waves don’t like to have kinks. The underlying dynamical reason for this, in the case of a wave on a string, is that a kink is pointlike, so the portion of the string at the kink is infinitesimal in size, and therefore has essentially zero mass.
If the transverse forces acting on it differed by some finite amount, then its acceleration would be infinite, which is not possible. The difference between the two forces is $T_{s_1} - T_{s_2}$, so $s_1 = s_2$. But this relies on the assumption that $T$ is the same on both sides of the boundary. Now this is true, because we can’t put different amounts of tension on two ropes that are tied together end to end. Any excess tension applied to one rope is distributed equally to the other. For other types of waves, however, we cannot make a similar argument, and therefore it need not be true that $s_1 = s_2$.

A more detailed analysis shows that in general we have not $\alpha = v_2/v_1$ but $\alpha = z_2/z_1$, where $z$ is a quantity called impedance which is defined for this purpose. In a great many examples, as for the waves on a string, it is true that $v_2/v_1 = z_2/z_1$, but this is not a universal fact. Most of the exceptions are rather specialized and technical, such as the reflection of light waves when the media have magnetic properties, but one fairly common and important example is the case of sound waves, for which $z = \rho v$ depends not just on the wave velocity $v$ but also on the density $\rho$. A practical example occurs in medical ultrasound scans, where the contrast of the image is made possible because of the very large differences in impedance between different types of tissue. The speed of sound in various tissues such as bone and muscle varies by about a factor of 2, which is not a particularly huge factor, but there are also large variations in density. The lung, for example, is basically a sponge or sack filled with air. For this reason, the acoustic impedances of the tissues show a huge amount of variation, with, e.g., $z_{\text{bone}}/z_{\text{lung}} \approx 40$.

Inverted and uninverted reflections in general

For waves on a string, reflections back into a faster medium are inverted, while those back into a slower medium are uninverted. Is this true for all types of waves? The rather subtle answer is that it depends on what property of the wave you are discussing.

Let’s start by considering wave disturbances of freeway traffic. Anyone who has driven frequently on crowded freeways has observed the phenomenon in which one driver taps the brakes, starting a chain reaction that travels backward down the freeway as each person in turn exercises caution in order to avoid rear-ending anyone. The reason why this type of wave is relevant is that it gives a simple, easily visualized example of how our description of a wave depends on which aspect of the wave we have in mind. In steadily flowing freeway traffic, both the density of cars and their velocity are constant all along the road. Since there is no disturbance in this pattern of constant velocity and density, we say that there is no wave. Now if a wave is touched off by a person tapping the brakes, we can either describe it as a region of high density or as a region of decreasing velocity.

The freeway traffic wave is in fact a good model of a sound wave,
and a sound wave can likewise be described either by the density (or pressure) of the air or by its speed. Likewise many other types of waves can be described by either of two functions, one of which is often the derivative of the other with respect to position.

Now let's consider reflections. If we observe the freeway wave in a mirror, the high-density area will still appear high in density, but velocity in the opposite direction will now be described by a negative number. A person observing the mirror image will draw the same density graph, but the velocity graph will be flipped across the $x$ axis, and its original region of negative slope will now have positive slope. Although I don't know any physical situation that would correspond to the reflection of a traffic wave, we can immediately apply the same reasoning to sound waves, which often do get reflected, and determine that a reflection can either be density-inverting and velocity-noninverting or density-noninverting and velocity-inverting.

This same type of situation will occur over and over as one encounters new types of waves, and to apply the analogy we need only determine which quantities, like velocity, become negated in a mirror image and which, like density, stay the same.

