CHAPTER 15
Thermodynamics

FIGURE 15.1 A steam engine uses heat transfer to do work. (credit: Gerald Friedrich, Pixabay)

CHAPTER OUTLINE

INTRODUCTION TO THERMODYNAMICS Heat transfer is energy in transit, and it can be used to do work. It can also be converted to any other form of energy. A car engine, for example, burns fuel for heat transfer into a gas. Work is done by the gas as it exerts a force through a distance, converting its energy into a variety of other forms—into the car's kinetic or gravitational potential energy; into electrical energy to run the spark plugs, radio, and lights; and back into stored energy in the car's battery. But most of the heat transfer produced from burning fuel in the engine does not do work on the gas. Rather, the energy is released into the environment, implying that the engine is quite inefficient.

It is often said that modern gasoline engines cannot be made to be significantly more efficient. We hear the same about heat transfer to electrical energy in large power stations, whether they are coal, oil, natural gas, or nuclear powered. Why is that the case? Is the inefficiency caused by design problems that could be solved with better engineering and superior materials? Is it part of some money-making conspiracy by those who sell energy? Actually, the truth is more interesting, and reveals much about the nature of heat transfer.

Basic physical laws govern how heat transfer for doing work takes place and place insurmountable limits onto its efficiency. This chapter will explore these laws as well as many applications and concepts associated with them. These topics are part of *thermodynamics*—the study of heat transfer and its relationship to doing work.

15.1 The First Law of Thermodynamics

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the first law of thermodynamics.
- Describe how conservation of energy relates to the first law of thermodynamics.
- Identify instances of the first law of thermodynamics working in everyday situations, including biological metabolism.
- Calculate changes in the internal energy of a system, after accounting for heat transfer and work done.

FIGURE 15.2 This boiling tea kettle represents energy in motion. The water in the kettle is turning to water vapor because heat is being transferred from the stove to the kettle. As the entire system gets hotter, work is done—from the evaporation of the water to the whistling of the kettle. (credit: Gina Hamilton)

If we are interested in how heat transfer is converted into doing work, then the conservation of energy principle is important. The first law of thermodynamics applies the conservation of energy principle to systems where heat transfer and doing work are the methods of transferring energy into and out of the system. The **first law of thermodynamics** states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system. In equation form, the first law of thermodynamics is

$$\Delta U = Q - W. \tag{15.1}$$

Here ΔU is the *change in internal energy* U of the system. Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfer into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. We use the following sign conventions: if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system and positive W takes energy from the system. Thus $\Delta U = Q - W$. Note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this—heat transfer into them takes place so that they can do work. (See Figure 15.3.) We will now examine Q, W, and ΔU further.

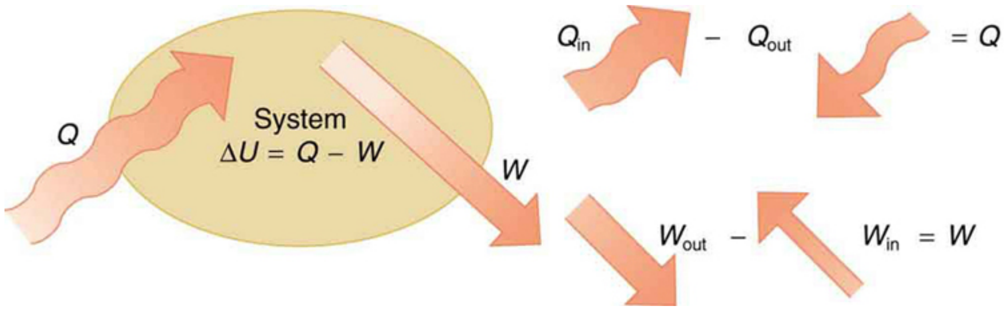


FIGURE 15.3 The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. Q is positive for net heat transfer *into* the system. W is the total work done on and by the system. W is positive when more work is done *by* the system than on it. The change in the internal energy of the system, ΔU, is related to heat and work by the first law of thermodynamics, $\Delta U = Q - W$.

Making Connections: Law of Thermodynamics and Law of Conservation of Energy

The first law of thermodynamics is actually the law of conservation of energy stated in a form most useful in thermodynamics. The first law gives the relationship between heat transfer, work done, and the change in internal energy of a system.

Heat *Q* and Work *W*

Heat transfer (Q) and doing work (W) are the two everyday means of bringing energy into or taking energy out of a system. The processes are quite different. Heat transfer, a less organized process, is driven by temperature differences. Work, a quite organized process, involves a macroscopic force exerted through a distance. Nevertheless, heat and work can produce identical results.For example, both can cause a temperature increase. Heat transfer into a system, such as when the Sun warms the air in a bicycle tire, can increase its temperature, and so can work done on the system, as when the bicyclist pumps air into the tire. Once the temperature increase has occurred, it is impossible to tell whether it was caused by heat transfer or by doing work. This uncertainty is an important point. Heat transfer and work are both energy in transit—neither is stored as such in a system. However, both can change the internal energy U of a system. Internal energy is a form of energy completely different from either heat or work.

Internal Energy *U*

We can think about the internal energy of a system in two different but consistent ways. The first is the atomic and molecular view, which examines the system on the atomic and molecular scale. The **internal energy** U of a system is the sum of the kinetic and potential energies of its atoms and molecules. Recall that kinetic plus potential energy is called mechanical energy. Thus internal energy is the sum of atomic and molecular mechanical energy. Because it is impossible to keep track of all individual atoms and molecules, we must deal with averages and distributions. A second way to view the internal energy of a system is in terms of its macroscopic characteristics, which are very similar to atomic and molecular average values.

Macroscopically, we define the change in internal energy ΔU to be that given by the first law of thermodynamics:

$$\Delta U = Q - W. \tag{15.2}$$

Many detailed experiments have verified that $\Delta U = Q - W$, where ΔU is the change in total kinetic and potential energy of all atoms and molecules in a system. It has also been determined experimentally that the internal energy U of a system depends only on the state of the system and *not how it reached that state*. More specifically, U is found to be a function of a few macroscopic quantities (pressure, volume, and temperature, for example), independent of past history such as whether there has been heat transfer or work done. This independence means that if we know the state of a system, we can calculate changes in its internal energy U from a few macroscopic variables.

Making Connections: Macroscopic and Microscopic

In thermodynamics, we often use the macroscopic picture when making calculations of how a system behaves, while the atomic and molecular picture gives underlying explanations in terms of averages and distributions. We shall see this again in later sections of this chapter. For example, in the topic of entropy, calculations will be made using the atomic and molecular view.

To get a better idea of how to think about the internal energy of a system, let us examine a system going from State 1 to State 2. The system has internal energy U_1 in State 1, and it has internal energy U_2 in State 2, no matter how it got to either state. So the change in internal energy $\Delta U = U_2 - U_1$ is independent of what caused the change. In other words, ΔU *is independent of path*. By path, we mean the method of getting from the starting point to the ending point. Why is this independence important? Note that $\Delta U = Q - W$. Both Q and W *depend on path*, but ΔU does not. This path independence means that internal energy U is easier to consider than either heat transfer or work done.

EXAMPLE 15.1

Calculating Change in Internal Energy: The Same Change in U is Produced by Two Different Processes

(a) Suppose there is heat transfer of 40.00 J to a system, while the system does 10.00 J of work. Later, there is heat transfer of 25.00 J out of the system while 4.00 J of work is done on the system. What is the net change in internal energy of the system?

(b) What is the change in internal energy of a system when a total of 150.00 J of heat transfer occurs out of (from) the system and 159.00 J of work is done on the system? (See Figure 15.4).

Strategy

In part (a), we must first find the net heat transfer and net work done from the given information. Then the first law of thermodynamics $(\Delta U = Q - W)$ can be used to find the change in internal energy. In part (b), the net heat transfer and work done are given, so the equation can be used directly.

Solution for (a)

The net heat transfer is the heat transfer into the system minus the heat transfer out of the system, or

$$Q = 40.00\text{ J} - 25.00\text{ J} = 15.00\text{ J}. \quad (15.3)$$

Similarly, the total work is the work done by the system minus the work done on the system, or

$$W = 10.00\text{ J} - 4.00\text{ J} = 6.00\text{ J}. \quad (15.4)$$

Thus the change in internal energy is given by the first law of thermodynamics:

$$\Delta U = Q - W = 15.00\text{ J} - 6.00\text{ J} = 9.00\text{ J}. \quad (15.5)$$

We can also find the change in internal energy for each of the two steps. First, consider 40.00 J of heat transfer in and 10.00 J of work out, or

$$\Delta U_1 = Q_1 - W_1 = 40.00\text{ J} - 10.00\text{ J} = 30.00\text{ J}. \quad (15.6)$$

Now consider 25.00 J of heat transfer out and 4.00 J of work in, or

$$\Delta U_2 = Q_2 - W_2 = -25.00\text{ J} - (-4.00\text{ J}) = -21.00\text{ J}. \quad (15.7)$$

The total change is the sum of these two steps, or

$$\Delta U = \Delta U_1 + \Delta U_2 = 30.00\text{ J} + (-21.00\text{ J}) = 9.00\text{ J}. \quad (15.8)$$

Discussion on (a)

No matter whether you look at the overall process or break it into steps, the change in internal energy is the same.

Solution for (b)

Here the net heat transfer and total work are given directly to be $Q = -150.00\text{ J}$ and $W = -159.00\text{ J}$, so that

$$\Delta U = Q - W = -150.00\text{ J} - (-159.00\text{ J}) = 9.00\text{ J}. \tag{15.9}$$

Discussion on (b)

A very different process in part (b) produces the same 9.00-J change in internal energy as in part (a). Note that the change in the system in both parts is related to ΔU and not to the individual Qs or Ws involved. The system ends up in the *same* state in both (a) and (b). Parts (a) and (b) present two different paths for the system to follow between the same starting and ending points, and the change in internal energy for each is the same—it is independent of path.

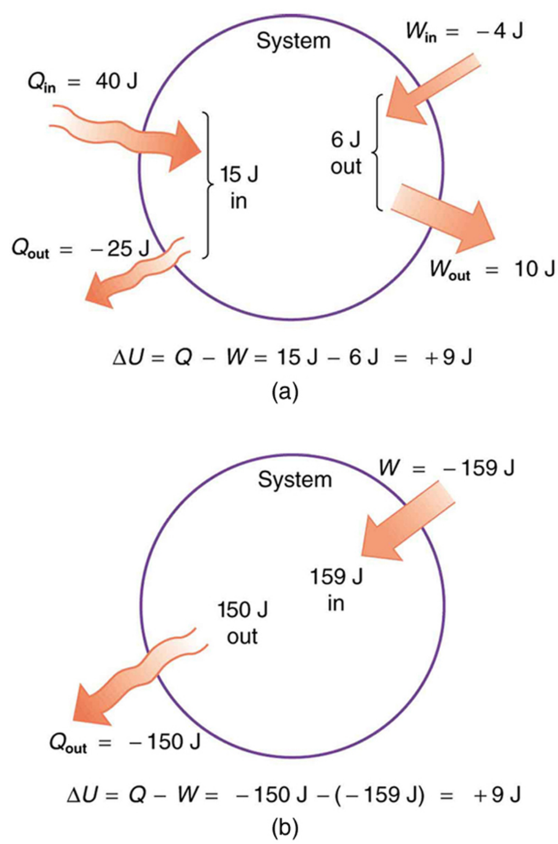


FIGURE 15.4 Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is $\Delta U = Q - W = 9.00\text{ J}$. (b) Heat transfer removes 150.00 J from the system while work puts 159.00 J into it, producing an increase of 9.00 J in internal energy. If the system starts out in the same state in (a) and (b), it will end up in the same final state in either case—its final state is related to internal energy, not how that energy was acquired.

Human Metabolism and the First Law of Thermodynamics

Human metabolism is the conversion of food into heat transfer, work, and stored fat. Metabolism is an interesting example of the first law of thermodynamics in action. We now take another look at these topics via the first law of thermodynamics. Considering the body as the system of interest, we can use the first law to examine heat transfer, doing work, and internal energy in activities ranging from sleep to heavy exercise. What are some of the major characteristics of heat transfer, doing work, and energy in the body? For one, body temperature is normally kept constant by heat transfer to the surroundings. This means Q is negative. Another fact is that the body usually does work on the outside world. This means W is positive. In such situations, then, the body loses internal energy, since $\Delta U = Q - W$ is negative.

Now consider the effects of eating. Eating increases the internal energy of the body by adding chemical potential energy (this is an unromantic view of a good steak). The body *metabolizes* all the food we consume. Basically, metabolism is an oxidation process in which the chemical potential energy of food is released. This implies that food input is in the form of work. Food energy is reported in a special unit, known as the Calorie. This energy is measured by burning food in a calorimeter, which is how the units are determined.

In chemistry and biochemistry, one calorie (spelled with a *lowercase* c) is defined as the energy (or heat transfer) required to raise the temperature of one gram of pure water by one degree Celsius. Nutritionists and weight-watchers tend to use the *dietary* calorie, which is frequently called a Calorie (spelled with a *capital* C). One food Calorie is the energy needed to raise the temperature of one *kilogram* of water by one degree Celsius. This means that one dietary Calorie is equal to one kilocalorie for the chemist, and one must be careful to avoid confusion between the two.

Again, consider the internal energy the body has lost. There are three places this internal energy can go—to heat transfer, to doing work, and to stored fat (a tiny fraction also goes to cell repair and growth). Heat transfer and doing work take internal energy out of the body, and food puts it back. If you eat just the right amount of food, then your average internal energy remains constant. Whatever you lose to heat transfer and doing work is replaced by food, so that, in the long run, $\Delta U = 0$. If you overeat repeatedly, then ΔU is always positive, and your body stores this extra internal energy as fat. The reverse is true if you eat too little. If ΔU is negative for a few days, then the body metabolizes its own fat to maintain body temperature and do work that takes energy from the body. This process is how dieting produces weight loss.

Life is not always this simple, as any dieter knows. The body stores fat or metabolizes it only if energy intake changes for a period of several days. Once you have been on a major diet, the next one is less successful because your body alters the way it responds to low energy intake. Your basal metabolic rate (BMR) is the rate at which food is converted into heat transfer and work done while the body is at complete rest. The body adjusts its basal metabolic rate to partially compensate for over-eating or under-eating. The body will decrease the metabolic rate rather than eliminate its own fat to replace lost food intake. You will chill more easily and feel less energetic as a result of the lower metabolic rate, and you will not lose weight as fast as before. Exercise helps to lose weight, because it produces both heat transfer from your body and work, and raises your metabolic rate even when you are at rest. Weight loss is also aided by the quite low efficiency of the body in converting internal energy to work, so that the loss of internal energy resulting from doing work is much greater than the work done.It should be noted, however, that living systems are not in thermalequilibrium.

The body provides us with an excellent indication that many thermodynamic processes are *irreversible*. An irreversible process can go in one direction but not the reverse, under a given set of conditions. For example, although body fat can be converted to do work and produce heat transfer, work done on the body and heat transfer into it cannot be converted to body fat. Otherwise, we could skip lunch by sunning ourselves or by walking down stairs. Another example of an irreversible thermodynamic process is photosynthesis. This process is the intake of one form of energy—light—by plants and its conversion to chemical potential energy. Both applications of the first law of thermodynamics are illustrated in Figure 15.5. One great advantage of conservation laws such as the first law of thermodynamics is that they accurately describe the beginning and ending points of complex processes, such as metabolism and photosynthesis, without regard to the complications in between. Table 15.1 presents a summary of terms relevant to the first law of thermodynamics.

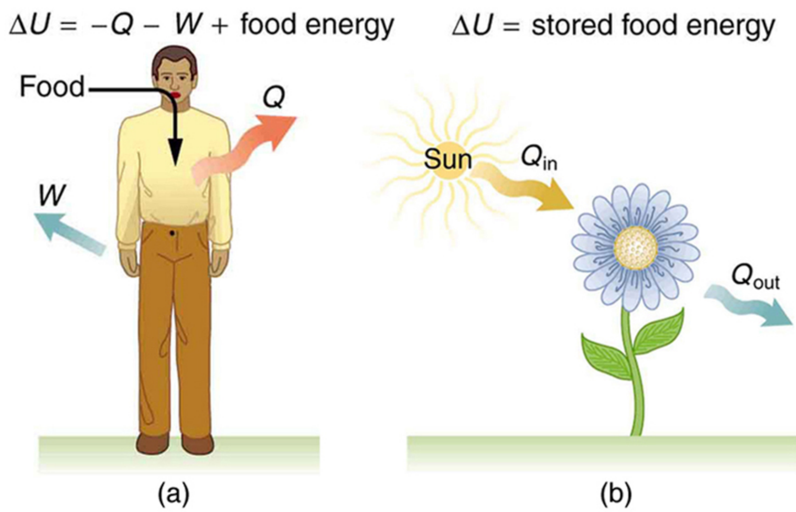


FIGURE 15.5 (a) The first law of thermodynamics applied to metabolism. Heat transferred out of the body (Q) and work done by the body (W) remove internal energy, while food intake replaces it. (Food intake may be considered as work done on the body.) (b) Plants convert part of the radiant heat transfer in sunlight to stored chemical energy, a process called photosynthesis.

Term	Definition
U	Internal energy—the sum of the kinetic and potential energies of a system's atoms and molecules. Can be divided into many subcategories, such as thermal and chemical energy. Depends only on the state of a system (such as its P, V, and T), not on how the energy entered the system. Change in internal energy is path independent.
Q	Heat—energy transferred because of a temperature difference. Characterized by random molecular motion. Highly dependent on path. Q entering a system is positive.
W	Work—energy transferred by a force moving through a distance. An organized, orderly process. Path dependent. W done by a system (either against an external force or to increase the volume of the system) is positive.

TABLE 15.1 Summary of Terms for the First Law of Thermodynamics, $\Delta U=Q-W$

15.2 The First Law of Thermodynamics and Some Simple Processes

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the processes of a simple heat engine.
- Explain the differences among the simple thermodynamic processes—isobaric, isochoric, isothermal, and adiabatic.
- Calculate total work done in a cyclical thermodynamic process.

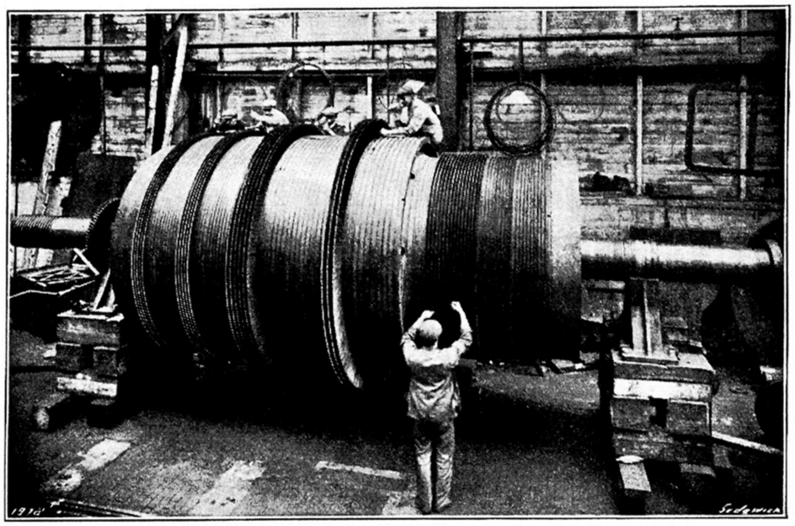


FIGURE 15.6 Beginning with the Industrial Revolution, humans have harnessed power through the use of the first law of thermodynamics, before we even understood it completely. This photo, of a steam engine at the Turbinia Works, dates from 1911, a mere 61 years after the first explicit statement of the first law of thermodynamics by Rudolph Clausius. (credit: public domain; author unknown)

One of the most important things we can do with heat transfer is to use it to do work for us. Such a device is called a **heat engine**. Car engines and steam turbines that generate electricity are examples of heat engines. Figure 15.7 shows schematically how the first law of thermodynamics applies to the typical heat engine.

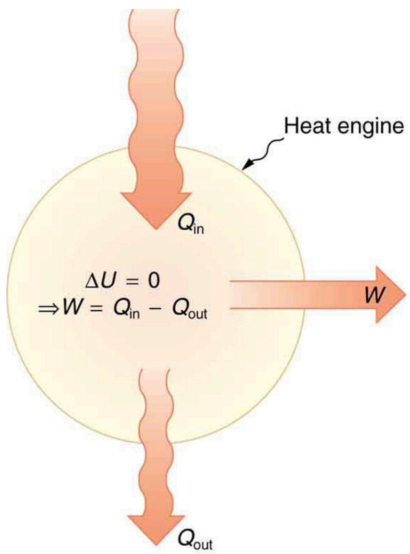


FIGURE 15.7 Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics. It is impossible to devise a system where $Q_{\text{out}} = 0$, that is, in which no heat transfer occurs to the environment.

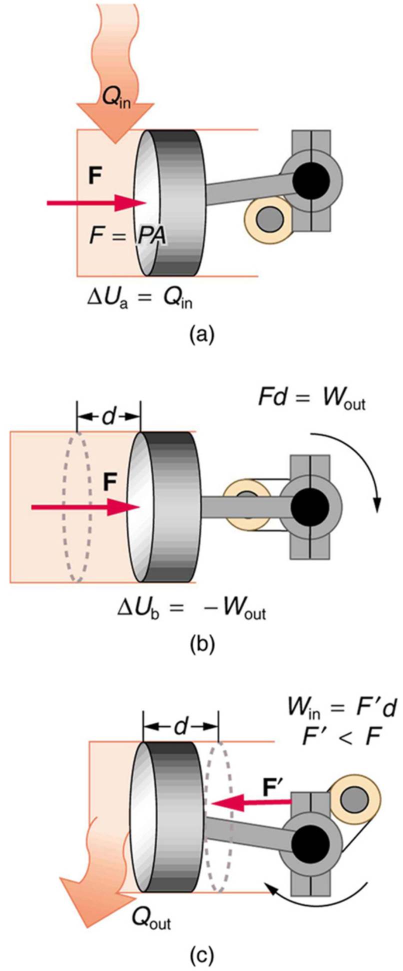


FIGURE 15.8 (a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating that the gas's internal energy has been decreased by doing work. (c) Heat transfer to the environment further reduces pressure in the gas so that the piston can be more easily returned to its starting position.

The illustrations above show one of the ways in which heat transfer does work. Fuel combustion produces heat transfer to a gas in a cylinder, increasing the pressure of the gas and thereby the force it exerts on a movable piston. The gas does work on the outside world, as this force moves the piston through some distance. Heat transfer to the gas cylinder results in work being done. To repeat this process, the piston needs to be returned to its starting point. Heat transfer now occurs from the gas to the surroundings so that its pressure decreases, and a force is exerted by the surroundings to push the piston back through some distance. Variations of this process are employed daily in hundreds of millions of heat engines. We will examine heat engines in detail in the next section. In this section, we consider some of the simpler underlying processes on which heat engines are based.

PV Diagrams and their Relationship to Work Done on or by a Gas

A process by which a gas does work on a piston at constant pressure is called an **isobaric process**. Since the pressure is constant, the force exerted is constant and the work done is given as

$$P\Delta V. \tag{15.10}$$

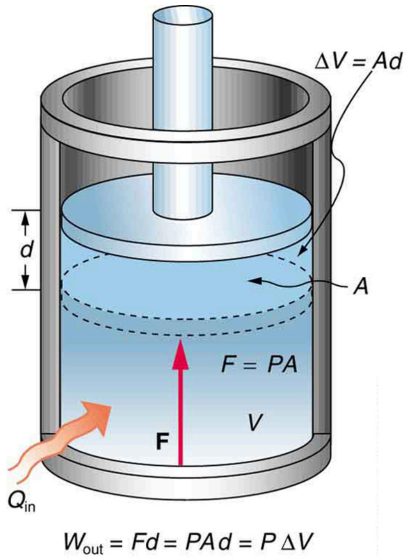


FIGURE 15.9 An isobaric expansion of a gas requires heat transfer to keep the pressure constant. Since pressure is constant, the work done is $P\Delta V$.

$$W = Fd \tag{15.11}$$

See the symbols as shown in Figure 15.9. Now $F = PA$, and so

$$W = PAd. \tag{15.12}$$

Because the volume of a cylinder is its cross-sectional area A times its length d, we see that $Ad = \Delta V$, the change in volume; thus,

$$W = P\Delta V \text{ (isobaric process).} \tag{15.13}$$

Note that if ΔV is positive, then W is positive, meaning that work is done *by* the gas on the outside world.

(Note that the pressure involved in this work that we've called P is the pressure of the gas *inside* the tank. If we call the pressure outside the tank P_{ext}, an expanding gas would be working *against* the external pressure; the work done would therefore be $W = -P_{\text{ext}}\Delta V$ (isobaric process). Many texts use this definition of work, and not the definition based on internal pressure, as the basis of the First Law of Thermodynamics. This definition reverses the sign conventions for work, and results in a statement of the first law that becomes $\Delta U = Q + W$.)

It is not surprising that $W = P\Delta V$, since we have already noted in our treatment of fluids that pressure is a type of potential energy per unit volume and that pressure in fact has units of energy divided by volume. We also noted in our discussion of the ideal gas law that PV has units of energy. In this case, some of the energy associated with pressure becomes work.

Figure 15.10 shows a graph of pressure versus volume (that is, a PV diagram for an isobaric process. You can see in the figure that the work done is the area under the graph. This property of PV diagrams is very useful and broadly applicable: *the work done on or by a system in going from one state to another equals the area under the curve on a* PV *diagram.*

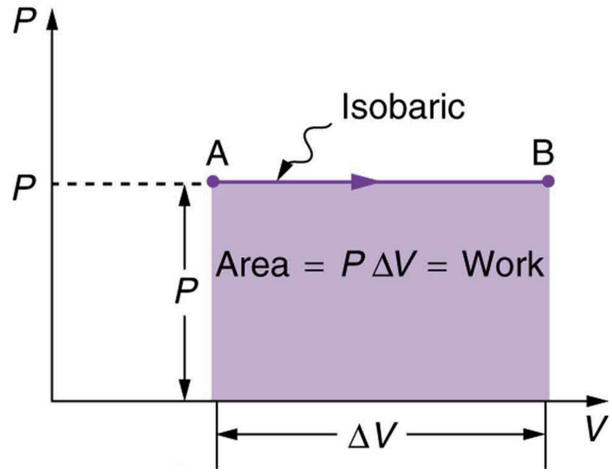


FIGURE 15.10 A graph of pressure versus volume for a constant-pressure, or isobaric, process, such as the one shown in Figure 15.9. The area under the curve equals the work done by the gas, since $W = P\Delta V$.

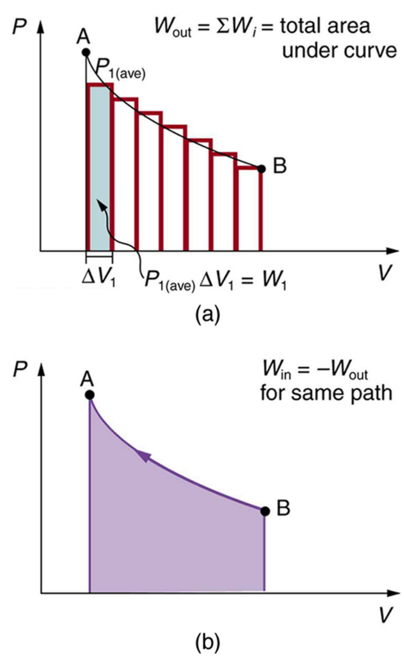


FIGURE 15.11 (a) A PV diagram in which pressure varies as well as volume. The work done for each interval is its average pressure times the change in volume, or the area under the curve over that interval. Thus the total area under the curve equals the total work done. (b) Work must be done on the system to follow the reverse path. This is interpreted as a negative area under the curve.

We can see where this leads by considering Figure 15.11(a), which shows a more general process in which both pressure and volume change. The area under the curve is closely approximated by dividing it into strips, each having an average constant pressure $P_{i(\text{ave})}$. The work done is $W_i = P_{i(\text{ave})}\Delta V_i$ for each strip, and the total work done is the sum of the W_i. Thus the total work done is the total area under the curve. If the path is reversed, as in Figure 15.11(b), then work is done on the system. The area under the curve in that case is negative, because ΔV is negative.

PV diagrams clearly illustrate that *the work done depends on the path taken and not just the endpoints*. This path dependence is seen in Figure 15.12(a), where more work is done in going from A to C by the path via point B than by the path via point D. The vertical paths, where volume is constant, are called **isochoric** processes. Since volume is constant, $\Delta V = 0$, and no work is done in an isochoric process. Now, if the system follows the cyclical path ABCDA, as in Figure 15.12(b), then the total work done is the area inside the loop. The negative area below path CD subtracts, leaving only the area inside the rectangle. In fact, the work done in any cyclical process (one that returns to its starting point) is the area inside the loop it forms on a PV diagram, as Figure 15.12(c) illustrates for a general cyclical process. Note that the loop must be traversed in the clockwise direction for work to be positive—that is, for there to be a net work output.

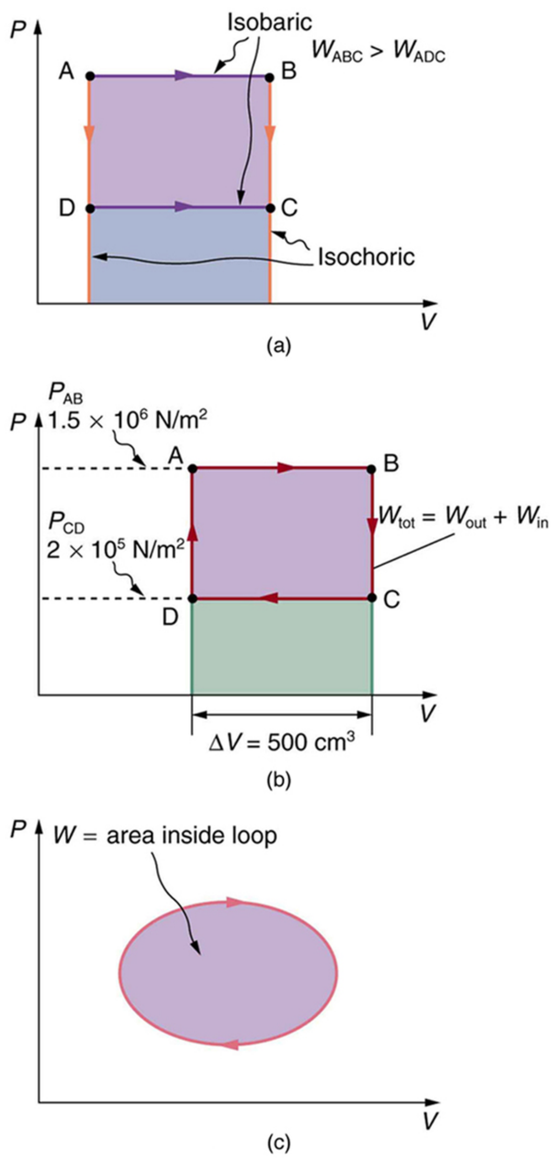


FIGURE 15.12 (a) The work done in going from A to C depends on path. The work is greater for the path ABC than for the path ADC, because the former is at higher pressure. In both cases, the work done is the area under the path. This area is greater for path ABC. (b) The total work done in the cyclical process ABCDA is the area inside the loop, since the negative area below CD subtracts out, leaving just the area inside the rectangle. (The values given for the pressures and the change in volume are intended for use in the example below.) (c) The area inside any closed loop is the work done in the cyclical process. If the loop is traversed in a clockwise direction, W is positive—it is work done on the outside environment. If the loop is traveled in a counter-clockwise direction, W is negative—it is work that is done to the system.

EXAMPLE 15.2

Total Work Done in a Cyclical Process Equals the Area Inside the Closed Loop on a *PV* Diagram

Calculate the total work done in the cyclical process ABCDA shown in Figure 15.12(b) by the following two methods to verify that work equals the area inside the closed loop on the PV diagram. (Take the data in the figure to be precise to three significant figures.) (a) Calculate the work done along each segment of the path and add these values to get the total work. (b) Calculate the area inside the rectangle ABCDA.

Strategy

To find the work along any path on a PV diagram, you use the fact that work is pressure times change in volume, or $W = P\Delta V$. So in part (a), this value is calculated for each leg of the path around the closed loop.

Solution for (a)

The work along path AB is

$$\begin{aligned} W_{\text{AB}} &= P_{\text{AB}} \Delta V_{\text{AB}} \\ &= (1.50\times 10^6\ \text{N/m}^2)(5.00\times 10^{-4}\ \text{m}^3) = 750\ \text{J}. \end{aligned} \tag{15.14}$$

Since the path BC is isochoric, $\Delta V_{\text{BC}} = 0$, and so $W_{\text{BC}} = 0$. The work along path CD is negative, since ΔV_{CD} is negative (the volume decreases). The work is

$$\begin{aligned} W_{\text{CD}} &= P_{\text{CD}} \Delta V_{\text{CD}} \\ &= (2.00\times 10^5\ \text{N/m}^2)(-5.00\times 10^{-4}\ \text{m}^3) = -100\ \text{J}. \end{aligned} \tag{15.15}$$

Again, since the path DA is isochoric, $\Delta V_{\text{DA}} = 0$, and so $W_{\text{DA}} = 0$. Now the total work is

$$\begin{aligned} W &= W_{\text{AB}} + W_{\text{BC}} + W_{\text{CD}} + W_{\text{DA}} \\ &= 750\ \text{J} + 0 + (-100\text{J}) + 0 = 650\ \text{J}. \end{aligned} \tag{15.16}$$

Solution for (b)

The area inside the rectangle is its height times its width, or

$$\begin{aligned} \text{area} &= (P_{\text{AB}} - P_{\text{CD}})\Delta V \\ &= \left[(1.50\times 10^6\ \text{N/m}^2) - (2.00\times 10^5\ \text{N/m}^2)\right](5.00\times 10^{-4}\ \text{m}^3) \\ &= 650\ \text{J}. \end{aligned} \tag{15.17}$$

Thus,

$$\text{area} = 650\ \text{J} = W. \tag{15.18}$$

Discussion

The result, as anticipated, is that the area inside the closed loop equals the work done. The area is often easier to calculate than is the work done along each path. It is also convenient to visualize the area inside different curves on PV diagrams in order to see which processes might produce the most work. Recall that work can be done to the system, or by the system, depending on the sign of W. A positive W is work that is done by the system on the outside environment; a negative W represents work done by the environment on the system.

Figure 15.13(a) shows two other important processes on a PV diagram. For comparison, both are shown starting from the same point A. The upper curve ending at point B is an **isothermal** process—that is, one in which temperature is kept constant. If the gas behaves like an ideal gas, as is often the case, and if no phase change occurs, then $PV = nRT$. Since T is constant, PV is a constant for an isothermal process. We ordinarily expect the temperature of a gas to decrease as it expands, and so we correctly suspect that heat transfer must occur from the surroundings to the gas to keep the temperature constant during an isothermal expansion. To show this more rigorously for the special case of a monatomic ideal gas, we note that the average kinetic energy of an atom in such a gas is given by

$$\frac{1}{2}m\bar{v}^2 = \frac{3}{2}kT. \tag{15.19}$$

The kinetic energy of the atoms in a monatomic ideal gas is its only form of internal energy, and so its total internal energy U is

$$U = N\frac{1}{2}m\bar{v}^2 = \frac{3}{2}NkT,\ \text{(monatomic ideal gas)}, \tag{15.20}$$

where N is the number of atoms in the gas. This relationship means that the internal energy of an ideal monatomic gas is constant during an isothermal process—that is, $\Delta U = 0$. If the internal energy does not change, then the net heat transfer into the gas must equal the net work done by the gas. That is, because $\Delta U = Q - W = 0$ here, $Q = W$. We must have just enough heat transfer to replace the work done. An isothermal process is inherently slow, because heat transfer occurs continuously to keep the gas temperature constant at all times and must be allowed to spread through the gas so that there are no hot or cold regions.

Note that diatomic gases, such as those in air, have different formulas for average kinetic energy of an atom and total internal energy.

Also shown in Figure 15.13(a) is a curve AC for an **adiabatic** process, defined to be one in which there is no heat transfer—that is, $Q = 0$. Processes that are nearly adiabatic can be achieved either by using very effective insulation or by performing the process so fast that there is little time for heat transfer. Temperature must decrease during an adiabatic expansion process, since work is done at the expense of internal energy:

$$U = \frac{3}{2}NkT. \tag{15.21}$$

(You might have noted that a gas released into atmospheric pressure from a pressurized cylinder is substantially colder than the gas in the cylinder.) In fact, because $Q = 0$, $\Delta U =- W$ for an adiabatic process. Lower temperature results in lower pressure along the way, so that curve AC is lower than curve AB, and less work is done. If the path ABCA could be followed by cooling the gas from B to C at constant volume (isochorically), Figure 15.13(b), there would be a net work output.

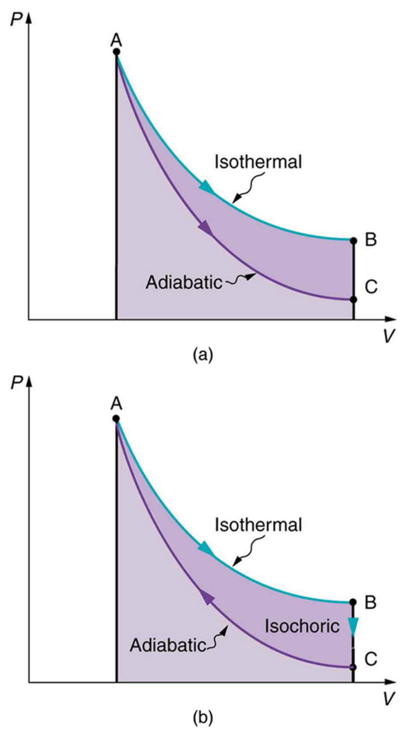


FIGURE 15.13 (a) The upper curve is an isothermal process ($\Delta T = 0$), whereas the lower curve is an adiabatic process ($Q = 0$). Both start from the same point A, but the isothermal process does more work than the adiabatic because heat transfer into the gas takes place to keep its temperature constant. This keeps the pressure higher all along the isothermal path than along the adiabatic path, producing more work. The adiabatic path thus ends up with a lower pressure and temperature at point C, even though the final volume is the same as for the isothermal process. (b) The cycle ABCA produces a net work output.

Reversible Processes

Both isothermal and adiabatic processes such as shown in Figure 15.13 are reversible in principle. A **reversible process** is one in which both the system and its environment can return to exactly the states they were in by following the reverse path. The reverse isothermal and adiabatic paths are BA and CA, respectively. Real macroscopic processes are never exactly reversible. In the previous examples, our system is a gas (like that in Figure 15.9), and its environment is the piston, cylinder, and the rest of the universe. If there are any energy-dissipating mechanisms, such as friction or turbulence, then heat transfer to the environment occurs for either direction of the piston. So, for example, if the path BA is followed and there is friction, then the gas will be returned to its original state but the environment will not—it will have been heated in both directions. Reversibility requires the direction of heat transfer to reverse for the reverse path. Since dissipative mechanisms cannot be completely eliminated, real processes cannot be reversible.

There must be reasons that real macroscopic processes cannot be reversible. We can imagine them going in reverse. For example, heat transfer occurs spontaneously from hot to cold and never spontaneously the reverse. Yet it would not violate the first law of thermodynamics for this to happen. In fact, all spontaneous processes, such as bubbles bursting, never go in reverse. There is a second thermodynamic law that forbids them from going in reverse. When we study this law, we will learn something about nature and also find that such a law limits the efficiency of heat engines. We will find that heat engines with the greatest possible theoretical efficiency would have to use reversible processes, and even they cannot convert all heat transfer into doing work. Table 15.2 summarizes the simpler thermodynamic processes and their definitions.

Isobaric	Constant pressure $W = P\Delta V$
Isochoric	Constant volume $W = 0$
Isothermal	Constant temperature $Q = W$
Adiabatic	No heat transfer $Q = 0$

TABLE 15.2 Summary of Simple Thermodynamic Processes

PHET EXPLORATIONS

States of Matter

Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.

Click to view content (https://openstax.org/books/college-physics-2e/pages/15-2-the-first-law-of-thermodynamics-and-some-simple-processes)

15.3 Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- State the expressions of the second law of thermodynamics.
- Calculate the efficiency and carbon dioxide emission of a coal-fired electricity plant, using second law characteristics.
- Describe and define the Otto cycle.

FIGURE 15.14 These ice floes melt during the Arctic summer. Some of them refreeze in the winter, but the second law of thermodynamics predicts that it would be extremely unlikely for the water molecules contained in these particular floes to reform the distinctive alligator-like shape they formed when the picture was taken in the summer of 2009. (credit: Patrick Kelley, U.S. Coast Guard, U.S. Geological Survey)

The second law of thermodynamics deals with the direction taken by spontaneous processes. Many processes occur spontaneously in one direction only—that is, they are irreversible, under a given set of conditions. Although irreversibility is seen in day-to-day life—a broken glass does not resume its original state, for instance—complete irreversibility is a statistical statement that cannot be seen during the lifetime of the universe. More precisely, an **irreversible process** is one that depends on path. If the process can go in only one direction, then the reverse path differs fundamentally and the process cannot be reversible. For example, as noted in the previous section, heat involves the transfer of energy from higher to lower temperature. A cold object in contact with a hot one never gets colder, transferring heat to the hot object and making it hotter. Furthermore, mechanical energy, such as kinetic energy, can be completely converted to thermal energy by friction, but the reverse is impossible. A hot stationary object never spontaneously cools off and starts moving. Yet another example is the expansion of a puff of gas introduced into one corner of a vacuum chamber. The gas expands to fill the chamber, but it never regroups in the corner. The random motion of the gas molecules could take them all back to the corner, but this is never observed to happen. (See Figure 15.15.)

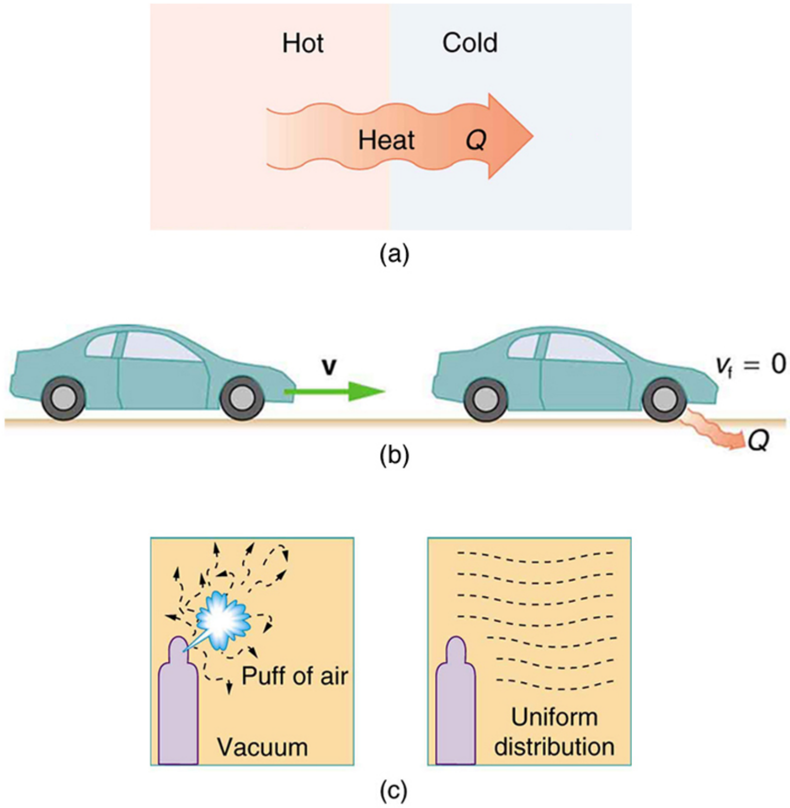


FIGURE 15.15 Examples of one-way processes in nature. (a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will never return them to the corner.

The fact that certain processes never occur suggests that there is a law forbidding them to occur. The first law of thermodynamics would allow them to occur—none of those processes violate conservation of energy. The law that forbids these processes is called the second law of thermodynamics. We shall see that the second law can be stated in many ways that may seem different, but which in fact are equivalent. Like all natural laws, the second law of thermodynamics gives insights into nature, and its several statements imply that it is broadly applicable, fundamentally affecting many apparently disparate processes.

The already familiar direction of heat transfer from hot to cold is the basis of our first version of the **second law of thermodynamics**.

The Second Law of Thermodynamics (first expression)

Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction.

Another way of stating this: It is impossible for any process to have as its sole result heat transfer from a cooler to a hotter object.

Heat Engines

Now let us consider a device that uses heat transfer to do work. As noted in the previous section, such a device is called a heat engine, and one is shown schematically in Figure 15.16(b). Gasoline and diesel engines, jet engines, and steam turbines are all heat engines that do work by using part of the heat transfer from some source. Heat transfer from the hot object (or hot reservoir) is denoted as Q_{h}, while heat transfer into the cold object (or cold reservoir) is Q_{c}, and the work done by the engine is W. The temperatures of the hot and cold reservoirs are T_{h} and T_{c}, respectively.

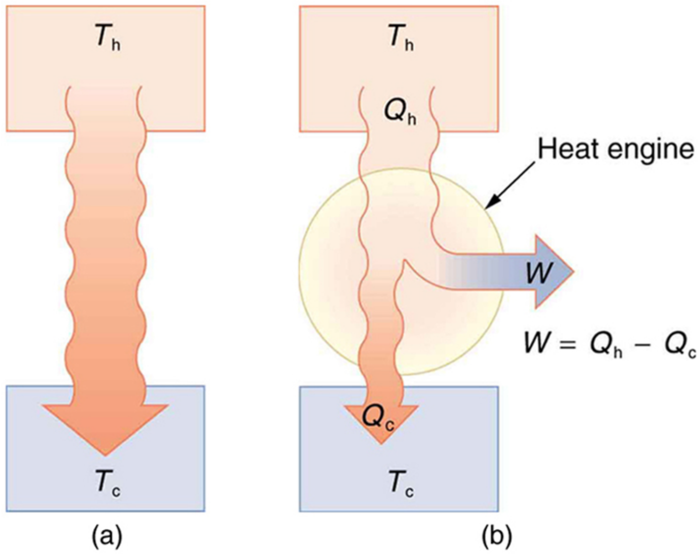


FIGURE 15.16 (a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs. Q_{h} is the heat transfer out of the hot reservoir, W is the work output, and Q_{c} is the heat transfer into the cold reservoir.