A light wave, for instance, consists of a traveling pattern of electric and magnetic fields. All you need to know in order to analyze the reflection of light waves is how electric and magnetic fields behave under reflection; you don’t need to know any of the detailed physics of electricity and magnetism. An electric field can be detected, for example, by the way one’s hair stands on end. The direction of the hair indicates the direction of the electric field. In a mirror image, the hair points the other way, so the electric field is apparently reversed in a mirror image. The behavior of magnetic fields, however, is a little tricky. The magnetic properties of a bar magnet, for instance, are caused by the aligned rotation of the outermost orbiting electrons of the atoms. In a mirror image, the direction of rotation is reversed, say from clockwise to counterclockwise, and so the magnetic field is reversed twice: once simply because the whole picture is flipped and once because of the reversed rotation of the electrons. In other words, magnetic fields do not reverse themselves in a mirror image. We can thus predict that there will be two possible types of reflection of light waves. In one, the electric field is inverted and the magnetic field uninverted (example 23, p. 728). In the other, the electric field is uninverted and the magnetic field inverted.
Problems

The symbols √, □, etc. are explained on page 396.

1. The musical note middle C has a frequency of 262 Hz. What are its period and wavelength?

2. The following is a graph of the height of a water wave as a function of position, at a certain moment in time.

Trace this graph onto another piece of paper, and then sketch below it the corresponding graphs that would be obtained if

(a) the amplitude and frequency were doubled while the velocity remained the same;

(b) the frequency and velocity were both doubled while the amplitude remained unchanged;

(c) the wavelength and amplitude were reduced by a factor of three while the velocity was doubled.

Explain all your answers. [Problem by Arnold Arons.]

3. (a) The graph shows the height of a water wave pulse as a function of position. Draw a graph of height as a function of time for a specific point on the water. Assume the pulse is traveling to the right.

(b) Repeat part a, but assume the pulse is traveling to the left.

(c) Now assume the original graph was of height as a function of time, and draw a graph of height as a function of position, assuming the pulse is traveling to the right.

(d) Repeat part c, but assume the pulse is traveling to the left.

Explain all your answers. [Problem by Arnold Arons.]

4. At a particular moment in time, a wave on a string has a shape described by $y = 3.5 \cos(0.73\pi x + 0.45\pi t + 0.37\pi)$. The stuff inside the cosine is in radians. Assume that the units of the numerical constants are such that $x, y,$ and $t$ are in SI units. ▶ Hint, p. 1036

(a) Is the wave moving in the positive $x$ or the negative $x$ direction?

(b) Find the wave’s period, frequency, wavelength.

(c) Find the wave’s velocity.

(d) Find the maximum velocity of any point on the string, and compare with the magnitude and direction of the wave’s velocity.
5 The figure shows one wavelength of a steady sinusoidal wave traveling to the right along a string. Define a coordinate system in which the positive $x$ axis points to the right and the positive $y$ axis up, such that the flattened string would have $y = 0$. Copy the figure, and label with $y = 0$ all the appropriate parts of the string. Similarly, label with $v = 0$ all parts of the string whose velocities are zero, and with $a = 0$ all parts whose accelerations are zero. There is more than one point whose velocity is of the greatest magnitude. Pick one of these, and indicate the direction of its velocity vector. Do the same for a point having the maximum magnitude of acceleration. Explain all your answers.

[Problem by Arnold Arons.]

6 (a) Find an equation for the relationship between the Doppler-shifted frequency of a wave and the frequency of the original wave, for the case of a stationary observer and a source moving directly toward or away from the observer.

(b) Check that the units of your answer make sense.

(c) Check that the dependence on $v_s$ makes sense.

7 Suggest a quantitative experiment to look for any deviation from the principle of superposition for surface waves in water. Try to make your experiment simple and practical.

8 The simplest trick with a lasso is to spin a flat loop in a horizontal plane. The whirling loop of a lasso is kept under tension mainly due to its own rotation. Although the spoke’s force on the loop has an inward component, we’ll ignore it. The purpose of this problem, which is based on one by A.P. French, is to prove a cute fact about wave disturbances moving around the loop. As far as I know, this fact has no practical implications for trick roping! Let the loop have radius $r$ and mass per unit length $\mu$, and let its angular velocity be $\omega$.

(a) Find the tension, $T$, in the loop in terms of $r$, $\mu$, and $\omega$. Assume the loop is a perfect circle, with no wave disturbances on it yet.

(b) Find the velocity of a wave pulse traveling around the loop. Discuss what happens when the pulse moves is in the same direction as the rotation, and when it travels contrary to the rotation.