Because the hot reservoir is heated externally, which is energy intensive, it is important that the work is done as efficiently as possible. In fact, we would like W to equal Q_{h}, and for there to be no heat transfer to the environment ($Q_{\mathrm{c}} = 0$). Unfortunately, this is impossible. The **second law of thermodynamics** also states, with regard to using heat transfer to do work (the second expression of the second law):

The Second Law of Thermodynamics (second expression)

It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.

A **cyclical process** brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. Most heat engines, such as reciprocating piston engines and rotating turbines, use cyclical processes. The second law, just stated in its second form, clearly states that such engines cannot have perfect conversion of heat transfer into work done. Before going into the underlying reasons for the limits on converting heat transfer into work, we need to explore the relationships among W, Q_{h}, and Q_{c}, and to define the efficiency of a cyclical heat engine. As noted, a cyclical process brings the system back to its original condition at the end of every cycle. Such a system's internal energy U is the same at the beginning and end of every cycle—that is, $\Delta U = 0$. The first law of thermodynamics states that

$$\Delta U = Q - W, \tag{15.22}$$

where Q is the *net* heat transfer during the cycle ($Q = Q_{\mathrm{h}} - Q_{\mathrm{c}}$) and W is the net work done by the system. Since $\Delta U = 0$ for a complete cycle, we have

$$0 = Q - W, \tag{15.23}$$

so that

$$W = Q. \tag{15.24}$$

Thus the net work done by the system equals the net heat transfer into the system, or

$$W = Q_{\mathrm{h}} - Q_{\mathrm{c}} \text{ (cyclical process)}, \tag{15.25}$$

just as shown schematically in Figure 15.16(b). The problem is that in all processes, there is some heat transfer Q_{c} to the environment—and usually a very significant amount at that.

In the conversion of energy to work, we are always faced with the problem of getting less out than we put in. We define *conversion efficiency* Eff to be the ratio of useful work output to the energy input (or, in other words, the ratio of what we get to what we spend). In that spirit, we define the efficiency of a heat engine to be its net work output W divided by heat transfer to the engine Q_h; that is,

$$Eff = \frac{W}{Q_h}. \quad (15.26)$$

Since $W = Q_h - Q_c$ in a cyclical process, we can also express this as

$$Eff = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h} \text{ (cyclical process)}, \quad (15.27)$$

making it clear that an efficiency of 1, or 100%, is possible only if there is no heat transfer to the environment ($Q_c = 0$). Note that all Qs are positive. The direction of heat transfer is indicated by a plus or minus sign. For example, Q_c is out of the system and so is preceded by a minus sign.

EXAMPLE 15.3

Daily Work Done by a Coal-Fired Power Station, Its Efficiency and Carbon Dioxide Emissions

A coal-fired power station is a huge heat engine. It uses heat transfer from burning coal to do work to turn turbines, which are used to generate electricity. In a single day, a large coal power station has 2.50×10^{14} J of heat transfer from coal and 1.48×10^{14} J of heat transfer into the environment. (a) What is the work done by the power station? (b) What is the efficiency of the power station? (c) In the combustion process, the following chemical reaction occurs: $C + O_2 \rightarrow CO_2$. This implies that every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of carbon dioxide into the atmosphere. Assuming that 1 kg of coal can provide 2.5×10^6 J of heat transfer upon combustion, how much CO_2 is emitted per day by this power plant?

Strategy for (a)

We can use $W = Q_h - Q_c$ to find the work output W, assuming a cyclical process is used in the power station. In this process, water is boiled under pressure to form high-temperature steam, which is used to run steam turbine-generators, and then condensed back to water to start the cycle again.

Solution for (a)

Work output is given by:

$$W = Q_h - Q_c. \quad (15.28)$$

Substituting the given values:

$$\begin{aligned} W &= 2.50 \times 10^{14}\text{ J} - 1.48 \times 10^{14}\text{ J} \\ &= 1.02 \times 10^{14}\text{ J}. \end{aligned} \quad (15.29)$$

Strategy for (b)

The efficiency can be calculated with $Eff = \frac{W}{Q_h}$ since Q_h is given and work W was found in the first part of this example.

Solution for (b)

Efficiency is given by: $Eff = \frac{W}{Q_h}$. The work W was just found to be 1.02×10^{14} J, and Q_h is given, so the efficiency is

$$\begin{aligned} Eff &= \frac{1.02 \times 10^{14}\text{ J}}{2.50 \times 10^{14}\text{ J}} \\ &= 0.408, \text{ or } 40.8\% \end{aligned} \quad (15.30)$$

Strategy for (c)

The daily consumption of coal is calculated using the information that each day there is 2.50×10^{14} J of heat transfer from coal. In the combustion process, we have $C + O_2 \rightarrow CO_2$. So every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of CO_2 into the atmosphere.

Solution for (c)

The daily coal consumption is

$$\frac{2.50\times10^{14}\text{ J}}{2.50\times10^{6}\text{ J/kg}} = 1.0\times10^{8}\text{ kg}. \quad (15.31)$$

Assuming that the coal is pure and that all the coal goes toward producing carbon dioxide, the carbon dioxide produced per day is

$$1.0\times10^{8}\text{ kg coal}\times\frac{44\text{ kg }CO_2}{12\text{ kg coal}} = 3.7\times10^{8}\text{ kg }CO_2. \quad (15.32)$$

This is 370,000 metric tons of CO_2 produced every day.

Discussion

If all the work output is converted to electricity in a period of one day, the average power output is 1180 MW (this is left to you as an end-of-chapter problem). This value is about the size of a large-scale conventional power plant. The efficiency found is acceptably close to the value of 42% given for coal power stations. It means that fully 59.2% of the energy is heat transfer to the environment, which usually results in warming lakes, rivers, or the ocean near the power station, and is implicated in a warming planet generally. While the laws of thermodynamics limit the efficiency of such plants—including plants fired by nuclear fuel, oil, and natural gas—the heat transfer to the environment could be, and sometimes is, used for heating homes or for industrial processes. The generally low cost of energy has not made it economical to make better use of the waste heat transfer from most heat engines. Coal-fired power plants produce the greatest amount of CO_2 per unit energy output (compared to natural gas or oil), making coal the least efficient fossil fuel.

With the information given in Example 15.3, we can find characteristics such as the efficiency of a heat engine without any knowledge of how the heat engine operates, but looking further into the mechanism of the engine will give us greater insight. Figure 15.17 illustrates the operation of the common four-stroke gasoline engine. The four steps shown complete this heat engine's cycle, bringing the gasoline-air mixture back to its original condition.

The **Otto cycle** shown in Figure 15.18(a) is used in four-stroke internal combustion engines, although in fact the true Otto cycle paths do not correspond exactly to the strokes of the engine.

The adiabatic process AB corresponds to the nearly adiabatic compression stroke of the gasoline engine. In both cases, work is done on the system (the gas mixture in the cylinder), increasing its temperature and pressure. Along path BC of the Otto cycle, heat transfer Q_h into the gas occurs at constant volume, causing a further increase in pressure and temperature. This process corresponds to burning fuel in an internal combustion engine, and takes place so rapidly that the volume is nearly constant. Path CD in the Otto cycle is an adiabatic expansion that does work on the outside world, just as the power stroke of an internal combustion engine does in its nearly adiabatic expansion. The work done by the system along path CD is greater than the work done on the system along path AB, because the pressure is greater, and so there is a net work output. Along path DA in the Otto cycle, heat transfer Q_c from the gas at constant volume reduces its temperature and pressure, returning it to its original state. In an internal combustion engine, this process corresponds to the exhaust of hot gases and the intake of an air-gasoline mixture at a considerably lower temperature. In both cases, heat transfer into the environment occurs along this final path.

The net work done by a cyclical process is the area inside the closed path on a PV diagram, such as that inside path ABCDA in Figure 15.18. Note that in every imaginable cyclical process, it is absolutely necessary for heat transfer from the system to occur in order to get a net work output. In the Otto cycle, heat transfer occurs along path DA. If no heat transfer occurs, then the return path is the same, and the net work output is zero. The lower the temperature on the path AB, the less work has to be done to compress the gas. The area inside the closed path is then greater, and so the engine does more work and is thus more efficient. Similarly, the higher the temperature along path CD, the more work output there is. (See Figure 15.19.) So efficiency is related to the temperatures of the hot and cold reservoirs. In the next section, we shall see what the absolute limit to the efficiency of a heat engine is, and how it is related to temperature.

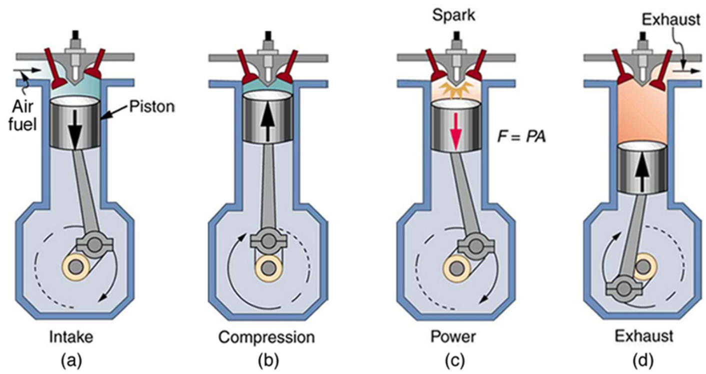


FIGURE 15.17 In the four-stroke internal combustion gasoline engine, heat transfer into work takes place in the cyclical process shown here. The piston is connected to a rotating crankshaft, which both takes work out of and does work on the gas in the cylinder. (a) Air is mixed with fuel during the intake stroke. (b) During the compression stroke, the air-fuel mixture is rapidly compressed in a nearly adiabatic process, as the piston rises with the valves closed. Work is done on the gas. (c) The power stroke has two distinct parts. First, the air-fuel mixture is ignited, converting chemical potential energy into thermal energy almost instantaneously, which leads to a great increase in pressure. Then the piston descends, and the gas does work by exerting a force through a distance in a nearly adiabatic process. (d) The exhaust stroke expels the hot gas to prepare the engine for another cycle, starting again with the intake stroke.

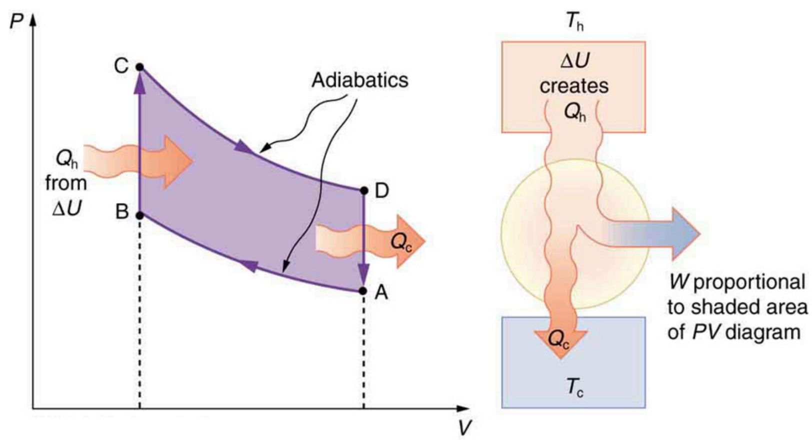


FIGURE 15.18 *PV* diagram for a simplified Otto cycle, analogous to that employed in an internal combustion engine. Point A corresponds to the start of the compression stroke of an internal combustion engine. Paths AB and CD are adiabatic and correspond to the compression and power strokes of an internal combustion engine, respectively. Paths BC and DA are isochoric and accomplish similar results to the ignition and exhaust-intake portions, respectively, of the internal combustion engine's cycle. Work is done on the gas along path AB, but more work is done by the gas along path CD, so that there is a net work output.

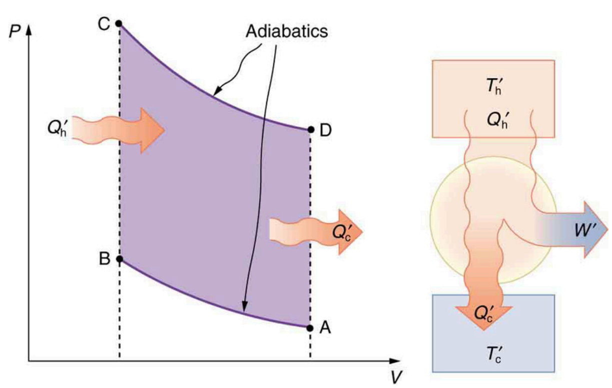


FIGURE 15.19 This Otto cycle produces a greater work output than the one in Figure 15.18, because the starting temperature of path CD is higher and the starting temperature of path AB is lower. The area inside the loop is greater, corresponding to greater net work output.

15.4 Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify a Carnot cycle.
- Calculate maximum theoretical efficiency of a nuclear reactor.
- Explain how dissipative processes affect the ideal Carnot engine.

FIGURE 15.20 This novelty toy, known as the drinking bird, is an example of Carnot's engine. It contains methylene chloride (mixed with a dye) in the abdomen, which boils at a very low temperature—about $100^{\circ}F$. To operate, one gets the bird's head wet. As the water evaporates, fluid moves up into the head, causing the bird to become top-heavy and dip forward back into the water. This cools down the methylene chloride in the head, and it moves back into the abdomen, causing the bird to become bottom heavy and tip up. Except for a very small input of energy—the original head-wetting—the bird becomes a perpetual motion machine of sorts. (credit: Arabesk.nl, Wikimedia Commons)

We know from the second law of thermodynamics that a heat engine cannot be 100% efficient, since there must always be some heat transfer Q_c to the environment, which is often called waste heat. How efficient, then, can a heat engine be? This question was answered at a theoretical level in 1824 by a young French engineer, Sadi Carnot (1796–1832), in his study of the then-emerging heat engine technology crucial to the Industrial Revolution. He devised a theoretical cycle, now called the **Carnot cycle**, which is the most efficient cyclical process possible. The second law of thermodynamics can be restated in terms of the Carnot cycle, and so what Carnot actually discovered was this fundamental law. Any heat engine employing the Carnot cycle is called a **Carnot engine**.

What is crucial to the Carnot cycle—and, in fact, defines it—is that only reversible processes are used. Irreversible processes involve dissipative factors, such as friction and turbulence. This increases heat transfer Q_c to the environment and reduces the efficiency of the engine. Obviously, then, reversible processes are superior.

Carnot Engine

Stated in terms of reversible processes, the **second law of thermodynamics** has a third form:

A Carnot engine operating between two given temperatures has the greatest possible efficiency of any heat engine operating between these two temperatures. Furthermore, all engines employing only reversible processes have this same maximum efficiency when operating between the same given temperatures.

Figure 15.21 shows the PV diagram for a Carnot cycle. The cycle comprises two isothermal and two adiabatic processes. Recall that both isothermal and adiabatic processes are, in principle, reversible.

Carnot also determined the efficiency of a perfect heat engine—that is, a Carnot engine. It is always true that the efficiency of a cyclical heat engine is given by:

$$Eff = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h}. \quad (15.33)$$

What Carnot found was that for a perfect heat engine, the ratio Q_c/Q_h equals the ratio of the absolute temperatures of the heat reservoirs. That is, $Q_c/Q_h = T_c/T_h$ for a Carnot engine, so that the maximum or **Carnot efficiency** Eff_C is given by

$$Eff_C = 1 - \frac{T_c}{T_h}, \quad (15.34)$$

where T_h and T_c are in kelvins (or any other absolute temperature scale). No real heat engine can do as well as the Carnot efficiency—an actual efficiency of about 0.7 of this maximum is usually the best that can be accomplished. But the ideal Carnot engine, like the drinking bird above, while a fascinating novelty, has zero power. This makes it unrealistic for any applications.

Carnot's interesting result implies that 100% efficiency would be possible only if $T_c = 0\text{ K}$—that is, only if the cold reservoir were at absolute zero, a practical and theoretical impossibility. But the physical implication is this—the only way to have all heat transfer go into doing work is to remove *all* thermal energy, and this requires a cold reservoir at absolute zero.

It is also apparent that the greatest efficiencies are obtained when the ratio T_c/T_h is as small as possible. Just as discussed for the Otto cycle in the previous section, this means that efficiency is greatest for the highest possible temperature of the hot reservoir and lowest possible temperature of the cold reservoir. (This setup increases the area inside the closed loop on the PV diagram; also, it seems reasonable that the greater the temperature difference, the easier it is to divert the heat transfer to work.) The actual reservoir temperatures of a heat engine are usually related to the type of heat source and the temperature of the environment into which heat transfer occurs. Consider the following example.

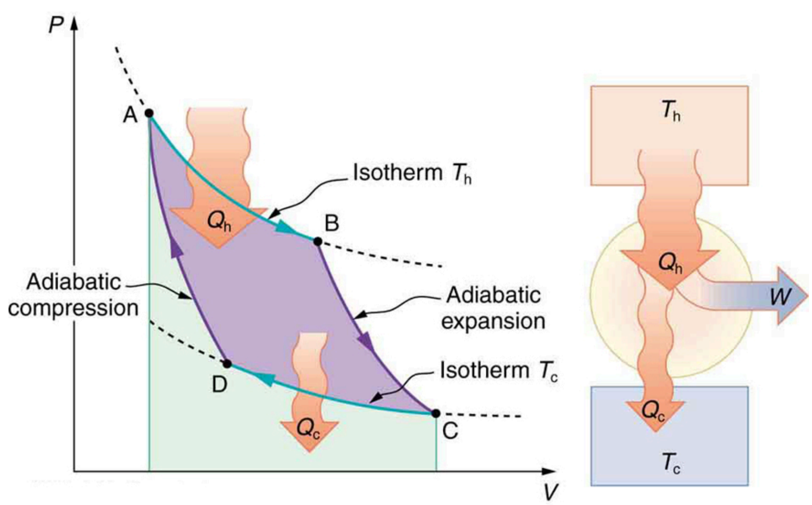


FIGURE 15.21 PV diagram for a Carnot cycle, employing only reversible isothermal and adiabatic processes. Heat transfer Q_h occurs into the working substance during the isothermal path AB, which takes place at constant temperature T_h. Heat transfer Q_c occurs out of the working substance during the isothermal path CD, which takes place at constant temperature T_c. The net work output W equals the area inside the path ABCDA. Also shown is a schematic of a Carnot engine operating between hot and cold reservoirs at temperatures T_h and T_c. Any heat engine using reversible processes and operating between these two temperatures will have the same maximum efficiency as the Carnot engine.

EXAMPLE 15.4

Maximum Theoretical Efficiency for a Nuclear Reactor

A nuclear power reactor has pressurized water at 300°C. (Higher temperatures are theoretically possible but practically not, due to limitations with materials used in the reactor.) Heat transfer from this water is a complex process (see Figure 15.22). Steam, produced in the steam generator, is used to drive the turbine-generators. Eventually the steam is condensed to water at $27^{\circ}\mathrm{C}$ and then heated again to start the cycle over. Calculate the maximum theoretical efficiency for a heat engine operating between these two temperatures.

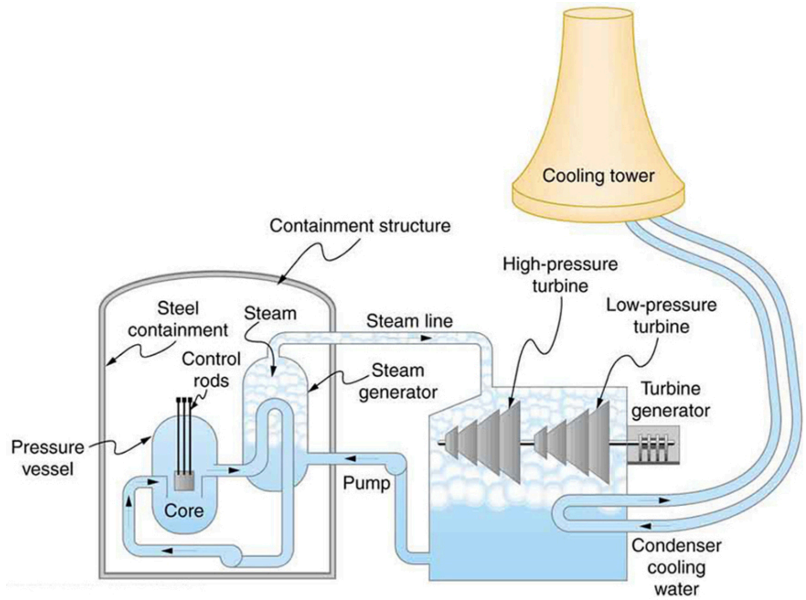


FIGURE 15.22 Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than operating a single generator that produces the same amount of electrical energy. The steam is condensed to liquid before being returned to the heat exchanger, to keep exit steam pressure low and aid the flow of steam through the turbines (equivalent to using a lower-temperature cold reservoir). The considerable energy associated with condensation must be dissipated into the local environment; in this example, a cooling tower is used so there is no direct heat transfer to an aquatic environment. (Note that the water going to the cooling tower does not come into contact with the steam flowing over the turbines.)

Strategy

Since temperatures are given for the hot and cold reservoirs of this heat engine, $Eff_{\mathrm{C}} = 1 - \frac{T_{\mathrm{c}}}{T_{\mathrm{h}}}$ can be used to calculate the Carnot (maximum theoretical) efficiency. Those temperatures must first be converted to kelvins.

Solution

The hot and cold reservoir temperatures are given as $300^{\circ}\mathrm{C}$ and $27.0^{\circ}\mathrm{C}$, respectively. In kelvins, then, $T_{\mathrm{h}} = 573\ \mathrm{K}$ and $T_{\mathrm{c}} = 300\ \mathrm{K}$, so that the maximum efficiency is

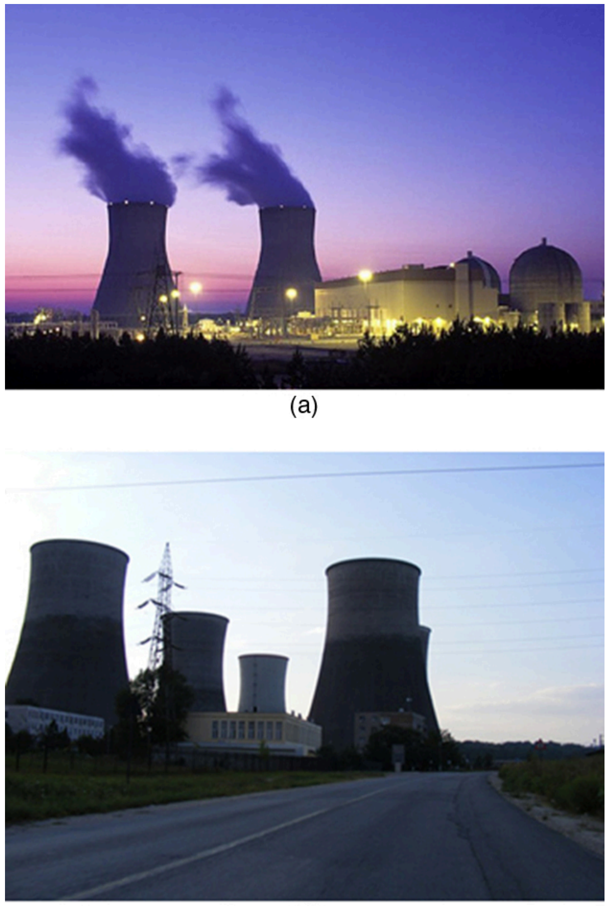
$$Eff_{\mathrm{C}} = 1 - \frac{T_{\mathrm{c}}}{T_{\mathrm{h}}}. \tag{15.35}$$

Thus,

$$\begin{aligned} Eff_{\mathrm{C}} &= 1 - \frac{300\ \mathrm{K}}{573\ \mathrm{K}} \\ &= 0.476, \text{ or } 47.6\%. \end{aligned} \tag{15.36}$$

Discussion

A typical nuclear power station’s actual efficiency is about 35%, a little better than 0.7 times the maximum possible value, a tribute to superior engineering. Electrical power stations fired by coal, oil, and natural gas have greater actual efficiencies (about 42%), because their boilers can reach higher temperatures and pressures. The cold reservoir temperature in any of these power stations is limited by the local environment. Figure 15.23 shows (a) the exterior of a nuclear power station and (b) the exterior of a coal-fired power station. Both have cooling towers into which water from the condenser enters the tower near the top and is sprayed downward, cooled by evaporation.



(a)

(b)

FIGURE 15.23 (a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station. Both have cooling towers in which water evaporates into the environment, representing Q_c. The nuclear reactor, which supplies Q_h, is housed inside the dome-shaped containment buildings. (credit: Robert & Mihaela Vicol, publicphoto.org)

Since all real processes are irreversible, the actual efficiency of a heat engine can never be as great as that of a Carnot engine, as illustrated in Figure 15.24(a). Even with the best heat engine possible, there are always dissipative processes in peripheral equipment, such as electrical transformers or car transmissions. These further reduce the overall efficiency by converting some of the engine's work output back into heat transfer, as shown in Figure 15.24(b).

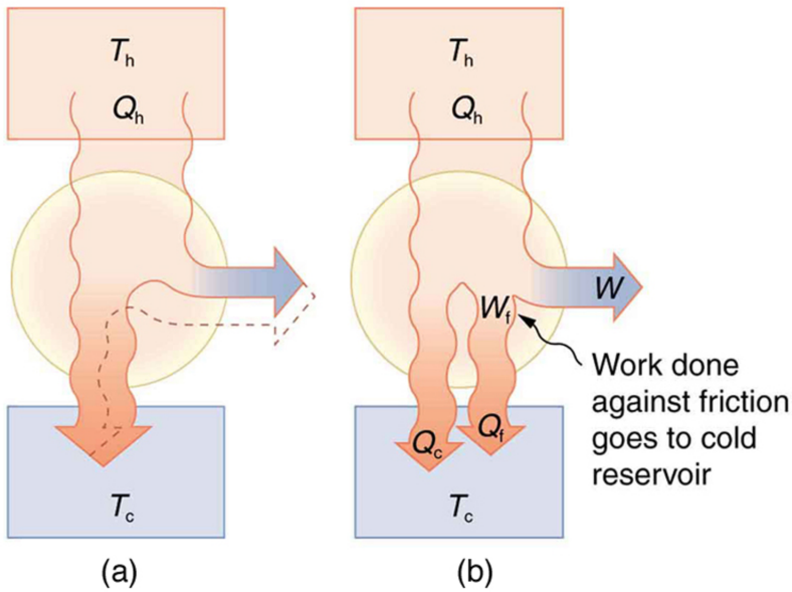


FIGURE 15.24 Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some of its work output into heat transfer to the environment.

15.5 Applications of Thermodynamics: Heat Pumps and Refrigerators

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the use of heat engines in heat pumps and refrigerators.
- Demonstrate how a heat pump works to warm an interior space.
- Explain the differences between heat pumps and refrigerators.
- Calculate a heat pump's coefficient of performance.

FIGURE 15.25 Almost every home contains a refrigerator. Most people don't realize they are also sharing their homes with a heat pump. (credit: Id1337x, Wikimedia Commons)

Heat pumps, air conditioners, and refrigerators utilize heat transfer from cold to hot. They are heat engines run backward. We say backward, rather than reverse, because except for Carnot engines, all heat engines, though they can be run backward, cannot truly be reversed. Heat transfer occurs from a cold reservoir Q_c and into a hot one. This requires work input W, which is also converted to heat transfer. Thus the heat transfer to the hot reservoir is $Q_h = Q_c + W$. (Note that Q_h, Q_c, and W are positive, with their directions indicated on schematics rather than by sign.) A heat pump's mission is for heat transfer Q_h to occur into a warm environment, such as a home in the winter. The mission of air conditioners and refrigerators is for heat transfer Q_c to occur from a cool environment, such as chilling a room or keeping food at lower temperatures than the environment. (Actually, a heat pump can be used both to heat and cool a space. It is essentially an air conditioner and a heating unit all in one. In this section we will concentrate on its heating mode.)

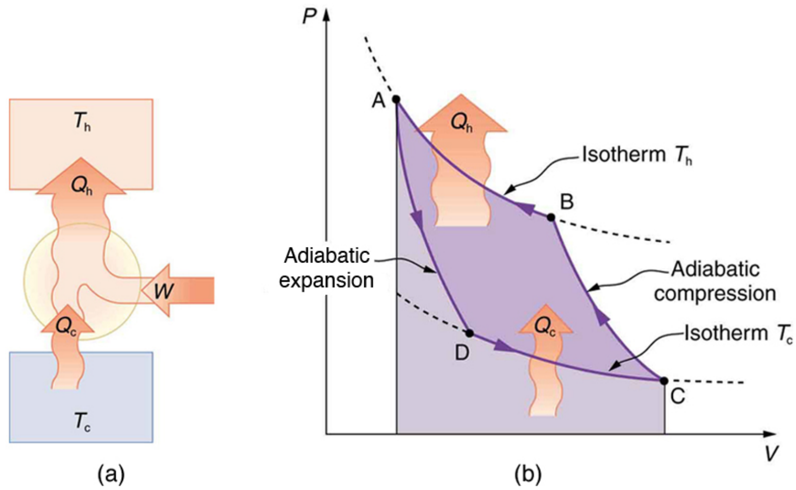


FIGURE 15.26 Heat pumps, air conditioners, and refrigerators are heat engines operated backward. The one shown here is based on a Carnot (reversible) engine. (a) Schematic diagram showing heat transfer from a cold reservoir to a warm reservoir with a heat pump. The directions of W , Q_h, and Q_c are opposite what they would be in a heat engine. (b) PV diagram for a Carnot cycle similar to that in Figure 15.27 but reversed, following path ADCBA. The area inside the loop is negative, meaning there is a net work input. There is heat transfer Q_c into the system from a cold reservoir along path DC, and heat transfer Q_h out of the system into a hot reservoir along path BA.

Heat Pumps

The great advantage of using a heat pump to keep your home warm, rather than just burning fuel, is that a heat pump supplies $Q_h = Q_c + W$. Heat transfer is from the outside air, even at a temperature below freezing, to the indoor space. You only pay for W, and you get an additional heat transfer of Q_c from the outside at no cost; in many cases, at least twice as much energy is transferred to the heated space as is used to run the heat pump. When you burn fuel to keep warm, you pay for all of it. The disadvantage is that the work input (required by the second law of thermodynamics) is sometimes more expensive than simply burning fuel, especially if the work is done by electrical energy.

The basic components of a heat pump in its heating mode are shown in Figure 15.27. A working fluid such as a non-CFC refrigerant is used. In the outdoor coils (the evaporator), heat transfer Q_c occurs to the working fluid from the cold outdoor air, turning it into a gas.

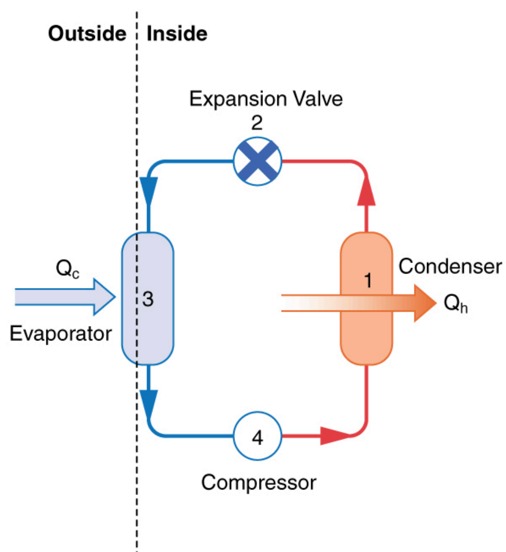


FIGURE 15.27 A simple heat pump has four basic components: (1) condenser, (2) expansion valve, (3) evaporator, and (4) compressor. In the heating mode, heat transfer Q_c occurs to the working fluid in the evaporator (3) from the colder outdoor air, turning it into a gas. The electrically driven compressor (4) increases the temperature and pressure of the gas and forces it into the condenser coils (1) inside the heated space. Because the temperature of the gas is higher than the temperature in the room, heat transfer from the gas to the room

occurs as the gas condenses to a liquid. The working fluid is then cooled as it flows back through an expansion valve (2) to the outdoor evaporator coils.

The electrically driven compressor (work input W) raises the temperature and pressure of the gas and forces it into the condenser coils that are inside the heated space. Because the temperature of the gas is higher than the temperature inside the room, heat transfer to the room occurs and the gas condenses to a liquid. The liquid then flows back through a pressure-reducing valve to the outdoor evaporator coils, being cooled through expansion. (In a cooling cycle, the evaporator and condenser coils exchange roles and the flow direction of the fluid is reversed.)

The quality of a heat pump is judged by how much heat transfer Q_h occurs into the warm space compared with how much work input W is required. In the spirit of taking the ratio of what you get to what you spend, we define a **heat pump's coefficient of performance** (COP_hp) to be

$$COP_\text{hp} = \frac{Q_\text{h}}{W}. \qquad 15.37$$

Since the efficiency of a heat engine is $Eff = W/Q_\text{h}$, we see that $COP_\text{hp} = 1/Eff$, an important and interesting fact. First, since the efficiency of any heat engine is less than 1, it means that COP_hp is always greater than 1—that is, a heat pump always has more heat transfer Q_h than work put into it. Second, it means that heat pumps work best when temperature differences are small. The efficiency of a perfect, or Carnot, engine is $Eff_\text{C} = 1 - (T_\text{c}/T_\text{h})$; thus, the smaller the temperature difference, the smaller the efficiency and the greater the COP_hp (because $COP_\text{hp} = 1/Eff$). In other words, heat pumps do not work as well in very cold climates as they do in more moderate climates.

Friction and other irreversible processes reduce heat engine efficiency, but they do *not* benefit the operation of a heat pump—instead, they reduce the work input by converting part of it to heat transfer back into the cold reservoir before it gets into the heat pump.

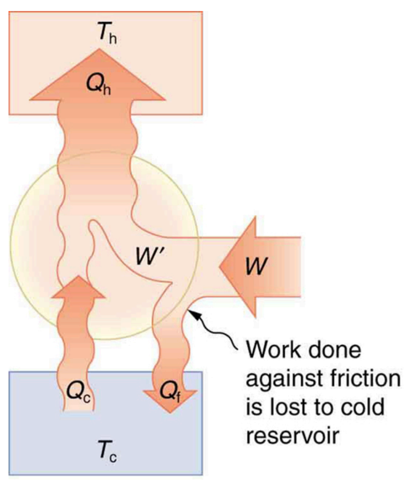


FIGURE 15.28 When a real heat engine is run backward, some of the intended work input (W) goes into heat transfer before it gets into the heat engine, thereby reducing its coefficient of performance COP_hp. In this figure, W' represents the portion of W that goes into the heat pump, while the remainder of W is lost in the form of frictional heat (Q_f) to the cold reservoir. If all of W had gone into the heat pump, then Q_h would have been greater. The best heat pump uses adiabatic and isothermal processes, since, in theory, there would be no dissipative processes to reduce the heat transfer to the hot reservoir.

EXAMPLE 15.5

The Best COP_hp of a Heat Pump for Home Use

A heat pump used to warm a home must employ a cycle that produces a working fluid at temperatures greater than typical indoor temperature so that heat transfer to the inside can take place. Similarly, it must produce a working fluid at temperatures that are colder than the outdoor temperature so that heat transfer occurs from outside. Its hot and cold reservoir temperatures therefore cannot be too close, placing a limit on its COP_hp. (See Figure 15.29.)

What is the best coefficient of performance possible for such a heat pump, if it has a hot reservoir temperature of $45.0°\text{C}$ and a cold reservoir temperature of $-15.0°\text{C}$?

Strategy

A Carnot engine reversed will give the best possible performance as a heat pump. As noted above, $COP_{\text{hp}} = 1/Eff$, so that we need to first calculate the Carnot efficiency to solve this problem.

Solution

Carnot efficiency in terms of absolute temperature is given by**:**

$$Eff_{\text{C}} = 1 - \frac{T_{\text{c}}}{T_{\text{h}}}. \quad (15.38)$$

The temperatures in kelvins are $T_{\text{h}} = 318\text{ K}$ and $T_{\text{c}} = 258\text{ K}$, so that

$$Eff_{\text{C}} = 1 - \frac{258\text{ K}}{318\text{ K}} = 0.1887. \quad (15.39)$$

Thus, from the discussion above,

$$COP_{\text{hp}} = \frac{1}{Eff} = \frac{1}{0.1887} = 5.30, \quad (15.40)$$

or

$$COP_{\text{hp}} = \frac{Q_{\text{h}}}{W} = 5.30, \quad (15.41)$$

so that

$$Q_{\text{h}} = 5.30\text{ W}. \quad (15.42)$$

Discussion

This result means that the heat transfer by the heat pump is 5.30 times as much as the work put into it. It would cost 5.30 times as much for the same heat transfer by an electric room heater as it does for that produced by this heat pump. This is not a violation of conservation of energy. Cold ambient air provides 4.3 J per 1 J of work from the electrical outlet.

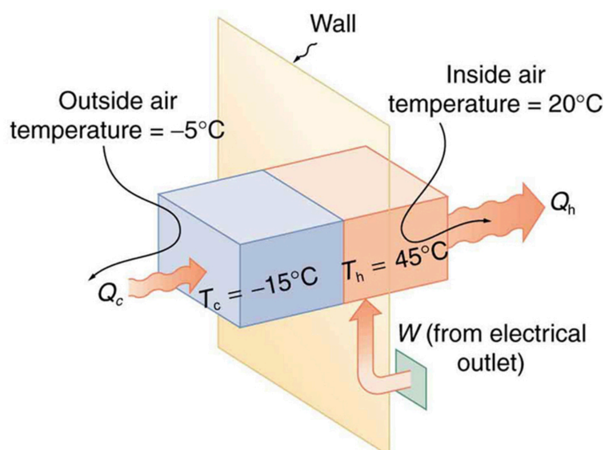


FIGURE 15.29 Heat transfer from the outside to the inside, along with work done to run the pump, takes place in the heat pump of the example above**.** Note that the cold temperature produced by the heat pump is lower than the outside temperature, so that heat transfer into the working fluid occurs. The pump's compressor produces a temperature greater than the indoor temperature in order for heat transfer into the house to occur.

Real heat pumps do not perform quite as well as the ideal one in the previous example; their values of COP_{hp} range from about 2 to 4. This range means that the heat transfer Q_{h} from the heat pumps is 2 to 4 times as great as the work W put into them. Their economical feasibility is still limited, however, since W is usually supplied by electrical energy that costs more per joule than heat transfer by burning fuels like natural gas. Furthermore, the initial cost of a heat pump is greater than that of many furnaces, so that a heat pump must last longer for its cost to be recovered. Heat pumps are most likely to be economically superior where winter temperatures are mild, electricity is relatively cheap, and other fuels are relatively expensive. Also, since they can cool as well as heat a space, they have advantages where cooling in summer months is also desired. Thus some of the best locations for heat pumps are in warm summer climates with cool winters. Figure 15.30 shows a heat pump, called a "*reverse cycle*" or "*split-system cooler*" in some countries.

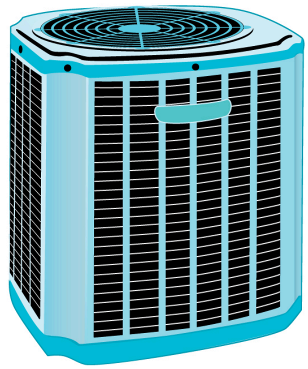


FIGURE 15.30 In hot weather, heat transfer occurs from air inside the room to air outside, cooling the room. In cool weather, heat transfer occurs from air outside to air inside, warming the room. This switching is achieved by reversing the direction of flow of the working fluid.

Air Conditioners and Refrigerators

Air conditioners and refrigerators are designed to cool something down in a warm environment. As with heat pumps, work input is required for heat transfer from cold to hot, and this is expensive. The quality of air conditioners and refrigerators is judged by how much heat transfer Q_c occurs from a cold environment compared with how much work input W is required. What is considered the benefit in a heat pump is considered waste heat in a refrigerator. We thus define the **coefficient of performance** (COP_{ref}) of an air conditioner or refrigerator to be

$$COP_{ref} = \frac{Q_c}{W}. \tag{15.43}$$

Noting again that $Q_h = Q_c + W$, we can see that an air conditioner will have a lower coefficient of performance than a heat pump, because $COP_{hp} = Q_h/W$ and Q_h is greater than Q_c. In this module's Problems and Exercises, you will show that

$$COP_{ref} = COP_{hp} - 1 \tag{15.44}$$

for a heat engine used as either an air conditioner or a heat pump operating between the same two temperatures. Real air conditioners and refrigerators typically do remarkably well, having values of COP_{ref} ranging from 2 to 6. These numbers are better than the COP_{hp} values for the heat pumps mentioned above, because the temperature differences are smaller, but they are less than those for Carnot engines operating between the same two temperatures.

A type of COP rating system called the "energy efficiency rating" (EER) has been developed. This rating is an example where non-SI units are still used and relevant to consumers. To make it easier for the consumer, Australia, Canada, New Zealand, and the U.S. use an Energy Star Rating out of 5 stars—the more stars, the more energy efficient the appliance. EERs are expressed in mixed units of British thermal units (Btu) per hour of heating or cooling divided by the power input in watts. Room air conditioners are readily available with EERs ranging from 6 to 12. Although not the same as the COPs just described, these EERs are good for comparison purposes—the greater the EER, the cheaper an air conditioner is to operate (but the higher its purchase price is likely to be).

The EER of an air conditioner or refrigerator can be expressed as

$$EER = \frac{Q_c/t_1}{W/t_2}, \tag{15.45}$$

where Q_c is the amount of heat transfer from a cold environment in British thermal units, t_1 is time in hours, W is the work input in joules, and t_2 is time in seconds.

Problem-Solving Strategies for Thermodynamics

1. *Examine the situation to determine whether heat, work, or internal energy are involved.* Look for any system where the primary methods of transferring energy are heat and work. Heat engines, heat pumps, refrigerators, and air conditioners are examples of such systems.
2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful. Maximum efficiency means a Carnot engine is involved. Efficiency is not the same as the coefficient of performance.
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).* Be sure to distinguish heat transfer into a system from heat transfer out of the system, as well as work input from work output. In many situations, it is useful to determine the type of process, such as isothermal or adiabatic.
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. *Substitute the known quantities along with their units into the appropriate equation and obtain numerical solutions complete with units.*
7. *Check the answer to see if it is reasonable: Does it make sense?* For example, efficiency is always less than 1, whereas coefficients of performance are greater than 1.

15.6 Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define entropy and calculate the increase of entropy in a system with reversible and irreversible processes.
- Explain the expected fate of the universe in entropic terms.
- Calculate the increasing disorder of a system.

FIGURE 15.31 The ice in this drink is slowly melting. Eventually the liquid will reach thermal equilibrium, as predicted by the second law of thermodynamics. (credit: Jon Sullivan, PDPhoto.org)

There is yet another way of expressing the second law of thermodynamics. This version relates to a concept called **entropy**. By examining it, we shall see that the directions associated with the second law—heat transfer from hot to cold, for example—are related to the tendency in nature for systems to become disordered and for less energy to be available for use as work. The entropy of a system can in fact be shown to be a measure of its disorder and of the unavailability of energy to do work.

Making Connections: Entropy, Energy, and Work

Recall that the simple definition of energy is the ability to do work. Entropy is a measure of how much energy is not available to do work. Although all forms of energy are interconvertible, and all can be used to do work, it is not always possible, even in principle, to convert the entire available energy into work. That unavailable energy is of interest in thermodynamics, because the field of thermodynamics arose from efforts to convert heat to work.

We can see how entropy is defined by recalling our discussion of the Carnot engine. We noted that for a Carnot cycle, and hence for any reversible processes, $Q_c/Q_h = T_c/T_h$. Rearranging terms yields

$$\frac{Q_c}{T_c} = \frac{Q_h}{T_h} \tag{15.46}$$

for any reversible process. Q_c and Q_h are absolute values of the heat transfer at temperatures T_c and T_h, respectively. This ratio of Q/T is defined to be the **change in entropy** ΔS for a reversible process,

$$\Delta S = \left(\frac{Q}{T}\right)_{\text{rev}}, \tag{15.47}$$

where Q is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin (J/K). If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature, avoiding the need to use integral calculus to find ΔS.

The definition of ΔS is strictly valid only for reversible processes, such as used in a Carnot engine. However, we can find ΔS precisely even for real, irreversible processes. The reason is that the entropy s of a system, like internal energy U, depends only on the state of the system and not how it reached that condition. Entropy is a property of state. Thus the change in entropy ΔS of a system between state 1 and state 2 is the same no matter how the change occurs. We just need to find or imagine a reversible process that takes us from state 1 to state 2 and calculate ΔS for that process. That will be the change in entropy for any process going from state 1 to state 2. (See Figure 15.32.)

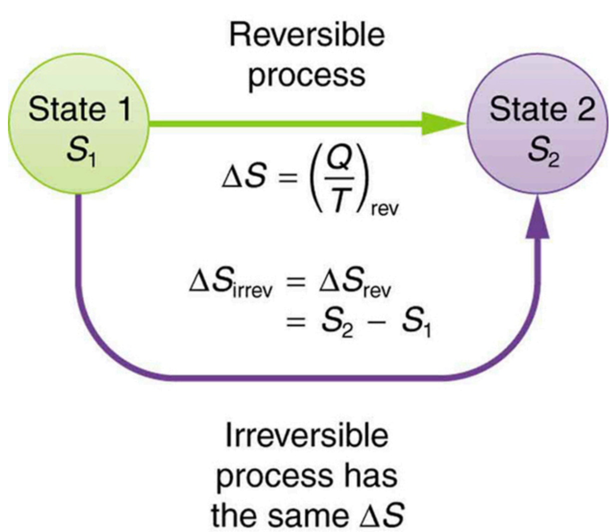


FIGURE 15.32 When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS, whether a hypothetical reversible path is followed or a real irreversible path is taken.