9 A string hangs vertically, free at the bottom and attached at the top.

(a) Find the velocity of waves on the string as a function of the distance from the bottom.

(b) Find the acceleration of waves on the string.

(c) Interpret your answers to parts a and b for the case where a pulse comes down and reaches the end of the string. What happens next? Check your answer against experiment and conservation of energy.
10 Singing that is off-pitch by more than about 1% sounds bad. How fast would a singer have to be moving relative to the rest of a band to make this much of a change in pitch due to the Doppler effect?

11 Light travels faster in warmer air. On a sunny day, the sun can heat a road and create a layer of hot air above it. Let’s model this layer as a uniform one with a sharp boundary separating it from the cooler air above. Use this model to explain the formation of a mirage appearing like the shiny surface of a pool of water.

12 (a) Compute the amplitude of light that is reflected back into air at an air-water interface, relative to the amplitude of the incident wave. Assume that the light arrives in the direction directly perpendicular to the surface. The speeds of light in air and water are $3.0 \times 10^8$ and $2.2 \times 10^8$ m/s, respectively.

(b) Find the energy of the reflected wave as a fraction of the incident energy.

13 A concert flute produces its lowest note, at about 262 Hz, when half of a wavelength fits inside its tube. Compute the length of the flute.

14 (a) A good tenor saxophone player can play all of the following notes without changing her fingering, simply by altering the tightness of her lips: E♭ (150 Hz), E♭ (300 Hz), B♭ (450 Hz), and E♭ (600 Hz). How is this possible? (I’m not asking you to analyze the coupling between the lips, the reed, the mouthpiece, and the air column, which is very complicated.)

(b) Some saxophone players are known for their ability to use this technique to play “freak notes,” i.e., notes above the normal range of the instrument. Why isn’t it possible to play notes below the normal range using this technique?

15 The table gives the frequencies of the notes that make up the key of F major, starting from middle C and going up through all seven notes.

(a) Calculate the first four or five harmonics of C and G, and determine whether these two notes will be consonant or dissonant. (Recall that harmonics that differ by about 1-10% cause dissonance.)

(b) Do the same for C and B♭.
A Fabry-Perot interferometer, shown in the figure being used to measure the diameter of a thin filament, consists of two glass plates with an air gap between them. As the top plate is moved up or down with a screw, the light passing through the plates goes through a cycle of constructive and destructive interference, which is mainly due to interference between rays that pass straight through and those that are reflected twice back into the air gap. (Although the dimensions in this drawing are distorted for legibility, the glass plates would really be much thicker than the length of the wave-trains of light, so no interference effects would be observed due to reflections within the glass.)

(a) If the top plate is cranked down so that the thickness, \( d \), of the air gap is much less than the wavelength \( \lambda \) of the light, i.e., in the limit \( d \to 0 \), what is the phase relationship between the two rays? (Recall that the phase can be inverted by a reflection.) Is the interference constructive, or destructive?

(b) If \( d \) is now slowly increased, what is the first value of \( d \) for which the interference is the same as at \( d \to 0 \)? Express your answer in terms of \( \lambda \).

(c) Suppose the apparatus is first set up as shown in the figure. The filament is then removed, and \( n \) cycles of brightening and dimming are counted while the top plate is brought down to \( d = 0 \). What is the thickness of the filament, in terms of \( n \) and \( \lambda \)?

Based on a problem by D.J. Raymond.

(a) A wave pulse moves into a new medium, where its velocity is greater by a factor \( \alpha \). Find an expression for the fraction, \( f \), of the wave energy that is transmitted, in terms of \( \alpha \). Note that, as discussed in the text, you cannot simply find \( f \) by squaring the amplitude of the transmitted wave.  