Now let us take a look at the change in entropy of a Carnot engine and its heat reservoirs for one full cycle. The hot reservoir has a loss of entropy $\Delta S_h = -Q_h/T_h$, because heat transfer occurs out of it (remember that when heat transfers out, then Q has a negative sign). The cold reservoir has a gain of entropy $\Delta S_c = Q_c/T_c$, because heat transfer occurs into it. (We assume the reservoirs are sufficiently large that their temperatures are constant.) So the total change in entropy is

$$\Delta S_{\text{tot}} = \Delta S_{\text{h}} + \Delta S_{\text{c}}. \quad 15.48$$

Thus, since we know that $Q_{\text{h}}/T_{\text{h}} = Q_{\text{c}}/T_{\text{c}}$ for a Carnot engine,

$$\Delta S_{\text{tot}} = -\frac{Q_{\text{h}}}{T_{\text{h}}} + \frac{Q_{\text{c}}}{T_{\text{c}}} = 0. \quad 15.49$$

This result, which has general validity, means that *the total change in entropy for a system in any reversible process is zero.*

The entropy of various parts of the system may change, but the total change is zero. Furthermore, the system does not affect the entropy of its surroundings, since heat transfer between them does not occur. Thus the reversible process changes neither the total entropy of the system nor the entropy of its surroundings. Sometimes this is stated as follows: *Reversible processes do not affect the total entropy of the universe.* Real processes are not reversible, though, and they do change total entropy. We can, however, use hypothetical reversible processes to determine the value of entropy in real, irreversible processes. The following example illustrates this point.

EXAMPLE 15.6

Entropy Increases in an Irreversible (Real) Process

Spontaneous heat transfer from hot to cold is an irreversible process. Calculate the total change in entropy if 4000 J of heat transfer occurs from a hot reservoir at $T_{\text{h}} = 600\text{ K}(327^{\circ}\text{ C})$ to a cold reservoir at $T_{\text{c}} = 250\text{ K}(-23^{\circ}\text{ C})$, assuming there is no temperature change in either reservoir. (See Figure 15.33.)

Strategy

How can we calculate the change in entropy for an irreversible process when $\Delta S_{\text{tot}} = \Delta S_{\text{h}} + \Delta S_{\text{c}}$ is valid only for reversible processes? Remember that the total change in entropy of the hot and cold reservoirs will be the same whether a reversible or irreversible process is involved in heat transfer from hot to cold. So we can calculate the change in entropy of the hot reservoir for a hypothetical reversible process in which 4000 J of heat transfer occurs from it; then we do the same for a hypothetical reversible process in which 4000 J of heat transfer occurs to the cold reservoir. This produces the same changes in the hot and cold reservoirs that would occur if the heat transfer were allowed to occur irreversibly between them, and so it also produces the same changes in entropy.

Solution

We now calculate the two changes in entropy using $\Delta S_{\text{tot}} = \Delta S_{\text{h}} + \Delta S_{\text{c}}$. First, for the heat transfer from the hot reservoir,

$$\Delta S_{\text{h}} = \frac{-Q_{\text{h}}}{T_{\text{h}}} = \frac{-4000\text{ J}}{600\text{ K}} = -6.67\text{ J/K}. \quad 15.50$$

And for the cold reservoir,

$$\Delta S_{\text{c}} = \frac{Q_{\text{c}}}{T_{\text{c}}} = \frac{4000\text{ J}}{250\text{ K}} = 16.0\text{ J/K}. \quad 15.51$$

Thus the total is

$$\begin{aligned} \Delta S_{\text{tot}} &= \Delta S_{\text{h}} + \Delta S_{\text{c}} \\ &= (-6.67 + 16.0)\text{ J/K} \\ &= 9.33\text{ J/K}. \end{aligned} \quad 15.52$$

Discussion

There is an *increase* in entropy for the system of two heat reservoirs undergoing this irreversible heat transfer. We will see that this means there is a loss of ability to do work with this transferred energy. Entropy has increased, and energy has become unavailable to do work.

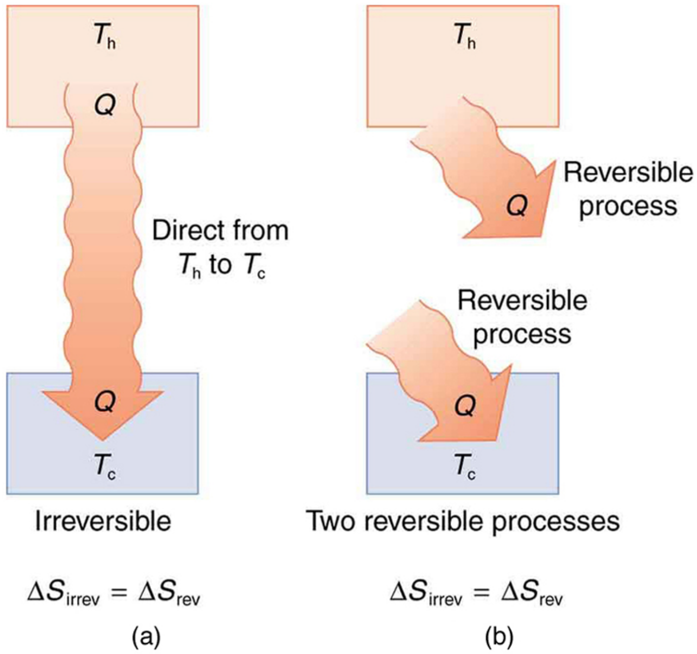


FIGURE 15.33 (a) Heat transfer from a hot object to a cold one is an irreversible process that produces an overall increase in entropy. (b) The same final state and, thus, the same change in entropy is achieved for the objects if reversible heat transfer processes occur between the two objects whose temperatures are the same as the temperatures of the corresponding objects in the irreversible process.

It is reasonable that entropy increases for heat transfer from hot to cold. Since the change in entropy is Q/T, there is a larger change at lower temperatures. The decrease in entropy of the hot object is therefore less than the increase in entropy of the cold object, producing an overall increase, just as in the previous example. This result is very general:

There is an increase in entropy for any system undergoing an irreversible process.

With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process. There is a fourth version of **the second law of thermodynamics stated in terms of entropy**:

The total entropy of a system either increases or remains constant in any process; it never decreases.

For example, heat transfer cannot occur spontaneously from cold to hot, because entropy would decrease.

Entropy is very different from energy. Entropy is *not* conserved but increases in all real processes. Reversible processes (such as in Carnot engines) are the processes in which the most heat transfer to work takes place and are also the ones that keep entropy constant. Thus we are led to make a connection between entropy and the availability of energy to do work.

Entropy and the Unavailability of Energy to Do Work

What does a change in entropy mean, and why should we be interested in it? One reason is that entropy is directly related to the fact that not all heat transfer can be converted into work. The next example gives some indication of how an increase in entropy results in less heat transfer into work.

EXAMPLE 15.7

Less Work is Produced by a Given Heat Transfer When Entropy Change is Greater

(a) Calculate the work output of a Carnot engine operating between temperatures of 600 K and 100 K for 4000 J of heat transfer to the engine. (b) Now suppose that the 4000 J of heat transfer occurs first from the 600 K reservoir to a 250 K reservoir (without doing any work, and this produces the increase in entropy calculated above) before transferring into a Carnot engine operating between 250 K and 100 K. What work output is produced? (See Figure 15.34.)

Strategy

In both parts, we must first calculate the Carnot efficiency and then the work output.

Solution (a)

The Carnot efficiency is given by

$$Eff_{\mathrm{C}} = 1 - \frac{T_{\mathrm{c}}}{T_{\mathrm{h}}}. \tag{15.53}$$

Substituting the given temperatures yields

$$Eff_{\mathrm{C}} = 1 - \frac{100\ \mathrm{K}}{600\ \mathrm{K}} = 0.833. \tag{15.54}$$

Now the work output can be calculated using the definition of efficiency for any heat engine as given by

$$Eff = \frac{W}{Q_{\mathrm{h}}}. \tag{15.55}$$

Solving for W and substituting known terms gives

$$\begin{aligned} W &= Eff_{\mathrm{C}} Q_{\mathrm{h}} \\ &= (0.833)(4000\ \mathrm{J}) = 3333\ \mathrm{J}. \end{aligned} \tag{15.56}$$

Solution (b)

Similarly,

$$Eff'_{\mathrm{C}} = 1 - \frac{T_{\mathrm{c}}}{T'_{\mathrm{c}}} = 1 - \frac{100\ \mathrm{K}}{250\ \mathrm{K}} = 0.600, \tag{15.57}$$

so that

$$\begin{aligned} W &= Eff'_{\mathrm{C}} Q_h \\ &= (0.600)(4000\ \mathrm{J}) = 2400\ \mathrm{J}. \end{aligned} \tag{15.58}$$

Discussion

There is 933 J less work from the same heat transfer in the second process. This result is important. The same heat transfer into two perfect engines produces different work outputs, because the entropy change differs in the two cases. In the second case, entropy is greater and less work is produced. Entropy is associated with the *un*availability of energy to do work.

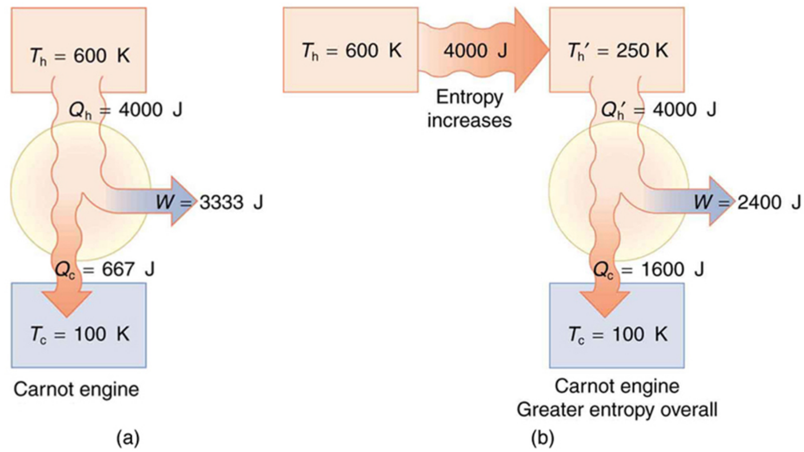


FIGURE 15.34 (a) A Carnot engine working at between 600 K and 100 K has 4000 J of heat transfer and performs 3333 J of work. (b) The 4000 J of heat transfer occurs first irreversibly to a 250 K reservoir and then goes into a Carnot engine. The increase in entropy caused by the heat transfer to a colder reservoir results in a smaller work output of 2400 J. There is a permanent loss of 933 J of energy for the purpose of doing work.

When entropy increases, a certain amount of energy becomes *permanently* unavailable to do work. The energy is not lost, but its character is changed, so that some of it can never be converted to doing work—that is, to an organized force acting through a distance. For instance, in the previous example, 933 J less work was done after an increase in entropy of 9.33 J/K occurred in the 4000 J heat transfer from the 600 K reservoir to the 250 K reservoir. It can be shown that the amount of energy that becomes unavailable for work is

$$W_{\text{unavail}} = \Delta S \cdot T_0, \tag{15.59}$$

where T_0 is the lowest temperature utilized. In the previous example,

$$W_{\text{unavail}} = (9.33\text{ J/K})(100\text{ K}) = 933\text{ J} \tag{15.60}$$

as found.

Heat Death of the Universe: An Overdose of Entropy

In the early, energetic universe, all matter and energy were easily interchangeable and identical in nature. Gravity played a vital role in the young universe. Although it may have *seemed* disorderly, and therefore, superficially entropic, in fact, there was enormous potential energy available to do work—all the future energy in the universe.

As the universe matured, temperature differences arose, which created more opportunity for work. Stars are hotter than planets, for example, which are warmer than icy asteroids, which are warmer still than the vacuum of the space between them.

Most of these are cooling down from their usually violent births, at which time they were provided with energy of their own—nuclear energy in the case of stars, volcanic energy on Earth and other planets, and so on. Without additional energy input, however, their days are numbered.

As entropy increases, less and less energy in the universe is available to do work. On Earth, we still have great stores of energy such as fossil and nuclear fuels; large-scale temperature differences, which can provide wind energy; geothermal energies due to differences in temperature in Earth's layers; and tidal energies owing to our abundance of liquid water. As these are used, a certain fraction of the energy they contain can never be converted into doing work. Eventually, all fuels will be exhausted, all temperatures will equalize, and it will be impossible for heat engines to function, or for work to be done.

Entropy increases in a closed system, such as the universe. But in parts of the universe, for instance, in the Solar system, it is not a locally closed system. Energy flows from the Sun to the planets, replenishing Earth's stores of energy. The Sun will continue to supply us with energy for about another five billion years. We will enjoy direct solar energy, as well as side effects of solar energy, such as wind power and biomass energy from photosynthetic plants. The energy from the Sun will keep our water at the liquid state, and the Moon's gravitational pull will continue to provide tidal energy. But Earth's geothermal energy will slowly run down and won't be replenished.

But in terms of the universe, and the very long-term, very large-scale picture, the entropy of the universe is increasing, and so the availability of energy to do work is constantly decreasing. Eventually, when all stars have died, all forms of potential energy have been utilized, and all temperatures have equalized (depending on the mass of the universe, either at a very high temperature following a universal contraction, or a very low one, just before all activity ceases) there will be no possibility of doing work.

Either way, the universe is destined for thermodynamic equilibrium—maximum entropy. This is often called the *heat death of the universe*, and will mean the end of all activity. However, whether the universe contracts and heats up, or continues to expand and cools down, the end is not near. Calculations of black holes suggest that entropy can easily continue for at least 10^{100} years.

Order to Disorder

Entropy is related not only to the unavailability of energy to do work—it is also a measure of disorder. This notion was initially postulated by Ludwig Boltzmann in the 1800s. For example, melting a block of ice means taking a highly structured and orderly system of water molecules and converting it into a disorderly liquid in which molecules have no fixed positions. (See Figure 15.35.) There is a large increase in entropy in the process, as seen in the following example.

EXAMPLE 15.8

Entropy Associated with Disorder

Find the increase in entropy of 1.00 kg of ice originally at 0^o C that is melted to form water at 0^o C.

Strategy

As before, the change in entropy can be calculated from the definition of ΔS once we find the energy Q needed to melt the ice.

Solution

The change in entropy is defined as:

$$\Delta S = \frac{Q}{T}. \quad (15.61)$$

Here Q is the heat transfer necessary to melt 1.00 kg of ice and is given by

$$Q = mL_\text{f}, \quad (15.62)$$

where m is the mass and L_f is the latent heat of fusion. $L_\text{f} = 334\text{ kJ/kg}$ for water, so that

$$Q = (1.00\text{ kg})(334\text{ kJ/kg}) = 3.34 \times 10^5\text{ J}. \quad (15.63)$$

Now the change in entropy is positive, since heat transfer occurs into the ice to cause the phase change; thus,

$$\Delta S = \frac{Q}{T} = \frac{3.34 \times 10^5\text{ J}}{T}. \quad (15.64)$$

T is the melting temperature of ice. That is, $T = 0^\text{o}\text{C} = 273\text{ K}$. So the change in entropy is

$$\begin{aligned}\Delta S &= \frac{3.34 \times 10^5\text{ J}}{273\text{ K}} \\ &= 1.22 \times 10^3\text{ J/K}.\end{aligned} \quad (15.65)$$

Discussion

This is a significant increase in entropy accompanying an increase in disorder.

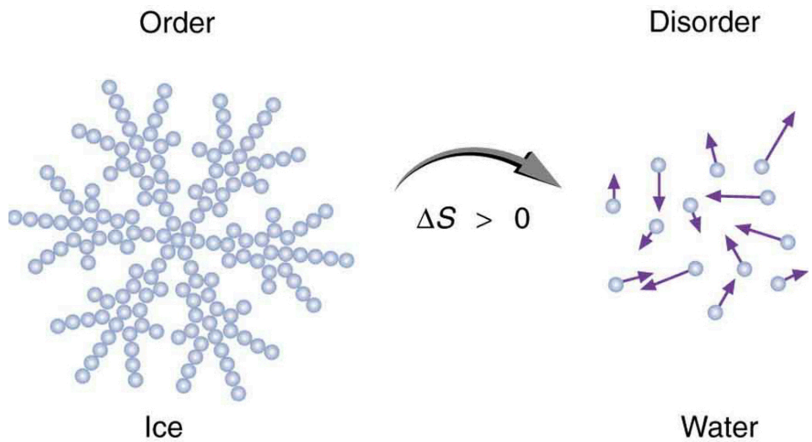


FIGURE 15.35 When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without fixed locations or orientations. Its entropy increases because heat transfer occurs into it. Entropy is a measure of disorder.

In another easily imagined example, suppose we mix equal masses of water originally at two different temperatures, say 20.0^o C and 40.0^o C. The result is water at an intermediate temperature of 30.0^o C. Three outcomes have resulted: entropy has increased, some energy has become unavailable to do work, and the system has become less orderly. Let us think about each of these results.

First, entropy has increased for the same reason that it did in the example above. Mixing the two bodies of water has the same effect as heat transfer from the hot one and the same heat transfer into the cold one. The mixing decreases the entropy of the hot water but increases the entropy of the cold water by a greater amount, producing an overall increase in entropy.

Second, once the two masses of water are mixed, there is only one temperature—you cannot run a heat engine with them. The energy that could have been used to run a heat engine is now unavailable to do work.

Third, the mixture is less orderly, or to use another term, less structured. Rather than having two masses at different temperatures and with different distributions of molecular speeds, we now have a single mass with a uniform temperature.

These three results—entropy, unavailability of energy, and disorder—are not only related but are in fact essentially equivalent.

Life, Evolution, and the Second Law of Thermodynamics

Some people misunderstand the second law of thermodynamics, stated in terms of entropy, to say that the process of the evolution of life violates this law. Over time, complex organisms evolved from much simpler ancestors, representing a large decrease in entropy of the Earth's biosphere. It is a fact that living organisms have evolved to be highly structured, and much lower in entropy than the substances from which they grow. But it is *always* possible for the entropy of one part of the universe to decrease, provided the total change in entropy of the universe increases. In equation form, we can write this as

$$\Delta S_\text{tot} = \Delta S_\text{syst} + \Delta S_\text{envir} > 0. \tag{15.66}$$

Thus ΔS_syst can be negative as long as ΔS_envir is positive and greater in magnitude.

How is it possible for a system to decrease its entropy? Energy transfer is necessary. If I pick up marbles that are scattered about the room and put them into a cup, my work has decreased the entropy of that system. If I gather iron ore from the ground and convert it into steel and build a bridge, my work has decreased the entropy of that system. Energy coming from the Sun can decrease the entropy of local systems on Earth—that is, ΔS_syst is negative. But the overall entropy of the rest of the universe increases by a greater amount—that is, ΔS_envir is positive and greater in magnitude. Thus, $\Delta S_\text{tot} = \Delta S_\text{syst} + \Delta S_\text{envir} > 0$, and the second law of thermodynamics is *not* violated.

Every time a plant stores some solar energy in the form of chemical potential energy, or an updraft of warm air lifts a soaring bird, the Earth can be viewed as a heat engine operating between a hot reservoir supplied by the Sun and a cold reservoir supplied by dark outer space—a heat engine of high complexity, causing local decreases in entropy as it uses part of the heat transfer from the Sun into deep space. There is a large total increase in entropy resulting from this massive heat transfer. A small part of this heat transfer is stored in structured systems on Earth, producing much smaller local decreases in entropy. (See Figure 15.36.)

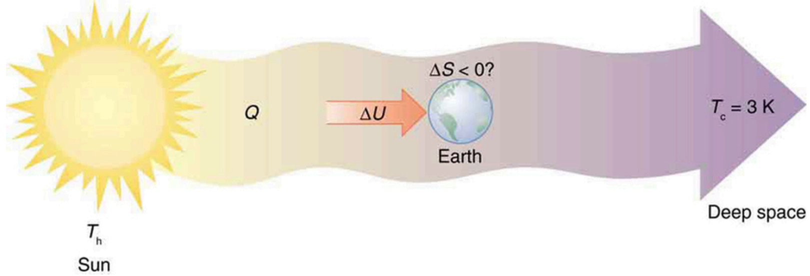


FIGURE 15.36 Earth's entropy may decrease in the process of intercepting a small part of the heat transfer from the Sun into deep space. Entropy for the entire process increases greatly while Earth becomes more structured with living systems and stored energy in various forms.

PHET EXPLORATIONS

Reversible Reactions

Watch a reaction proceed over time. How does total energy affect a reaction rate? Vary temperature, barrier height, and potential energies. Record concentrations and time in order to extract rate coefficients. Do temperature dependent studies to extract Arrhenius parameters. This simulation is best used with teacher guidance because it presents an analogy of chemical reactions.

Click to view content (https://openstax.org/l/21reversereact).

15.7 Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify probabilities in entropy.
- Analyze statistical probabilities in entropic systems.

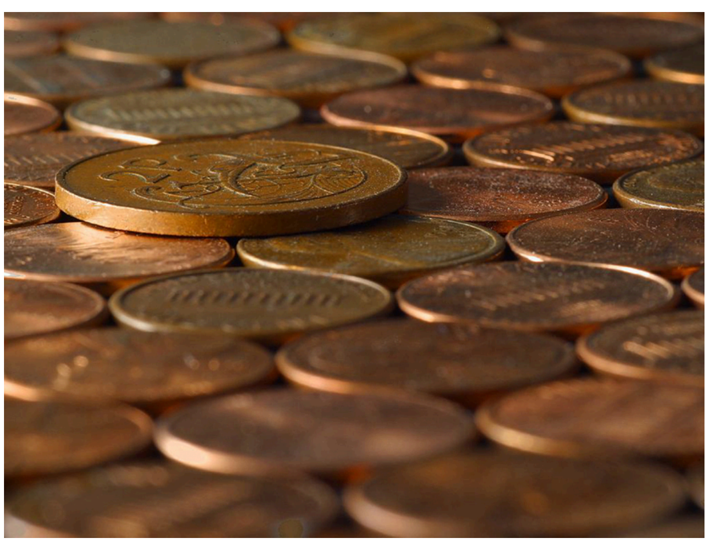


FIGURE 15.37 When you toss a coin a large number of times, heads and tails tend to come up in roughly equal numbers. Why doesn't heads come up 100, 90, or even 80% of the time? (credit: Jon Sullivan, PDPhoto.org)

The various ways of formulating the second law of thermodynamics tell what happens rather than why it happens.

Why should heat transfer occur only from hot to cold? Why should energy become ever less available to do work? Why should the universe become increasingly disorderly? The answer is that it is a matter of overwhelming probability. Disorder is simply vastly more likely than order.

When you watch an emerging rain storm begin to wet the ground, you will notice that the drops fall in a disorganized manner both in time and in space. Some fall close together, some far apart, but they never fall in straight, orderly rows. It is not impossible for rain to fall in an orderly pattern, just highly unlikely, because there are many more disorderly ways than orderly ones. To illustrate this fact, we will examine some random processes, starting with coin tosses.

Coin Tosses

What are the possible outcomes of tossing 5 coins? Each coin can land either heads or tails. On the large scale, we are concerned only with the total heads and tails and not with the order in which heads and tails appear. The following possibilities exist:

$$\begin{gathered}\text{5 heads, 0 tails}\\ \text{4 heads, 1 tail}\\ \text{3 heads, 2 tails}\\ \text{2 heads, 3 tails}\\ \text{1 head, 4 tails}\\ \text{0 head, 5 tails}\end{gathered} \tag{15.67}$$

These are what we call macrostates. A **macrostate** is an overall property of a system. It does not specify the details of the system, such as the order in which heads and tails occur or which coins are heads or tails.

Using this nomenclature, a system of 5 coins has the 6 possible macrostates just listed. Some macrostates are more likely to occur than others. For instance, there is only one way to get 5 heads, but there are several ways to get 3 heads and 2 tails, making the latter macrostate more probable. Table 15.3 lists of all the ways in which 5 coins can be tossed, taking into account the order in which heads and tails occur. Each sequence is called a **microstate**—a detailed description of every element of a system.

	Individual microstates	Number of microstates
5 heads, 0 tails	HHHHH	1
4 heads, 1 tail	HHHHT, HHHTH, HHTHH, HTHHH, THHHH	5
3 heads, 2 tails	HTHTH, THTHH, HTHHT, THHTH, THHHT HTHTH, THTHH, HTHHT, THHTH, THHHT	10
2 heads, 3 tails	TTTHH, TTHHT, THHTT, HHTTT, TTHTH, THTHT, HTHTT, THTTH, HTTHT, HTTTH	10
1 head, 4 tails	TTTTH, TTTHT, TTHTT, THTTT, HTTTT	5

TABLE 15.3 5-Coin Toss

	Individual microstates	Number of microstates
0 heads, 5 tails	TTTTT	1
		Total: 32

TABLE 15.3 5-Coin Toss

The macrostate of 3 heads and 2 tails can be achieved in 10 ways and is thus 10 times more probable than the one having 5 heads. Not surprisingly, it is equally probable to have the reverse, 2 heads and 3 tails. Similarly, it is equally probable to get 5 tails as it is to get 5 heads. Note that all of these conclusions are based on the crucial assumption that each microstate is equally probable. With coin tosses, this requires that the coins not be asymmetric in a way that favors one side over the other, as with loaded dice. With any system, the assumption that all microstates are equally probable must be valid, or the analysis will be erroneous.

The two most orderly possibilities are 5 heads or 5 tails. (They are more structured than the others.) They are also the least likely, only 2 out of 32 possibilities. The most disorderly possibilities are 3 heads and 2 tails and its reverse. (They are the least structured.) The most disorderly possibilities are also the most likely, with 20 out of 32 possibilities for the 3 heads and 2 tails and its reverse. If we start with an orderly array like 5 heads and toss the coins, it is very likely that we will get a less orderly array as a result, since 30 out of the 32 possibilities are less orderly. So even if you start with an orderly state, there is a strong tendency to go from order to disorder, from low entropy to high entropy. The reverse can happen, but it is unlikely.

Macrostate		Number of microstates
Heads	Tails	(*W*)
100	0	1
99	1	1.0×10^{2}
95	5	7.5×10^{7}
90	10	1.7×10^{13}
75	25	2.4×10^{23}
60	40	1.4×10^{28}
55	45	6.1×10^{28}
51	49	9.9×10^{28}
50	50	1.0×10^{29}

TABLE 15.4 100-Coin Toss

Macrostate		Number of microstates
49	51	9.9×10^{28}
45	55	6.1×10^{28}
40	60	1.4×10^{28}
25	75	2.4×10^{23}
10	90	1.7×10^{13}
5	95	7.5×10^{7}
1	99	1.0×10^{2}
0	100	1
		Total: 1.27×10^{30}

TABLE 15.4 100-Coin Toss

This result becomes dramatic for larger systems. Consider what happens if you have 100 coins instead of just 5. The most orderly arrangements (most structured) are 100 heads or 100 tails. The least orderly (least structured) is that of 50 heads and 50 tails. There is only 1 way (1 microstate) to get the most orderly arrangement of 100 heads. There are 100 ways (100 microstates) to get the next most orderly arrangement of 99 heads and 1 tail (also 100 to get its reverse). And there are 1.0×10^{29} ways to get 50 heads and 50 tails, the least orderly arrangement. Table 15.4 is an abbreviated list of the various macrostates and the number of microstates for each macrostate. The total number of microstates—the total number of different ways 100 coins can be tossed—is an impressively large 1.27×10^{30}. Now, if we start with an orderly macrostate like 100 heads and toss the coins, there is a virtual certainty that we will get a less orderly macrostate. If we keep tossing the coins, it is possible, but exceedingly unlikely, that we will ever get back to the most orderly macrostate. If you tossed the coins once each second, you could expect to get either 100 heads or 100 tails once in 2×10^{22} years! This period is 1 trillion (10^{12}) times longer than the age of the universe, and so the chances are essentially zero. In contrast, there is an 8% chance of getting 50 heads, a 73% chance of getting from 45 to 55 heads, and a 96% chance of getting from 40 to 60 heads. Disorder is highly likely.

Disorder in a Gas

The fantastic growth in the odds favoring disorder that we see in going from 5 to 100 coins continues as the number of entities in the system increases. Let us now imagine applying this approach to perhaps a small sample of gas. Because counting microstates and macrostates involves statistics, this is called **statistical analysis**. The macrostates of a gas correspond to its macroscopic properties, such as volume, temperature, and pressure; and its microstates correspond to the detailed description of the positions and velocities of its atoms. Even a small amount of gas has a huge number of atoms: $1.0\ \text{cm}^3$ of an ideal gas at 1.0 atm and 0° C has 2.7×10^{19} atoms. So each macrostate has an immense number of microstates. In plain language, this means that there are an immense number of ways in which the atoms in a gas can be arranged, while still having the same pressure, temperature, and so on.

The most likely conditions (or macrostates) for a gas are those we see all the time—a random distribution of atoms in space with a Maxwell-Boltzmann distribution of speeds in random directions, as predicted by kinetic theory. This is the most disorderly and least structured condition we can imagine. In contrast, one type of very orderly and structured macrostate has all of the atoms in one corner of a container with identical velocities. There are very few ways to accomplish this (very few microstates corresponding to it), and so it is exceedingly unlikely ever to occur. (See Figure 15.38(b).) Indeed, it is so unlikely that we have a law saying that it is impossible, which has never been observed to be violated—the second law of thermodynamics.

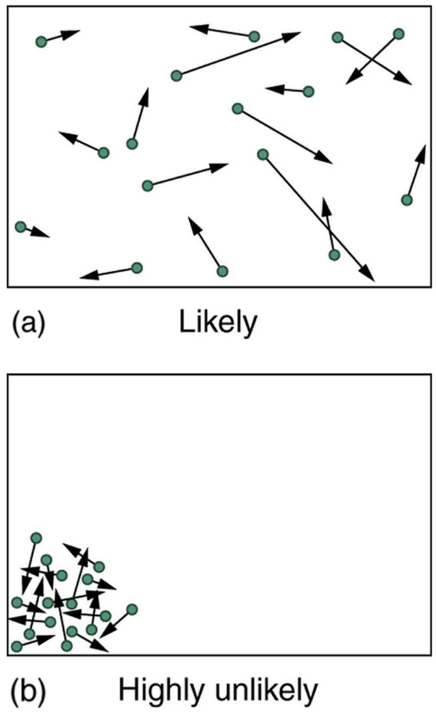


FIGURE 15.38 (a) The ordinary state of gas in a container is a disorderly, random distribution of atoms or molecules with a Maxwell-Boltzmann distribution of speeds. It is so unlikely that these atoms or molecules would ever end up in one corner of the container that it might as well be impossible. (b) With energy transfer, the gas can be forced into one corner and its entropy greatly reduced. But left alone, it will spontaneously increase its entropy and return to the normal conditions, because they are immensely more likely.

The disordered condition is one of high entropy, and the ordered one has low entropy. With a transfer of energy from another system, we could force all of the atoms into one corner and have a local decrease in entropy, but at the cost of an overall increase in entropy of the universe. If the atoms start out in one corner, they will quickly disperse and become uniformly distributed and will never return to the orderly original state (Figure 15.38(b)). Entropy will increase. With such a large sample of atoms, it is possible—but unimaginably unlikely—for entropy to decrease. Disorder is vastly more likely than order.

The arguments that disorder and high entropy are the most probable states are quite convincing. The great Austrian physicist Ludwig Boltzmann (1844–1906)—who, along with Maxwell, made so many contributions to kinetic theory—proved that the entropy of a system in a given state (a macrostate) can be written as

$$S = k \ln W, \quad 15.68$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate. W is proportional to the probability that the macrostate will occur. Thus entropy is directly related to the probability of a state—the more likely the state, the greater its entropy. Boltzmann proved that this expression for s is equivalent to the definition $\Delta S = Q/T$, which we have used extensively.

Thus the second law of thermodynamics is explained on a very basic level: entropy either remains the same or increases in every process. This phenomenon is due to the extraordinarily small probability of a decrease, based on the extraordinarily larger number of microstates in systems with greater entropy. Entropy *can* decrease, but for any macroscopic system, this outcome is so unlikely that it will never be observed.

EXAMPLE 15.9

Entropy Increases in a Coin Toss

Suppose you toss 100 coins starting with 60 heads and 40 tails, and you get the most likely result, 50 heads and 50 tails. What is the change in entropy?

Strategy

Noting that the number of microstates is labeled W in Table 15.4 for the 100-coin toss, we can use $\Delta S = S_\text{f} - S_\text{i} = k\ln W_\text{f} - k\ln W_\text{i}$ to calculate the change in entropy.

Solution

The change in entropy is

$$\Delta S = S_\text{f} - S_\text{i} = k\ln W_\text{f} - k\ln W_\text{i}, \tag{15.69}$$

where the subscript i stands for the initial 60 heads and 40 tails state, and the subscript f for the final 50 heads and 50 tails state. Substituting the values for W from Table 15.4 gives

$$\begin{aligned} \Delta S &= (1.38 \times 10^{-23}\ \text{J/K})[\ln(1.0 \times 10^{29}) - \ln(1.4 \times 10^{28})] \\ &= 2.7 \times 10^{-23}\ \text{J/K} \end{aligned} \tag{15.70}$$

Discussion

This increase in entropy means we have moved to a less orderly situation. It is not impossible for further tosses to produce the initial state of 60 heads and 40 tails, but it is less likely. There is about a 1 in 90 chance for that decrease in entropy (-2.7×10^{-23} J/K) to occur. If we calculate the decrease in entropy to move to the most orderly state, we get $\Delta S = -92 \times 10^{-23}$ J/K. There is about a $1 \text{ in } 10^{30}$ chance of this change occurring. So while very small decreases in entropy are unlikely, slightly greater decreases are impossibly unlikely. These probabilities imply, again, that for a macroscopic system, a decrease in entropy is impossible. For example, for heat transfer to occur spontaneously from 1.00 kg of 0°C ice to its 0°C environment, there would be a decrease in entropy of 1.22×10^3 J/K. Given that a ΔS of 10^{-21} J/K corresponds to about a $1 \text{ in } 10^{30}$ chance, a decrease of this size (10^3 J/K) is an *utter* impossibility. Even for a milligram of melted ice to spontaneously refreeze is impossible.

Problem-Solving Strategies for Entropy

1. *Examine the situation to determine if entropy is involved.*
2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful.
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).* You must carefully identify the heat transfer, if any, and the temperature at which the process takes place. It is also important to identify the initial and final states.
5. *Solve the appropriate equation for the quantity to be determined (the unknown).* Note that the change in entropy can be determined between any states by calculating it for a reversible process.
6. *Substitute the known value along with their units into the appropriate equation, and obtain numerical solutions complete with units.*
7. *To see if it is reasonable: Does it make sense?* For example, total entropy should increase for any real process or be constant for a reversible process. Disordered states should be more probable and have greater entropy than ordered states.

Glossary

adiabatic process a process in which no heat transfer takes place

Carnot cycle a cyclical process that uses only reversible processes, the adiabatic and isothermal processes

Carnot efficiency the maximum theoretical efficiency for a heat engine

Carnot engine a heat engine that uses a Carnot cycle

change in entropy the ratio of heat transfer to temperature Q/T

coefficient of performance for a heat pump, it is the ratio of heat transfer at the output (the hot reservoir) to the work supplied; for a refrigerator or air conditioner, it is the ratio of heat transfer from the cold reservoir to the work supplied

cyclical process a process in which the path returns to its original state at the end of every cycle

entropy a measurement of a system's disorder and its inability to do work in a system

first law of thermodynamics states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system

heat engine a machine that uses heat transfer to do work

heat pump a machine that generates heat transfer from cold to hot

human metabolism conversion of food into heat transfer, work, and stored fat

internal energy the sum of the kinetic and potential energies of a system's atoms and molecules

irreversible process any process that depends on path direction

isobaric process constant-pressure process in which a gas does work

isochoric process a constant-volume process

isothermal process a constant-temperature process

macrostate an overall property of a system

microstate each sequence within a larger macrostate

Otto cycle a thermodynamic cycle, consisting of a pair of adiabatic processes and a pair of isochoric processes, that converts heat into work, e.g., the four-stroke engine cycle of intake, compression, ignition, and exhaust

reversible process a process in which both the heat engine system and the external environment theoretically can be returned to their original states

second law of thermodynamics heat transfer flows from a hotter to a cooler object, never the reverse, and some heat energy in any process is lost to available work in a cyclical process

second law of thermodynamics stated in terms of entropy the total entropy of a system either increases or remains constant; it never decreases

statistical analysis using statistics to examine data, such as counting microstates and macrostates

Section Summary

15.1 The First Law of Thermodynamics

- The first law of thermodynamics is given as $\Delta U = Q - W$, where ΔU is the change in internal energy of a system, Q is the net heat transfer (the sum of all heat transfer into and out of the system), and W is the net work done (the sum of all work done on or by the system).
- Both Q and W are energy in transit; only ΔU represents an independent quantity capable of being stored.
- The internal energy U of a system depends only on the state of the system and not how it reached that state.
- Metabolism of living organisms, and photosynthesis of plants, are specialized types of heat transfer, doing work, and internal energy of systems.

15.2 The First Law of Thermodynamics and Some Simple Processes

- One of the important implications of the first law of thermodynamics is that machines can be harnessed to do work that humans previously did by hand or by external energy supplies such as running water or the heat of the Sun. A machine that uses heat transfer to do work is known as a heat engine.
- There are several simple processes, used by heat engines, that flow from the first law of thermodynamics. Among them are the isobaric, isochoric, isothermal and adiabatic processes.
- These processes differ from one another based on how they affect pressure, volume, temperature, and heat transfer.
- If the work done is performed on the outside environment, work (W) will be a positive value. If the work done is done to the heat engine system, work (W) will be a negative value.
- Some thermodynamic processes, including isothermal and adiabatic processes, are reversible in theory; that is, both the thermodynamic system and the environment can be returned to their initial states. However, because of loss of energy

owing to the second law of thermodynamics, complete reversibility does not work in practice.

15.3 Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

- The two expressions of the second law of thermodynamics are: (i) Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction; and (ii) It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.
- Irreversible processes depend on path and do not return to their original state. Cyclical processes are processes that return to their original state at the end of every cycle.
- In a cyclical process, such as a heat engine, the net work done by the system equals the net heat transfer into the system, or $W = Q_{\mathrm{h}} - Q_{\mathrm{c}}$, where Q_{h} is the heat transfer from the hot object (hot reservoir), and Q_{c} is the heat transfer into the cold object (cold reservoir).
- Efficiency can be expressed as $Eff = \frac{W}{Q_{\mathrm{h}}}$, the ratio of work output divided by the amount of energy input.
- The four-stroke gasoline engine is often explained in terms of the Otto cycle, which is a repeating sequence of processes that convert heat into work.

15.4 Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated

- The Carnot cycle is a theoretical cycle that is the most efficient cyclical process possible. Any engine using the Carnot cycle, which uses only reversible processes (adiabatic and isothermal), is known as a Carnot engine.
- Any engine that uses the Carnot cycle enjoys the maximum theoretical efficiency.
- While Carnot engines are ideal engines, in reality, no engine achieves Carnot's theoretical maximum efficiency, since dissipative processes, such as friction, play a role. Carnot cycles without heat loss may be possible at absolute zero, but this has never been seen in nature.

15.5 Applications of Thermodynamics: Heat Pumps and Refrigerators

- An artifact of the second law of thermodynamics is the ability to heat an interior space using a heat pump. Heat pumps compress cold ambient air and, in so doing, heat it to room temperature without violation of conservation principles.
- To calculate the heat pump's coefficient of performance, use the equation $COP_{\mathrm{hp}} = \frac{Q_{\mathrm{h}}}{W}$.
- A refrigerator is a heat pump; it takes warm ambient air and expands it to chill it.

15.6 Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

- Entropy is the loss of energy available to do work.
- Another form of the second law of thermodynamics states that the total entropy of a system either increases or remains constant; it never decreases.
- Entropy is zero in a reversible process; it increases in an irreversible process.
- The ultimate fate of the universe is likely to be thermodynamic equilibrium, where the universal temperature is constant and no energy is available to do work.
- Entropy is also associated with the tendency toward disorder in a closed system.

15.7 Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

- Disorder is far more likely than order, which can be seen statistically.
- The entropy of a system in a given state (a macrostate) can be written as
  $S = k \ln W$,
  where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate.

Conceptual Questions

15.1 The First Law of Thermodynamics

1. Describe the photo of the tea kettle at the beginning of this section in terms of heat transfer, work done, and internal energy. How is heat being transferred? What is the work done and what is doing it? How does the kettle maintain its internal energy?
2. The first law of thermodynamics and the conservation of energy, as discussed in Conservation of Energy, are clearly related. How do they differ in the types of energy considered?
3. Heat transfer Q and work done W are always energy in transit, whereas internal energy U is energy stored in a system. Give an example of each type of energy, and state specifically how it is either in transit or resides in a system.
4. How do heat transfer and internal energy differ? In particular, which can be stored as such in a system and which cannot?
5. If you run down some stairs and stop, what happens to your kinetic energy and your initial gravitational potential energy?
6. Give an explanation of how food energy (calories) can be viewed as molecular potential energy (consistent with the atomic and molecular definition of internal energy).
7. Identify the type of energy transferred to your body in each of the following as either internal energy, heat transfer, or doing work: (a) basking in sunlight; (b) eating food; (c) riding an elevator to a higher floor.

15.2 The First Law of Thermodynamics and Some Simple Processes

8. A great deal of effort, time, and money has been spent in the quest for the so-called perpetual-motion machine, which is defined as a hypothetical machine that operates or produces useful work indefinitely and/or a hypothetical machine that produces more work or energy than it consumes. Explain, in terms of heat engines and the first law of thermodynamics, why or why not such a machine is likely to be constructed.
9. One method of converting heat transfer into doing work is for heat transfer into a gas to take place, which expands, doing work on a piston, as shown in the figure below. (a) Is the heat transfer converted directly to work in an isobaric process, or does it go through another form first? Explain your answer. (b) What about in an isothermal process? (c) What about in an adiabatic process (where heat transfer occurred prior to the adiabatic process)?

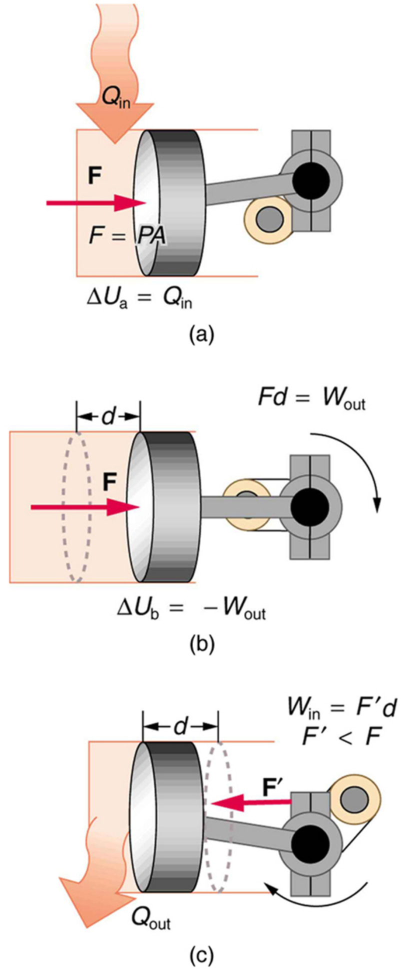


FIGURE 15.39

10. Would the previous question make any sense for an isochoric process? Explain your answer.
11. We ordinarily say that $\Delta U = 0$ for an isothermal process. Does this assume no phase change takes place? Explain your answer.
12. The temperature of a rapidly expanding gas decreases. Explain why in terms of the first law of thermodynamics. (Hint: Consider whether the gas does work and whether heat transfer occurs rapidly into the gas through conduction.)

13. Which cyclical process represented by the two closed loops, ABCFA and ABDEA, on the PV diagram in the figure below produces the greatest *net* work? Is that process also the one with the smallest work input required to return it to point A? Explain your responses.

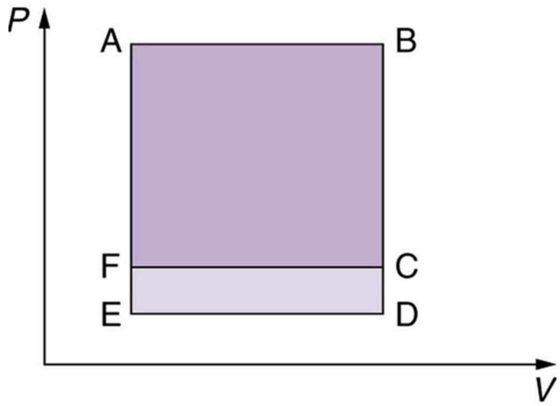


FIGURE 15.40 The two cyclical processes shown on this PV diagram start with and return the system to the conditions at point A, but they follow different paths and produce different amounts of work.

14. A real process may be nearly adiabatic if it occurs over a very short time. How does the short time span help the process to be adiabatic?

15. It is unlikely that a process can be isothermal unless it is a very slow process. Explain why. Is the same true for isobaric and isochoric processes? Explain your answer.

15.3 Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

16. Imagine you are driving a car up Pike's Peak in Colorado. To raise a car weighing 1000 kilograms a distance of 100 meters would require about a million joules. You could raise a car 12.5 kilometers with the energy in a gallon of gas. Driving up Pike's Peak (a mere 3000-meter climb) should consume a little less than a quart of gas. But other considerations have to be taken into account. Explain, in terms of efficiency, what factors may keep you from realizing your ideal energy use on this trip.

17. Is a temperature difference necessary to operate a heat engine? State why or why not.

18. Definitions of efficiency vary depending on how energy is being converted. Compare the definitions of efficiency for the human body and heat engines. How does the definition of efficiency in each relate to the type of energy being converted into doing work?

19. Why—other than the fact that the second law of thermodynamics says reversible engines are the most efficient—should heat engines employing reversible processes be more efficient than those employing irreversible processes? Consider that dissipative mechanisms are one cause of irreversibility.

15.4 Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated

20. Think about the drinking bird at the beginning of this section (Figure 15.20). Although the bird enjoys the theoretical maximum efficiency possible, if left to its own devices over time, the bird will cease "drinking." What are some of the dissipative processes that might cause the bird's motion to cease?

21. Can improved engineering and materials be employed in heat engines to reduce heat transfer into the environment? Can they eliminate heat transfer into the environment entirely?

22. Does the second law of thermodynamics alter the conservation of energy principle?

15.5 Applications of Thermodynamics: Heat Pumps and Refrigerators

23. Explain why heat pumps do not work as well in very cold climates as they do in milder ones. Is the same true of refrigerators?

24. In some Northern European nations, homes are being built without heating systems of any type. They are very well insulated and are kept warm by the body heat of the residents. However, when the residents are not at home, it is still warm in these houses. What is a possible explanation?

25. Why do refrigerators, air conditioners, and heat pumps operate most cost-effectively for cycles with a small difference between T_h and T_c? (Note that the temperatures of the cycle employed are crucial to its COP.)

26. Grocery store managers contend that there is *less* total energy consumption in the summer if the store is kept at a *low* temperature. Make arguments to support or refute this claim, taking into account that there are numerous refrigerators and freezers in the store.

27. Can you cool a kitchen by leaving the refrigerator door open?

15.6 Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

28. A woman shuts her summer cottage up in September and returns in June. No one has entered the cottage in the meantime. Explain what she is likely to find, in terms of the second law of thermodynamics.
29. Consider a system with a certain energy content, from which we wish to extract as much work as possible. Should the system's entropy be high or low? Is this orderly or disorderly? Structured or uniform? Explain briefly.
30. Does a gas become more orderly when it liquefies? Does its entropy change? If so, does the entropy increase or decrease? Explain your answer.
31. Explain how water's entropy can decrease when it freezes without violating the second law of thermodynamics. Specifically, explain what happens to the entropy of its surroundings.
32. Is a uniform-temperature gas more or less orderly than one with several different temperatures? Which is more structured? In which can heat transfer result in work done without heat transfer from another system?
33. Give an example of a spontaneous process in which a system becomes less ordered and energy becomes less available to do work. What happens to the system's entropy in this process?
34. What is the change in entropy in an adiabatic process? Does this imply that adiabatic processes are reversible? Can a process be precisely adiabatic for a macroscopic system?
35. Does the entropy of a star increase or decrease as it radiates? Does the entropy of the space into which it radiates (which has a temperature of about 3 K) increase or decrease? What does this do to the entropy of the universe?
36. Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

15.7 Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

37. Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problems & Exercises

15.1 The First Law of Thermodynamics

1. What is the change in internal energy of a car if you put 12.0 gal of gasoline into its tank? The energy content of gasoline is $1.3 \times 10^8\ \text{J/gal}$. All other factors, such as the car's temperature, are constant.
2. How much heat transfer occurs from a system, if its internal energy decreased by 150 J while it was doing 30.0 J of work?
3. A system does $1.80 \times 10^8\ \text{J}$ of work while $7.50 \times 10^8\ \text{J}$ of heat transfer occurs to the environment. What is the change in internal energy of the system assuming no other changes (such as in temperature or by the addition of fuel)?
4. What is the change in internal energy of a system which does $4.50 \times 10^5\ \text{J}$ of work while $3.00 \times 10^6\ \text{J}$ of heat transfer occurs into the system, and $8.00 \times 10^6\ \text{J}$ of heat transfer occurs to the environment?
5. Suppose a woman does 500 J of work and 9500 J of heat transfer occurs into the environment in the process. (a) What is the decrease in her internal energy, assuming no change in temperature or consumption of food? (That is, there is no other energy transfer.) (b) What is her efficiency?
6. (a) How much food energy will a man metabolize in the process of doing 35.0 kJ of work with an efficiency of 5.00%? (b) How much heat transfer occurs to the environment to keep his temperature constant? Explicitly show how you follow the steps in the Problem-Solving Strategy for thermodynamics found in Problem-Solving Strategies for Thermodynamics.
7. (a) What is the average metabolic rate in watts of a man who metabolizes 10,500 kJ of food energy in one day? (b) What is the maximum amount of work in joules he can do without breaking down fat, assuming a maximum efficiency of 20.0%? (c) Compare his work output with the daily output of a 187-W (0.250-horsepower) motor.

8. (a) How long will the energy in a 1470-kJ (350-kcal) cup of yogurt last in a woman doing work at the rate of 150 W with an efficiency of 20.0% (such as in leisurely climbing stairs)? (b) Does the time found in part (a) imply that it is easy to consume more food energy than you can reasonably expect to work off with exercise?
9. (a) A woman climbing the Washington Monument metabolizes $6.00 \times 10^2\ \text{kJ}$ of food energy. If her efficiency is 18.0%, how much heat transfer occurs to the environment to keep her temperature constant? (b) Discuss the amount of heat transfer found in (a). Is it consistent with the fact that you quickly warm up when exercising?

15.2 The First Law of Thermodynamics and Some Simple Processes

10. A car tire contains $0.0380\ \text{m}^3$ of nitrogen at a gauge pressure of $2.20 \times 10^5\ \text{N/m}^2$ (about 32 psi). How much more internal energy does this gas have than the same volume has at zero gauge pressure (which is equivalent to normal atmospheric pressure)? (Use $U = \frac{5}{2}NkT$ for the internal energy of nitrogen, a diatomic gas.)
11. A helium-filled rigid ball has a gauge pressure of 0.200 atm and a volume of 10.0 L. How much greater is the internal energy of the helium in the ball than it would be at zero gauge pressure?
12. Steam to drive an old-fashioned steam locomotive is supplied at a constant gauge pressure of $1.75 \times 10^6\ \text{N/m}^2$ (about 250 psi) to a piston with a 0.200-m radius. (a) By calculating $P\Delta V$, find the work done by the steam when the piston moves 0.800 m. Note that this is the net work output, since gauge pressure is used. (b) Now find the amount of work by calculating the force exerted times the distance traveled. Is the answer the same as in part (a)?
13. A hand-driven tire pump has a piston with a 2.50-cm diameter and a maximum stroke of 30.0 cm. (a) How much work do you do in one stroke if the average gauge pressure is $2.40 \times 10^5\ \text{N/m}^2$ (about 35 psi)? (b) What average force do you exert on the piston, neglecting friction and gravitational force?
14. Calculate the net work output of a heat engine following path ABCDA in the figure below.

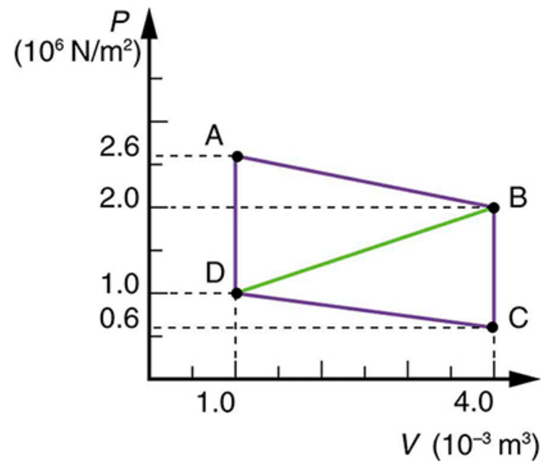


FIGURE 15.41

15. What is the net work output of a heat engine that follows path ABDA in the figure above, with a straight line from B to D? Why is the work output less than for path ABCDA? Explicitly show how you follow the steps in the Problem-Solving Strategies for Thermodynamics.
16. Unreasonable Results
What is wrong with the claim that a cyclical heat engine does 4.00 kJ of work on an input of 24.0 kJ of heat transfer while 16.0 kJ of heat transfers to the environment?
17. (a) A cyclical heat engine, operating between temperatures of $450^\circ\ \text{C}$ and $150^\circ\ \text{C}$ produces 4.00 MJ of work on a heat transfer of 5.00 MJ into the engine. How much heat transfer occurs to the environment? (b) What is unreasonable about the engine? (c) Which premise is unreasonable?
18. Construct Your Own Problem
Consider a car's gasoline engine. Construct a problem in which you calculate the maximum efficiency this engine can have. Among the things to consider are the effective hot and cold reservoir temperatures. Compare your calculated efficiency with the actual efficiency of car engines.
19. Construct Your Own Problem
Consider a car trip into the mountains. Construct a problem in which you calculate the overall efficiency of the car for the trip as a ratio of kinetic and potential energy gained to fuel consumed. Compare this efficiency to the thermodynamic efficiency quoted for gasoline engines and discuss why the thermodynamic efficiency is so much greater. Among the factors to be considered are the gain in altitude and speed, the mass of the car, the distance traveled, and typical fuel economy.

15.3 Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

20. A certain heat engine does 10.0 kJ of work and 8.50 kJ of heat transfer occurs to the environment in a cyclical process. (a) What was the heat transfer into this engine? (b) What was the engine's efficiency?

21. With $2.56 \times 10^{6}\text{ J}$ of heat transfer into this engine, a given cyclical heat engine can do only $1.50 \times 10^{5}\text{ J}$ of work. (a) What is the engine's efficiency? (b) How much heat transfer to the environment takes place?

22. (a) What is the work output of a cyclical heat engine having a 22.0% efficiency and $6.00 \times 10^{9}\text{ J}$ of heat transfer into the engine? (b) How much heat transfer occurs to the environment?

23. (a) What is the efficiency of a cyclical heat engine in which 75.0 kJ of heat transfer occurs to the environment for every 95.0 kJ of heat transfer into the engine? (b) How much work does it produce for 100 kJ of heat transfer into the engine?

24. The engine of a large ship does $2.00 \times 10^{8}\text{ J}$ of work with an efficiency of 5.00%. (a) How much heat transfer occurs to the environment? (b) How many barrels of fuel are consumed, if each barrel produces $6.00 \times 10^{9}\text{ J}$ of heat transfer when burned?

25. (a) How much heat transfer occurs to the environment by an electrical power station that uses $1.25 \times 10^{14}\text{ J}$ of heat transfer into the engine with an efficiency of 42.0%? (b) What is the ratio of heat transfer to the environment to work output? (c) How much work is done?

26. Assume that the turbines at a coal-powered power plant were upgraded, resulting in an improvement in efficiency of 3.32%. Assume that prior to the upgrade the power station had an efficiency of 36% and that the heat transfer into the engine in one day is still the same at $2.50 \times 10^{14}\text{ J}$. (a) How much more electrical energy is produced due to the upgrade? (b) How much less heat transfer occurs to the environment due to the upgrade?

27. This problem compares the energy output and heat transfer to the environment by two different types of nuclear power stations—one with the normal efficiency of 34.0%, and another with an improved efficiency of 40.0%. Suppose both have the same heat transfer into the engine in one day, $2.50 \times 10^{14}\text{ J}$. (a) How much more electrical energy is produced by the more efficient power station? (b) How much less heat transfer occurs to the environment by the more efficient power station? (One type of more efficient nuclear power station, the gas-cooled reactor, has not been reliable enough to be economically feasible in spite of its greater efficiency.)

15.4 Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated

28. A certain gasoline engine has an efficiency of 30.0%. What would the hot reservoir temperature be for a Carnot engine having that efficiency, if it operates with a cold reservoir temperature of 200°C?

29. A gas-cooled nuclear reactor operates between hot and cold reservoir temperatures of 700°C and 27.0°C. (a) What is the maximum efficiency of a heat engine operating between these temperatures? (b) Find the ratio of this efficiency to the Carnot efficiency of a standard nuclear reactor (found in Example 15.4).

30. (a) What is the hot reservoir temperature of a Carnot engine that has an efficiency of 42.0% and a cold reservoir temperature of 27.0°C? (b) What must the hot reservoir temperature be for a real heat engine that achieves 0.700 of the maximum efficiency, but still has an efficiency of 42.0% (and a cold reservoir at 27.0°C)? (c) Does your answer imply practical limits to the efficiency of car gasoline engines?

31. Steam locomotives have an efficiency of 17.0% and operate with a hot steam temperature of 425°C. (a) What would the cold reservoir temperature be if this were a Carnot engine? (b) What would the maximum efficiency of this steam engine be if its cold reservoir temperature were 150°C?

32. Practical steam engines utilize $450^\circ C$ steam, which is later exhausted at $270^\circ C$. (a) What is the maximum efficiency that such a heat engine can have? (b) Since $270^\circ C$ steam is still quite hot, a second steam engine is sometimes operated using the exhaust of the first. What is the maximum efficiency of the second engine if its exhaust has a temperature of $150^\circ C$? (c) What is the overall efficiency of the two engines? (d) Show that this is the same efficiency as a single Carnot engine operating between $450^\circ C$ and $150^\circ C$. Explicitly show how you follow the steps in the Problem-Solving Strategies for Thermodynamics.
33. A coal-fired electrical power station has an efficiency of 38%. The temperature of the steam leaving the boiler is $550^\circ C$. What percentage of the maximum efficiency does this station obtain? (Assume the temperature of the environment is $20^\circ C$.)
34. Would you be willing to financially back an inventor who is marketing a device that she claims has 25 kJ of heat transfer at 600 K, has heat transfer to the environment at 300 K, and does 12 kJ of work? Explain your answer.
35. Unreasonable Results
(a) Suppose you want to design a steam engine that has heat transfer to the environment at $270^\circ C$ and has a Carnot efficiency of 0.800. What temperature of hot steam must you use? (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?
36. Unreasonable Results
Calculate the cold reservoir temperature of a steam engine that uses hot steam at $450^\circ C$ and has a Carnot efficiency of 0.700. (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

15.5 Applications of Thermodynamics: Heat Pumps and Refrigerators

37. What is the coefficient of performance of an ideal heat pump that has heat transfer from a cold temperature of $-25.0^\circ C$ to a hot temperature of $40.0^\circ C$?
38. Suppose you have an ideal refrigerator that cools an environment at $-20.0^\circ C$ and has heat transfer to another environment at $50.0^\circ C$. What is its coefficient of performance?
39. What is the best coefficient of performance possible for a hypothetical refrigerator that could make liquid nitrogen at $-200^\circ C$ and has heat transfer to the environment at $35.0^\circ C$?
40. In a very mild winter climate, a heat pump has heat transfer from an environment at $5.00^\circ C$ to one at $35.0^\circ C$. What is the best possible coefficient of performance for these temperatures? Explicitly show how you follow the steps in the Problem-Solving Strategies for Thermodynamics.
41. (a) What is the best coefficient of performance for a heat pump that has a hot reservoir temperature of $50.0^\circ C$ and a cold reservoir temperature of $-20.0^\circ C$? (b) How much heat transfer occurs into the warm environment if 3.60×10^7 J of work $(10.0 \text{kW} \cdot \text{h})$ is put into it? (c) If the cost of this work input is $10.0 \text{ cents/kW} \cdot \text{h}$, how does its cost compare with the direct heat transfer achieved by burning natural gas at a cost of 85.0 cents per therm. (A therm is a common unit of energy for natural gas and equals 1.055×10^8 J.)
42. (a) What is the best coefficient of performance for a refrigerator that cools an environment at $-30.0^\circ C$ and has heat transfer to another environment at $45.0^\circ C$? (b) How much work in joules must be done for a heat transfer of 4186 kJ from the cold environment? (c) What is the cost of doing this if the work costs 10.0 cents per 3.60×10^6 J (a kilowatt-hour)? (d) How many kJ of heat transfer occurs into the warm environment? (e) Discuss what type of refrigerator might operate between these temperatures.
43. Suppose you want to operate an ideal refrigerator with a cold temperature of $-10.0^\circ C$, and you would like it to have a coefficient of performance of 7.00. What is the hot reservoir temperature for such a refrigerator?
44. An ideal heat pump is being considered for use in heating an environment with a temperature of $22.0^\circ C$. What is the cold reservoir temperature if the pump is to have a coefficient of performance of 12.0?
45. A 4-ton air conditioner removes 5.06×10^7 J (48,000 British thermal units) from a cold environment in 1.00 h. (a) What energy input in joules is necessary to do this if the air conditioner has an energy efficiency rating (EER) of 12.0? (b) What is the cost of doing this if the work costs 10.0 cents per 3.60×10^6 J (one kilowatt-hour)?
(c) Discuss whether this cost seems realistic. Note that the energy efficiency rating (EER) of an air conditioner or refrigerator is defined to be the number of British thermal units of heat transfer from a cold environment per hour divided by the watts of power input.

46. Show that the coefficients of performance of refrigerators and heat pumps are related by $COP_{\text{ref}} = COP_{\text{hp}} - 1$.
Start with the definitions of the COPs and the conservation of energy relationship between Q_{h}, Q_{c}, and W.

15.6 Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

47. (a) On a winter day, a certain house loses 5.00×10^{8} J of heat to the outside (about 500,000 Btu). What is the total change in entropy due to this heat transfer alone, assuming an average indoor temperature of 21.0° C and an average outdoor temperature of 5.00° C? (b) This large change in entropy implies a large amount of energy has become unavailable to do work. Where do we find more energy when such energy is lost to us?

48. On a hot summer day, 4.00×10^{6} J of heat transfer into a parked car takes place, increasing its temperature from 35.0° C to 45.0° C. What is the increase in entropy of the car due to this heat transfer alone?

49. A hot rock ejected from a volcano's lava fountain cools from 1100° C to 40.0° C, and its entropy decreases by 950 J/K. How much heat transfer occurs from the rock?

50. When 1.60×10^{5} J of heat transfer occurs into a meat pie initially at 20.0° C, its entropy increases by 480 J/K. What is its final temperature?

51. The Sun radiates energy at the rate of 3.80×10^{26} W from its 5500° C surface into dark empty space (a negligible fraction radiates onto Earth and the other planets). The effective temperature of deep space is -270° C. (a) What is the increase in entropy in one day due to this heat transfer? (b) How much work is made unavailable?

52. (a) In reaching equilibrium, how much heat transfer occurs from 1.00 kg of water at 40.0° C when it is placed in contact with 1.00 kg of 20.0° C water in reaching equilibrium? (b) What is the change in entropy due to this heat transfer? (c) How much work is made unavailable, taking the lowest temperature to be 20.0° C? Explicitly show how you follow the steps in the Problem-Solving Strategies for Entropy.

53. What is the decrease in entropy of 25.0 g of water that condenses on a bathroom mirror at a temperature of 35.0° C, assuming no change in temperature and given the latent heat of vaporization to be 2450 kJ/kg?

54. Find the increase in entropy of 1.00 kg of liquid nitrogen that starts at its boiling temperature, boils, and warms to 20.0° C at constant pressure.

55. A large electrical power station generates 1000 MW of electricity with an efficiency of 35.0%. (a) Calculate the heat transfer to the power station, Q_{h}, in one day. (b) How much heat transfer Q_{c} occurs to the environment in one day? (c) If the heat transfer in the cooling towers is from 35.0° C water into the local air mass, which increases in temperature from 18.0° C to 20.0° C, what is the total increase in entropy due to this heat transfer? (d) How much energy becomes unavailable to do work because of this increase in entropy, assuming an 18.0° C lowest temperature? (Part of Q_{c} could be utilized to operate heat engines or for simply heating the surroundings, but it rarely is.)

56. (a) How much heat transfer occurs from 20.0 kg of 90.0° C water placed in contact with 20.0 kg of 10.0° C water, producing a final temperature of 50.0° C? (b) How much work could a Carnot engine do with this heat transfer, assuming it operates between two reservoirs at constant temperatures of 90.0° C and 10.0° C? (c) What increase in entropy is produced by mixing 20.0 kg of 90.0° C water with 20.0 kg of 10.0° C water? (d) Calculate the amount of work made unavailable by this mixing using a low temperature of 10.0° C, and compare it with the work done by the Carnot engine. Explicitly show how you follow the steps in the Problem-Solving Strategies for Entropy. (e) Discuss how everyday processes make increasingly more energy unavailable to do work, as implied by this problem.

15.7 Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

57. Using Table 15.4, verify the contention that if you toss 100 coins each second, you can expect to get 100 heads or 100 tails once in 2×10^{22} years; calculate the time to two-digit accuracy.

58. What percent of the time will you get something in the range from 60 heads and 40 tails through 40 heads and 60 tails when tossing 100 coins? The total number of microstates in that range is 1.22×10^{30}. (Consult Table 15.4.)

59. (a) If tossing 100 coins, how many ways (microstates) are there to get the three most likely macrostates of 49 heads and 51 tails, 50 heads and 50 tails, and 51 heads and 49 tails? (b) What percent of the total possibilities is this? (Consult Table 15.4.)

60. (a) What is the change in entropy if you start with 100 coins in the 45 heads and 55 tails macrostate, toss them, and get 51 heads and 49 tails? (b) What if you get 75 heads and 25 tails? (c) How much more likely is 51 heads and 49 tails than 75 heads and 25 tails? (d) Does either outcome violate the second law of thermodynamics?

61. (a) What is the change in entropy if you start with 10 coins in the 5 heads and 5 tails macrostate, toss them, and get 2 heads and 8 tails? (b) How much more likely is 5 heads and 5 tails than 2 heads and 8 tails? (Take the ratio of the number of microstates to find out.) (c) If you were betting on 2 heads and 8 tails would you accept odds of 252 to 45? Explain why or why not.

Macrostate		Number of Microstates (W)
Heads	Tails	
10	0	1
9	1	10
8	2	45
7	3	120
6	4	210
5	5	252
4	6	210
3	7	120
2	8	45
1	9	10
0	10	1
		Total: 1024

TABLE 15.5 10-Coin Toss

TABLE 15.5 10-Coin Toss

62. (a) If you toss 10 coins, what percent of the time will you get the three most likely macrostates (6 heads and 4 tails, 5 heads and 5 tails, 4 heads and 6 tails)? (b) You can realistically toss 10 coins and count the number of heads and tails about twice a minute. At that rate, how long will it take on average to get either 10 heads and 0 tails or 0 heads and 10 tails?

63. (a) Construct a table showing the macrostates and all of the individual microstates for tossing 6 coins. (Use Table 15.5 as a guide.) (b) How many macrostates are there? (c) What is the total number of microstates? (d) What percent chance is there of tossing 5 heads and 1 tail? (e) How much more likely are you to toss 3 heads and 3 tails than 5 heads and 1 tail? (Take the ratio of the number of microstates to find out.)

64. In an air conditioner, 12.65 MJ of heat transfer occurs from a cold environment in 1.00 h. (a) What mass of ice melting would involve the same heat transfer? (b) How many hours of operation would be equivalent to melting 900 kg of ice? (c) If ice costs 20 cents per kg, do you think the air conditioner could be operated more cheaply than by simply using ice? Describe in detail how you evaluate the relative costs.

CHAPTER 16
Oscillatory Motion and Waves

FIGURE 16.1 There are at least four types of waves in this picture—only the water waves are evident. There are also sound waves, light waves, and waves on the guitar strings. (credit: John Norton)

CHAPTER OUTLINE

INTRODUCTION TO OSCILLATORY MOTION AND WAVES What do an ocean buoy, a child in a swing, the cone inside a speaker, a guitar, atoms in a crystal, the motion of chest cavities, and the beating of hearts all have in common? They all **oscillate**—-that is, they move back and forth between two points. Many systems oscillate, and they have certain characteristics in common. All oscillations involve force and energy. You push a child in a swing to get the motion started. The energy of atoms vibrating in a crystal can be increased with heat. You put energy into a guitar string when you pluck it.

Some oscillations create **waves**. A guitar creates sound waves. You can make water waves in a swimming pool by slapping the water with your hand. You can no doubt think of other types of waves. Some, such as water waves, are visible. Some, such as sound waves, are not. But *every wave is a disturbance that moves from its source and carries energy*. Other examples of waves include earthquakes and visible light. Even subatomic particles, such as electrons, can behave like waves.

By studying oscillatory motion and waves, we shall find that a small number of underlying principles describe all of them and that wave phenomena are more common than you have ever imagined. We begin by studying the type of force that underlies the simplest oscillations and waves. We will then expand our exploration of oscillatory motion and waves to include concepts such as simple harmonic motion, uniform circular motion, and damped harmonic motion. Finally, we will explore what happens when two or more waves share the same space, in the phenomena known as superposition and interference.

Click to view content (https://openstax.org/books/college-physics-2e/pages/16-introduction-to-oscillatory-motion-and-waves)

16.1 Hooke's Law: Stress and Strain Revisited

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain Newton's third law of motion with respect to stress and deformation.
- Describe the restoration of force and displacement.
- Calculate the energy in Hooke's Law of deformation, and the stored energy in a spring.

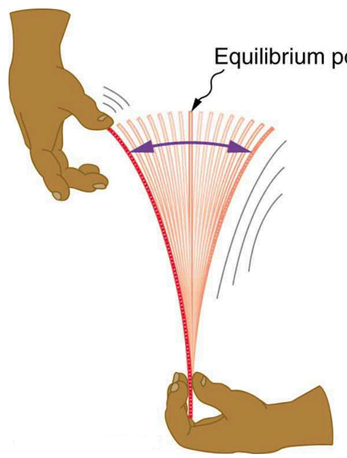


FIGURE 16.2 When displaced from its vertical equilibrium position, this plastic ruler oscillates back and forth because of the restoring force opposing displacement. When the ruler is on the left, there is a force to the right, and vice versa.

Newton's first law implies that an object oscillating back and forth is experiencing forces. Without force, the object would move in a straight line at a constant speed rather than oscillate. Consider, for example, plucking a plastic ruler to the left as shown in Figure 16.2. The deformation of the ruler creates a force in the opposite direction, known as a **restoring force**. Once released, the restoring force causes the ruler to move back toward its stable equilibrium position, where the net force on it is zero. However, by the time the ruler gets there, it gains momentum and continues to move to the right, producing the opposite deformation. It is then forced to the left, back through equilibrium, and the process is repeated until dissipative forces dampen the motion. These forces remove mechanical energy from the system, gradually reducing the motion until the ruler comes to rest.

The simplest oscillations occur when the restoring force is directly proportional to displacement. When stress and strain were covered in Newton's Third Law of Motion, the name was given to this relationship between force and displacement was Hooke's law:

$$F = -kx. \quad (16.1)$$

Here, F is the restoring force, x is the displacement from equilibrium or **deformation**, and k is a constant related to the difficulty in deforming the system. The minus sign indicates the restoring force is in the direction opposite to the displacement.

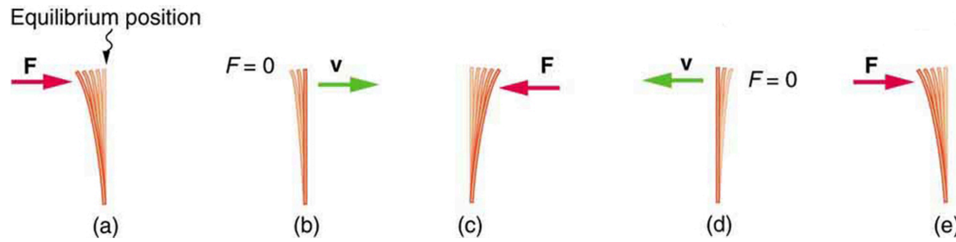
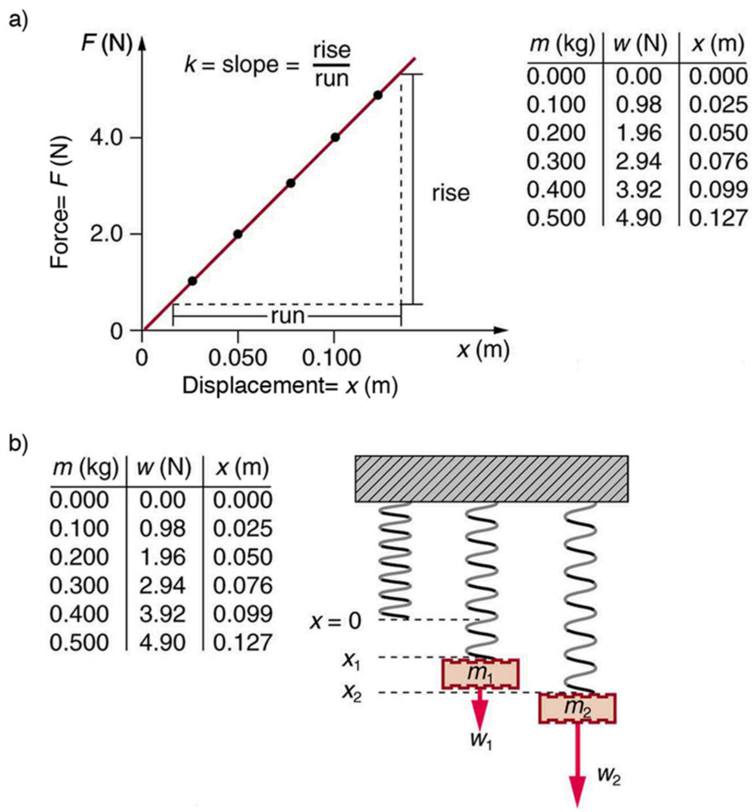


FIGURE 16.3 (a) The plastic ruler has been released, and the restoring force is returning the ruler to its equilibrium position. (b) The net force is zero at the equilibrium position, but the ruler has momentum and continues to move to the right. (c) The restoring force is in the opposite direction. It stops the ruler and moves it back toward equilibrium again. (d) Now the ruler has momentum to the left. (e) In the absence of damping (caused by frictional forces), the ruler reaches its original position. From there, the motion will repeat itself.

The **force constant** k is related to the rigidity (or stiffness) of a system—the larger the force constant, the greater the restoring force, and the stiffer the system. The units of k are newtons per meter (N/m). For example, k is directly related to Young's modulus when we stretch a string. Figure 16.4 shows a graph of the absolute value of the restoring force versus the displacement for a system that can be described by Hooke's law—a simple spring in this case. The slope of the graph equals the force constant k in newtons per meter. A common physics laboratory exercise is to measure restoring forces created by springs, determine if they follow Hooke's law, and calculate their force constants if they do.



m (kg)	w (N)	x (m)
0.000	0.00	0.000
0.100	0.98	0.025
0.200	1.96	0.050
0.300	2.94	0.076
0.400	3.92	0.099
0.500	4.90	0.127

m (kg)	w (N)	x (m)
0.000	0.00	0.000
0.100	0.98	0.025
0.200	1.96	0.050
0.300	2.94	0.076
0.400	3.92	0.099
0.500	4.90	0.127

FIGURE 16.4 (a) A graph of absolute value of the restoring force versus displacement is displayed. The fact that the graph is a straight line means that the system obeys Hooke's law. The slope of the graph is the force constant k. (b) The data in the graph were generated by measuring the displacement of a spring from equilibrium while supporting various weights. The restoring force equals the weight supported, if the mass is stationary.

EXAMPLE 16.1

How Stiff Are Car Springs?

FIGURE 16.5 The mass of a car increases due to the introduction of a passenger. This affects the displacement of the car on its suspension system. (credit: exfordy on Flickr)

What is the force constant for the suspension system of a car that settles 1.20 cm when an 80.0-kg person gets in?

Strategy

Consider the car to be in its equilibrium position $x = 0$ before the person gets in. The car then settles down 1.20 cm, which means it is displaced to a position $x = -1.20 \times 10^{-2}$ m. At that point, the springs supply a restoring force F equal to the person's weight $w = mg = (80.0 \text{ kg})(9.80 \text{ m/s}^2) = 784 \text{ N}$. We take this force to be F in Hooke's law. Knowing F and x, we can then solve the force constant k.

Solution

1. Solve Hooke's law, $F = -kx$, for k:

$$k = -\frac{F}{x}. \quad \text{16.2}$$

Substitute known values and solve k:

$$\begin{aligned} k &= -\frac{784 \text{ N}}{-1.20\times10^{-2} \text{ m}} \\ &= 6.53 \times 10^4 \text{ N/m.} \end{aligned} \quad \text{16.3}$$

Discussion

Note that F and x have opposite signs because they are in opposite directions—the restoring force is up, and the displacement is down. Also, note that the car would oscillate up and down when the person got in if it were not for damping (due to frictional forces) provided by shock absorbers. Bouncing cars are a sure sign of bad shock absorbers.

Energy in Hooke's Law of Deformation

In order to produce a deformation, work must be done. That is, a force must be exerted through a distance, whether you pluck a guitar string or compress a car spring. If the only result is deformation, and no work goes into thermal, sound, or kinetic energy, then all the work is initially stored in the deformed object as some form of potential energy. The potential energy stored in a spring is $\text{PE}_{\text{el}} = \frac{1}{2}kx^2$. Here, we generalize the idea to elastic potential energy for a deformation of any system that can be described by Hooke's law. Hence,

$$\mathrm{PE_{el}} = \frac{1}{2}kx^2, \quad (16.4)$$

where $\mathrm{PE_{el}}$ is the **elastic potential energy** stored in any deformed system that obeys Hooke's law and has a displacement x from equilibrium and a force constant k.

It is possible to find the work done in deforming a system in order to find the energy stored. This work is performed by an applied force F_{app}. The applied force is exactly opposite to the restoring force (action-reaction), and so $F_{\text{app}} = kx$. Figure 16.6 shows a graph of the applied force versus deformation x for a system that can be described by Hooke's law. Work done on the system is force multiplied by distance, which equals the area under the curve or $(1/2)kx^2$ (Method A in the figure). Another way to determine the work is to note that the force increases linearly from 0 to kx, so that the average force is $(1/2)\,kx$, the distance moved is x, and thus $W = F_{\text{app}}d = [(1/2)kx](x) = (1/2)kx^2$ (Method B in the figure).

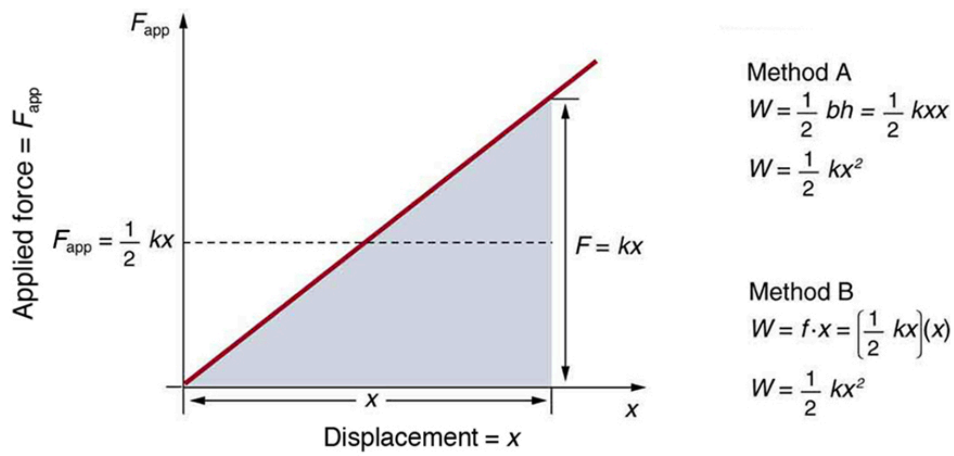


FIGURE 16.6 A graph of applied force versus distance for the deformation of a system that can be described by Hooke's law is displayed. The work done on the system equals the area under the graph or the area of the triangle, which is half its base multiplied by its height, or $W = (1/2)kx^2$.

EXAMPLE 16.2

Calculating Stored Energy: A Toy Gun Spring

We can use a toy gun's spring mechanism to ask and answer two simple questions: (a) How much energy is stored in the spring of a toy gun that has a force constant of 50.0 N/m and is compressed 0.150 m? (b) If you neglect friction and the mass of the spring, at what speed will a 2.00-g projectile be ejected from the gun?

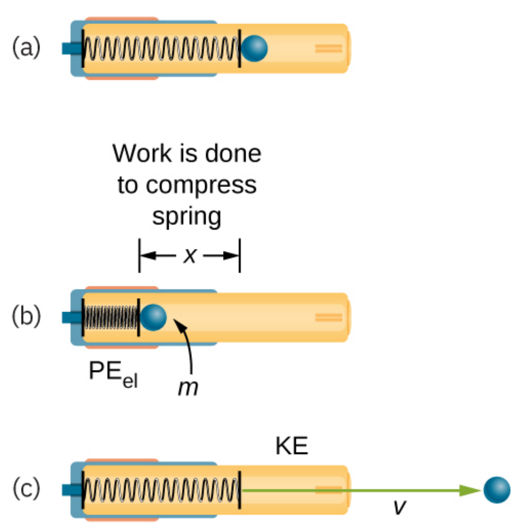


FIGURE 16.7 (a) In this image of the gun, the spring is uncompressed before being cocked. (b) The spring has been compressed a distance

x, and the projectile is in place. (c) When released, the spring converts elastic potential energy PE_{el} into kinetic energy.

Strategy for a

(a): The energy stored in the spring can be found directly from elastic potential energy equation, because k and x are given.

Solution for a

Entering the given values for k and x yields

$$\begin{aligned} PE_{el} &= \frac{1}{2}kx^2 = \frac{1}{2}(50.0\text{ N/m})(0.150\text{ m})^2 = 0.563\text{ N}\cdot\text{m} \\ &= 0.563\text{ J} \end{aligned} \quad (16.5)$$

Strategy for b

Because there is no friction, the potential energy is converted entirely into kinetic energy. The expression for kinetic energy can be solved for the projectile's speed.

Solution for b

1. Identify known quantities:

$$KE_f = PE_{el} \text{ or } 1/2\, mv^2 = (1/2)\, kx^2 = PE_{el} = 0.563\text{ J} \quad (16.6)$$

2. Solve for v:

$$v = \left[\frac{2PE_{el}}{m}\right]^{1/2} = \left[\frac{2(0.563\text{ J})}{0.002\text{ kg}}\right]^{1/2} = 23.7(\text{J/kg})^{1/2} \quad (16.7)$$

3. Convert units: 23.7 m/s

Discussion

(a) and (b): This projectile speed is impressive for a toy gun (more than 80 km/h). The numbers in this problem seem reasonable. The force needed to compress the spring is small enough for an adult to manage, and the energy imparted to the dart is small enough to limit the damage it might do, especially because the darts in many of these guns are made of soft material with a rubber tip. Yet, the speed of the dart is great enough for it to travel an acceptable distance.

CHECK YOUR UNDERSTANDING

Envision holding the end of a ruler with one hand and deforming it with the other. When you let go, you can see the oscillations of the ruler. In what way could you modify this simple experiment to increase the rigidity of the system?

Solution

You could hold the ruler at its midpoint so that the part of the ruler that oscillates is half as long as in the original experiment.

CHECK YOUR UNDERSTANDING

If you apply a deforming force on an object and let it come to equilibrium, what happened to the work you did on the system?

Solution

It was stored in the object as potential energy.

16.2 Period and Frequency in Oscillations

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Observe the vibrations of a guitar string.
- Determine the frequency of oscillations.

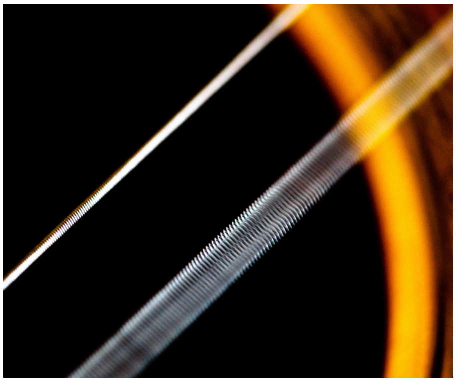


FIGURE 16.8 The strings on this guitar vibrate at regular time intervals. (credit: JAR)

When you pluck a guitar string, the resulting sound has a steady tone and lasts a long time. Each successive vibration of the string takes the same time as the previous one. We define **periodic motion** to be a motion that repeats itself at regular time intervals, such as exhibited by the guitar string or by an object on a spring moving up and down. The time to complete one oscillation remains constant and is called the **period** T. Its units are usually seconds, but may be any convenient unit of time. The word period refers to the time for some event whether repetitive or not; but we shall be primarily interested in periodic motion, which is by definition repetitive. A concept closely related to period is the frequency of an event. For example, if you get a paycheck twice a month, the frequency of payment is two per month and the period between checks is half a month. **Frequency** f is defined to be the number of events per unit time. For periodic motion, frequency is the number of oscillations per unit time. The relationship between frequency and period is

$$f = \frac{1}{T}. \tag{16.8}$$

The SI unit for frequency is the *cycle per second*, which is defined to be a *hertz* (Hz):

$$1\ \text{Hz} = 1\frac{\text{cycle}}{\text{s}} \text{ or } 1\ \text{Hz} = \frac{1}{\text{s}} \tag{16.9}$$

A cycle is one complete oscillation. Note that a vibration can be a single or multiple event, whereas oscillations are usually repetitive for a significant number of cycles.

EXAMPLE 16.3

Determine the Frequency of Two Oscillations: Medical Ultrasound and the Period of Middle C

We can use the formulas presented in this module to determine both the frequency based on known oscillations and the oscillation based on a known frequency. Let’s try one example of each. (a) A medical imaging device produces ultrasound by oscillating with a period of 0.400 µs. What is the frequency of this oscillation? (b) The frequency of middle C on a typical musical instrument is 264 Hz. What is the time for one complete oscillation?

Strategy

Both questions (a) and (b) can be answered using the relationship between period and frequency. In question (a), the period T is given and we are asked to find frequency f. In question (b), the frequency f is given and we are asked to find the period T.

Solution a

1. Substitute $0.400\ \mu\text{s}$ for T in $f = \frac{1}{T}$:

$$f = \frac{1}{T} = \frac{1}{0.400 \times 10^{-6}\ \text{s}}. \tag{16.10}$$

Solve to find

$$f = 2.50 \times 10^{6}\ \text{Hz}. \tag{16.11}$$

Discussion a

The frequency of sound found in (a) is much higher than the highest frequency that humans can hear and, therefore, is called ultrasound. Appropriate oscillations at this frequency generate ultrasound used for noninvasive medical diagnoses, such as observations of a fetus in the womb.

Solution b

1. Identify the known values:
   The time for one complete oscillation is the period T:

$$f = \frac{1}{T}. \tag{16.12}$$

2. Solve for T:

$$T = \frac{1}{f}. \tag{16.13}$$

3. Substitute the given value for the frequency into the resulting expression:

$$T = \frac{1}{f} = \frac{1}{264\text{ Hz}} = \frac{1}{264\text{ cycles/s}} = 3.79 \times 10^{-3}\text{ s} = 3.79\text{ ms}. \tag{16.14}$$

Discussion b

The period found in (b) is the time per cycle, but this value is often quoted as simply the time in convenient units (ms or milliseconds in this case).

CHECK YOUR UNDERSTANDING

Identify an event in your life (such as receiving a paycheck) that occurs regularly. Identify both the period and frequency of this event.

Solution

I visit my parents for dinner every other Sunday. The frequency of my visits is 26 per calendar year. The period is two weeks.

16.3 Simple Harmonic Motion: A Special Periodic Motion

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe a simple harmonic oscillator.
- Explain the link between simple harmonic motion and waves.

The oscillations of a system in which the net force can be described by Hooke's law are of special importance, because they are very common. They are also the simplest oscillatory systems. **Simple Harmonic Motion** (SHM) is the name given to oscillatory motion for a system where the net force can be described by Hooke's law, and such a system is called a **simple harmonic oscillator**. If the net force can be described by Hooke's law and there is no *damping* (by friction or other non-conservative forces), then a simple harmonic oscillator will oscillate with equal displacement on either side of the equilibrium position, as shown for an object on a spring in Figure 16.9. The maximum displacement from equilibrium is called the **amplitude** X. The units for amplitude and displacement are the same, but depend on the type of oscillation. For the object on the spring, the units of amplitude and displacement are meters; whereas for sound oscillations, they have units of pressure (and other types of oscillations have yet other units). Because amplitude is the maximum displacement, it is related to the energy in the oscillation.

Take-Home Experiment: SHM and the Marble

Find a bowl or basin that is shaped like a hemisphere on the inside. Place a marble inside the bowl and tilt the bowl periodically so the marble rolls from the bottom of the bowl to equally high points on the sides of the bowl.

Get a feel for the force required to maintain this periodic motion. What is the restoring force and what role does the force you apply play in the simple harmonic motion (SHM) of the marble?

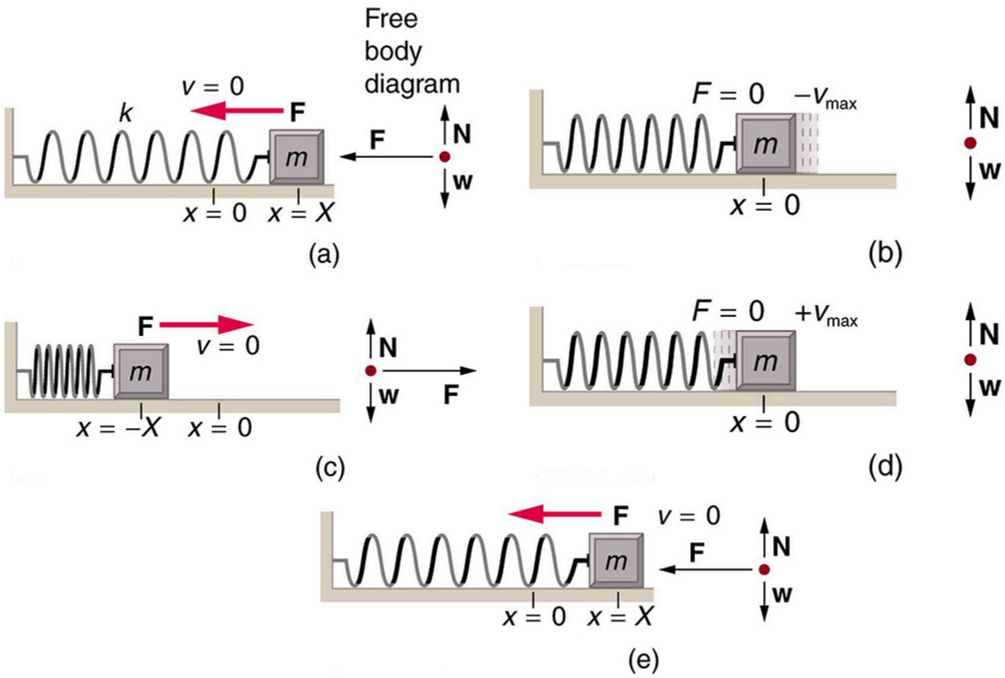


FIGURE 16.9 An object attached to a spring sliding on a frictionless surface is an uncomplicated simple harmonic oscillator. When displaced from equilibrium, the object performs simple harmonic motion that has an amplitude X and a period T. The object's maximum speed occurs as it passes through equilibrium. The stiffer the spring is, the smaller the period T. The greater the mass of the object is, the greater the period T.

What is so significant about simple harmonic motion? One special thing is that the period T and frequency f of a simple harmonic oscillator are independent of amplitude. The string of a guitar, for example, will oscillate with the same frequency whether plucked gently or hard. Because the period is constant, a simple harmonic oscillator can be used as a clock.

Two important factors do affect the period of a simple harmonic oscillator. The period is related to how stiff the system is. A very stiff object has a large force constant k, which causes the system to have a smaller period. For example, you can adjust a diving board's stiffness—the stiffer it is, the faster it vibrates, and the shorter its period. Period also depends on the mass of the oscillating system. The more massive the system is, the longer the period. For example, a heavy person on a diving board bounces up and down more slowly than a light one.

In fact, the mass m and the force constant k are the *only* factors that affect the period and frequency of simple harmonic motion.

Period of Simple Harmonic Oscillator

The *period of a simple harmonic oscillator* is given by

$$T = 2\pi\sqrt{\frac{m}{k}} \tag{16.15}$$

and, because $f = 1/T$, the *frequency of a simple harmonic oscillator* is

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}. \tag{16.16}$$

Note that neither T nor f has any dependence on amplitude.