(b) Suppose we wish to transmit a pulse from one medium to another, maximizing the fraction of the wave energy transmitted. To do so, we sandwich another layer in between them, so that the wave moves from the initial medium, where its velocity is \( v_1 \), through the intermediate layer, where it is \( v_2 \), and on into the final layer, where it becomes \( v_3 \). What is the optimal value of \( v_2 \)? (Assume that the middle layer is thicker than the length of the pulse, so there are no interference effects. Also, although there will be later echoes that are transmitted after multiple reflections back and forth across the middle layer, you are only to optimize the strength of the transmitted pulse that is first to emerge. In other words, it’s simply a matter of applying your answer from part a twice to find the amount that finally gets through.)
The expressions for the amplitudes of reflected and transmitted waves depend on the unitless ratio $v_2/v_1$ (or, more generally, on the ratio of the impedances). Call this ratio $\alpha$. (a) Show that changing $\alpha$ to $1/\alpha$ (e.g., by interchanging the roles of the two media) has an effect on the reflected amplitude that can be expressed in a simple way, and discuss what this means in terms of inversion and energy. (b) Find the two values of $\alpha$ for which $|R| = 1/2$. 

Key to symbols:

- easy
- typical
- challenging
- difficult
- very difficult

✓ An answer check is available at www.lightandmatter.com.
Chapter 7
Relativity

7.1 Time is not absolute
When Einstein first began to develop the theory of relativity, around 1905, the only real-world observations he could draw on were ambiguous and indirect. Today, the evidence is part of everyday life. For example, every time you use a GPS receiver, you’re using Einstein’s theory of relativity. Somewhere between 1905 and today, technology became good enough to allow conceptually simple experiments that students in the early 20th century could only discuss in terms like “Imagine that we could…” A good jumping-on point is 1971. In that year, J.C. Hafele and R.E. Keating brought atomic clocks aboard commercial airliners, and went around the world, once from east to west and once from west to east. Hafele and Keating observed that there was a discrepancy between the times measured by the traveling clocks and the times measured by similar clocks that stayed home at the U.S. Naval Observatory in Washington. The east-going clock lost time, ending up off by $-59 \pm 10$ nanoseconds, while the west-going one gained $273 \pm 7$ ns.

7.1.1 The correspondence principle
This establishes that time doesn’t work the way Newton believed it did when he wrote that “Absolute, true, and mathematical time, of itself, and from its own nature flows equably without regard to anything external…” We are used to thinking of time as absolute and universal, so it is disturbing to find that it can flow at a different rate for observers in different frames of reference. Nevertheless, the effects that Hafele and Keating observed were small. This makes sense: Newton’s laws have already been thoroughly tested by experiments under a wide variety of conditions, so a new theory like relativity must agree with Newton’s to a good approximation, within the Newtonian theory’s realm of applicability. This requirement of backward-compatibility is known as the correspondence principle.

7.1.2 Causality
It’s also reassuring that the effects on time were small compared to the three-day lengths of the plane trips. There was therefore no opportunity for paradoxical scenarios such as one in which the east-going experimenter arrived back in Washington before he left and then convinced himself not to take the trip. A theory that maintains
this kind of orderly relationship between cause and effect is said to satisfy causality.

Causality is like a water-hungry front-yard lawn in Los Angeles: we know we want it, but it’s not easy to explain why. Even in plain old Newtonian physics, there is no clear distinction between past and future. In figure c, number 18 throws the football to number 25, and the ball obeys Newton’s laws of motion. If we took a video of the pass and played it backward, we would see the ball flying from 25 to 18, and Newton’s laws would still be satisfied. Nevertheless, we have a strong psychological impression that there is a forward arrow of time. I can remember what the stock market did last year, but I can’t remember what it will do next year. Joan of Arc’s military victories against England caused the English to burn her at the stake; it’s hard to accept that Newton’s laws provide an equally good description of a process in which her execution in 1431 caused her to win a battle in 1429. There is no consensus at this point among physicists on the origin and significance of time’s arrow, and for our present purposes we don’t need to solve this mystery. Instead, we merely note the empirical fact that, regardless of what causality really means and where it really comes from, its behavior is consistent. Specifically, experiments show that if an observer in a certain frame of reference observes that event A causes event B, then observers in other frames agree that A causes B, not the other way around. This is merely a generalization about a large body of experimental results, not a logically necessary assumption. If Keating had gone around the world and arrived back in Washington before he left, it would have disproved this statement about causality.