Take-Home Experiment: Mass and Ruler Oscillations

Find two identical wooden or plastic rulers. Tape one end of each ruler firmly to the edge of a table so that the length of each ruler that protrudes from the table is the same. On the free end of one ruler tape a heavy object such as a few large coins. Pluck the ends of the rulers at the same time and observe which one undergoes more cycles in a time period, and measure the period of oscillation of each of the rulers.

EXAMPLE 16.4

Calculate the Frequency and Period of Oscillations: Bad Shock Absorbers in a Car

If the shock absorbers in a car go bad, then the car will oscillate at the least provocation, such as when going over bumps in the road and after stopping (See Figure 16.10). Calculate the frequency and period of these oscillations for such a car if the car's mass (including its load) is 900 kg and the force constant (k) of the suspension system is $6.53 \times 10^4\ \text{N/m}$.

Strategy

The frequency of the car's oscillations will be that of a simple harmonic oscillator as given in the equation $f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$. The mass and the force constant are both given.

Solution

1. Enter the known values of k and m:

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{2\pi}\sqrt{\frac{6.53 \times 10^4\ \text{N/m}}{900\ \text{kg}}}. \quad (16.17)$$

2. Calculate the frequency:

$$\frac{1}{2\pi}\sqrt{72.6/\text{s}^2} = 1.3656/\text{s}^1 \approx 1.36/\text{s}^1 = 1.36\text{Hz}. \quad (16.18)$$

3. You could use $T = 2\pi\sqrt{\frac{m}{k}}$ to calculate the period, but it is simpler to use the relationship $T = 1/f$ and substitute the value just found for f:

$$T = \frac{1}{f} = \frac{1}{1.356\ \text{Hz}} = 0.738\ \text{s}. \quad (16.19)$$

Discussion

The values of T and f both seem about right for a bouncing car. You can observe these oscillations if you push down hard on the end of a car and let go.

The Link between Simple Harmonic Motion and Waves

If a time-exposure photograph of the bouncing car were taken as it drove by, the headlight would make a wavelike streak, as shown in Figure 16.10. Similarly, Figure 16.11 shows an object bouncing on a spring as it leaves a wavelike "trace" of its position on a moving strip of paper. Both waves are sine functions. All simple harmonic motion is intimately related to sine and cosine waves.

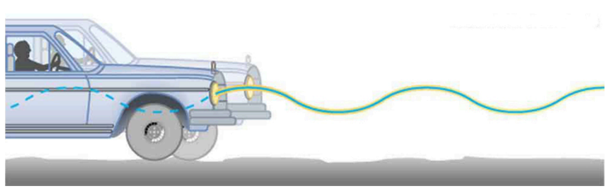


FIGURE 16.10 The bouncing car makes a wavelike motion. If the restoring force in the suspension system can be described only by Hooke's law, then the wave is a sine function. (The wave is the trace produced by the headlight as the car moves to the right.)

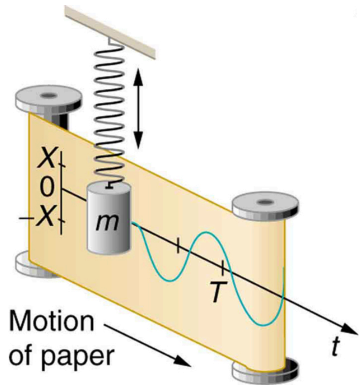


FIGURE 16.11 The vertical position of an object bouncing on a spring is recorded on a strip of moving paper, leaving a sine wave.

The displacement as a function of time *t* in any simple harmonic motion—that is, one in which the net restoring force can be described by Hooke's law, is given by

$$x(t)= X \cos\frac{2\pi t}{T}, \quad (16.20)$$

where X is amplitude. At $t = 0$, the initial position is $x_0 = X$, and the displacement oscillates back and forth with a period T. (When $t = T$, we get $x = X$ again because $\cos 2\pi = 1$.). Furthermore, from this expression for x, the velocity v as a function of time is given by:

$$v(t)= -v_{\text{max}} \sin\left(\frac{2\pi t}{T}\right), \quad (16.21)$$

where $v_{\text{max}} = 2\pi X/T = X\sqrt{k/m}$. The object has zero velocity at maximum displacement—for example, $v = 0$ when $t = 0$, and at that time $x = X$. The minus sign in the first equation for $v(t)$ gives the correct direction for the velocity. Just after the start of the motion, for instance, the velocity is negative because the system is moving back toward the equilibrium point. Finally, we can get an expression for acceleration using Newton's second law. [Then we have $x(t)$, $v(t)$, t, and $a(t)$, the quantities needed for kinematics and a description of simple harmonic motion.] According to Newton's second law, the acceleration is $a = F/m = kx/m$. So, $a(t)$ is also a cosine function:

$$a(t)= -\frac{kX}{m}\cos\frac{2\pi t}{T}. \quad (16.22)$$

Hence, $a(t)$ is directly proportional to and in the opposite direction to $x(t)$.

Figure 16.12 shows the simple harmonic motion of an object on a spring and presents graphs of $x(t), v(t)$, and $a(t)$ versus time.

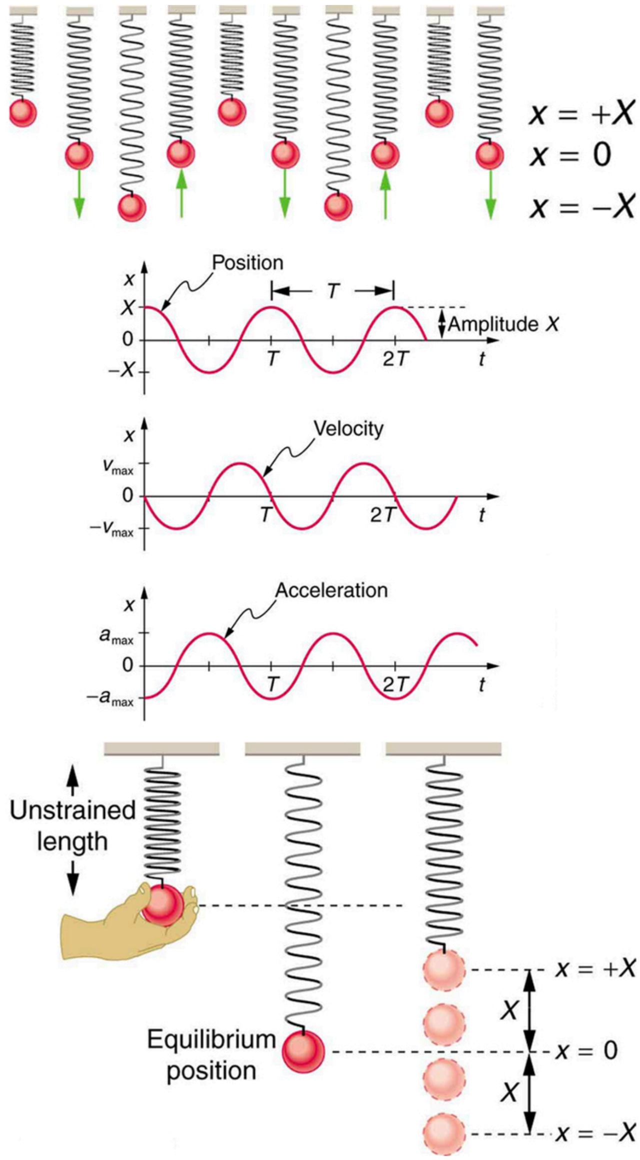


FIGURE 16.12 Graphs of $x(t)$, $v(t)$, and $a(t)$ versus t for the motion of an object on a spring. The net force on the object can be described by Hooke's law, and so the object undergoes simple harmonic motion. Note that the initial position has the vertical displacement at its maximum value X; v is initially zero and then negative as the object moves down; and the initial acceleration is negative, back toward the equilibrium position and becomes zero at that point.

The most important point here is that these equations are mathematically straightforward and are valid for all simple harmonic motion. They are very useful in visualizing waves associated with simple harmonic motion, including visualizing how waves add with one another.

CHECK YOUR UNDERSTANDING

Suppose you pluck a banjo string. You hear a single note that starts out loud and slowly quiets over time. Describe what happens to the sound waves in terms of period, frequency and amplitude as the sound decreases in volume.

Solution

Frequency and period remain essentially unchanged. Only amplitude decreases as volume decreases.

CHECK YOUR UNDERSTANDING

A babysitter is pushing a child on a swing. At the point where the swing reaches X, where would the corresponding point on a wave of this motion be located?

Solution

X is the maximum deformation, which corresponds to the amplitude of the wave. The point on the wave would either be at the very top or the very bottom of the curve.

PHET EXPLORATIONS

Masses and Springs

A realistic mass and spring laboratory. Hang masses from springs and adjust the spring stiffness and damping. You can even slow time. Transport the lab to different planets. A chart shows the kinetic, potential, and thermal energy for each spring.

Click to view content (https://openstax.org/books/college-physics-2e/pages/16-3-simple-harmonic-motion-a-special-periodic-motion)

16.4 The Simple Pendulum

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Measure acceleration due to gravity.

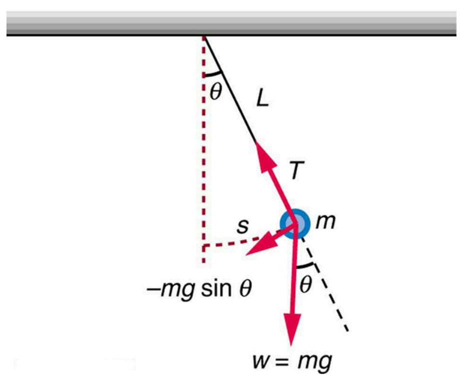


FIGURE 16.13 A simple pendulum has a small-diameter bob and a string that has a very small mass but is strong enough not to stretch appreciably. The linear displacement from equilibrium is s, the length of the arc. Also shown are the forces on the bob, which result in a net force of $-mg\ \sin\theta$ toward the equilibrium position—that is, a restoring force.

Pendulums are in common usage. Some have crucial uses, such as in clocks; some are for fun, such as a child's swing; and some are just there, such as the sinker on a fishing line. For small displacements, a pendulum is a simple harmonic oscillator. A **simple pendulum** is defined to have an object that has a small mass, also known as the pendulum bob, which is suspended from a light wire or string, such as shown in Figure 16.13. Exploring the simple pendulum a bit further, we can discover the conditions under which it performs simple harmonic motion, and we can derive an interesting expression for its period.

We begin by defining the displacement to be the arc length s. We see from Figure 16.13 that the net force on the bob is tangent to the arc and equals $-mg\ \sin\ \theta$. (The weight mg has components $mg\ \cos\ \theta$ along the string and $mg\ \sin\ \theta$ tangent to the arc.) Tension in the string exactly cancels the component $mg\ \cos\ \theta$ parallel to the string. This leaves a *net* restoring force back toward the equilibrium position at $\theta = 0$.

Now, if we can show that the restoring force is directly proportional to the displacement, then we have a simple harmonic oscillator. In trying to determine if we have a simple harmonic oscillator, we should note that for small angles (less than about 15^{o}), $\sin\ \theta \approx\ \theta$ ($\sin\ \theta$ and θ differ by about 1% or less at smaller angles). Thus, for angles less than about 15^{o}, the restoring force F is

$$F \approx -mg\theta. \qquad (16.23)$$

The displacement s is directly proportional to θ. When θ is expressed in radians, the arc length in a circle is related to its radius (L in this instance) by:

$$s = L\theta, \tag{16.24}$$

so that

$$\theta = \frac{s}{L}. \tag{16.25}$$

For small angles, then, the expression for the restoring force is:

$$F \approx -\frac{mg}{L}s \tag{16.26}$$

This expression is of the form:

$$F = -kx, \tag{16.27}$$

where the force constant is given by $k = mg/L$ and the displacement is given by $x = s$. For angles less than about 15^{o}, the restoring force is directly proportional to the displacement, and the simple pendulum is a simple harmonic oscillator.

Using this equation, we can find the period of a pendulum for amplitudes less than about 15^{o}. For the simple pendulum:

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{m}{mg/L}}. \tag{16.28}$$

Thus,

$$T = 2\pi\sqrt{\frac{L}{g}} \tag{16.29}$$

for the period of a simple pendulum. This result is interesting because of its simplicity. The only things that affect the period of a simple pendulum are its length and the acceleration due to gravity. The period is completely independent of other factors, such as mass. As with simple harmonic oscillators, the period T for a pendulum is nearly independent of amplitude, especially if θ is less than about 15^{o}. Even simple pendulum clocks can be finely adjusted and accurate.

Note the dependence of T on g. If the length of a pendulum is precisely known, it can actually be used to measure the acceleration due to gravity. Consider the following example.

EXAMPLE 16.5

Measuring Acceleration due to Gravity: The Period of a Pendulum

What is the acceleration due to gravity in a region where a simple pendulum having a length 75.000 cm has a period of 1.7357 s?

Strategy

We are asked to find g given the period T and the length L of a pendulum. We can solve $T = 2\pi\sqrt{\frac{L}{g}}$ for g, assuming only that the angle of deflection is less than 15^{o}.

Solution

1. Square $T = 2\pi\sqrt{\frac{L}{g}}$ and solve for g:

$$g = 4\pi^2\frac{L}{T^2}. \tag{16.30}$$

2. Substitute known values into the new equation:

$$g = 4\pi^2\frac{0.75000\text{ m}}{(1.7357\text{ s})^2}. \tag{16.31}$$

3. Calculate to find g:

$$g = 9.8281\ \text{m/s}^2. \tag{16.32}$$

Discussion

This method for determining g can be very accurate. This is why length and period are given to five digits in this example. For the precision of the approximation $\sin\theta \approx \theta$ to be better than the precision of the pendulum length and period, the maximum displacement angle should be kept below about 0.5°.

Making Career Connections

Knowing g can be important in geological exploration; for example, a map of g over large geographical regions aids the study of plate tectonics and helps in the search for oil fields and large mineral deposits.

Take Home Experiment: Determining g

Use a simple pendulum to determine the acceleration due to gravity g in your own locale. Cut a piece of a string or dental floss so that it is about 1 m long. Attach a small object of high density to the end of the string (for example, a metal nut or a car key). Starting at an angle of less than 10°, allow the pendulum to swing and measure the pendulum's period for 10 oscillations using a stopwatch. Calculate g. How accurate is this measurement? How might it be improved?

CHECK YOUR UNDERSTANDING

An engineer builds two simple pendula. Both are suspended from small wires secured to the ceiling of a room. Each pendulum hovers 2 cm above the floor. Pendulum 1 has a bob with a mass of $10\ \text{kg}$. Pendulum 2 has a bob with a mass of $100\ \text{kg}$. Describe how the motion of the pendula will differ if the bobs are both displaced by 12°.

Solution

The movement of the pendula will not differ at all because the mass of the bob has no effect on the motion of a simple pendulum. The pendula are only affected by the period (which is related to the pendulum's length) and by the acceleration due to gravity.

PHET EXPLORATIONS

Pendulum Lab

Play with one or two pendulums and discover how the period of a simple pendulum depends on the length of the string, the mass of the pendulum bob, and the amplitude of the swing. It's easy to measure the period using the photogate timer. You can vary friction and the strength of gravity. Use the pendulum to find the value of g on planet X. Notice the anharmonic behavior at large amplitude.

Click to view content (https://openstax.org/books/college-physics-2e/pages/16-4-the-simple-pendulum)

16.5 Energy and the Simple Harmonic Oscillator

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Determine the maximum speed of an oscillating system.

To study the energy of a simple harmonic oscillator, we first consider all the forms of energy it can have We know from Hooke's Law: Stress and Strain Revisited that the energy stored in the deformation of a simple harmonic oscillator is a form of potential energy given by:

$$\mathrm{PE_{el}} = \frac{1}{2}kx^2. \qquad 16.33$$

Because a simple harmonic oscillator has no dissipative forces, the other important form of energy is kinetic energy KE. Conservation of energy for these two forms is:

$$\mathrm{KE} + \mathrm{PE_{el}} = \text{constant} \qquad 16.34$$

or

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}. \qquad 16.35$$

This statement of conservation of energy is valid for *all* simple harmonic oscillators, including ones where the gravitational force plays a role

Namely, for a simple pendulum we replace the velocity with $v = L\omega$, the spring constant with $k = mg/L$, and the displacement term with $x = L\theta$. Thus

$$\frac{1}{2}mL^2\omega^2 + \frac{1}{2}mgL\theta^2 = \text{constant}. \qquad 16.36$$

In the case of undamped simple harmonic motion, the energy oscillates back and forth between kinetic and potential, going completely from one to the other as the system oscillates. So for the simple example of an object on a frictionless surface attached to a spring, as shown again in Figure 16.14, the motion starts with all of the energy stored in the spring. As the object starts to move, the elastic potential energy is converted to kinetic energy, becoming entirely kinetic energy at the equilibrium position. It is then converted back into elastic potential energy by the spring, the velocity becomes zero when the kinetic energy is completely converted, and so on. This concept provides extra insight here and in later applications of simple harmonic motion, such as alternating current circuits.

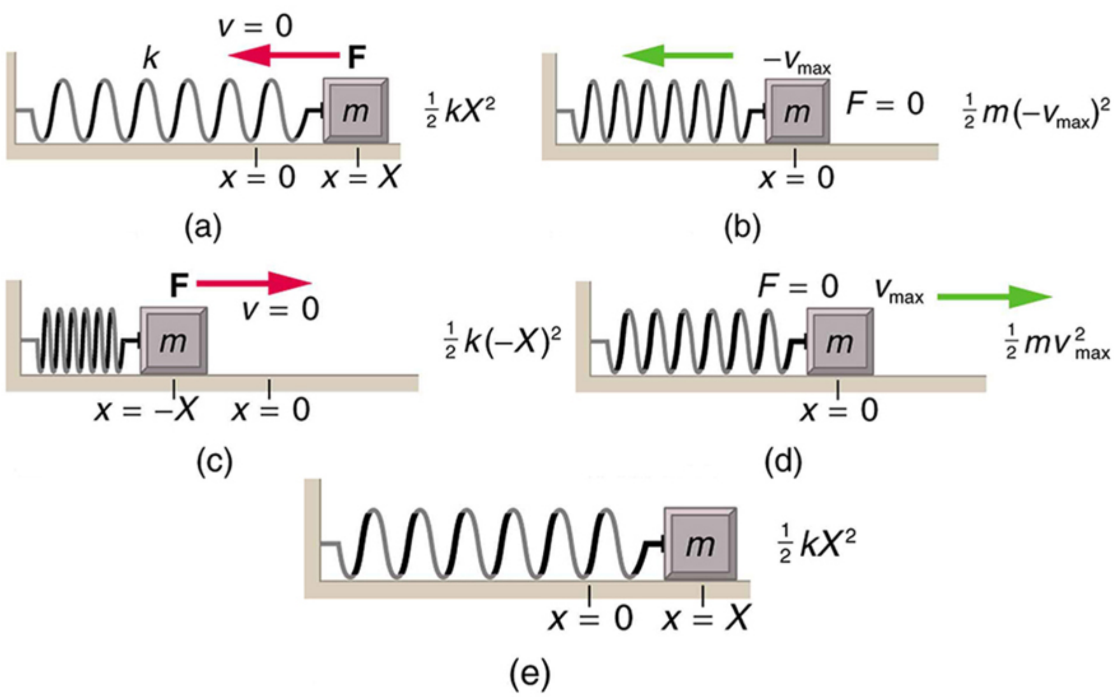


FIGURE 16.14 The transformation of energy in simple harmonic motion is illustrated for an object attached to a spring on a frictionless surface.

The conservation of energy principle can be used to derive an expression for velocity v. If we start our simple harmonic motion with zero velocity and maximum displacement ($x = X$), then the total energy is

$$\frac{1}{2}kX^2. \qquad 16.37$$

This total energy is constant and is shifted back and forth between kinetic energy and potential energy, at most times being shared by each. The conservation of energy for this system in equation form is thus:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kX^2. \qquad 16.38$$

Solving this equation for v yields:

$$v = \pm\sqrt{\frac{k}{m}\left(X^2 - x^2\right)}. \qquad 16.39$$

Manipulating this expression algebraically gives:

$$v = \pm\sqrt{\frac{k}{m}}X\sqrt{1 - \frac{x^2}{X^2}} \qquad 16.40$$

and so

$$v = \pm v_{\text{max}}\sqrt{1 - \frac{x^2}{X^2}}, \qquad 16.41$$

where

$$v_{\text{max}} = \sqrt{\frac{k}{m}}X. \qquad 16.42$$

From this expression, we see that the velocity is a maximum (v_{max}) at $x = 0$, as stated earlier in $v(t) = -v_{\text{max}}\sin\frac{2\pi t}{T}$. Notice that the maximum velocity depends on three factors. Maximum velocity is directly proportional to amplitude. As you might guess, the greater the maximum displacement the greater the maximum velocity. Maximum velocity is also greater for stiffer systems, because they exert greater force for the same displacement. This observation is seen in the expression for v_{max}; it is proportional to the square root of the force constant k. Finally, the maximum velocity is smaller for objects that have larger masses, because the maximum velocity is inversely proportional to the square root of m. For a given force, objects that have large masses accelerate more slowly.

A similar calculation for the simple pendulum produces a similar result, namely:

$$\omega_{\text{max}} = \sqrt{\frac{g}{L}}\theta_{\text{max}}. \qquad 16.43$$

EXAMPLE 16.6

Determine the Maximum Speed of an Oscillating System: A Bumpy Road

Suppose that a car is 900 kg and has a suspension system that has a force constant $k = 6.53 \times 10^4\ \text{N/m}$. The car hits a bump and bounces with an amplitude of 0.100 m. What is its maximum vertical velocity if you assume no damping occurs?

Strategy

We can use the expression for v_{max} given in $v_{\text{max}} = \sqrt{\frac{k}{m}}X$ to determine the maximum vertical velocity. The variables m and k are given in the problem statement, and the maximum displacement X is 0.100 m.

Solution

1. Identify known.
2. Substitute known values into $v_{\text{max}} = \sqrt{\frac{k}{m}}X$:

$$v_{max} = \sqrt{\frac{6.53 \times 10^4 \text{ N/m}}{900 \text{ kg}}}(0.100 \text{ m}). \quad \textbf{16.44}$$

3. Calculate to find $v_{max} = 0.852$ m/s.

Discussion

This answer seems reasonable for a bouncing car. There are other ways to use conservation of energy to find v_{max}. We could use it directly, as was done in the example featured in Hooke's Law: Stress and Strain Revisited.

The small vertical displacement y of an oscillating simple pendulum, starting from its equilibrium position, is given as

$$y(t) = a \sin \omega t, \quad \textbf{16.45}$$

where a is the amplitude, ω is the angular velocity and t is the time taken. Substituting $\omega = \frac{2\pi}{T}$, we have

$$y(t) = a \sin\left(\frac{2\pi t}{T}\right). \quad \textbf{16.46}$$

Thus, the displacement of pendulum is a function of time as shown above.

Also the velocity of the pendulum is given by

$$v(t) = \frac{2a\pi}{T} \cos\left(\frac{2\pi t}{T}\right), \quad \textbf{16.47}$$

so the motion of the pendulum is a function of time.

CHECK YOUR UNDERSTANDING

Why does it hurt more if your hand is snapped with a ruler than with a loose spring, even if the displacement of each system is equal?

Solution

The ruler is a stiffer system, which carries greater force for the same amount of displacement. The ruler snaps your hand with greater force, which hurts more.

CHECK YOUR UNDERSTANDING

You are observing a simple harmonic oscillator. Identify one way you could decrease the maximum velocity of the system.

Solution

You could increase the mass of the object that is oscillating.

16.6 Uniform Circular Motion and Simple Harmonic Motion

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Compare simple harmonic motion with uniform circular motion.

FIGURE 16.15 The horses on this merry-go-round exhibit uniform circular motion. (credit: Wonderlane, Flickr)

There is an easy way to produce simple harmonic motion by using uniform circular motion. Figure 16.16 shows one way of using this method. A ball is attached to a uniformly rotating vertical turntable, and its shadow is projected on the floor as shown. The shadow undergoes simple harmonic motion. Hooke's law usually describes uniform circular motions (ω constant) rather than systems that have large visible displacements. So observing the projection of uniform circular motion, as in Figure 16.16, is often easier than observing a precise large-scale simple harmonic oscillator. If studied in sufficient depth, simple harmonic motion produced in this manner can give considerable insight into many aspects of oscillations and waves and is very useful mathematically. In our brief treatment, we shall indicate some of the major features of this relationship and how they might be useful.

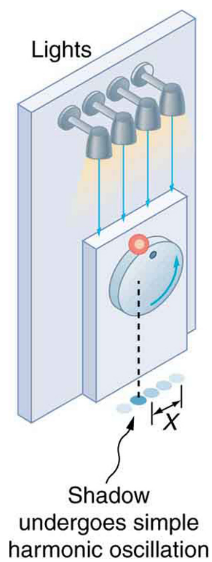


FIGURE 16.16 The shadow of a ball rotating at constant angular velocity ω on a turntable goes back and forth in precise simple harmonic motion.

Figure 16.17 shows the basic relationship between uniform circular motion and simple harmonic motion. The point P travels around the circle at constant angular velocity ω. The point P is analogous to an object on the merry-go-round. The projection of the position of P onto a fixed axis undergoes simple harmonic motion and is analogous to the shadow of the object. At the time shown in the figure, the projection has position x and moves to the left with velocity v. The velocity of the point P around the circle equals $\bar{v}_{\text{max}}$.The projection of $\bar{v}_{\text{max}}$ on the x-axis is the velocity v of the simple harmonic motion along the x-axis.

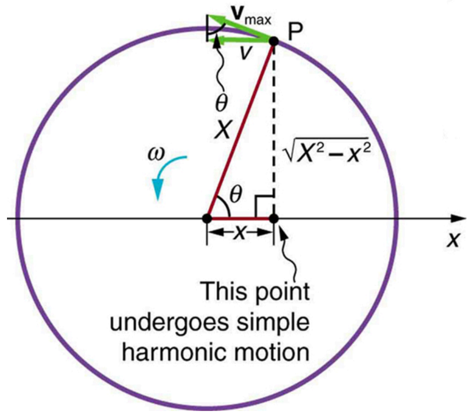


FIGURE 16.17 A point P moving on a circular path with a constant angular velocity ω is undergoing uniform circular motion. Its projection on the x-axis undergoes simple harmonic motion. Also shown is the velocity of this point around the circle, $\overline{v}_{\text{max}}$, and its projection, which is v. Note that these velocities form a similar triangle to the displacement triangle.

To see that the projection undergoes simple harmonic motion, note that its position x is given by

$$x = X\cos\theta, \tag{16.48}$$

where $\theta = \omega t$, ω is the constant angular velocity, and X is the radius of the circular path. Thus,

$$x = X\cos\omega t. \tag{16.49}$$

The angular velocity ω is in radians per unit time; in this case 2π radians is the time for one revolution T. That is, $\omega = 2\pi/T$. Substituting this expression for ω, we see that the position x is given by:

$$x(t) = \cos\left(\frac{2\pi t}{T}\right). \tag{16.50}$$

This expression is the same one we had for the position of a simple harmonic oscillator in Simple Harmonic Motion: A Special Periodic Motion. If we make a graph of position versus time as in Figure 16.18, we see again the wavelike character (typical of simple harmonic motion) of the projection of uniform circular motion onto the x-axis.

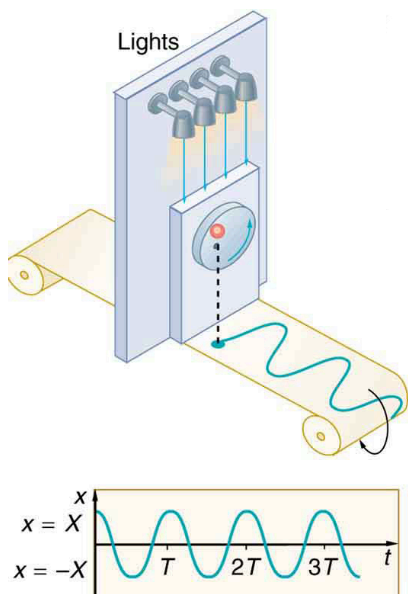


FIGURE 16.18 The position of the projection of uniform circular motion performs simple harmonic motion, as this wavelike graph of x versus t indicates.

Now let us use Figure 16.17 to do some further analysis of uniform circular motion as it relates to simple harmonic motion. The triangle formed by the velocities in the figure and the triangle formed by the displacements (X, x, and $\sqrt{X^2 - x^2}$) are similar right triangles. Taking ratios of similar sides, we see that

$$\frac{v}{v_{\max}} = \frac{\sqrt{X^2 - x^2}}{X} = \sqrt{1 - \frac{x^2}{X^2}}. \quad \mathbf{16.51}$$

We can solve this equation for the speed v or

$$v = v_{\max}\sqrt{1 - \frac{x^2}{X^2}}. \quad \mathbf{16.52}$$

This expression for the speed of a simple harmonic oscillator is exactly the same as the equation obtained from conservation of energy considerations in Energy and the Simple Harmonic Oscillator.You can begin to see that it is possible to get all of the characteristics of simple harmonic motion from an analysis of the projection of uniform circular motion.

Finally, let us consider the period T of the motion of the projection. This period is the time it takes the point P to complete one revolution. That time is the circumference of the circle $2\pi X$ divided by the velocity around the circle, $v_{\max}$. Thus, the period T is

$$T = \frac{2\pi X}{v_{\max}}. \quad \mathbf{16.53}$$

We know from conservation of energy considerations that

$$v_{\max} = \sqrt{\frac{k}{m}}X. \quad \mathbf{16.54}$$

Solving this equation for $X/v_{\max}$ gives

$$\frac{X}{v_{\max}} = \sqrt{\frac{m}{k}}. \quad \mathbf{16.55}$$

Substituting this expression into the equation for T yields

$$T = 2\pi\sqrt{\frac{m}{k}}. \quad \mathbf{16.56}$$

Thus, the period of the motion is the same as for a simple harmonic oscillator. We have determined the period for any simple harmonic oscillator using the relationship between uniform circular motion and simple harmonic motion.

Some modules occasionally refer to the connection between uniform circular motion and simple harmonic motion. Moreover, if you carry your study of physics and its applications to greater depths, you will find this relationship useful. It can, for example, help to analyze how waves add when they are superimposed.

CHECK YOUR UNDERSTANDING

Identify an object that undergoes uniform circular motion. Describe how you could trace the simple harmonic motion of this object as a wave.

Solution

A record player undergoes uniform circular motion. You could attach dowel rod to one point on the outside edge of the turntable and attach a pen to the other end of the dowel. As the record player turns, the pen will move. You can drag a long piece of paper under the pen, capturing its motion as a wave.

16.7 Damped Harmonic Motion

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Compare and discuss underdamped and overdamped oscillating systems.
- Explain critically damped system.

FIGURE 16.19 In order to counteract dampening forces, this mom needs to keep pushing the swing. (credit: Erik A. Johnson, Flickr)

A guitar string stops oscillating a few seconds after being plucked. To keep a child happy on a swing, you must keep pushing. Although we can often make friction and other non-conservative forces negligibly small, completely undamped motion is rare. In fact, we may even want to damp oscillations, such as with car shock absorbers.

For a system that has a small amount of damping, the period and frequency are nearly the same as for simple harmonic motion, but the amplitude gradually decreases as shown in Figure 16.20. This occurs because the non-conservative damping force removes energy from the system, usually in the form of thermal energy. In general, energy removal by non-conservative forces is described as

$$W_{\text{nc}} = \Delta(\text{KE} + \text{PE}), \tag{16.57}$$

where W_{nc} is work done by a non-conservative force (here the damping force). For a damped harmonic oscillator, W_{nc} is negative because it removes mechanical energy (KE + PE) from the system.

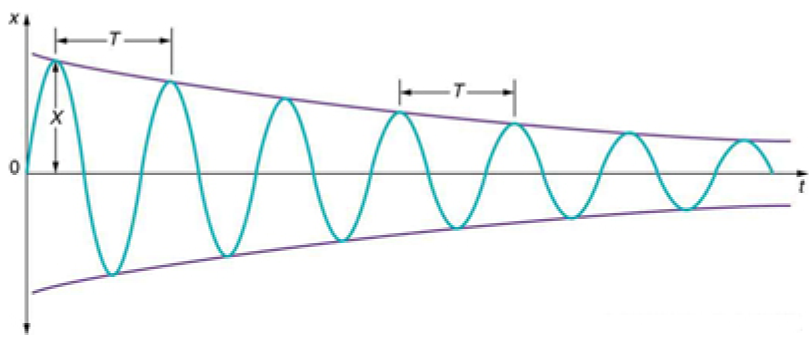


FIGURE 16.20 In this graph of displacement versus time for a harmonic oscillator with a small amount of damping, the amplitude slowly decreases, but the period and frequency are nearly the same as if the system were completely undamped.

If you gradually *increase* the amount of damping in a system, the period and frequency begin to be affected, because damping opposes and hence slows the back and forth motion. (The net force is smaller in both directions.) If there is very large damping, the system does not even oscillate—it slowly moves toward equilibrium. Figure 16.21 shows the displacement of a harmonic oscillator for different amounts of damping. When we want to damp out oscillations, such as in the suspension of a car, we may want the system to return to equilibrium as quickly as possible **Critical damping** is defined as the condition in which the damping of an oscillator results in it returning as quickly as possible to its equilibrium position The critically damped system may overshoot the equilibrium position, but if it does, it will do so only once. Critical damping is represented by Curve A in Figure 16.21. With less-than critical damping, the system will return to equilibrium faster but will overshoot and cross over one or more times. Such a system is **underdamped**; its displacement is represented by the curve in Figure 16.20. Curve B in Figure 16.21 represents an **overdamped** system. As with critical damping, it too may overshoot the equilibrium position, but will reach equilibrium over a longer period of time.

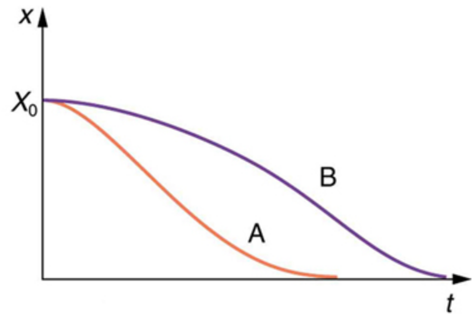


FIGURE 16.21 Displacement versus time for a critically damped harmonic oscillator (A) and an overdamped harmonic oscillator (B). The critically damped oscillator returns to equilibrium at $X = 0$ in the smallest time possible without overshooting.

Critical damping is often desired, because such a system returns to equilibrium rapidly and remains at equilibrium as well. In addition, a constant force applied to a critically damped system moves the system to a new equilibrium position in the shortest time possible without overshooting or oscillating about the new position. For example, when you stand on bathroom scales that have a needle gauge, the needle moves to its equilibrium position without oscillating. It would be quite inconvenient if the needle oscillated about the new equilibrium position for a long time before settling. Damping forces can vary greatly in character. Friction, for example, is sometimes independent of velocity (as assumed in most places in this text). But many damping forces depend on velocity—sometimes in complex ways, sometimes simply being proportional to velocity.

EXAMPLE 16.7

Damping an Oscillatory Motion: Friction on an Object Connected to a Spring

Damping oscillatory motion is important in many systems, and the ability to control the damping is even more so. This is generally attained using non-conservative forces such as the friction between surfaces, and viscosity for objects moving through fluids. The following example considers friction. Suppose a 0.200-kg object is connected to a spring as shown in Figure 16.22, but there is simple friction between the object and the surface, and the coefficient of friction μ_k is equal to 0.0800. (a) What is the frictional force between the surfaces? (b) What total distance does the object travel if it is released 0.100 m from equilibrium, starting at $v = 0$? The force constant of the spring is $k = 50.0 \text{ N/m}$.

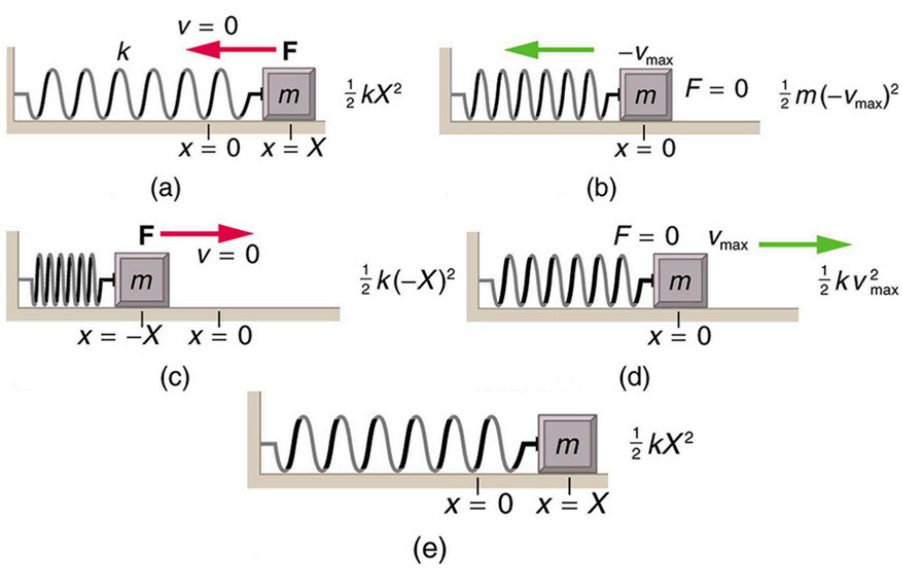


FIGURE 16.22 The transformation of energy in simple harmonic motion is illustrated for an object attached to a spring on a frictionless surface.

Strategy

This problem requires you to integrate your knowledge of various concepts regarding waves, oscillations, and damping. To solve an integrated concept problem, you must first identify the physical principles involved. Part (a) is about the frictional force. This is a topic involving the application of Newton's Laws. Part (b) requires an understanding of work and conservation of energy, as well as some understanding of horizontal oscillatory systems.

Now that we have identified the principles we must apply in order to solve the problems, we need to identify the knowns and unknowns for each part of the question, as well as the quantity that is constant in Part (a) and Part (b) of the question.

Solution a

1. Choose the proper equation: Friction is $f = \mu_k mg$.
2. Identify the known values.
3. Enter the known values into the equation:

$$f = (0.0800)(0.200 \text{ kg})(9.80 \text{ m/s}^2). \quad 16.58$$

4. Calculate and convert units: $f = 0.157 \text{ N}$.

Discussion a

The force here is small because the system and the coefficients are small.

Solution b

Identify the known:

- The system involves elastic potential energy as the spring compresses and expands, friction that is related to the work done, and the kinetic energy as the body speeds up and slows down.
- Energy is not conserved as the mass oscillates because friction is a non-conservative force.
- The motion is horizontal, so gravitational potential energy does not need to be considered.
- Because the motion starts from rest, the energy in the system is initially $\text{PE}_{\text{el,i}} = (1/2)kX^2$. This energy is removed by work done by friction $W_{\text{nc}} = -fd$, where d is the total distance traveled and $f = \mu_k mg$ is the force of friction. When the system stops moving, the friction force will balance the force exerted by the spring, so $\text{PE}_{\text{e1,f}} = (1/2)kx^2$ where x is the final position and is given by

$$\begin{aligned} F_{\text{el}} &= f \\ kx &= \mu_k mg \,. \\ x &= \frac{\mu_k mg}{k} \end{aligned} \quad 16.59$$

1. By equating the work done to the energy removed, solve for the distance d .
2. The work done by the non-conservative forces equals the initial, stored elastic potential energy. Identify the correct equation to use:

$$\text{W}_{\text{nc}} = \Delta(\text{KE} + \text{PE}) = \text{PE}_{\text{el,f}} - \text{PE}_{\text{el,i}} = \frac{1}{2}k\left(\left(\frac{\mu_k mg}{k}\right)^2 - X^2\right). \quad 16.60$$

3. Recall that $W_{\text{nc}} = -fd$.
4. Enter the friction as $f = \mu_k mg$ into $W_{\text{nc}} = -fd$, thus

$$W_{\text{nc}} = -\mu_k mgd. \quad 16.61$$

5. Combine these two equations to find

$$\frac{1}{2}k\left(\left(\frac{\mu_k mg}{k}\right)^2 - X^2\right) = -\mu_k mgd. \quad 16.62$$

6. Solve the equation for d :

$$d = \frac{k}{2\mu_k mg}\left(X^2 - \left(\frac{\mu_k mg}{k}\right)^2\right). \quad 16.63$$

7. Enter the known values into the resulting equation:

$$d = \frac{50.0\ \text{N/m}}{2(0.0800)(0.200\ \text{kg})(9.80\ \text{m/s}^2)}\left((0.100\ \text{m})^2 - \left(\frac{(0.0800)(0.200\ \text{kg})(9.80\ \text{m/s}^2)}{50.0\ \text{N/m}}\right)^2\right). \quad 16.64$$

8. Calculate d and convert units:

$$d = 1.59\ \text{m}. \quad 16.65$$

Discussion b

This is the total distance traveled back and forth across $x = 0$, which is the undamped equilibrium position. The number of oscillations about the equilibrium position will be more than $d/X = (1.59\ \text{m})/(0.100\ \text{m}) = 15.9$ because the amplitude of the oscillations is decreasing with time. At the end of the motion, this system will not return to $x = 0$ for this type of damping force, because static friction will exceed the restoring force. This system is underdamped. In contrast, an overdamped system with a simple constant damping force would not cross the equilibrium position $x = 0$ a single time. For example, if this system had a damping force 20 times greater, it would only move 0.0484 m toward the equilibrium position from its original 0.100-m position.

This worked example illustrates how to apply problem-solving strategies to situations that integrate the different concepts you have learned. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknowns using familiar problem-solving strategies. These are found throughout the text, and many worked examples show how to use them for single topics. In this integrated concepts example, you can see how to apply them across several topics. You will find these techniques useful in applications of physics outside a physics course, such as in your profession, in other science disciplines, and in everyday life.

CHECK YOUR UNDERSTANDING

Why are completely undamped harmonic oscillators so rare?

Solution

Friction often comes into play whenever an object is moving. Friction causes damping in a harmonic oscillator.

CHECK YOUR UNDERSTANDING

Describe the difference between overdamping, underdamping, and critical damping.

Solution

An overdamped system moves slowly toward equilibrium. An underdamped system moves quickly to equilibrium, but will oscillate about the equilibrium point as it does so. A critically damped system moves as quickly as possible toward equilibrium without oscillating about the equilibrium.

16.8 Forced Oscillations and Resonance

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Observe resonance of a paddle ball on a string.
- Observe amplitude of a damped harmonic oscillator.

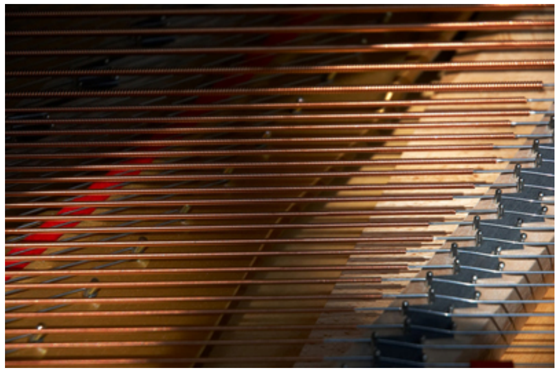


FIGURE 16.23 You can cause the strings in a piano to vibrate simply by producing sound waves from your voice. (credit: Matt Billings, Flickr)

Sit in front of a piano sometime and sing a loud brief note at it with the dampers off its strings. It will sing the same note back at you—the strings, having the same frequencies as your voice, are resonating in response to the forces from the sound waves that you sent to them. Your voice and a piano's strings is a good example of the fact that objects—in this case, piano strings—can be forced to oscillate but oscillate best at their natural frequency. In this section, we shall briefly explore applying a *periodic driving force* acting on a simple harmonic oscillator. The driving force puts energy into the system at a certain frequency, not necessarily the same as the natural frequency of the system. The **natural frequency** is the frequency at which a system would oscillate if there were no driving and no damping force.

Most of us have played with toys involving an object supported on an elastic band, something like the paddle ball suspended from a finger in Figure 16.24. Imagine the finger in the figure is your finger. At first you hold your finger steady, and the ball bounces up and down with a small amount of damping. If you move your finger up and down slowly, the ball will follow along without bouncing much on its own. As you increase the frequency at which you move your finger up and down, the ball will respond by oscillating with increasing amplitude. When you drive the ball at its natural frequency, the ball's oscillations increase in amplitude with each oscillation for as long as you drive it. The phenomenon of driving a system with a frequency equal to its natural frequency is called **resonance**. A system being driven at its natural frequency is said to **resonate**. As the driving frequency gets progressively higher than the resonant or natural frequency, the amplitude of the oscillations becomes smaller, until the oscillations nearly disappear and your finger simply moves up and down with little effect on the ball.

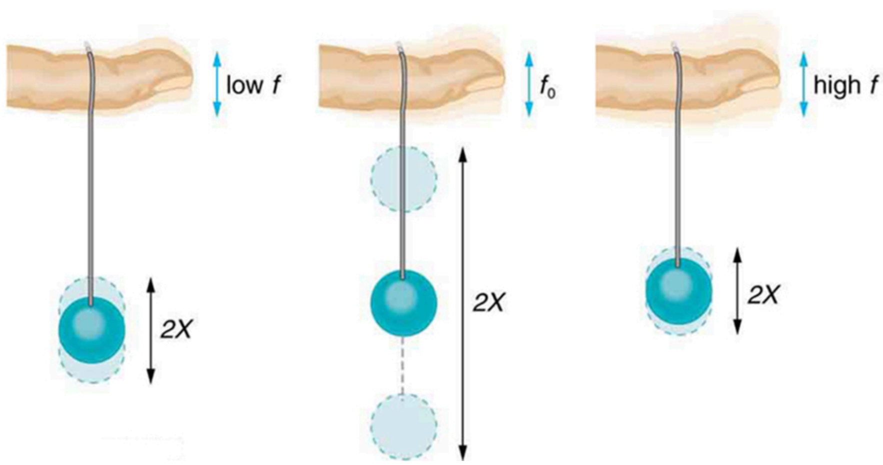


FIGURE 16.24 The paddle ball on its rubber band moves in response to the finger supporting it. If the finger moves with the natural frequency f_0 of the ball on the rubber band, then a resonance is achieved, and the amplitude of the ball's oscillations increases dramatically. At higher and lower driving frequencies, energy is transferred to the ball less efficiently, and it responds with lower-amplitude oscillations.

Figure 16.25 shows a graph of the amplitude of a damped harmonic oscillator as a function of the frequency of the periodic force driving it. There are three curves on the graph, each representing a different amount of damping. All three curves peak at the point where the frequency of the driving force equals the natural frequency of the harmonic oscillator. The highest peak, or greatest response, is for the least amount of damping, because less energy is removed by the damping force.

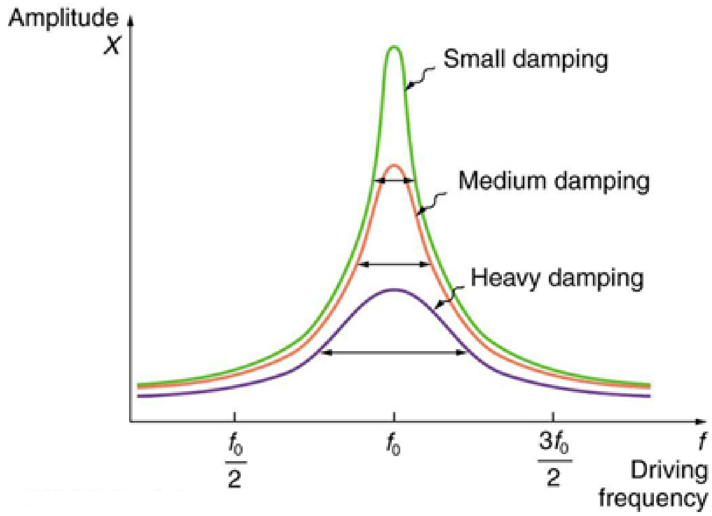


FIGURE 16.25 Amplitude of a harmonic oscillator as a function of the frequency of the driving force. The curves represent the same oscillator with the same natural frequency but with different amounts of damping. Resonance occurs when the driving frequency equals the natural frequency, and the greatest response is for the least amount of damping. The narrowest response is also for the least damping.

It is interesting that the widths of the resonance curves shown in Figure 16.25 depend on damping: the less the damping, the narrower the resonance. The message is that if you want a driven oscillator to resonate at a very specific frequency, you need as little damping as possible. Little damping is the case for piano strings and many other musical instruments. Conversely, if you want small-amplitude oscillations, such as in a car's suspension system, then you want heavy damping. Heavy damping reduces the amplitude, but the tradeoff is that the system responds at more frequencies.

These features of driven harmonic oscillators apply to a huge variety of systems. When you tune a radio, for example, you are adjusting its resonant frequency so that it only oscillates to the desired station's broadcast (driving) frequency. The more selective the radio is in discriminating between stations, the smaller its damping. Magnetic resonance imaging (MRI) is a widely used medical diagnostic tool in which atomic nuclei (mostly hydrogen nuclei) are made to resonate by incoming radio waves (on the order of 100 MHz). A child on a swing is driven by a parent at the swing's natural frequency to achieve maximum amplitude. In all of these cases, the efficiency of energy transfer from the driving force into the oscillator is best at resonance. Speed bumps and gravel roads prove that even a car's suspension system is not immune to resonance. In spite of finely engineered shock absorbers, which ordinarily convert mechanical energy to thermal energy almost as fast as it comes in, speed bumps still cause a large-amplitude oscillation. On gravel roads that are corrugated, you may have noticed that if you travel at the "wrong" speed, the bumps are very noticeable whereas at other speeds you may hardly feel the bumps at all. Figure 16.26 shows a photograph of a famous example (the Tacoma Narrows Bridge) of the destructive effects of a driven harmonic oscillation. The Millennium Bridge in London was closed for a short period of time for the same reason while inspections were carried out.

In our bodies, the chest cavity is a clear example of a system at resonance. The diaphragm and chest wall drive the oscillations of the chest cavity which result in the lungs inflating and deflating. The system is critically damped and the muscular diaphragm oscillates at the resonant value for the system, making it highly efficient.

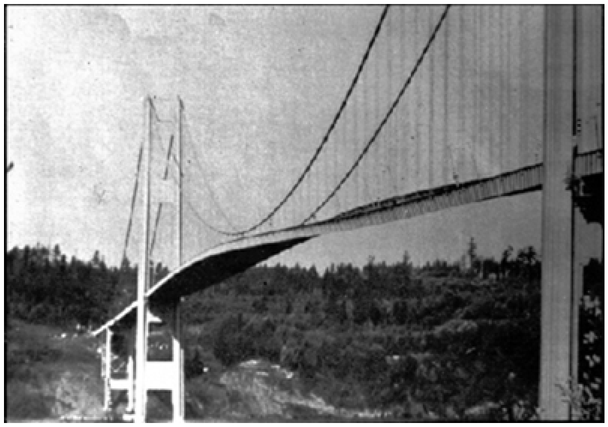


FIGURE 16.26 In 1940, the Tacoma Narrows Bridge in Washington state collapsed. Heavy cross winds drove the bridge into oscillations at its resonant frequency. Damping decreased when support cables broke loose and started to slip over the towers, allowing increasingly greater amplitudes until the structure failed (credit: PRI's *Studio 360*, via Flickr)

CHECK YOUR UNDERSTANDING

A famous magic trick involves a performer singing a note toward a crystal glass until the glass shatters. Explain why the trick works in terms of resonance and natural frequency.

Solution

The performer must be singing a note that corresponds to the natural frequency of the glass. As the sound wave is directed at the glass, the glass responds by resonating at the same frequency as the sound wave. With enough energy introduced into the system, the glass begins to vibrate and eventually shatters.

16.9 Waves

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- State the characteristics of a wave.
- Calculate the velocity of wave propagation.

FIGURE 16.27 Waves in the ocean behave similarly to all other types of waves. (credit: Steve Jurveston, Flickr)

What do we mean when we say something is a wave? The most intuitive and easiest wave to imagine is the familiar water wave. More precisely, a **wave** is a disturbance that propagates, or moves from the place it was created. For water waves, the disturbance is in the surface of the water, perhaps created by a rock thrown into a pond or by a swimmer splashing the surface repeatedly. For sound waves, the disturbance is a change in air pressure, perhaps created by the oscillating cone inside a speaker. For earthquakes, there are several types of disturbances, including disturbance of Earth’s surface and pressure disturbances under the surface. Even radio waves are most easily understood using an analogy with water waves. Visualizing water waves is useful because there is more to it than just a mental image. Water waves exhibit characteristics common to all waves, such as amplitude, period, frequency and energy. All wave characteristics can be described by a small set of underlying principles.

A wave is a disturbance that propagates, or moves from the place it was created. The simplest waves repeat themselves for several cycles and are associated with simple harmonic motion. Let us start by considering the simplified water wave in Figure 16.28. The wave is an up and down disturbance of the water surface. It causes a sea gull to move up and down in simple harmonic motion as the wave crests and troughs (peaks and valleys) pass under the bird. The time for one complete up and down motion is the wave's period T. The wave's frequency is $f = 1/T$, as usual. The wave itself moves to the right in the figure. This movement of the wave is actually the disturbance moving to the right, not the water itself (or the bird would move to the right). We define **wave velocity** v_{w} to be the speed at which the disturbance moves. Wave velocity is sometimes also called the *propagation velocity or propagation speed,* because the disturbance propagates from one location to another.

Misconception Alert

Many people think that water waves push water from one direction to another. In fact, the particles of water tend to stay in one location, save for moving up and down due to the energy in the wave. The energy moves forward through the water, but the water stays in one place. If you feel yourself pushed in an ocean, what you feel is the energy of the wave, not a rush of water.

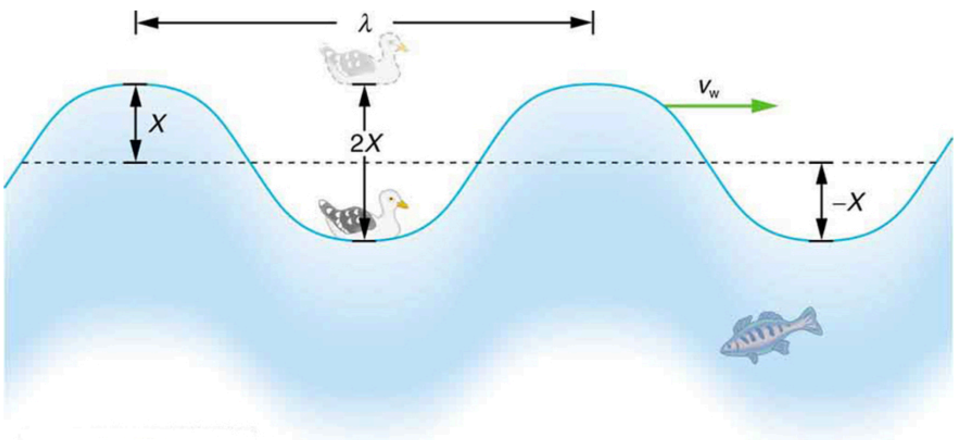


FIGURE 16.28 An idealized ocean wave passes under a sea gull that bobs up and down in simple harmonic motion. The wave has a wavelength λ, which is the distance between adjacent identical parts of the wave. The up and down disturbance of the surface propagates parallel to the surface at a speed v_{w}.

The water wave in the figure also has a length associated with it, called its **wavelength** λ, the distance between adjacent identical parts of a wave. (λ is the distance parallel to the direction of propagation.) The speed of propagation v_{w} is the distance the wave travels in a given time, which is one wavelength in the time of one period. In equation form, that is

$$v_{\mathrm{w}} = \frac{\lambda}{T} \tag{16.66}$$

or

$$v_{\mathrm{w}} = f\lambda. \tag{16.67}$$

This fundamental relationship holds for all types of waves. For water waves, v_{w} is the speed of a surface wave; for sound, v_{w} is the speed of sound; and for visible light, v_{w} is the speed of light, for example.

Take-Home Experiment: Waves in a Bowl

Fill a large bowl or basin with water and wait for the water to settle so there are no ripples. Gently drop a cork into the middle of the bowl. Estimate the wavelength and period of oscillation of the water wave that propagates away from the cork. Remove the cork from the bowl and wait for the water to settle again. Gently drop the cork at a height that is different from the first drop. Does the wavelength depend upon how high above the water the cork is dropped?

EXAMPLE 16.8

Calculate the Velocity of Wave Propagation: Gull in the Ocean

Calculate the wave velocity of the ocean wave in Figure 16.28 if the distance between wave crests is 10.0 m and the time for a sea gull to bob up and down is 5.00 s.

Strategy

We are asked to find v_{w}. The given information tells us that $\lambda = 10.0 \text{ m}$ and $T = 5.00 \text{ s}$. Therefore, we can use $v_{\text{w}} = \frac{\lambda}{T}$ to find the wave velocity.

Solution

1. Enter the known values into $v_{\text{w}} = \frac{\lambda}{T}$:

$$v_{\text{w}} = \frac{10.0 \text{ m}}{5.00 \text{ s}}. \tag{16.68}$$

2. Solve for v_{w} to find $v_{\text{w}} = 2.00 \text{ m/s}$.

Discussion

This slow speed seems reasonable for an ocean wave. Note that the wave moves to the right in the figure at this speed, not the varying speed at which the sea gull moves up and down.

Transverse and Longitudinal Waves

A simple wave consists of a periodic disturbance that propagates from one place to another. The wave in Figure 16.29 propagates in the horizontal direction while the surface is disturbed in the vertical direction. Such a wave is called a **transverse wave** or shear wave; in such a wave, the disturbance is perpendicular to the direction of propagation. In contrast, in a **longitudinal wave** or compressional wave, the disturbance is parallel to the direction of propagation. Figure 16.30 shows an example of a longitudinal wave. The size of the disturbance is its amplitude *X* and is completely independent of the speed of propagation v_{w}.

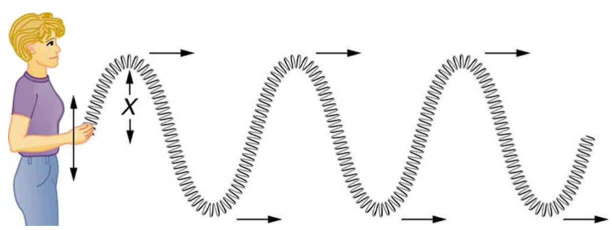


FIGURE 16.29 In this example of a transverse wave, the wave propagates horizontally, and the disturbance in the cord is in the vertical direction.

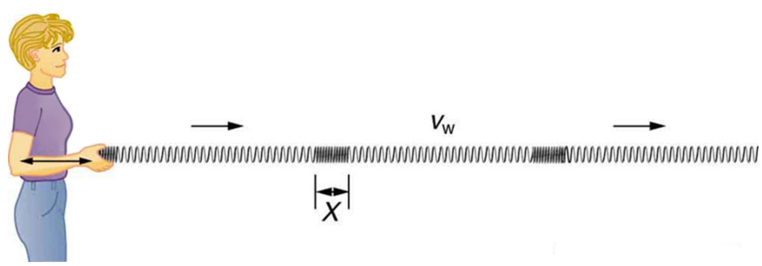


FIGURE 16.30 In this example of a longitudinal wave, the wave propagates horizontally, and the disturbance in the cord is also in the horizontal direction.

Waves may be transverse, longitudinal, or *a combination of the two*. (Water waves are actually a combination of transverse and longitudinal. The simplified water wave illustrated in Figure 16.28 shows no longitudinal motion of the bird.) The waves on the strings of musical instruments are transverse—so are electromagnetic waves, such as visible light.

Sound waves in air and water are longitudinal. Their disturbances are periodic variations in pressure that are transmitted in fluids. Fluids do not have appreciable shear strength, and thus the sound waves in them must be longitudinal or compressional. Sound in solids can be both longitudinal and transverse.

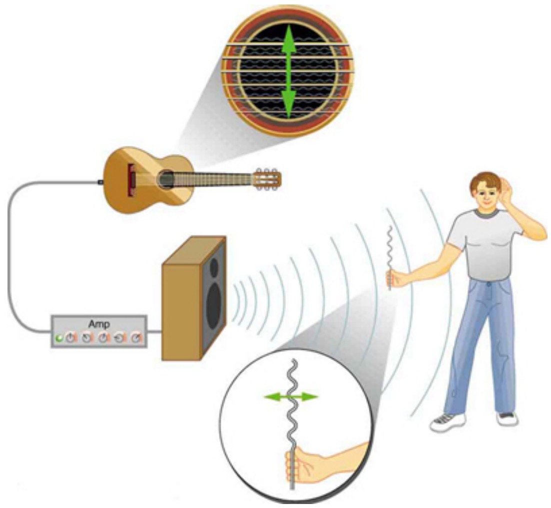


FIGURE 16.31 The wave on a guitar string is transverse. The sound wave rattles a sheet of paper in a direction that shows the sound wave is longitudinal.

Earthquake waves under Earth's surface also have both longitudinal and transverse components (called compressional or P-waves and shear or S-waves, respectively). These components have important individual characteristics—they propagate at different speeds, for example. Earthquakes also have surface waves that are similar to surface waves on water.

CHECK YOUR UNDERSTANDING

Why is it important to differentiate between longitudinal and transverse waves?

Solution

In the different types of waves, energy can propagate in a different direction relative to the motion of the wave. This is important to understand how different types of waves affect the materials around them.

PHET EXPLORATIONS

Wave on a String

Watch a string vibrate in slow motion. Wiggle the end of the string and make waves, or adjust the frequency and amplitude of an oscillator. Adjust the damping and tension. The end can be fixed, loose, or open.

Click to view content (https://openstax.org/books/college-physics-2e/pages/16-9-waves)

16.10 Superposition and Interference

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain standing waves.
- Describe the mathematical representation of overtones and beat frequency.

FIGURE 16.32 These waves result from the superposition of several waves from different sources, producing a complex pattern. (credit: waterborough, Wikimedia Commons)

Most waves do not look very simple. They look more like the waves in Figure 16.32 than like the simple water wave considered in Waves. (Simple waves may be created by a simple harmonic oscillation, and thus have a sinusoidal shape). Complex waves are more interesting, even beautiful, but they look formidable. Most waves appear complex because they result from several simple waves adding together. Luckily, the rules for adding waves are quite simple.

When two or more waves arrive at the same point, they superimpose themselves on one another. More specifically, the disturbances of waves are superimposed when they come together—a phenomenon called **superposition**. Each disturbance corresponds to a force, and forces add. If the disturbances are along the same line, then the resulting wave is a simple addition of the disturbances of the individual waves—that is, their amplitudes add. Figure 16.33 and Figure 16.34 illustrate superposition in two special cases, both of which produce simple results.

Figure 16.33 shows two identical waves that arrive at the same point exactly in phase. The crests of the two waves are precisely aligned, as are the troughs. This superposition produces pure **constructive interference**. Because the disturbances add, pure constructive interference produces a wave that has twice the amplitude of the individual waves, but has the same wavelength.

Figure 16.34 shows two identical waves that arrive exactly out of phase—that is, precisely aligned crest to trough—producing pure **destructive interference**. Because the disturbances are in the opposite direction for this superposition, the resulting amplitude is zero for pure destructive interference—the waves completely cancel.

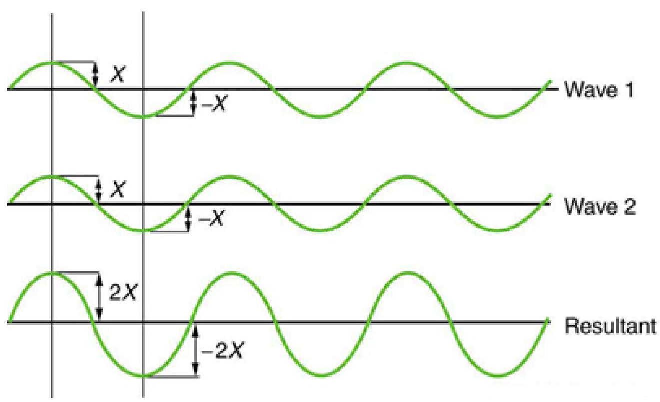


FIGURE 16.33 Pure constructive interference of two identical waves produces one with twice the amplitude, but the same wavelength.

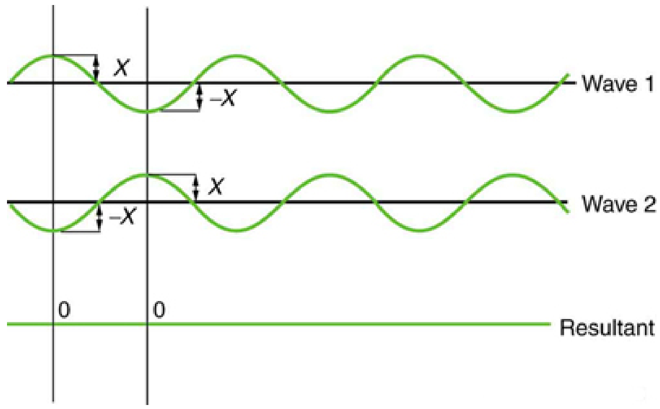


FIGURE 16.34 Pure destructive interference of two identical waves produces zero amplitude, or complete cancellation.

While pure constructive and pure destructive interference do occur, they require precisely aligned identical waves. The superposition of most waves produces a combination of constructive and destructive interference and can vary from place to place and time to time. Sound from a stereo, for example, can be loud in one spot and quiet in another. Varying loudness means the sound waves add partially constructively and partially destructively at different locations. A stereo has at least two speakers creating sound waves, and waves can reflect from walls. All these waves superimpose. An example of sounds that vary over time from constructive to destructive is found in the combined whine of airplane jets heard by a stationary passenger. The combined sound can fluctuate up and down in volume as the sound from the two engines varies in time from constructive to destructive. These examples are of waves that are similar.

An example of the superposition of two dissimilar waves is shown in Figure 16.35. Here again, the disturbances add and subtract, producing a more complicated looking wave.

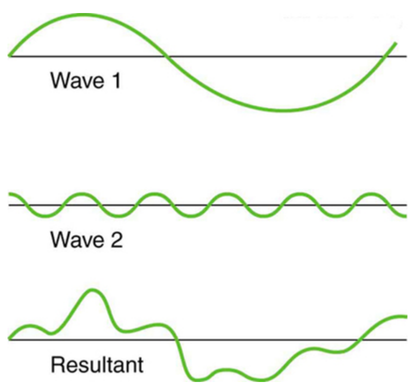


FIGURE 16.35 Superposition of non-identical waves exhibits both constructive and destructive interference.

Standing Waves

Sometimes waves do not seem to move; rather, they just vibrate in place. Unmoving waves can be seen on the surface of a glass of milk in a refrigerator, for example. Vibrations from the refrigerator motor create waves on the milk that oscillate up and down but do not seem to move across the surface. These waves are formed by the superposition of two or more moving waves, such as illustrated in Figure 16.36 for two identical waves moving in opposite directions. The waves move through each other with their disturbances adding as they go by. If the two waves have the same amplitude and wavelength, then they alternate between constructive and destructive interference. The resultant looks like a wave standing in place and, thus, is called a **standing wave**. Waves on the glass of milk are one example of standing waves. There are other standing waves, such as on guitar strings and in organ pipes. With the glass of milk, the two waves that produce standing waves may come from reflections from the side of the glass.

A closer look at earthquakes provides evidence for conditions appropriate for resonance, standing waves, and constructive and destructive interference. A building may be vibrated for several seconds with a driving frequency matching that of the natural frequency of vibration of the building—producing a resonance resulting in one building collapsing while neighboring buildings do not. Often buildings of a certain height are devastated while other taller buildings remain intact. The building height matches the condition for setting up a standing wave for that particular height. As the earthquake waves travel along the surface of Earth and reflect off denser rocks, constructive interference occurs at certain points. Often areas closer to the epicenter are not damaged while areas farther away are damaged.

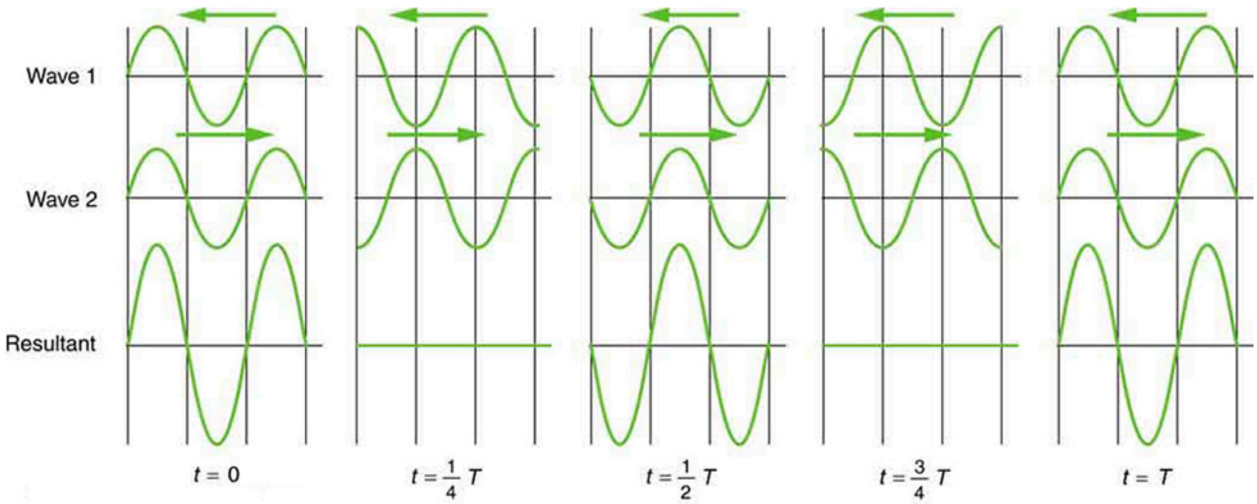


FIGURE 16.36 Standing wave created by the superposition of two identical waves moving in opposite directions. The oscillations are at fixed locations in space and result from alternately constructive and destructive interference.

Standing waves are also found on the strings of musical instruments and are due to reflections of waves from the ends of the string. Figure 16.37 and Figure 16.38 show three standing waves that can be created on a string that is fixed at both ends. **Nodes** are the points where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave. The fixed ends of strings must be nodes, too, because the string cannot move there. The word **antinode** is used to denote the location of maximum amplitude in standing waves. Standing waves on strings have a frequency that is related to the propagation speed v_{w} of the disturbance on the string. The wavelength λ is determined by the distance between the points where the string is fixed in place.

The lowest frequency, called the **fundamental frequency**, is thus for the longest wavelength, which is seen to be $\lambda_1 = 2L$. Therefore, the fundamental frequency is $f_1 = v_{\mathrm{w}}/\lambda_1 = v_{\mathrm{w}}/2L$. In this case, the **overtones** or harmonics are multiples of the fundamental frequency. As seen in Figure 16.38, the first harmonic can easily be calculated since $\lambda_2 = L$. Thus, $f_2 = v_{\mathrm{w}}/\lambda_2 = v_{\mathrm{w}}/2L = 2f_1$. Similarly, $f_3 = 3f_1$, and so on. All of these frequencies can be changed by adjusting the tension in the string. The greater the tension, the greater v_{w} is and the higher the frequencies. This observation is familiar to anyone who has ever observed a string instrument being tuned. We will see in later chapters that standing waves are crucial to many resonance phenomena, such as in sounding boxes on string instruments.

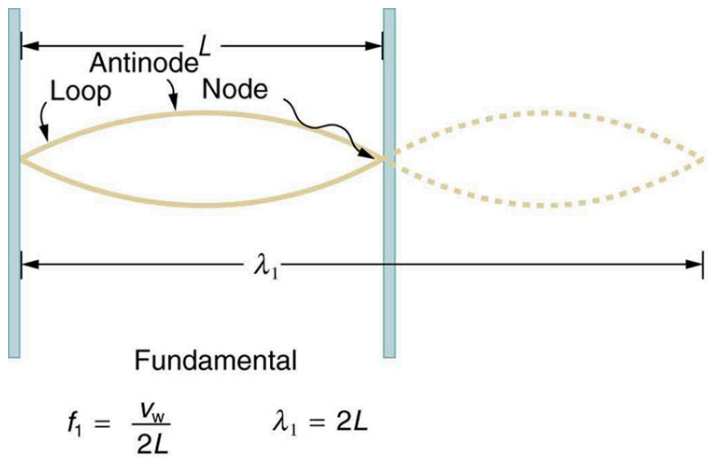


FIGURE 16.37 The figure shows a string oscillating at its fundamental frequency.

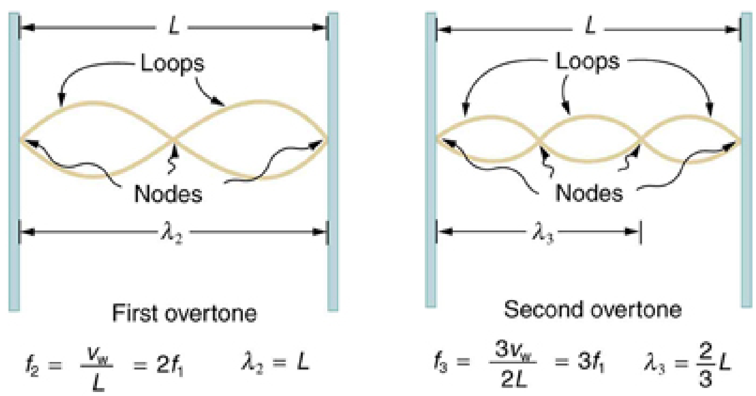


FIGURE 16.38 First and second overtones are shown.

Beats

Striking two adjacent keys on a piano produces a warbling combination usually considered to be unpleasant. The superposition of two waves of similar but not identical frequencies is the culprit. Another example is often noticeable in jet aircraft, particularly the two-engine variety, while taxiing. The combined sound of the engines goes up and down in loudness. This varying loudness happens because the sound waves have similar but not identical frequencies. The discordant warbling of the piano and the fluctuating loudness of the jet engine noise are both due to alternately constructive and destructive interference as the two waves go in and out of phase. Figure 16.39 illustrates this graphically.

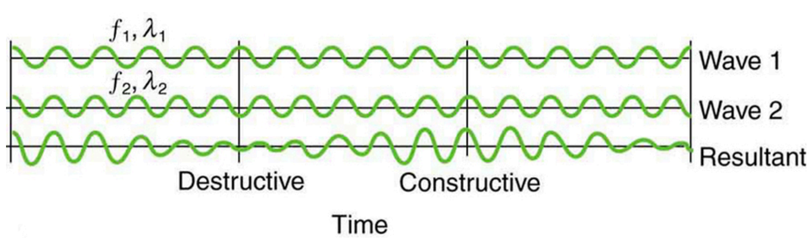


FIGURE 16.39 Beats are produced by the superposition of two waves of slightly different frequencies but identical amplitudes. The waves alternate in time between constructive interference and destructive interference, giving the resulting wave a time-varying amplitude.

The wave resulting from the superposition of two similar-frequency waves has a frequency that is the average of the two. This wave fluctuates in amplitude, or *beats*, with a frequency called the **beat frequency**. We can determine the beat frequency by adding two waves together mathematically. Note that a wave can be represented at one point in space as

$$x = X\cos\left(\frac{2\pi t}{T}\right) = X\cos(2\pi ft), \quad 16.69$$

where $f = 1/T$ is the frequency of the wave. Adding two waves that have different frequencies but identical amplitudes produces a resultant

$$x = x_1 + x_2. \quad 16.70$$

More specifically,

$$x = X\cos(2\pi f_1 t) + X\cos(2\pi f_2 t). \quad 16.71$$

Using a trigonometric identity, it can be shown that

$$x = 2X\cos(\pi f_B t)\cos(2\pi f_{ave} t), \quad 16.72$$

where

$$f_B = |f_1 - f_2| \quad 16.73$$

is the beat frequency, and f_{ave} is the average of f_1 and f_2. These results mean that the resultant wave has twice the amplitude and the average frequency of the two superimposed waves, but it also fluctuates in overall amplitude at the beat frequency f_B. The first cosine term in the expression effectively causes the amplitude to go up and down. The second cosine term is the wave with frequency f_{ave}. This result is valid for all types of waves. However, if it is a sound wave, providing the two frequencies are similar, then what we hear is an average frequency that gets louder and softer (or warbles) at the beat frequency.

Making Career Connections

Piano tuners use beats routinely in their work. When comparing a note with a tuning fork, they listen for beats and adjust the string until the beats go away (to zero frequency). For example, if the tuning fork has a 256 Hz frequency and two beats per second are heard, then the other frequency is either 254 or 258 Hz. Most keys hit multiple strings, and these strings are actually adjusted until they have nearly the same frequency and give a slow beat for richness. Twelve-string guitars and mandolins are also tuned using beats.

While beats may sometimes be annoying in audible sounds, we will find that beats have many applications. Observing beats is a very useful way to compare similar frequencies. There are applications of beats as apparently disparate as in ultrasonic imaging and radar speed traps.

CHECK YOUR UNDERSTANDING

Imagine you are holding one end of a jump rope, and your friend holds the other. If your friend holds her end still, you can move your end up and down, creating a transverse wave. If your friend then begins to move her end up and down, generating a wave in the opposite direction, what resultant wave forms would you expect to see in the jump rope?

Solution

The rope would alternate between having waves with amplitudes two times the original amplitude and reaching equilibrium with no amplitude at all. The wavelengths will result in both constructive and destructive interference

CHECK YOUR UNDERSTANDING

Define nodes and antinodes.

Solution

Nodes are areas of wave interference where there is no motion. Antinodes are areas of wave interference where the motion is at its maximum point.

CHECK YOUR UNDERSTANDING

You hook up a stereo system. When you test the system, you notice that in one corner of the room, the sounds seem dull. In another area, the sounds seem excessively loud. Describe how the sound moving about the room could result in these effects.

Solution

With multiple speakers putting out sounds into the room, and these sounds bouncing off walls, there is bound to be some wave interference. In the dull areas, the interference is probably mostly destructive. In the louder areas, the interference is probably mostly constructive.

PHET EXPLORATIONS

Wave Interference

Make waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern.

Click to view content (https://openstax.org/books/college-physics-2e/pages/16-10-superposition-and-interference)

16.11 Energy in Waves: Intensity

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate the intensity and the power of rays and waves.

FIGURE 16.40 The destructive effect of an earthquake is palpable evidence of the energy carried in these waves. The Richter scale rating of earthquakes is related to both their amplitude and the energy they carry. (credit: Petty Officer 2nd Class Candice Villarreal, U.S. Navy)

All waves carry energy. The energy of some waves can be directly observed. Earthquakes can shake whole cities to the ground, performing the work of thousands of wrecking balls.

Loud sounds pulverize nerve cells in the inner ear, causing permanent hearing loss. Ultrasound is used for deep-heat treatment of muscle strains. A laser beam can burn away a malignancy. Water waves chew up beaches.

The amount of energy in a wave is related to its amplitude. Large-amplitude earthquakes produce large ground displacements. Loud sounds have higher pressure amplitudes and come from larger-amplitude source vibrations than soft sounds. Large ocean breakers churn up the shore more than small ones. More quantitatively, a wave is a displacement that is resisted by a restoring force. The larger the displacement x, the larger the force $F = kx$ needed to create it. Because work W is related to force multiplied by distance (Fx) and energy is put into the wave by the work done to create it, the energy in a wave is related to amplitude. In fact, a wave's energy is directly proportional to its amplitude squared because

$$W \propto Fx = kx^2. \quad (16.74)$$

The energy effects of a wave depend on time as well as amplitude. For example, the longer deep-heat ultrasound is applied, the more energy it transfers. Waves can also be concentrated or spread out. Sunlight, for example, can be focused to burn wood. Earthquakes spread out, so they do less damage the farther they get from the source. In both cases, changing the area the waves cover has important effects. All these pertinent factors are included in the definition of **intensity** I as power per unit area:

$$I = \frac{P}{A} \quad (16.75)$$

where P is the power carried by the wave through area A. The definition of intensity is valid for any energy in transit, including that carried by waves. The SI unit for intensity is watts per square meter (W/m^2). For example, infrared and visible energy from the Sun impinge on Earth at an intensity of $1300\ \text{W/m}^2$ just above the atmosphere. There are other intensity-related units in use, too. The most common is the decibel. For example, a 90 decibel sound level corresponds to an intensity of $10^{-3}\ \text{W/m}^2$. (This quantity is not much power per unit area considering that 90 decibels is a relatively high sound level. Decibels will be discussed in some detail in a later chapter.

EXAMPLE 16.9

Calculating intensity and power: How much energy is in a ray of sunlight?

The average intensity of sunlight on Earth's surface is about $700\ \text{W/m}^2$.

(a) Calculate the amount of energy that falls on a solar collector having an area of $0.500\ \text{m}^2$ in $4.00\ \text{h}$.

(b) What intensity would such sunlight have if concentrated by a magnifying glass onto an area 200 times smaller than its own?

Strategy a

Because power is energy per unit time or $P = \frac{E}{t}$, the definition of intensity can be written as $I = \frac{P}{A} = \frac{E/t}{A}$, and this equation can be solved for E with the given information.

Solution a

1. Begin with the equation that states the definition of intensity:

$$I = \frac{P}{A}. \quad (16.76)$$

2. Replace P with its equivalent E/t:

$$I = \frac{E/t}{A}. \quad (16.77)$$

3. Solve for E:

$$E = IAt. \quad (16.78)$$

4. Substitute known values into the equation:

$$E = \left(700\ \text{W/m}^2\right)\left(0.500\ \text{m}^2\right)[(4.00\ \text{h})(3600\ \text{s/h})]. \quad (16.79)$$

5. Calculate to find E and convert units:

$$5.04 \times 10^6\ \text{J}, \quad (16.80)$$

Discussion a

The energy falling on the solar collector in 4 h in part is enough to be useful—for example, for heating a significant amount of water.

Strategy b

Taking a ratio of new intensity to old intensity and using primes for the new quantities, we will find that it depends on the ratio of the areas. All other quantities will cancel.

Solution b

1. Take the ratio of intensities, which yields:

$$\frac{I'}{I} = \frac{P'/A'}{P/A} = \frac{A}{A'} \left(\text{The powers cancel because } P' = P\right). \quad 16.81$$

2. Identify the knowns:

$$A = 200A', \quad 16.82$$

$$\frac{I'}{I} = 200. \quad 16.83$$

3. Substitute known quantities:

$$I' = 200I = 200\left(700 \text{ W/m}^2\right). \quad 16.84$$

4. Calculate to find I':

$$I' = 1.40 \times 10^5 \text{ W/m}^2. \quad 16.85$$

Discussion b

Decreasing the area increases the intensity considerably. The intensity of the concentrated sunlight could even start a fire.

EXAMPLE 16.10

Determine the combined intensity of two waves: Perfect constructive interference

If two identical waves are spatially separated, each having an intensity of 1.00 W/m^2, interfere perfectly constructively in a particular location, what is the intensity of the wave at this particular location?

Strategy

We know from Superposition and Interference that when two identical waves, which have equal amplitudes X, interfere perfectly constructively, the resulting wave has an amplitude of $2X$. Because a wave's intensity is proportional to amplitude squared, the intensity of the resulting wave is four times as great as in the individual waves.

Solution

1. Recall that intensity is proportional to amplitude squared.
2. Calculate the new amplitude:

$$I' \propto \left(X'\right)^2 = (2X)^2 = 4X^2. \quad 16.86$$

3. Recall that the intensity of the old amplitude was:

$$I \propto X^2. \quad 16.87$$

4. Take the ratio of new intensity to the old intensity. This gives:

$$\frac{I'}{I} = 4. \quad 16.88$$

5. Calculate to find I':

$$I' = 4I = 4.00 \text{ W/m}^2. \quad 16.89$$

Discussion

The intensity goes up by a factor of 4 when the amplitude doubles. This answer is a little disquieting. The two individual waves each have intensities of 1.00 W/m^2, yet their sum has an intensity of 4.00 W/m^2, which may appear to violate conservation of energy. This violation, of course, cannot happen. What does happen is intriguing. The area over which the intensity is 4.00 W/m^2 is much less than the area covered by the two waves before they interfered. There are other areas where the intensity is zero. The addition of waves is not as simple as our first look in Superposition and Interference suggested. We actually get a pattern of both constructive interference and destructive interference whenever two waves are added. For example, if we have two stereo speakers putting out $1.00\ \text{W/m}^2$ each, there will be places in the room where the intensity is $4.00\ \text{W/m}^2$, other places where the intensity is zero, and others in between. Figure 16.41 shows what this interference might look like. We will pursue interference patterns elsewhere in this text.

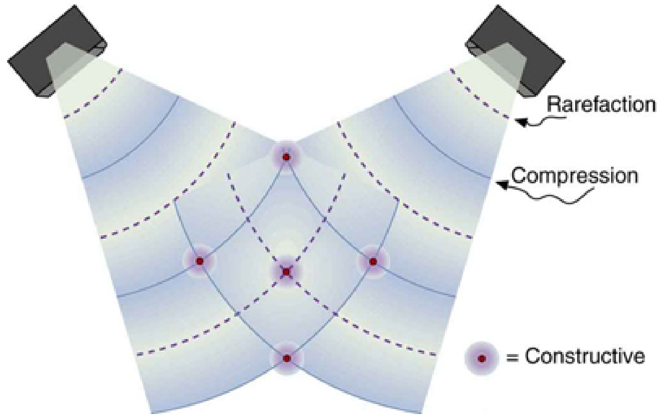


FIGURE 16.41 These stereo speakers produce both constructive interference and destructive interference in the room, a property common to the superposition of all types of waves. The shading is proportional to intensity.

CHECK YOUR UNDERSTANDING

Which measurement of a wave is most important when determining the wave's intensity?

Solution

Amplitude, because a wave’s energy is directly proportional to its amplitude squared.

Glossary

amplitude the maximum displacement from the equilibrium position of an object oscillating around the equilibrium position

antinode the location of maximum amplitude in standing waves

beat frequency the frequency of the amplitude fluctuations of a wave

constructive interference when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs

critical damping the condition in which the damping of an oscillator causes it to return as quickly as possible to its equilibrium position without oscillating back and forth about this position

deformation displacement from equilibrium

destructive interference when two identical waves arrive at the same point exactly out of phase; that is, precisely aligned crest to trough

elastic potential energy potential energy stored as a result of deformation of an elastic object, such as the stretching of a spring

force constant a constant related to the rigidity of a system: the larger the force constant, the more rigid the system; the force constant is represented by k

frequency number of events per unit of time

fundamental frequency the lowest frequency of a periodic waveform

intensity power per unit area

longitudinal wave a wave in which the disturbance is parallel to the direction of propagation

natural frequency the frequency at which a system would oscillate if there were no driving and no damping forces

nodes the points where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave

oscillate moving back and forth regularly between two points

over damping the condition in which damping of an oscillator causes it to return to equilibrium without oscillating; oscillator moves more slowly toward equilibrium than in the critically damped system

overtones multiples of the fundamental frequency of a sound

period time it takes to complete one oscillation

periodic motion motion that repeats itself at regular time intervals

resonance the phenomenon of driving a system with a frequency equal to the system's natural frequency

resonate a system being driven at its natural frequency

restoring force force acting in opposition to the force caused by a deformation

simple harmonic motion the oscillatory motion in a system where the net force can be described by Hooke's law

simple harmonic oscillator a device that implements Hooke's law, such as a mass that is attached to a spring, with the other end of the spring being connected to a rigid support such as a wall

simple pendulum an object with a small mass suspended from a light wire or string

superposition the phenomenon that occurs when two or more waves arrive at the same point

transverse wave a wave in which the disturbance is perpendicular to the direction of propagation

under damping the condition in which damping of an oscillator causes it to return to equilibrium with the amplitude gradually decreasing to zero; system returns to equilibrium faster but overshoots and crosses the equilibrium position one or more times

wave a disturbance that moves from its source and carries energy

wave velocity the speed at which the disturbance moves. Also called the propagation velocity or propagation speed

wavelength the distance between adjacent identical parts of a wave

Section Summary

16.1 Hooke's Law: Stress and Strain Revisited

- An oscillation is a back and forth motion of an object between two points of deformation.
- An oscillation may create a wave, which is a disturbance that propagates from where it was created.
- The simplest type of oscillations and waves are related to systems that can be described by

  Hooke's law:
  $F = -kx,$
  where F is the restoring force, x is the displacement from equilibrium or deformation, and k is the force constant of the system.
- Elastic potential energy PE_{el} stored in the deformation of a system that can be described by Hooke's law is given by
  $PE_{el} = (1/2)kx^2.$

16.2 Period and Frequency in Oscillations

- Periodic motion is a repetitious oscillation.
- The time for one oscillation is the period T.
- The number of oscillations per unit time is the frequency f.
- These quantities are related by
  $f = \frac{1}{T}$.

16.3 Simple Harmonic Motion: A Special Periodic Motion

- Simple harmonic motion is oscillatory motion for a system that can be described only by Hooke's law. Such a system is also called a simple harmonic oscillator.
- Maximum displacement is the amplitude X. The period T and frequency f of a simple harmonic oscillator are given by
  $T = 2\pi\sqrt{\frac{m}{k}}$ and $f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$, where m is the mass of the system.
- Displacement in simple harmonic motion as a function of time is given by $x(t) = X \cos \frac{2\pi t}{T}$.
- The velocity is given by $v(t) = -v_{\max} \sin \frac{2\pi t}{T}$, where $v_{\max} = \sqrt{k/m}X$.
- The acceleration is found to be
  $a(t) = -\frac{kX}{m} \cos \frac{2\pi t}{T}$.

16.4 The Simple Pendulum

- A mass m suspended by a wire of length L is a simple pendulum and undergoes simple harmonic motion for amplitudes less than about 15°.
  The period of a simple pendulum is

  $$T = 2\pi\sqrt{\frac{L}{g}},$$

  where L is the length of the string and g is the acceleration due to gravity.

16.5 Energy and the Simple Harmonic Oscillator

- Energy in the simple harmonic oscillator is shared between elastic potential energy and kinetic energy, with the total being constant:
  $\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}.$
- Maximum velocity depends on three factors: it is directly proportional to amplitude, it is greater for stiffer systems, and it is smaller for objects that have larger masses:
  $v_{\max} = \sqrt{\frac{k}{m}}X.$

16.6 Uniform Circular Motion and Simple Harmonic Motion

A projection of uniform circular motion undergoes simple harmonic oscillation.

16.7 Damped Harmonic Motion

- Damped harmonic oscillators have non-conservative forces that dissipate their energy.
- Critical damping returns the system to equilibrium as fast as possible without overshooting.
- An underdamped system will oscillate through the equilibrium position.
- An overdamped system moves more slowly toward equilibrium than one that is critically damped.

16.8 Forced Oscillations and Resonance

- A system's natural frequency is the frequency at which the system will oscillate if not affected by driving or damping forces.
- A periodic force driving a harmonic oscillator at its natural frequency produces resonance. The system is said to resonate.
- The less damping a system has, the higher the amplitude of the forced oscillations near resonance. The more damping a system has, the broader response it has to varying driving frequencies.

16.9 Waves

- A wave is a disturbance that moves from the point of creation with a wave velocity v_{w}.
- A wave has a wavelength λ, which is the distance between adjacent identical parts of the wave.
- Wave velocity and wavelength are related to the wave's frequency and period by $v_{\text{w}} = \frac{\lambda}{T}$ or $v_{\text{w}} = f\lambda$.
- A transverse wave has a disturbance perpendicular to its direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

16.10 Superposition and Interference

- Superposition is the combination of two waves at the same location.
- Constructive interference occurs when two identical waves are superimposed in phase.
- Destructive interference occurs when two identical waves are superimposed exactly out of phase.
- A standing wave is one in which two waves superimpose to produce a wave that varies in amplitude but does not propagate.
- Nodes are points of no motion in standing waves.

- An antinode is the location of maximum amplitude of a standing wave.
- Waves on a string are resonant standing waves with a fundamental frequency and can occur at higher multiples of the fundamental, called overtones or harmonics.
- Beats occur when waves of similar frequencies f_1 and f_2 are superimposed. The resulting amplitude oscillates with a beat frequency given by $f_B = |f_1 - f_2|$.

16.11 Energy in Waves: Intensity

Intensity is defined to be the power per unit area:

$I = \frac{P}{A}$ and has units of W/m^2.

Conceptual Questions

16.1 Hooke's Law: Stress and Strain Revisited

1. Describe a system in which elastic potential energy is stored.

16.3 Simple Harmonic Motion: A Special Periodic Motion

2. What conditions must be met to produce simple harmonic motion?

3. (a) If frequency is not constant for some oscillation, can the oscillation be simple harmonic motion? (b) Can you think of any examples of harmonic motion where the frequency may depend on the amplitude?

4. Give an example of a simple harmonic oscillator, specifically noting how its frequency is independent of amplitude.

5. Explain why you expect an object made of a stiff material to vibrate at a higher frequency than a similar object made of a spongy material.