7.1.3 Time distortion arising from motion and gravity

Hafele and Keating were testing specific quantitative predictions of relativity, and they verified them to within their experiment’s error bars. Let’s work backward instead, and inspect the empirical results for clues as to how time works.

The two traveling clocks experienced effects in opposite directions, and this suggests that the rate at which time flows depends on the motion of the observer. The east-going clock was moving in the same direction as the earth’s rotation, so its velocity relative to the earth’s center was greater than that of the clock that remained in Washington, while the west-going clock’s velocity was correspondingly reduced. The fact that the east-going clock fell behind, and the west-going one got ahead, shows that the effect of motion is to make time go more slowly. This effect of motion on time was predicted by Einstein in his original 1905 paper on relativity, written when he was 26.

If this had been the only effect in the Hafele-Keating experiment, then we would have expected to see effects on the two flying clocks that were equal in size. Making up some simple numbers to keep the
arithmetic transparent, suppose that the earth rotates from west to east at 1000 km/hr, and that the planes fly at 300 km/hr. Then the speed of the clock on the ground is 1000 km/hr, the speed of the clock on the east-going plane is 1300 km/hr, and that of the west-going clock 700 km/hr. Since the speeds of 700, 1000, and 1300 km/hr have equal spacing on either side of 1000, we would expect the discrepancies of the moving clocks relative to the one in the lab to be equal in size but opposite in sign.

In fact, the two effects are unequal in size: −59 ns and 273 ns. This implies that there is a second effect involved, simply due to the planes’ being up in the air. This was verified more directly in a 1978 experiment by Iijima and Fujiwara, figure e, in which identical atomic clocks were kept at rest at the top and bottom of a mountain near Tokyo. This experiment, unlike the Hafele-Keating one, isolates one effect on time, the gravitational one: time’s rate of flow increases with height in a gravitational field. Einstein didn’t figure out how to incorporate gravity into relativity until 1915, after much frustration and many false starts. The simpler version of the theory without gravity is known as special relativity, the full version as general relativity. We’ll restrict ourselves to special relativity.
The correspondence principle requires that the relativistic distortion of time become small for small velocities.

We can now see in more detail how to apply the correspondence principle. The behavior of the three clocks in the Hafele-Keating experiment shows that the amount of time distortion increases as the speed of the clock’s motion increases. Newton lived in an era when the fastest mode of transportation was a galloping horse, and the best pendulum clocks would accumulate errors of perhaps a minute over the course of several days. A horse is much slower than a jet plane, so the distortion of time would have had a relative size of only \(10^{-15}\) — much smaller than the clocks were capable of detecting. At the speed of a passenger jet, the effect is about \(10^{-12}\), and state-of-the-art atomic clocks in 1971 were capable of measuring that. A GPS satellite travels much faster than a jet airplane, and the effect on the satellite turns out to be \(10^{-10}\). The general idea here is that all physical laws are approximations, and approximations aren’t simply right or wrong in different situations. Approximations are better or worse in different situations, and the question is whether a particular approximation is good enough in a given situation to serve a particular purpose. The faster the motion, the worse the Newtonian approximation of absolute time. Whether the approximation is good enough depends on what you’re trying to accomplish. The correspondence principle says that the approximation must have been good enough to explain all the experiments done in the centuries before Einstein came up with relativity.

By the way, don’t get an inflated idea of the importance of the Hafele-Keating experiment. Special relativity had already been confirmed by a vast and varied body of experiments decades before 1971. The only reason I’m giving such a prominent role to this experiment, which was actually more important as a test of general relativity, is that it is conceptually very direct.

### 7.2 Distortion of space and time

#### 7.2.1 The Lorentz transformation

Relativity says that when two observers are in different frames of reference, each observer considers the other one’s perception of time to be distorted. We’ll also see that something similar happens to their observations of distances, so both space and time are distorted. What exactly is this distortion? How do we even conceptualize it?

The idea isn’t really as radical as it might seem at first. We can visualize the structure of space and time using a graph with position and time on its axes. These graphs are familiar by now, but we’re going to look at them in a slightly different way. Before, we used them to describe the motion of objects. The grid underlying the graph was merely the stage on which the actors played their