6. As you pass a freight truck with a trailer on a highway, you notice that its trailer is bouncing up and down slowly. Is it more likely that the trailer is heavily loaded or nearly empty? Explain your answer.

7. Some people modify cars to be much closer to the ground than when manufactured. Should they install stiffer springs? Explain your answer.

16.4 The Simple Pendulum

8. Pendulum clocks are made to run at the correct rate by adjusting the pendulum's length. Suppose you move from one city to another where the acceleration due to gravity is slightly greater, taking your pendulum clock with you, will you have to lengthen or shorten the pendulum to keep the correct time, other factors remaining constant? Explain your answer.

16.5 Energy and the Simple Harmonic Oscillator

9. Explain in terms of energy how dissipative forces such as friction reduce the amplitude of a harmonic oscillator. Also explain how a driving mechanism can compensate. (A pendulum clock is such a system.)

16.7 Damped Harmonic Motion

10. Give an example of a damped harmonic oscillator. (They are more common than undamped or simple harmonic oscillators.)

11. How would a car bounce after a bump under each of these conditions?
- overdamping
- underdamping
- critical damping

12. Most harmonic oscillators are damped and, if undriven, eventually come to a stop. How is this observation related to the second law of thermodynamics?

16.8 Forced Oscillations and Resonance

13. Why are soldiers in general ordered to "route step" (walk out of step) across a bridge?

16.9 Waves

14. Give one example of a transverse wave and another of a longitudinal wave, being careful to note the relative directions of the disturbance and wave propagation in each.

15. What is the difference between propagation speed and the frequency of a wave? Does one or both affect wavelength? If so, how?

16.10 Superposition and Interference

16. Speakers in stereo systems have two color-coded terminals to indicate how to hook up the wires. If the wires are reversed, the speaker moves in a direction opposite that of a properly connected speaker. Explain why it is important to have both speakers connected the same way.

16.11 Energy in Waves: Intensity

17. Two identical waves undergo pure constructive interference. Is the resultant intensity twice that of the individual waves? Explain your answer.

18. Circular water waves decrease in amplitude as they move away from where a rock is dropped. Explain why.

Problems & Exercises

16.1 Hooke's Law: Stress and Strain Revisited

1. Fish are hung on a spring scale to determine their mass (most fishermen feel no obligation to truthfully report the mass).
(a) What is the force constant of the spring in such a scale if it the spring stretches 8.00 cm for a 10.0 kg load?
(b) What is the mass of a fish that stretches the spring 5.50 cm?
(c) How far apart are the half-kilogram marks on the scale?

2. It is weigh-in time for the local under-85-kg rugby team. The bathroom scale used to assess eligibility can be described by Hooke's law and is depressed 0.75 cm by its maximum load of 120 kg. (a) What is the spring's effective spring constant? (b) A player stands on the scales and depresses it by 0.48 cm. Is he eligible to play on this under-85 kg team?

3. One type of BB gun uses a spring-driven plunger to blow the BB from its barrel. (a) Calculate the force constant of its plunger's spring if you must compress it 0.150 m to drive the 0.0500-kg plunger to a top speed of 20.0 m/s. (b) What force must be exerted to compress the spring?

4. (a) The springs of a pickup truck act like a single spring with a force constant of $1.30 \times 10^5\ \text{N/m}$. By how much will the truck be depressed by its maximum load of 1000 kg?
(b) If the pickup truck has four identical springs, what is the force constant of each?

5. When an 80.0-kg man stands on a pogo stick, the spring is compressed 0.120 m.
(a) What is the force constant of the spring? (b) Will the spring be compressed more when he hops down the road?

6. A spring has a length of 0.200 m when a 0.300-kg mass hangs from it, and a length of 0.750 m when a 1.95-kg mass hangs from it. (a) What is the force constant of the spring? (b) What is the unloaded length of the spring?

16.2 Period and Frequency in Oscillations

7. What is the period of $60.0\ \text{Hz}$ electrical power?

8. If your heart rate is 150 beats per minute during strenuous exercise, what is the time per beat in units of seconds?

9. Find the frequency of a tuning fork that takes $2.50 \times 10^{-3}\ \text{s}$ to complete one oscillation.

10. A stroboscope is set to flash every $8.00 \times 10^{-5}\ \text{s}$. What is the frequency of the flashes?

11. A tire has a tread pattern with a crevice every 2.00 cm. Each crevice makes a single vibration as the tire moves. What is the frequency of these vibrations if the car moves at 30.0 m/s?

12. Engineering Application
Each piston of an engine makes a sharp sound every other revolution of the engine. (a) How fast is a race car going if its eight-cylinder engine emits a sound of frequency 750 Hz, given that the engine makes 2000 revolutions per kilometer? (b) At how many revolutions per minute is the engine rotating?

16.3 Simple Harmonic Motion: A Special Periodic Motion

13. A type of cuckoo clock keeps time by having a mass bouncing on a spring, usually something cute like a cherub in a chair. What force constant is needed to produce a period of 0.500 s for a 0.0150-kg mass?

14. If the spring constant of a simple harmonic oscillator is doubled, by what factor will the mass of the system need to change in order for the frequency of the motion to remain the same?

15. A 0.500-kg mass suspended from a spring oscillates with a period of 1.50 s. How much mass must be added to the object to change the period to 2.00 s?

16. By how much leeway (both percentage and mass) would you have in the selection of the mass of the object in the previous problem if you did not wish the new period to be greater than 2.01 s or less than 1.99 s?

17. Suppose you attach the object with mass m to a vertical spring originally at rest, and let it bounce up and down. You release the object from rest at the spring's original rest length. (a) Show that the spring exerts an upward force of $2.00\ mg$ on the object at its lowest point. (b) If the spring has a force constant of $10.0\ \text{N/m}$ and a 0.25-kg-mass object is set in motion as described, find the amplitude of the oscillations. (c) Find the maximum velocity.

18. A diver on a diving board is undergoing simple harmonic motion. Her mass is 55.0 kg and the period of her motion is 0.800 s. The next diver is a male whose period of simple harmonic oscillation is 1.05 s. What is his mass if the mass of the board is negligible?

19. Suppose a diving board with no one on it bounces up and down in a simple harmonic motion with a frequency of 4.00 Hz. The board has an effective mass of 10.0 kg. What is the frequency of the simple harmonic motion of a 75.0-kg diver on the board?

20. 

FIGURE 16.42 This child's toy relies on springs to keep infants entertained. (credit: By Humboldthead, Flickr)

The device pictured in Figure 16.42 entertains infants while keeping them from wandering. The child bounces in a harness suspended from a door frame by a spring constant.
(a) If the spring stretches 0.250 m while supporting an 8.0-kg child, what is its spring constant?
(b) What is the time for one complete bounce of this child? (c) What is the child's maximum velocity if the amplitude of her bounce is 0.200 m?

21. A 90.0-kg skydiver hanging from a parachute bounces up and down with a period of 1.50 s. What is the new period of oscillation when a second skydiver, whose mass is 60.0 kg, hangs from the legs of the first, as seen in Figure 16.43.

FIGURE 16.43 The oscillations of one skydiver are about to be affected by a second skydiver. (credit: U.S. Army, www.army.mil)

16.4 The Simple Pendulum

As usual, the acceleration due to gravity in these problems is taken to be $g = 9.80\ \text{m/s}^2$, unless otherwise specified.

22. What is the length of a pendulum that has a period of 0.500 s?

23. Some people think a pendulum with a period of 1.00 s can be driven with "mental energy" or psycho kinetically, because its period is the same as an average heartbeat. True or not, what is the length of such a pendulum?

24. What is the period of a 1.00-m-long pendulum?

25. How long does it take a child on a swing to complete one swing if her center of gravity is 4.00 m below the pivot?

26. The pendulum on a cuckoo clock is 5.00 cm long. What is its frequency?

27. Two parakeets sit on a swing with their combined center of mass 10.0 cm below the pivot. At what frequency do they swing?

28. (a) A pendulum that has a period of 3.00000 s and that is located where the acceleration due to gravity is $9.79\ \text{m/s}^2$ is moved to a location where the acceleration due to gravity is $9.82\ \text{m/s}^2$. What is its new period? (b) Explain why so many digits are needed in the value for the period, based on the relation between the period and the acceleration due to gravity.

29. A pendulum with a period of 2.00000 s in one location $\left(g = 9.80 \text{ m/s}^2\right)$ is moved to a new location where the period is now 1.99796 s. What is the acceleration due to gravity at its new location?

30. (a) What is the effect on the period of a pendulum if you double its length?
(b) What is the effect on the period of a pendulum if you decrease its length by 5.00%?

31. Find the ratio of the new/old periods of a pendulum if the pendulum were transported from Earth to the Moon, where the acceleration due to gravity is 1.63 m/s^2.

32. At what rate will a pendulum clock run on the Moon, where the acceleration due to gravity is 1.63 m/s^2, if it keeps time accurately on Earth? That is, find the time (in hours) it takes the clock's hour hand to make one revolution on the Moon.

33. Suppose the length of a clock's pendulum is changed by 1.000%, exactly at noon one day. What time will it read 24.00 hours later, assuming it the pendulum has kept perfect time before the change? Note that there are two answers, and perform the calculation to four-digit precision.

34. If a pendulum-driven clock gains 5.00 s/day, what fractional change in pendulum length must be made for it to keep perfect time?

16.5 Energy and the Simple Harmonic Oscillator

35. The length of nylon rope from which a mountain climber is suspended has a force constant of $1.40 \times 10^4 \text{N/m}$
(a) What is the frequency at which he bounces, given his mass plus and the mass of his equipment are 90.0 kg? Ignore the change in gravitational potential energy after the cord begins to stretch.
(b) How much would this rope stretch to break the climber's fall if he free-falls 2.00 m before the rope runs out of slack? Hint: Use conservation of energy.
(c) Repeat both parts of this problem in the situation where twice this length of nylon rope is used.

36. Engineering Application
Near the top of the Citigroup Center building in New York City, there is an object with mass of $4.00 \times 10^5 \text{ kg}$ on springs that have adjustable force constants. Its function is to dampen wind-driven oscillations of the building by oscillating at the same frequency as the building is being driven—the driving force is transferred to the object, which oscillates instead of the entire building. (a) What effective force constant should the springs have to make the object oscillate with a period of 2.00 s? (b) What energy is stored in the springs for a 2.00-m displacement from equilibrium?

16.6 Uniform Circular Motion and Simple Harmonic Motion

37. (a)What is the maximum velocity of an 85.0-kg person bouncing on a bathroom scale having a force constant of $1.50 \times 10^6 \text{ N/m}$, if the amplitude of the bounce is 0.200 cm? (b)What is the maximum energy stored in the spring?

38. A novelty clock has a 0.0100-kg mass object bouncing on a spring that has a force constant of 1.25 N/m. What is the maximum velocity of the object if the object bounces 3.00 cm above and below its equilibrium position? (b) How many joules of kinetic energy does the object have at its maximum velocity?

39. At what positions is the speed of a simple harmonic oscillator half its maximum? That is, what values of x/X give $v = \pm v_{\text{max}}/2$, where X is the amplitude of the motion?

40. A ladybug sits 12.0 cm from the center of a Beatles music album spinning at 33.33 rpm. What is the maximum velocity of its shadow on the wall behind the turntable, if illuminated parallel to the record by the parallel rays of the setting Sun?

16.7 Damped Harmonic Motion

41. The amplitude of a lightly damped oscillator decreases by 3.0% during each cycle. What percentage of the mechanical energy of the oscillator is lost in each cycle?

16.8 Forced Oscillations and Resonance

42. How much energy must the shock absorbers of a 1200-kg car dissipate in order to damp a bounce that initially has a velocity of 0.800 m/s at the equilibrium position? Assume the car returns to its original vertical position.

43. If a car has a suspension system with a force constant of $5.00 \times 10^4\ \text{N/m}$, how much energy must the car's shocks remove to dampen an oscillation starting with a maximum displacement of 0.0750 m?

44. (a) How much will a spring that has a force constant of 40.0 N/m be stretched by an object with a mass of 0.500 kg when hung motionless from the spring? (b) Calculate the decrease in gravitational potential energy of the 0.500-kg object when it descends this distance. (c) Part of this gravitational energy goes into the spring. Calculate the energy stored in the spring by this stretch, and compare it with the gravitational potential energy. Explain where the rest of the energy might go.

45. Suppose you have a 0.750-kg object on a horizontal surface connected to a spring that has a force constant of 150 N/m. There is simple friction between the object and surface with a static coefficient of friction $\mu_\text{s} = 0.100$. (a) How far can the spring be stretched without moving the mass? (b) If the object is set into oscillation with an amplitude twice the distance found in part (a), and the kinetic coefficient of friction is $\mu_\text{k} = 0.0850$, what total distance does it travel before stopping? Assume it starts at the maximum amplitude.

46. Engineering Application: A suspension bridge oscillates with an effective force constant of $1.00 \times 10^8\ \text{N/m}$. (a) How much energy is needed to make it oscillate with an amplitude of 0.100 m? (b) If soldiers march across the bridge with a cadence equal to the bridge's natural frequency and impart $1.00 \times 10^4\ \text{J}$ of energy each second, how long does it take for the bridge's oscillations to go from 0.100 m to 0.500 m amplitude?

16.9 Waves

47. Storms in the South Pacific can create waves that travel all the way to the California coast, which are 12,000 km away. How long does it take them if they travel at 15.0 m/s?

48. Waves on a swimming pool propagate at 0.750 m/s. You splash the water at one end of the pool and observe the wave go to the opposite end, reflect, and return in 30.0 s. How far away is the other end of the pool?

49. Wind gusts create ripples on the ocean that have a wavelength of 5.00 cm and propagate at 2.00 m/s. What is their frequency?

50. How many times a minute does a boat bob up and down on ocean waves that have a wavelength of 40.0 m and a propagation speed of 5.00 m/s?

51. Scouts at a camp shake the rope bridge they have just crossed and observe the wave crests to be 8.00 m apart. If they shake the bridge twice per second, what is the propagation speed of the waves?

52. What is the wavelength of the waves you create in a swimming pool if you splash your hand at a rate of 2.00 Hz and the waves propagate at 0.800 m/s?

53. What is the wavelength of an earthquake that shakes you with a frequency of 10.0 Hz and gets to another city 84.0 km away in 12.0 s?

54. Radio waves transmitted through space at $3.00 \times 10^8\ \text{m/s}$ by the Voyager spacecraft have a wavelength of 0.120 m. What is their frequency?

55. Your ear is capable of differentiating sounds that arrive at the ear just 1.00 ms apart. What is the minimum distance between two speakers that produce sounds that arrive at noticeably different times on a day when the speed of sound is 340 m/s?

56. (a) Seismographs measure the arrival times of earthquakes with a precision of 0.100 s. To get the distance to the epicenter of the quake, they compare the arrival times of S- and P-waves, which travel at different speeds. Figure 16.44) If S- and P-waves travel at 4.00 and 7.20 km/s, respectively, in the region considered, how precisely can the distance to the source of the earthquake be determined? (b) Seismic waves from underground detonations of nuclear bombs can be used to locate the test site and detect violations of test bans. Discuss whether your answer to (a) implies a serious limit to such detection. (Note also that the uncertainty is greater if there is an uncertainty in the propagation speeds of the S- and P-waves.)

FIGURE 16.44 A seismograph as described in above problem.(credit: Oleg Alexandrov)

16.10 Superposition and Interference

57. A car has two horns, one emitting a frequency of 199 Hz and the other emitting a frequency of 203 Hz. What beat frequency do they produce?

58. The middle-C hammer of a piano hits two strings, producing beats of 1.50 Hz. One of the strings is tuned to 260.00 Hz. What frequencies could the other string have?

59. Two tuning forks having frequencies of 460 and 464 Hz are struck simultaneously. What average frequency will you hear, and what will the beat frequency be?

60. Twin jet engines on an airplane are producing an average sound frequency of 4100 Hz with a beat frequency of 0.500 Hz. What are their individual frequencies?

61. A wave traveling on a Slinky® that is stretched to 4 m takes 2.4 s to travel the length of the Slinky and back again. (a) What is the speed of the wave? (b) Using the same Slinky stretched to the same length, a standing wave is created which consists of three antinodes and four nodes. At what frequency must the Slinky be oscillating?

62. Three adjacent keys on a piano (F, F-sharp, and G) are struck simultaneously, producing frequencies of 349, 370, and 392 Hz. What beat frequencies are produced by this discordant combination?

16.11 Energy in Waves: Intensity

63. Medical Application
Ultrasound of intensity $1.50 \times 10^2\ \text{W/m}^2$ is produced by the rectangular head of a medical imaging device measuring 3.00 by 5.00 cm. What is its power output?

64. The low-frequency speaker of a stereo set has a surface area of $0.05\ \text{m}^2$ and produces 1W of acoustical power. What is the intensity at the speaker? If the speaker projects sound uniformly in all directions, at what distance from the speaker is the intensity $0.1\ \text{W/m}^2$?

65. To increase intensity of a wave by a factor of 50, by what factor should the amplitude be increased?

66. Engineering Application
A device called an insolation meter is used to measure the intensity of sunlight has an area of 100 cm^2 and registers 6.50 W. What is the intensity in W/m^2?

67. Astronomy Application
Energy from the Sun arrives at the top of the Earth's atmosphere with an intensity of $1.30\ \text{kW/m}^2$. How long does it take for $1.8 \times 10^9\ \text{J}$ to arrive on an area of $1.00\ \text{m}^2$?

68. Suppose you have a device that extracts energy from ocean breakers in direct proportion to their intensity. If the device produces 10.0 kW of power on a day when the breakers are 1.20 m high, how much will it produce when they are 0.600 m high?

69. Engineering Application
(a) A photovoltaic array of (solar cells) is 10.0% efficient in gathering solar energy and converting it to electricity. If the average intensity of sunlight on one day is $700\ \text{W/m}^2$, what area should your array have to gather energy at the rate of 100 W? (b) What is the maximum cost of the array if it must pay for itself in two years of operation averaging 10.0 hours per day? Assume that it earns money at the rate of 9.00 ¢ per kilowatt-hour.

70. A microphone receiving a pure sound tone feeds an oscilloscope, producing a wave on its screen. If the sound intensity is originally $2.00 \times 10^{-5}\ \text{W/m}^2$, but is turned up until the amplitude increases by 30.0%, what is the new intensity?

71. Medical Application
(a) What is the intensity in W/m^2 of a laser beam used to burn away cancerous tissue that, when 90.0% absorbed, puts 500 J of energy into a circular spot 2.00 mm in diameter in 4.00 s? (b) Discuss how this intensity compares to the average intensity of sunlight (about $700\ \text{W/m}^2$) and the implications that would have if the laser beam entered your eye. Note how the amount of damage depends on the time duration of the exposure.

CHAPTER 17

Physics of Hearing

FIGURE 17.1 This tree fell some time ago. When it fell, atoms in the air were disturbed. Physicists would call this disturbance sound whether someone was around to hear it or not. (credit: B.A. Bowen Photography)

CHAPTER OUTLINE

INTRODUCTION TO THE PHYSICS OF HEARING If a tree falls in the forest and no one is there to hear it, does it make a sound? The answer to this old philosophical question depends on how you define sound. If sound only exists when someone is around to perceive it, then there was no sound. However, if we define sound in terms of physics; that is, a disturbance of the atoms in matter transmitted from its origin outward (in other words, a wave), then there *was* a sound, even if nobody was around to hear it.

Such a wave is the physical phenomenon we call *sound.* Its perception is hearing. Both the physical phenomenon and its perception are interesting and will be considered in this text. We shall explore both sound and hearing; they are related, but are not the same thing. We will also explore the many practical uses of sound waves, such as in medical imaging.

Click to view content (https://openstax.org/books/college-physics-2e/pages/17-introduction-to-the-physics-of-hearing)

17.1 Sound

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define sound and hearing.
- Describe sound as a longitudinal wave.

FIGURE 17.2 This glass has been shattered by a high-intensity sound wave of the same frequency as the resonant frequency of the glass. While the sound is not visible, the effects of the sound prove its existence. (credit: ||read||, Flickr)

Sound can be used as a familiar illustration of waves. Because hearing is one of our most important senses, it is interesting to see how the physical properties of sound correspond to our perceptions of it. **Hearing** is the perception of sound, just as vision is the perception of visible light. But sound has important applications beyond hearing. Ultrasound, for example, is not heard but can be employed to form medical images and is also used in treatment.

The physical phenomenon of **sound** is defined to be a disturbance of matter that is transmitted from its source outward. Sound is a wave. On the atomic scale, it is a disturbance of atoms that is far more ordered than their thermal motions. In many instances, sound is a periodic wave, and the atoms undergo simple harmonic motion. In this text, we shall explore such periodic sound waves.

A vibrating string produces a sound wave as illustrated in Figure 17.3, Figure 17.4, and Figure 17.5. As the string oscillates back and forth, it transfers energy to the air, mostly as thermal energy created by turbulence. But a small part of the string's energy goes into compressing and expanding the surrounding air, creating slightly higher and lower local pressures. These compressions (high pressure regions) and rarefactions (low pressure regions) move out as longitudinal pressure waves having the same frequency as the string—they are the disturbance that is a sound wave. (Sound waves in air and most fluids are longitudinal, because fluids have almost no shear strength. In solids, sound waves can be both transverse and longitudinal.) Figure 17.5 shows a graph of gauge pressure versus distance from the vibrating string.

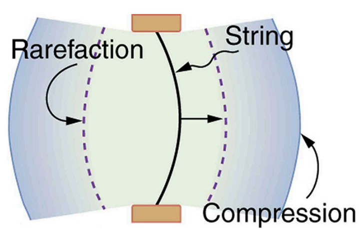


FIGURE 17.3 A vibrating string moving to the right compresses the air in front of it and expands the air behind it.

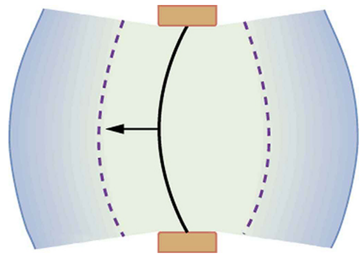


FIGURE 17.4 As the string moves to the left, it creates another compression and rarefaction as the ones on the right move away from the string.

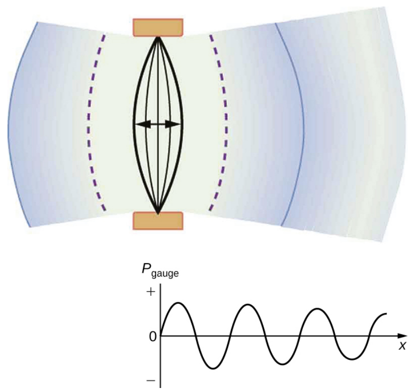


FIGURE 17.5 After many vibrations, there are a series of compressions and rarefactions moving out from the string as a sound wave. The graph shows gauge pressure versus distance from the source. Pressures vary only slightly from atmospheric for ordinary sounds.

The amplitude of a sound wave decreases with distance from its source, because the energy of the wave is spread over a larger and larger area. But it is also absorbed by objects, such as the eardrum in Figure 17.6, and converted to thermal energy by the viscosity of air. In addition, during each compression a little heat transfers to the air and during each rarefaction even less heat transfers from the air, so that the heat transfer reduces the organized disturbance into random thermal motions. (These processes can be viewed as a manifestation of the second law of thermodynamics presented in Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency.) Whether the heat transfer from compression to rarefaction is significant depends on how far apart they are—that is, it depends on wavelength. Wavelength, frequency, amplitude, and speed of propagation are important for sound, as they are for all waves.

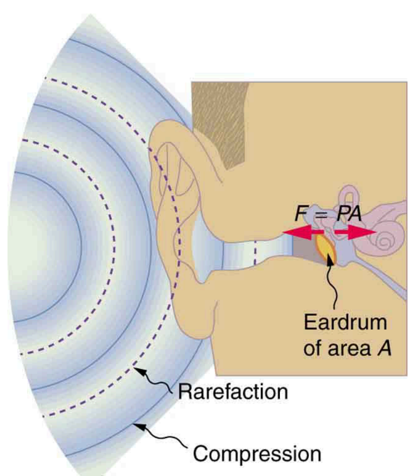


FIGURE 17.6 Sound wave compressions and rarefactions travel up the ear canal and force the eardrum to vibrate. There is a net force on the eardrum, since the sound wave pressures differ from the atmospheric pressure found behind the eardrum. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person.

PHET EXPLORATIONS

Wave Interference

Make waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern.

Click to view content (https://openstax.org/books/college-physics-2e/pages/17-1-sound)

17.2 Speed of Sound, Frequency, and Wavelength

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define pitch.
- Describe the relationship between the speed of sound, its frequency, and its wavelength.
- Describe the effects on the speed of sound as it travels through various media.
- Describe the effects of temperature on the speed of sound.

FIGURE 17.7 When a firework explodes, the light energy is perceived before the sound energy. Sound travels more slowly than light does. (credit: Dominic Alves, Flickr)

Sound, like all waves, travels at a certain speed and has the properties of frequency and wavelength. You can observe direct evidence of the speed of sound while watching a fireworks display. The flash of an explosion is seen well before its sound is heard, implying both that sound travels at a finite speed and that it is much slower than light. You can also directly sense the frequency of a sound. Perception of frequency is called **pitch**. The wavelength of sound is not directly sensed, but indirect evidence is found in the correlation of the size of musical instruments with their pitch. Small instruments, such as a piccolo, typically make high-pitch sounds, while large instruments, such as a tuba, typically make low-pitch sounds. High pitch means small wavelength, and the size of a musical instrument is directly related to the wavelengths of sound it produces. So a small instrument creates short-wavelength sounds. Similar arguments hold that a large instrument creates long-wavelength sounds.

The relationship of the speed of sound, its frequency, and wavelength is the same as for all waves:

$$v_{\text{w}} = f\lambda, \tag{17.1}$$

where v_{w} is the speed of sound, f is its frequency, and λ is its wavelength. The wavelength of a sound is the distance between adjacent identical parts of a wave—for example, between adjacent compressions as illustrated in Figure 17.8. The frequency is the same as that of the source and is the number of waves that pass a point per unit time.

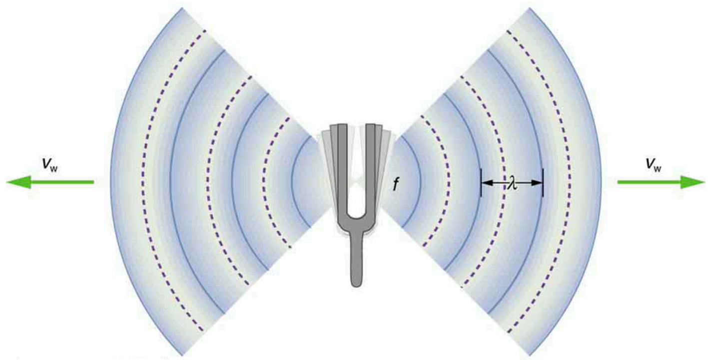


FIGURE 17.8 A sound wave emanates from a source vibrating at a frequency f, propagates at v_w, and has a wavelength λ.

Table 17.1 makes it apparent that the speed of sound varies greatly in different media. The speed of sound in a medium is determined by a combination of the medium's rigidity (or compressibility in gases) and its density. The more rigid (or less compressible) the medium, the faster the speed of sound. For materials that have similar rigidities, sound will travel faster through the one with the lower density because the sound energy is more easily transferred from particle to particle. The speed of sound in air is low, because air is compressible. Because liquids and solids are relatively rigid and very difficult to compress, the speed of sound in such media is generally greater than in gases.

Medium	v_w(m/s)
Gases at $0^\circ C$	
Air	331
Carbon dioxide	259
Oxygen	316
Helium	965
Hydrogen	1290
Liquids at $20^\circ C$	
Ethanol	1160
Mercury	1450
Water, fresh	1480
Sea water	1540
Human tissue	1540
Solids (longitudinal or bulk)	
Vulcanized rubber	54

TABLE 17.1 Speed of Sound in Various Media

Medium	v_w(m/s)
Polyethylene	920
Marble	3810
Glass, Pyrex	5640
Lead	1960
Aluminum	5120
Steel	5960

TABLE 17.1 Speed of Sound in Various Media

Earthquakes, essentially sound waves in Earth's crust, are an interesting example of how the speed of sound depends on the rigidity of the medium. Earthquakes have both longitudinal and transverse components, and these travel at different speeds. The bulk modulus of granite is greater than its shear modulus. For that reason, the speed of longitudinal or pressure waves (P-waves) in earthquakes in granite is significantly higher than the speed of transverse or shear waves (S-waves). Both components of earthquakes travel slower in less rigid material, such as sediments. P-waves have speeds of 4 to 7 km/s, and S-waves correspondingly range in speed from 2 to 5 km/s, both being faster in more rigid material. The P-wave gets progressively farther ahead of the S-wave as they travel through Earth's crust. The time between the P- and S-waves is routinely used to determine the distance to their source, the epicenter of the earthquake. The time and nature of these wave differences also provides the evidence for the nature of Earth's core. Through careful analysis of seismographic records of large earthquakes whose waves could be clearly detected around the world, Richard Dixon Oldham established that waves passing through the center of the Earth behaved as if they were moving through a different medium: a liquid. Later on, Inge Lehmann used more precise observations (partly based on a better coordinated network of seismographs she helped set up) to better define the nature of the core: that it was a solid inner core surrounded by a liquid outer core.

The speed of sound is affected by temperature in a given medium. For air at sea level, the speed of sound is given by

$$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{T}{273\ \mathrm{K}}}, \quad (17.2)$$

where the temperature (denoted as T) is in units of kelvin. The speed of sound in gases is related to the average speed of particles in the gas, v_{rms}, and that

$$v_{\mathrm{rms}} = \sqrt{\frac{3\,kT}{m}}, \quad (17.3)$$

where k is the Boltzmann constant (1.38×10^{-23} J/K) and m is the mass of each (identical) particle in the gas. So, it is reasonable that the speed of sound in air and other gases should depend on the square root of temperature. While not negligible, this is not a strong dependence. At $0^{\circ}\mathrm{C}$, the speed of sound is 331 m/s, whereas at $20.0^{\circ}\mathrm{C}$ it is 343 m/s, less than a 4% increase. Figure 17.9 shows a use of the speed of sound by a bat to sense distances. Echoes are also used in medical imaging.

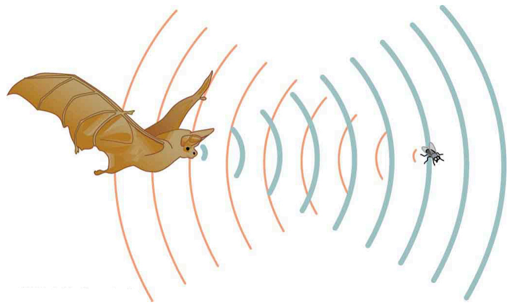


FIGURE 17.9 A bat uses sound echoes to find its way about and to catch prey. The time for the echo to return is directly proportional to the distance.

One of the more important properties of sound is that its speed is nearly independent of frequency. This independence is certainly true in open air for sounds in the audible range of 20 to 20,000 Hz. If this independence were not true, you would certainly notice it for music played by a marching band in a football stadium, for example. Suppose that high-frequency sounds traveled faster—then the farther you were from the band, the more the sound from the low-pitch instruments would lag that from the high-pitch ones. But the music from all instruments arrives in cadence independent of distance, and so all frequencies must travel at nearly the same speed. Recall that

$$v_{\mathrm{w}} = f\lambda. \tag{17.4}$$

In a given medium under fixed conditions, v_{w} is constant, so that there is a relationship between f and λ; the higher the frequency, the smaller the wavelength. See Figure 17.10 and consider the following example.

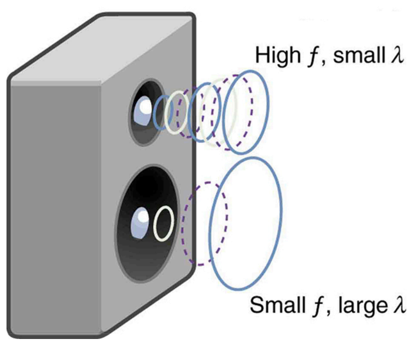


FIGURE 17.10 Because they travel at the same speed in a given medium, low-frequency sounds must have a greater wavelength than high-frequency sounds. Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, while the higher-frequency sounds are emitted by the small speaker, called a tweeter.

EXAMPLE 17.1

Calculating Wavelengths: What Are the Wavelengths of Audible Sounds?

Calculate the wavelengths of sounds at the extremes of the audible range, 20 and 20,000 Hz, in $30.0^{\circ}\mathrm{C}$ air. (Assume that the frequency values are accurate to two significant figures.)

Strategy

To find wavelength from frequency, we can use $v_{\mathrm{w}} = f\lambda$.

Solution

1. Identify knowns. The value for v_{w}, is given by

$$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{T}{273\ \mathrm{K}}}. \tag{17.5}$$

2. Convert the temperature into kelvin and then enter the temperature into the equation

$$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{303\ \mathrm{K}}{273\ \mathrm{K}}} = 348.7\ \mathrm{m/s}. \tag{17.6}$$

3. Solve the relationship between speed and wavelength for λ:

$$\lambda = \frac{v_w}{f}. \quad (17.7)$$

4. Enter the speed and the minimum frequency to give the maximum wavelength:

$$\lambda_{max} = \frac{348.7 \text{ m/s}}{20 \text{ Hz}} = 17 \text{ m}. \quad (17.8)$$

5. Enter the speed and the maximum frequency to give the minimum wavelength:

$$\lambda_{min} = \frac{348.7 \text{ m/s}}{20{,}000 \text{ Hz}} = 0.017 \text{ m} = 1.7 \text{ cm}. \quad (17.9)$$

Discussion

Because the product of f multiplied by λ equals a constant, the smaller f is, the larger λ must be, and vice versa.

The speed of sound can change when sound travels from one medium to another. However, the frequency usually remains the same because it is like a driven oscillation and has the frequency of the original source. If v_w changes and f remains the same, then the wavelength λ must change. That is, because $v_w = f\lambda$, the higher the speed of a sound, the greater its wavelength for a given frequency.

Making Connections: Take-Home Investigation—Voice as a Sound Wave

Suspend a sheet of paper so that the top edge of the paper is fixed and the bottom edge is free to move. You could tape the top edge of the paper to the edge of a table. Gently blow near the edge of the bottom of the sheet and note how the sheet moves. Speak softly and then louder such that the sounds hit the edge of the bottom of the paper, and note how the sheet moves. Explain the effects.

CHECK YOUR UNDERSTANDING

Imagine you observe two fireworks explode. You hear the explosion of one as soon as you see it. However, you see the other firework for several milliseconds before you hear the explosion. Explain why this is so.

Solution

Sound and light both travel at definite speeds. The speed of sound is slower than the speed of light. The first firework is probably very close by, so the speed difference is not noticeable. The second firework is farther away, so the light arrives at your eyes noticeably sooner than the sound wave arrives at your ears.

CHECK YOUR UNDERSTANDING

You observe two musical instruments that you cannot identify. One plays high-pitch sounds and the other plays low-pitch sounds. How could you determine which is which without hearing either of them play?

Solution

Compare their sizes. High-pitch instruments are generally smaller than low-pitch instruments because they generate a smaller wavelength.

17.3 Sound Intensity and Sound Level

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define intensity, sound intensity, and sound pressure level.
- Calculate sound intensity levels in decibels (dB).

FIGURE 17.11 Noise on crowded roadways like this one in Delhi makes it hard to hear others unless they shout. (credit: Lingaraj G J, Flickr)

In a quiet forest, you can sometimes hear a single leaf fall to the ground. After settling into bed, you may hear your blood pulsing through your ears. But when a passing motorist has his stereo turned up, you cannot even hear what the person next to you in your car is saying. We are all very familiar with the loudness of sounds and aware that they are related to how energetically the source is vibrating. In cartoons depicting a screaming person (or an animal making a loud noise), the cartoonist often shows an open mouth with a vibrating uvula, the hanging tissue at the back of the mouth, to suggest a loud sound coming from the throat Figure 17.12. High noise exposure is hazardous to hearing, and it is common for musicians to have hearing losses that are sufficiently severe that they interfere with the musicians' abilities to perform. The relevant physical quantity is sound intensity, a concept that is valid for all sounds whether or not they are in the audible range.

Intensity is defined to be the power per unit area carried by a wave. Power is the rate at which energy is transferred by the wave. In equation form, **intensity** I is

$$I = \frac{P}{A}, \quad \text{17.10}$$

where P is the power through an area A. The SI unit for I is $\mathrm{W/m^2}$. The intensity of a sound wave is related to its amplitude squared by the following relationship:

$$I = \frac{(\Delta p)^2}{2\rho v_{\mathrm{w}}}. \quad \text{17.11}$$

Here Δp is the pressure variation or pressure amplitude (half the difference between the maximum and minimum pressure in the sound wave) in units of pascals (Pa) or $\mathrm{N/m^2}$. (We are using a lower case p for pressure to distinguish it from power, denoted by P above.) The energy (as kinetic energy $\frac{mv^2}{2}$) of an oscillating element of air due to a traveling sound wave is proportional to its amplitude squared. In this equation, ρ is the density of the material in which the sound wave travels, in units of $\mathrm{kg/m^3}$, and v_{w} is the speed of sound in the medium, in units of m/s. The pressure variation is proportional to the amplitude of the oscillation, and so I varies as $(\Delta p)^2$ (Figure 17.12). This relationship is consistent with the fact that the sound wave is produced by some vibration; the greater its pressure amplitude, the more the air is compressed in the sound it creates.

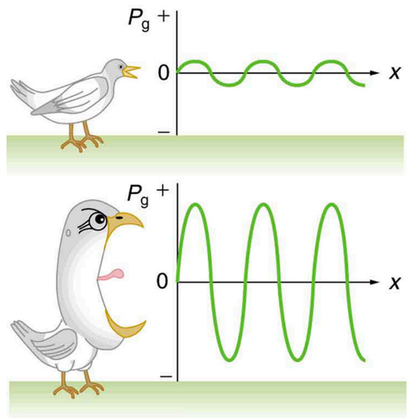


FIGURE 17.12 Graphs of the gauge pressures in two sound waves of different intensities. The more intense sound is produced by a source that has larger-amplitude oscillations and has greater pressure maxima and minima. Because pressures are higher in the greater-intensity sound, it can exert larger forces on the objects it encounters.

Sound intensity levels are quoted in decibels (dB) much more often than sound intensities in watts per meter squared. Decibels are the unit of choice in the scientific literature as well as in the popular media. The reasons for this choice of units are related to how we perceive sounds. How our ears perceive sound can be more accurately described by the logarithm of the intensity rather than directly to the intensity. The **sound intensity level** β in decibels of a sound having an intensity I in watts per meter squared is defined to be

$$\beta\ (\text{dB}) = 10\ \log_{10}\left(\frac{I}{I_0}\right), \qquad 17.12$$

where $I_0 = 10^{-12}\ \text{W/m}^2$ is a reference intensity. In particular, I_0 is the lowest or threshold intensity of sound a person with normal hearing can perceive at a frequency of 1000 Hz. Sound intensity level is not the same as intensity. Because β is defined in terms of a ratio, it is a unitless quantity telling you the *level* of the sound relative to a fixed standard ($10^{-12}\ \text{W/m}^2$, in this case). The units of decibels (dB) are used to indicate this ratio is multiplied by 10 in its definition. The bel, upon which the decibel is based, is named for Alexander Graham Bell, the inventor of the telephone.

Sound intensity level β (dB)	Intensity $I(\text{W/m}^2)$	Example/effect
0	1×10^{-12}	Threshold of hearing at 1000 Hz
10	1×10^{-11}	Rustle of leaves
20	1×10^{-10}	Whisper at 1 m distance
30	1×10^{-9}	Quiet home
40	1×10^{-8}	Average home
50	1×10^{-7}	Average office, soft music
60	1×10^{-6}	Normal conversation
70	1×10^{-5}	Noisy office, busy traffic
80	1×10^{-4}	Loud radio, classroom lecture

TABLE 17.2 Sound Intensity Levels and Intensities

Sound intensity level β (dB)	Intensity I(W/m^2)	Example/effect
90	1×10^{-3}	Inside a heavy truck; damage from prolonged exposure[1]
100	1×10^{-2}	Noisy factory, siren at 30 m; damage from 8 h per day exposure
110	1×10^{-1}	Damage from 30 min per day exposure
120	1	Loud rock concert, pneumatic chipper at 2 m; threshold of pain
140	1×10^{2}	Jet airplane at 30 m; severe pain, damage in seconds
160	1×10^{4}	Bursting of eardrums

TABLE 17.2 Sound Intensity Levels and Intensities

The decibel level of a sound having the threshold intensity of $10^{-12}\ \text{W/m}^2$ is $\beta = 0\ \text{dB}$, because $\log_{10} 1 = 0$. That is, the threshold of hearing is 0 decibels. Table 17.2 gives levels in decibels and intensities in watts per meter squared for some familiar sounds.

One of the more striking things about the intensities in Table 17.2 is that the intensity in watts per meter squared is quite small for most sounds. The ear is sensitive to as little as a trillionth of a watt per meter squared—even more impressive when you realize that the area of the eardrum is only about $1\ \text{cm}^2$, so that only 10^{-16} W falls on it at the threshold of hearing! Air molecules in a sound wave of this intensity vibrate over a distance of less than one molecular diameter, and the gauge pressures involved are less than 10^{-9} atm.

Another impressive feature of the sounds in Table 17.2 is their numerical range. Sound intensity varies by a factor of 10^{12} from threshold to a sound that causes damage in seconds. You are unaware of this tremendous range in sound intensity because how your ears respond can be described approximately as the logarithm of intensity. Thus, sound intensity levels in decibels fit your experience better than intensities in watts per meter squared. The decibel scale is also easier to relate to because most people are more accustomed to dealing with numbers such as 0, 53, or 120 than numbers such as 1.00×10^{-11}.

One more observation readily verified by examining Table 17.2 or using $I = \frac{(\Delta p)^2}{2\rho v_\text{w}}$ is that each factor of 10 in intensity corresponds to 10 dB. For example, a 90 dB sound compared with a 60 dB sound is 30 dB greater, or three factors of 10 (that is, 10^3 times) as intense. Another example is that if one sound is 10^7 as intense as another, it is 70 dB higher. See Table 17.3.

1 Several government agencies and health-related professional associations recommend that 85 dB not be exceeded for 8-hour daily exposures in the absence of hearing protection.

I_2/I_1	$\beta_2 - \beta_1$
2.0	3.0 dB
5.0	7.0 dB
10.0	10.0 dB

TABLE 17.3 Ratios of Intensities and Corresponding Differences in Sound Intensity Levels

EXAMPLE 17.2

Calculating Sound Intensity Levels: Sound Waves

Calculate the sound intensity level in decibels for a sound wave traveling in air at 0°C and having a pressure amplitude of 0.656 Pa.

Strategy

We are given Δp, so we can calculate I using the equation $I = (\Delta p)^2/\left(2\rho v_\text{w}\right)^2$. Using I, we can calculate β straight from its definition in $\beta\ (\text{dB}) = 10\ \log_{10}\left(I/I_0\right)$.

Solution

(1) Identify knowns:

Sound travels at 331 m/s in air at 0°C.

Air has a density of $1.29\ \text{kg/m}^3$ at atmospheric pressure and 0°C.

(2) Enter these values and the pressure amplitude into $I = (\Delta p)^2/\left(2\rho v_\text{w}\right)$:

$$I = \frac{(\Delta p)^2}{2\rho v_\text{w}} = \frac{(0.656\ \text{Pa})^2}{2\left(1.29\ \text{kg/m}^3\right)(331\ \text{m/s})} = 5.04 \times 10^{-4}\ \text{W/m}^2. \quad 17.13$$

(3) Enter the value for I and the known value for I_0 into $\beta\ (\text{dB}) = 10\ \log_{10}\left(I/I_0\right)$. Calculate to find the sound intensity level in decibels:

$$10\ \log_{10}\left(5.04 \times 10^8\right) = 10\ \left(8.70\right)\ \text{dB} = 87\ \text{dB}. \quad 17.14$$

Discussion

This 87 dB sound has an intensity five times as great as an 80 dB sound. So a factor of five in intensity corresponds to a difference of 7 dB in sound intensity level. This value is true for any intensities differing by a factor of five.

EXAMPLE 17.3

Change Intensity Levels of a Sound: What Happens to the Decibel Level?

Show that if one sound is twice as intense as another, it has a sound level about 3 dB higher.

Strategy

You are given that the ratio of two intensities is 2 to 1, and are then asked to find the difference in their sound levels in decibels. You can solve this problem using of the properties of logarithms.

Solution

(1) Identify knowns:

The ratio of the two intensities is 2 to 1, or:

$$\frac{I_2}{I_1} = 2.00. \quad (17.15)$$

We wish to show that the difference in sound levels is about 3 dB. That is, we want to show:

$$\beta_2 - \beta_1 = 3 \text{ dB}. \quad (17.16)$$

Note that:

$$\log_{10} b - \log_{10} a = \log_{10}\left(\frac{b}{a}\right). \quad (17.17)$$

(2) Use the definition of β to get:

$$\beta_2 - \beta_1 = 10 \log_{10}\left(\frac{I_2}{I_1}\right) = 10 \log_{10} 2.00 = 10\ (0.301) \text{ dB}. \quad (17.18)$$

Thus,

$$\beta_2 - \beta_1 = 3.01 \text{ dB}. \quad (17.19)$$

Discussion

This means that the two sound intensity levels differ by 3.01 dB, or about 3 dB, as advertised. Note that because only the ratio I_2/I_1 is given (and not the actual intensities), this result is true for any intensities that differ by a factor of two. For example, a 56.0 dB sound is twice as intense as a 53.0 dB sound, a 97.0 dB sound is half as intense as a 100 dB sound, and so on.

It should be noted at this point that there is another decibel scale in use, called the **sound pressure level**, based on the ratio of the pressure amplitude to a reference pressure. This scale is used particularly in applications where sound travels in water. It is beyond the scope of most introductory texts to treat this scale because it is not commonly used for sounds in air, but it is important to note that very different decibel levels may be encountered when sound pressure levels are quoted. For example, ocean noise pollution produced by ships may be as great as 200 dB expressed in the sound pressure level, where the more familiar sound intensity level we use here would be something under 140 dB for the same sound.

Take-Home Investigation: Feeling Sound

Find a CD player and a CD that has rock music. Place the player on a light table, insert the CD into the player, and start playing the CD. Place your hand gently on the table next to the speakers. Increase the volume and note the level when the table just begins to vibrate as the rock music plays. Increase the reading on the volume control until it doubles. What has happened to the vibrations?

CHECK YOUR UNDERSTANDING

Describe how amplitude is related to the loudness of a sound.

Solution

Amplitude is directly proportional to the experience of loudness. As amplitude increases, loudness increases.

CHECK YOUR UNDERSTANDING

Identify common sounds at the levels of 10 dB, 50 dB, and 100 dB.

Solution

10 dB: Running fingers through your hair.

50 dB: Inside a quiet home with no television or radio.

100 dB: Take-off of a jet plane.

17.4 Doppler Effect and Sonic Booms

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define Doppler effect, Doppler shift, and sonic boom.
- Calculate the frequency of a sound heard by someone observing Doppler shift.
- Describe the sounds produced by objects moving faster than the speed of sound.

The characteristic sound of a motorcycle buzzing by is an example of the **Doppler effect**. The high-pitch scream shifts dramatically to a lower-pitch roar as the motorcycle passes by a stationary observer. The closer the motorcycle brushes by, the more abrupt the shift. The faster the motorcycle moves, the greater the shift. We also hear this characteristic shift in frequency for passing race cars, airplanes, and trains. It is so familiar that it is used to imply motion and children often mimic it in play.

The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer. Although less familiar, this effect is easily noticed for a stationary source and moving observer. For example, if you ride a train past a stationary warning bell, you will hear the bell's frequency shift from high to low as you pass by. The actual change in frequency due to relative motion of source and observer is called a **Doppler shift**. The Doppler effect and Doppler shift are named for the Austrian physicist and mathematician Christian Johann Doppler (1803–1853), who did experiments with both moving sources and moving observers. Doppler, for example, had musicians play on a moving open train car and also play standing next to the train tracks as a train passed by. Their music was observed both on and off the train, and changes in frequency were measured.

What causes the Doppler shift? Figure 17.13, Figure 17.14, and Figure 17.15 compare sound waves emitted by stationary and moving sources in a stationary air mass. Each disturbance spreads out spherically from the point where the sound was emitted. If the source is stationary, then all of the spheres representing the air compressions in the sound wave centered on the same point, and the stationary observers on either side see the same wavelength and frequency as emitted by the source, as in Figure 17.13. If the source is moving, as in Figure 17.14, then the situation is different. Each compression of the air moves out in a sphere from the point where it was emitted, but the point of emission moves. This moving emission point causes the air compressions to be closer together on one side and farther apart on the other. Thus, the wavelength is shorter in the direction the source is moving (on the right in Figure 17.14), and longer in the opposite direction (on the left in Figure 17.14). Finally, if the observers move, as in Figure 17.15, the frequency at which they receive the compressions changes. The observer moving toward the source receives them at a higher frequency, and the person moving away from the source receives them at a lower frequency.

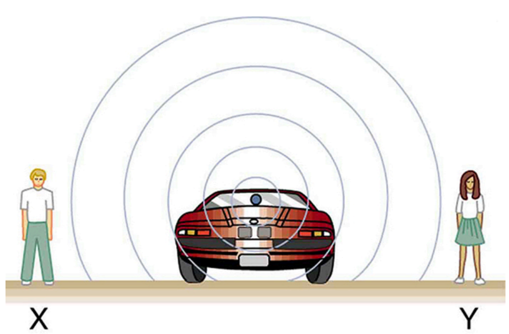


FIGURE 17.13 Sounds emitted by a source spread out in spherical waves. Because the source, observers, and air are stationary, the wavelength and frequency are the same in all directions and to all observers.

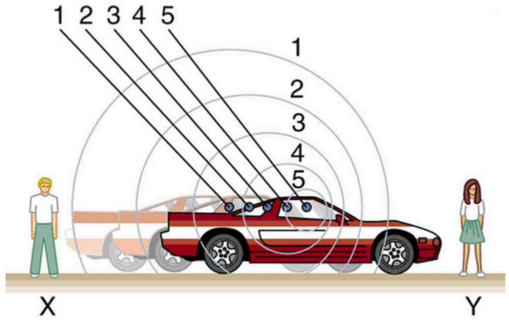


FIGURE 17.14 Sounds emitted by a source moving to the right spread out from the points at which they were emitted. The wavelength is reduced and, consequently, the frequency is increased in the direction of motion, so that the observer on the right hears a higher-pitch sound. The opposite is true for the observer on the left, where the wavelength is increased and the frequency is reduced.

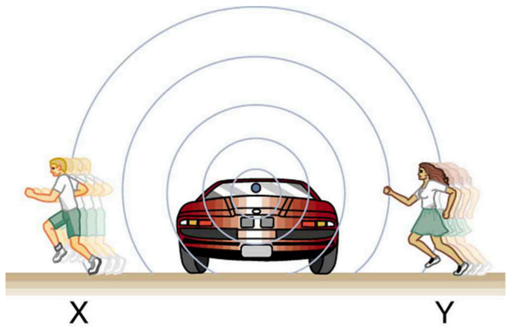


FIGURE 17.15 The same effect is produced when the observers move relative to the source. Motion toward the source increases frequency as the observer on the right passes through more wave crests than she would if stationary. Motion away from the source decreases frequency as the observer on the left passes through fewer wave crests than he would if stationary.

We know that wavelength and frequency are related by $v_{\text{w}} = f\lambda$, where v_{w} is the fixed speed of sound. The sound moves in a medium and has the same speed v_{w} in that medium whether the source is moving or not. Thus f multiplied by λ is a constant. Because the observer on the right in Figure 17.14 receives a shorter wavelength, the frequency she receives must be higher. Similarly, the observer on the left receives a longer wavelength, and hence he hears a lower frequency. The same thing happens in Figure 17.15. A higher frequency is received by the observer moving toward the source, and a lower frequency is received by an observer moving away from the source. In general, then, relative motion of source and observer toward one another increases the received frequency. Relative motion apart decreases frequency. The greater the relative speed is, the greater the effect.

The Doppler Effect

The Doppler effect occurs not only for sound but for any wave when there is relative motion between the observer and the source. There are Doppler shifts in the frequency of sound, light, and water waves, for example. Doppler shifts can be used to determine velocity, such as when ultrasound is reflected from blood in a medical diagnostic. The recession of galaxies is determined by the shift in the frequencies of light received from them and has implied much about the origins of the universe. Modern physics has been profoundly affected by observations of Doppler shifts.

For a stationary observer and a moving source, the frequency f_{obs} received by the observer can be shown to be

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right), \quad \textbf{17.20}$$

where f_{s} is the frequency of the source, v_{s} is the speed of the source along a line joining the source and observer, and v_{w} is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away from the observer, producing the appropriate shifts up and down in frequency. Note that the greater the speed of the source, the greater the effect. Similarly, for a stationary source and moving observer, the frequency received by the observer f_{obs} is given by

$$f_{obs} = f_s\left(\frac{v_w \pm v_{obs}}{v_w}\right), \quad 17.21$$

where v_{obs} is the speed of the observer along a line joining the source and observer. Here the plus sign is for motion toward the source, and the minus is for motion away from the source.

EXAMPLE 17.4

Calculate Doppler Shift: A Train Horn

Suppose a train that has a 150-Hz horn is moving at 35.0 m/s in still air on a day when the speed of sound is 340 m/s.

(a) What frequencies are observed by a stationary person at the side of the tracks as the train approaches and after it passes?

(b) What frequency is observed by the train's engineer traveling on the train?

Strategy

To find the observed frequency in (a), $f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right)$, must be used because the source is moving. The minus sign is used for the approaching train, and the plus sign for the receding train. In (b), there are two Doppler shifts—one for a moving source and the other for a moving observer.

Solution for (a)

(1) Enter known values into $f_{obs} = f_s\left(\frac{v_w}{v_w - v_s}\right)$.

$$f_{obs} = f_s\left(\frac{v_w}{v_w - v_s}\right) = (150\text{ Hz})\left(\frac{340\text{ m/s}}{340\text{ m/s} - 35.0\text{ m/s}}\right) \quad 17.22$$

(2) Calculate the frequency observed by a stationary person as the train approaches.

$$f_{obs} = (150\text{ Hz})(1.11) = 167\text{ Hz} \quad 17.23$$

(3) Use the same equation with the plus sign to find the frequency heard by a stationary person as the train recedes.

$$f_{obs} = f_s\left(\frac{v_w}{v_w + v_s}\right) = (150\text{ Hz})\left(\frac{340\text{ m/s}}{340\text{ m/s} + 35.0\text{ m/s}}\right) \quad 17.24$$

(4) Calculate the second frequency.

$$f_{obs} = (150\text{ Hz})(0.907) = 136\text{ Hz} \quad 17.25$$

Discussion on (a)

The numbers calculated are valid when the train is far enough away that the motion is nearly along the line joining train and observer. In both cases, the shift is significant and easily noticed. Note that the shift is 17.0 Hz for motion toward and 14.0 Hz for motion away. The shifts are not symmetric.

Solution for (b)

(1) Identify knowns:

- It seems reasonable that the engineer would receive the same frequency as emitted by the horn, because the relative velocity between them is zero.
- Relative to the medium (air), the speeds are $v_s = v_{obs} = 35.0\text{ m/s}$.
- The first Doppler shift is for the moving observer; the second is for the moving source.

(2) Use the following equation:

$$f_{obs} = \left[f_s\left(\frac{v_w \pm v_{obs}}{v_w}\right)\right]\left(\frac{v_w}{v_w \pm v_s}\right). \quad 17.26$$

The quantity in the square brackets is the Doppler-shifted frequency due to a moving observer. The factor on the right is the effect of the moving source.

(3) Because the train engineer is moving in the direction toward the horn, we must use the plus sign for v_{obs}; however, because the horn is also moving in the direction away from the engineer, we also use the plus sign for v_s. But the train is carrying both the engineer and the horn at the same velocity, so $v_s = v_{obs}$. As a result, everything but f_s cancels, yielding

$$f_{obs} = f_s. \tag{17.27}$$

Discussion for (b)

We may expect that there is no change in frequency when source and observer move together because it fits your experience. For example, there is no Doppler shift in the frequency of conversations between driver and passenger on a motorcycle. People talking when a wind moves the air between them also observe no Doppler shift in their conversation. The crucial point is that source and observer are not moving relative to each other.

Sonic Booms to Bow Wakes

What happens to the sound produced by a moving source, such as a jet airplane, that approaches or even exceeds the speed of sound? The answer to this question applies not only to sound but to all other waves as well.

Suppose a jet airplane is coming nearly straight at you, emitting a sound of frequency f_s. The greater the plane's speed v_s, the greater the Doppler shift and the greater the value observed for f_{obs}. Now, as v_s approaches the speed of sound, f_{obs} approaches infinity, because the denominator in $f_{obs} = f_s \left(\frac{v_w}{v_w \pm v_s} \right)$ approaches zero. At the speed of sound, this result means that in front of the source, each successive wave is superimposed on the previous one because the source moves forward at the speed of sound. The observer gets them all at the same instant, and so the frequency is infinite. (Before airplanes exceeded the speed of sound, some people argued it would be impossible because such constructive superposition would produce pressures great enough to destroy the airplane.) If the source exceeds the speed of sound, no sound is received by the observer until the source has passed, so that the sounds from the approaching source are mixed with those from it when receding. This mixing appears messy, but something interesting happens—a sonic boom is created. (See Figure 17.16.)

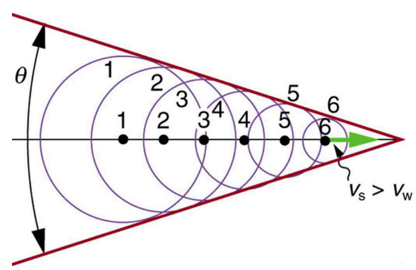


FIGURE 17.16 Sound waves from a source that moves faster than the speed of sound spread spherically from the point where they are emitted, but the source moves ahead of each. Constructive interference along the lines shown (actually a cone in three dimensions) creates a shock wave called a sonic boom. The faster the speed of the source, the smaller the angle θ.

There is constructive interference along the lines shown (a cone in three dimensions) from similar sound waves arriving there simultaneously. This superposition forms a disturbance called a **sonic boom**, a constructive interference of sound created by an object moving faster than sound. Inside the cone, the interference is mostly destructive, and so the sound intensity there is much less than on the shock wave. An aircraft creates two sonic booms, one from its nose and one from its tail. (See Figure 17.17.) During television coverage of space shuttle landings, two distinct booms could often be heard. These were separated by exactly the time it would take the shuttle to pass by a point. Observers on the ground often do not see the aircraft creating the sonic boom, because it has passed by before the shock wave reaches them, as seen in Figure 17.17. If the aircraft flies close by at low altitude, pressures in the sonic boom can be destructive and break windows as well as rattle nerves. Because of how destructive sonic booms can be, supersonic flights are banned over populated areas of the United States.

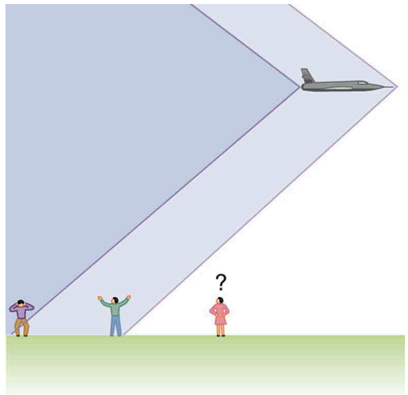


FIGURE 17.17 Two sonic booms, created by the nose and tail of an aircraft, are observed on the ground after the plane has passed by.

Sonic booms are one example of a broader phenomenon called bow wakes. A **bow wake**, such as the one in Figure 17.18, is created when the wave source moves faster than the wave propagation speed. Water waves spread out in circles from the point where created, and the bow wake is the familiar V-shaped wake trailing the source. A more exotic bow wake is created when a subatomic particle travels through a medium faster than the speed of light travels in that medium. (In a vacuum, the maximum speed of light will be $c = 3.00 \times 10^8\ \text{m/s}$; in the medium of water, the speed of light is closer to $0.75c$. If the particle creates light in its passage, that light spreads on a cone with an angle indicative of the speed of the particle, as illustrated in Figure 17.19. Such a bow wake is called Cerenkov radiation and is commonly observed in particle physics.

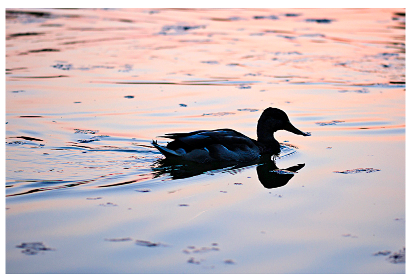


FIGURE 17.18 Bow wake created by a duck. Constructive interference produces the rather structured wake, while there is relatively little wave action inside the wake, where interference is mostly destructive. (credit: Horia Varlan, Flickr)

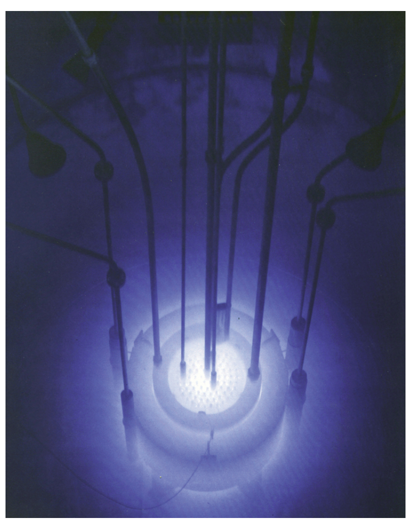


FIGURE 17.19 The blue glow in this research reactor pool is Cerenkov radiation caused by subatomic particles traveling faster than the speed of light in water. (credit: U.S. Nuclear Regulatory Commission)

Doppler shifts and sonic booms are interesting sound phenomena that occur in all types of waves. They can be of considerable use. For example, the Doppler shift in ultrasound can be used to measure blood velocity, while police use the Doppler shift in radar (a microwave) to measure car velocities. In meteorology, the Doppler shift is used to track the motion of storm clouds; such “Doppler Radar” can give velocity and direction and rain or snow potential of imposing weather fronts. In astronomy, we can examine the light emitted from distant galaxies and determine their speed relative to ours. As galaxies move away from us, their light is shifted to a lower frequency, and so to a longer wavelength—the so-called red shift. Such information from galaxies far, far away has allowed us to estimate the age of the universe (from the Big Bang) as about 14 billion years.

CHECK YOUR UNDERSTANDING

Why did scientist Christian Doppler observe musicians both on a moving train and also from a stationary point not on the train?

Solution

Doppler needed to compare the perception of sound when the observer is stationary and the sound source moves, as well as when the sound source and the observer are both in motion.

CHECK YOUR UNDERSTANDING

Describe a situation in your life when you might rely on the Doppler shift to help you either while driving a car or walking near traffic.

Solution

If I am driving and I hear Doppler shift in an ambulance siren, I would be able to tell when it was getting closer and also if it has passed by. This would help me to know whether I needed to pull over and let the ambulance through.

17.5 Sound Interference and Resonance: Standing Waves in Air Columns

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define antinode, node, fundamental, overtones, and harmonics.
- Identify instances of sound interference in everyday situations.
- Describe how sound interference occurring inside open and closed tubes changes the characteristics of the sound, and how this applies to sounds produced by musical instruments.
- Calculate the length of a tube using sound wave measurements.

FIGURE 17.20 Some types of headphones use the phenomena of constructive and destructive interference to cancel out outside noises. (credit: JVC America, Flickr)

Interference is the hallmark of waves, all of which exhibit constructive and destructive interference exactly analogous to that seen for water waves. In fact, one way to prove something "is a wave" is to observe interference effects. So, sound being a wave, we expect it to exhibit interference; we have already mentioned a few such effects, such as the beats from two similar notes played simultaneously.

Figure 17.21 shows a clever use of sound interference to cancel noise. Larger-scale applications of active noise reduction by destructive interference are contemplated for entire passenger compartments in commercial aircraft. To obtain destructive interference, a fast electronic analysis is performed, and a second sound is introduced with its maxima and minima exactly reversed from the incoming noise. Sound waves in fluids are pressure waves and consistent with Pascal's principle; pressures from two different sources add and subtract like simple numbers; that is, positive and negative gauge pressures add to a much smaller pressure, producing a lower-intensity sound. Although completely destructive interference is possible only under the simplest conditions, it is possible to reduce noise levels by 30 dB or more using this technique.

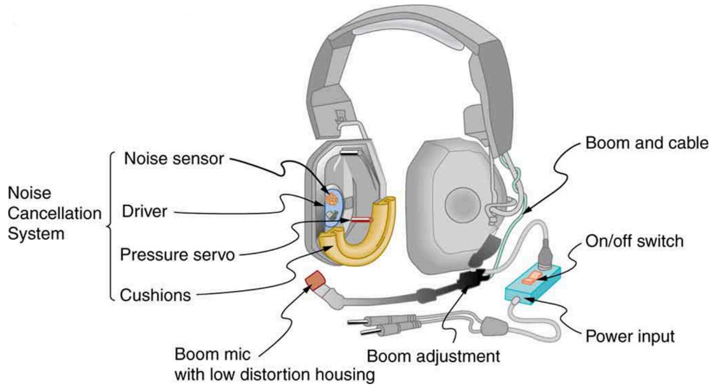


FIGURE 17.21 Headphones designed to cancel noise with destructive interference create a sound wave exactly opposite to the incoming sound. These headphones can be more effective than the simple passive attenuation used in most ear protection. Such headphones were used on the record-setting, around the world nonstop flight of the Voyager aircraft to protect the pilots' hearing from engine noise.

Where else can we observe sound interference? All sound resonances, such as in musical instruments, are due to constructive and destructive interference. Only the resonant frequencies interfere constructively to form standing waves, while others interfere destructively and are absent. From the toot made by blowing over a bottle, to the characteristic flavor of a violin's sounding box, to the recognizability of a great singer's voice, resonance and standing waves play a vital role.

Interference

Interference is such a fundamental aspect of waves that observing interference is proof that something is a wave. The wave nature of light was established by experiments showing interference. Similarly, when electrons scattered from crystals exhibited interference, their wave nature was confirmed to be exactly as predicted by symmetry with certain wave characteristics of light.

Suppose we hold a tuning fork near the end of a tube that is closed at the other end, as shown in Figure 17.22, Figure 17.23, Figure 17.24, and Figure 17.25. If the tuning fork has just the right frequency, the air column in the tube resonates loudly, but at most frequencies it vibrates very little. This observation just means that the air column has only certain natural frequencies. The figures show how a resonance at the lowest of these natural frequencies is formed. A disturbance travels down the tube at the speed of sound and bounces off the closed end. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes constructively with the continuing sound produced by the tuning fork. The incoming and reflected sounds form a standing wave in the tube as shown.

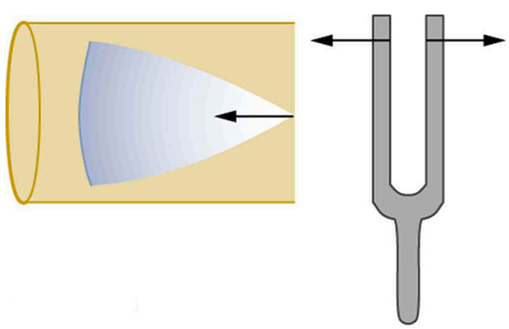


FIGURE 17.22 Resonance of air in a tube closed at one end, caused by a tuning fork. A disturbance moves down the tube.

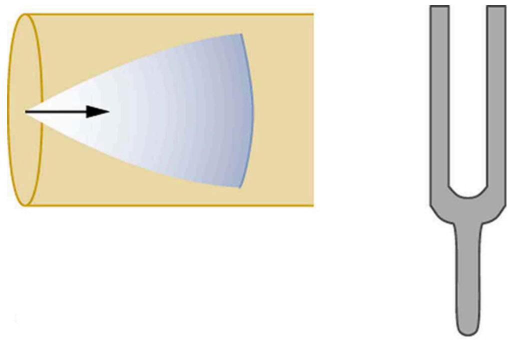


FIGURE 17.23 Resonance of air in a tube closed at one end, caused by a tuning fork. The disturbance reflects from the closed end of the tube.

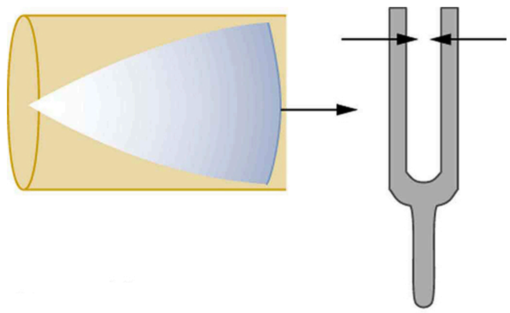


FIGURE 17.24 Resonance of air in a tube closed at one end, caused by a tuning fork. If the length of the tube L is just right, the disturbance gets back to the tuning fork half a cycle later and interferes constructively with the continuing sound from the tuning fork. This interference forms a standing wave, and the air column resonates.

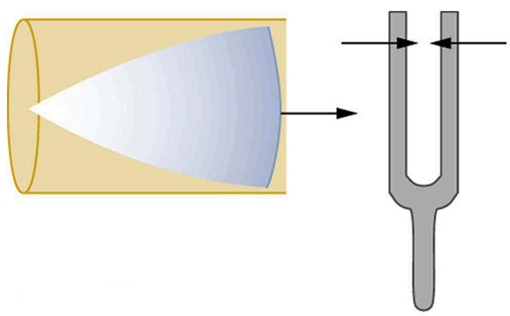


FIGURE 17.25 Resonance of air in a tube closed at one end, caused by a tuning fork. A graph of air displacement along the length of the tube shows none at the closed end, where the motion is constrained, and a maximum at the open end. This standing wave has one-fourth of its wavelength in the tube, so that $\lambda = 4L$.

The standing wave formed in the tube has its maximum air displacement (an **antinode**) at the open end, where motion is unconstrained, and no displacement (a **node**) at the closed end, where air movement is halted. The distance from a node to an antinode is one-fourth of a wavelength, and this equals the length of the tube; thus, $\lambda = 4L$. This same resonance can be produced by a vibration introduced at or near the closed end of the tube, as shown in Figure 17.26. It is best to consider this a natural vibration of the air column independently of how it is induced.

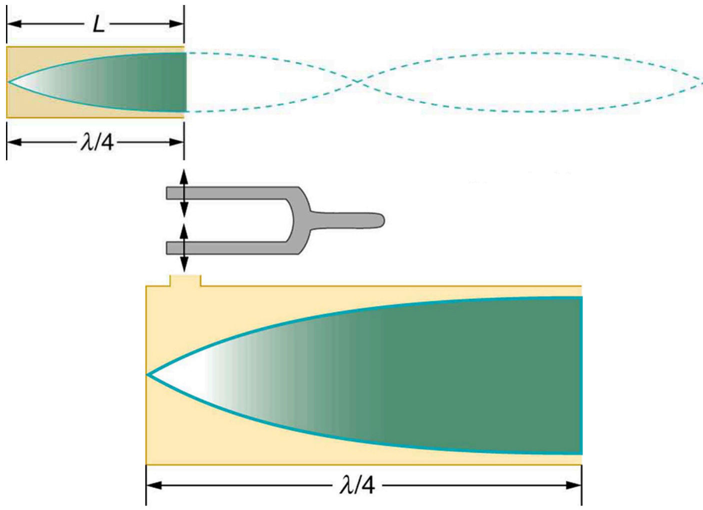


FIGURE 17.26 The same standing wave is created in the tube by a vibration introduced near its closed end.

Given that maximum air displacements are possible at the open end and none at the closed end, there are other, shorter wavelengths that can resonate in the tube, such as the one shown in Figure 17.27. Here the standing wave has three-fourths of its wavelength in the tube, or $L = (3/4)\lambda'$, so that $\lambda' = 4L/3$. Continuing this process reveals a whole series of shorter-wavelength and higher-frequency sounds that resonate in the tube. We use specific terms for the resonances in any system. The lowest resonant frequency is called the **fundamental**, while all higher resonant frequencies are called **overtones**. All resonant frequencies are integral multiples of the fundamental, and they are collectively called **harmonics**. The fundamental is the first harmonic, the first overtone is the second harmonic, and so on. Figure 17.28 shows the fundamental and the first three overtones (the first four harmonics) in a tube closed at one end.

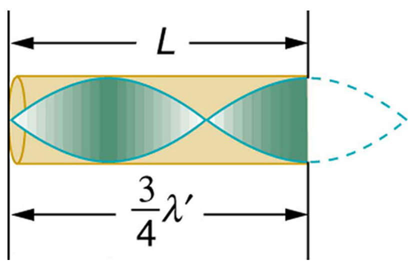


FIGURE 17.27 Another resonance for a tube closed at one end. This has maximum air displacements at the open end, and none at the closed end. The wavelength is shorter, with three-fourths λ' equaling the length of the tube, so that $\lambda' = 4L/3$. This higher-frequency vibration is the first overtone.

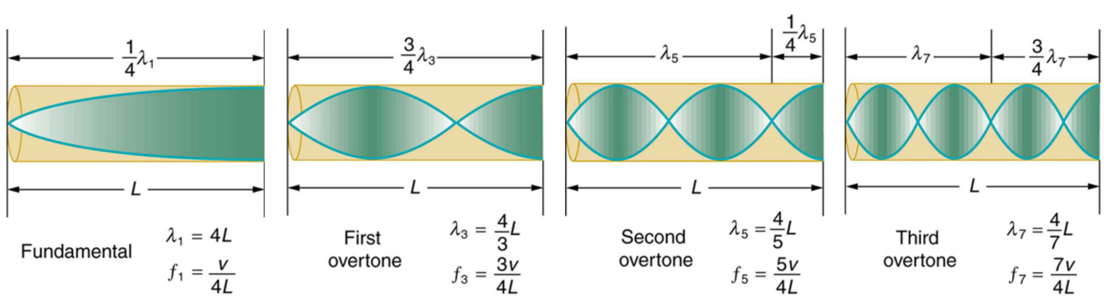


FIGURE 17.28 The fundamental and three lowest overtones for a tube closed at one end. All have maximum air displacements at the open end and none at the closed end.

The fundamental and overtones can be present simultaneously in a variety of combinations. For example, middle C on a trumpet has a sound distinctively different from middle C on a clarinet, both instruments being modified versions of a tube closed at one end. The fundamental frequency is the same (and usually the most intense), but the overtones and their mix of intensities are different and subject to shading by the musician. This mix is what gives various musical instruments (and human voices) their distinctive characteristics, whether they have air columns, strings, sounding boxes, or drumheads. In fact, much of our speech is determined by shaping the cavity formed by the throat and mouth and positioning the tongue to adjust the fundamental and combination of overtones. Simple resonant cavities can be made to resonate with the sound of the vowels, for example. (See Figure 17.29.) In males, at puberty, the larynx grows and the shape of the resonant cavity changes giving rise to the difference in predominant frequencies in speech between different sexes.

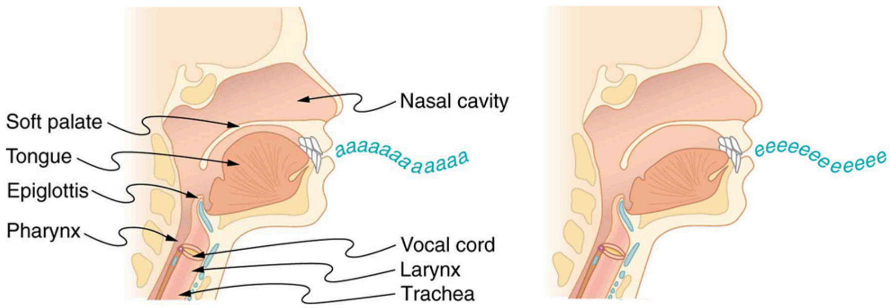


FIGURE 17.29 The throat and mouth form an air column closed at one end that resonates in response to vibrations in the voice box. The spectrum of overtones and their intensities vary with mouth shaping and tongue position to form different sounds. The voice box can be replaced with a mechanical vibrator, and understandable speech is still possible. Variations in basic shapes make different voices recognizable.

Now let us look for a pattern in the resonant frequencies for a simple tube that is closed at one end. The fundamental has $\lambda = 4L$, and frequency is related to wavelength and the speed of sound as given by:

$$v_{\mathrm{w}} = f\lambda. \quad (17.28)$$

Solving for f in this equation gives

$$f = \frac{v_{\mathrm{w}}}{\lambda} = \frac{v_{\mathrm{w}}}{4L}, \quad (17.29)$$

where v_{w} is the speed of sound in air. Similarly, the first overtone has $\lambda' = 4L/3$ (see Figure 17.28), so that

$$f' = 3\frac{v_{\mathrm{w}}}{4L} = 3f. \quad (17.30)$$

Because $f' = 3f$, we call the first overtone the third harmonic. Continuing this process, we see a pattern that can be generalized in a single expression. The resonant frequencies of a tube closed at one end are

$$f_n = n\frac{v_{\mathrm{w}}}{4L}, n = 1,3,5, \quad (17.31)$$

where f_1 is the fundamental, f_3 is the first overtone, and so on. It is interesting that the resonant frequencies depend on the speed of sound and, hence, on temperature. This dependence poses a noticeable problem for organs in old unheated cathedrals, and it is also the reason why musicians commonly bring their wind instruments to room temperature before playing them.

EXAMPLE 17.5

Find the Length of a Tube with a 128 Hz Fundamental

(a) What length should a tube closed at one end have on a day when the air temperature, is $22.0^{\circ}\mathrm{C}$, if its fundamental frequency is to be 128 Hz (C below middle C)?

(b) What is the frequency of its fourth overtone?

Strategy

The length L can be found from the relationship in $f_n = n\frac{v_{\mathrm{w}}}{4L}$, but we will first need to find the speed of sound v_{w}.

Solution for (a)

(1) Identify knowns:

- the fundamental frequency is 128 Hz

- the air temperature is $22.0^{\circ}\mathrm{C}$

(2) Use $f_n = n\frac{v_w}{4L}$ to find the fundamental frequency ($n = 1$).

$$f_1 = \frac{v_w}{4L} \tag{17.32}$$

(3) Solve this equation for length.

$$L = \frac{v_w}{4f_1} \tag{17.33}$$

(4) Find the speed of sound using $v_w = (331\text{ m/s})\sqrt{\frac{T}{273\text{ K}}}$.

$$v_w = (331\text{ m/s})\sqrt{\frac{295\text{ K}}{273\text{ K}}} = 344\text{ m/s} \tag{17.34}$$

(5) Enter the values of the speed of sound and frequency into the expression for L.

$$L = \frac{v_w}{4f_1} = \frac{344\text{ m/s}}{4(128\text{ Hz})} = 0.672\text{ m} \tag{17.35}$$

Discussion on (a)

Many wind instruments are modified tubes that have finger holes, valves, and other devices for changing the length of the resonating air column and hence, the frequency of the note played. Horns producing very low frequencies, such as tubas, require tubes so long that they are coiled into loops.

Solution for (b)

(1) Identify knowns:

- the first overtone has $n = 3$
- the second overtone has $n = 5$
- the third overtone has $n = 7$
- the fourth overtone has $n = 9$

(2) Enter the value for the fourth overtone into $f_n = n\frac{v_w}{4L}$.

$$f_9 = 9\frac{v_w}{4L} = 9f_1 = 1.15\text{ kHz} \tag{17.36}$$

Discussion on (b)

Whether this overtone occurs in a simple tube or a musical instrument depends on how it is stimulated to vibrate and the details of its shape. The trombone, for example, does not produce its fundamental frequency and only makes overtones.

Another type of tube is one that is *open* at both ends. Examples are some organ pipes, flutes, and oboes. The resonances of tubes open at both ends can be analyzed in a very similar fashion to those for tubes closed at one end. The air columns in tubes open at both ends have maximum air displacements at both ends, as illustrated in Figure 17.30. Standing waves form as shown.

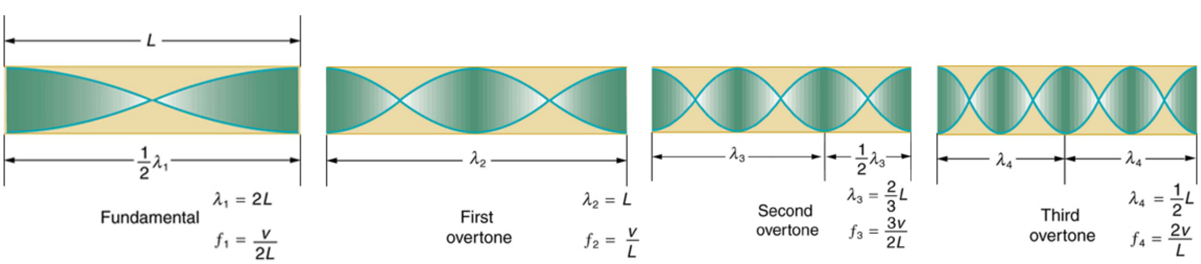


FIGURE 17.30 The resonant frequencies of a tube open at both ends are shown, including the fundamental and the first three overtones. In

all cases the maximum air displacements occur at both ends of the tube, giving it different natural frequencies than a tube closed at one end.

Based on the fact that a tube open at both ends has maximum air displacements at both ends, and using Figure 17.30 as a guide, we can see that the resonant frequencies of a tube open at both ends are:

$$f_n = n\frac{v_w}{2L},\ n = 1, 2, 3...,\qquad 17.37$$

where f_1 is the fundamental, f_2 is the first overtone, f_3 is the second overtone, and so on. Note that a tube open at both ends has a fundamental frequency twice what it would have if closed at one end. It also has a different spectrum of overtones than a tube closed at one end. So if you had two tubes with the same fundamental frequency but one was open at both ends and the other was closed at one end, they would sound different when played because they have different overtones. Middle C, for example, would sound richer played on an open tube, because it has even multiples of the fundamental as well as odd. A closed tube has only odd multiples.

Real-World Applications: Resonance in Everyday Systems

Resonance occurs in many different systems, including strings, air columns, and atoms. Resonance is the driven or forced oscillation of a system at its natural frequency. At resonance, energy is transferred rapidly to the oscillating system, and the amplitude of its oscillations grows until the system can no longer be described by Hooke’s law. An example of this is the distorted sound intentionally produced in certain types of rock music.

Wind instruments use resonance in air columns to amplify tones made by lips or vibrating reeds. Other instruments also use air resonance in clever ways to amplify sound. Figure 17.31 shows a violin and a guitar, both of which have sounding boxes but with different shapes, resulting in different overtone structures. The vibrating string creates a sound that resonates in the sounding box, greatly amplifying the sound and creating overtones that give the instrument its characteristic flavor. The more complex the shape of the sounding box, the greater its ability to resonate over a wide range of frequencies. The marimba, like the one shown in Figure 17.32 uses pots or gourds below the wooden slats to amplify their tones. The resonance of the pot can be adjusted by adding water.

FIGURE 17.31 String instruments such as violins and guitars use resonance in their sounding boxes to amplify and enrich the sound created by their vibrating strings. The bridge and supports couple the string vibrations to the sounding boxes and air within. (credits: guitar, Feliciano Guimares, Fotopedia; violin, Steve Snodgrass, Flickr)

FIGURE 17.32 Resonance has been used in musical instruments since prehistoric times. This marimba uses gourds as resonance chambers to amplify its sound. (credit: APC Events, Flickr)

We have emphasized sound applications in our discussions of resonance and standing waves, but these ideas apply to any system that has wave characteristics. Vibrating strings, for example, are actually resonating and have fundamentals and overtones similar to those for air columns. More subtle are the resonances in atoms due to the wave character of their electrons. Their orbitals can be viewed as standing waves, which have a fundamental (ground state) and overtones (excited states). It is fascinating that wave characteristics apply to such a wide range of physical systems.

CHECK YOUR UNDERSTANDING

Describe how noise-canceling headphones differ from standard headphones used to block outside sounds.

Solution

Regular headphones only block sound waves with a physical barrier. Noise-canceling headphones use destructive interference to reduce the loudness of outside sounds.

CHECK YOUR UNDERSTANDING

How is it possible to use a standing wave's node and antinode to determine the length of a closed-end tube?

Solution

When the tube resonates at its natural frequency, the wave's node is located at the closed end of the tube, and the antinode is located at the open end. The length of the tube is equal to one-fourth of the wavelength of this wave. Thus, if we know the wavelength of the wave, we can determine the length of the tube.

17.6 Hearing

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define hearing, pitch, loudness, timbre, note, tone, phon, ultrasound, and infrasound.
- Compare loudness to frequency and intensity of a sound.
- Identify structures of the inner ear and explain how they relate to sound perception.

FIGURE 17.33 Hearing allows this vocalist, his band, and his fans to enjoy music. (credit: West Point Public Affairs, Flickr)

The human ear has a tremendous range and sensitivity. It can give us a wealth of simple information—such as pitch, loudness, and direction. And from its input we can detect musical quality and nuances of voiced emotion. How is our hearing related to the physical qualities of sound, and how does the hearing mechanism work?

Hearing is the perception of sound. (Perception is commonly defined to be awareness through the senses, a typically circular definition of higher-level processes in living organisms.) Normal human hearing encompasses frequencies from 20 to 20,000 Hz, an impressive range. Sounds below 20 Hz are called **infrasound**, whereas those above 20,000 Hz are **ultrasound**. Neither is perceived by the ear, although infrasound can sometimes be felt as vibrations. When we do hear low-frequency vibrations, such as the sounds of a diving board, we hear the individual vibrations only because there are higher-frequency sounds in each. Other animals have hearing ranges different from that of humans. Dogs can hear sounds as high as 30,000 Hz, whereas bats and dolphins can hear up to 100,000-Hz sounds. You may have noticed that dogs respond to the sound of a dog whistle which produces sound out of the range of human hearing. Elephants are known to respond to frequencies below 20 Hz.

The perception of frequency is called **pitch**. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different frequency from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500.0 and 501.5 Hz are noticeably different. Pitch perception is directly related to frequency and is not greatly affected by other physical quantities such as intensity. Musical **notes** are particular sounds that can be produced by most instruments and in Western music have particular names. Combinations of notes constitute music. Some people can identify musical notes, such as A-sharp, C, or E-flat, just by listening to them. This uncommon ability is called perfect pitch.

The ear is remarkably sensitive to low-intensity sounds. The lowest audible intensity or threshold is about $10^{-12}\ \text{W/m}^2$ or 0 dB. Sounds as much as 10^{12} more intense can be briefly tolerated. Very few measuring devices are capable of observations over a range of a trillion. The perception of intensity is called **loudness**. At a given frequency, it is possible to discern differences of about 1 dB, and a change of 3 dB is easily noticed. But loudness is not related to intensity alone. Frequency has a major effect on how loud a sound seems. The ear has its maximum sensitivity to frequencies in the range of 2000 to 5000 Hz, so that sounds in this range are perceived as being louder than, say, those at 500 or 10,000 Hz, even when they all have the same intensity. Sounds near the high- and low-frequency extremes of the hearing range seem even less loud, because the ear is even less sensitive at those frequencies. Table 17.4 gives the dependence of certain human hearing perceptions on physical quantities.

Perception	Physical quantity
Pitch	Frequency
Loudness	Intensity and Frequency
Timbre	Number and relative intensity of multiple frequencies. Subtle craftsmanship leads to non-linear effects and more detail.

TABLE 17.4 Sound Perceptions

Perception	Physical quantity
Note	Basic unit of music with specific names, combined to generate tunes
Tone	Number and relative intensity of multiple frequencies.

TABLE 17.4 Sound Perceptions

When a violin plays middle C, there is no mistaking it for a piano playing the same note. The reason is that each instrument produces a distinctive set of frequencies and intensities. We call our perception of these combinations of frequencies and intensities **tone** quality, or more commonly the **timbre** of the sound. It is more difficult to correlate timbre perception to physical quantities than it is for loudness or pitch perception. Timbre is more subjective. Terms such as dull, brilliant, warm, cold, pure, and rich are employed to describe the timbre of a sound. So the consideration of timbre takes us into the realm of perceptual psychology, where higher-level processes in the brain are dominant. This is true for other perceptions of sound, such as music and noise. We shall not delve further into them; rather, we will concentrate on the question of loudness perception.

A unit called a **phon** is used to express loudness numerically. Phons differ from decibels because the phon is a unit of loudness perception, whereas the decibel is a unit of physical intensity. Figure 17.34 shows the relationship of loudness to intensity (or intensity level) and frequency for persons with normal hearing. The curved lines are equal-loudness curves. Each curve is labeled with its loudness in phons. Any sound along a given curve will be perceived as equally loud by the average person. The curves were determined by having large numbers of people compare the loudness of sounds at different frequencies and sound intensity levels. At a frequency of 1000 Hz, phons are taken to be numerically equal to decibels. The following example helps illustrate how to use the graph:

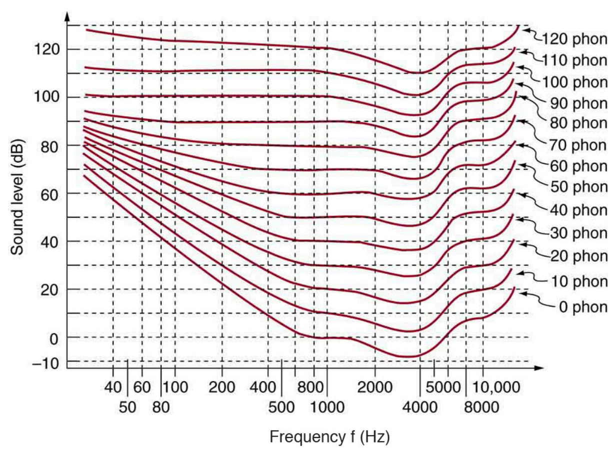


FIGURE 17.34 The relationship of loudness in phons to intensity level (in decibels) and intensity (in watts per meter squared) for persons with normal hearing. The curved lines are equal-loudness curves—all sounds on a given curve are perceived as equally loud. Phons and decibels are defined to be the same at 1000 Hz.

EXAMPLE 17.6

Measuring Loudness: Loudness Versus Intensity Level and Frequency

(a) What is the loudness in phons of a 100-Hz sound that has an intensity level of 80 dB? (b) What is the intensity level in decibels of a 4000-Hz sound having a loudness of 70 phons? (c) At what intensity level will an 8000-Hz sound have the same loudness as a 200-Hz sound at 60 dB?

Strategy for (a)

The graph in Figure 17.34 should be referenced in order to solve this example. To find the loudness of a given sound, you must know its frequency and intensity level and locate that point on the square grid, then interpolate between loudness curves to get the loudness in phons.

Solution for (a)

(1) Identify knowns:

- The square grid of the graph relating phons and decibels is a plot of intensity level versus frequency—both physical quantities.
- 100 Hz at 80 dB lies halfway between the curves marked 70 and 80 phons.

(2) Find the loudness: 75 phons.

Strategy for (b)

The graph in Figure 17.34 should be referenced in order to solve this example. To find the intensity level of a sound, you must have its frequency and loudness. Once that point is located, the intensity level can be determined from the vertical axis.

Solution for (b)

(1) Identify knowns:

- Values are given to be 4000 Hz at 70 phons.

(2) Follow the 70-phon curve until it reaches 4000 Hz. At that point, it is below the 70 dB line at about 67 dB.

(3) Find the intensity level:

67 dB

Strategy for (c)

The graph in Figure 17.34 should be referenced in order to solve this example.

Solution for (c)

(1) Locate the point for a 200 Hz and 60 dB sound.

(2) Find the loudness: This point lies just slightly above the 50-phon curve, and so its loudness is 51 phons.

(3) Look for the 51-phon level is at 8000 Hz: 63 dB.

Discussion

These answers, like all information extracted from Figure 17.34, have uncertainties of several phons or several decibels, partly due to difficulties in interpolation, but mostly related to uncertainties in the equal-loudness curves.

Further examination of the graph in Figure 17.34 reveals some interesting facts about human hearing. First, sounds below the 0-phon curve are not perceived by most people. So, for example, a 60 Hz sound at 40 dB is inaudible. The 0-phon curve represents the threshold of normal hearing. We can hear some sounds at intensity levels below 0 dB. For example, a 3-dB, 5000-Hz sound is audible, because it lies above the 0-phon curve. The loudness curves all have dips in them between about 2000 and 5000 Hz. These dips mean the ear is most sensitive to frequencies in that range. For example, a 15-dB sound at 4000 Hz has a loudness of 20 phons, the same as a 20-dB sound at 1000 Hz. The curves rise at both extremes of the frequency range, indicating that a greater-intensity level sound is needed at those frequencies to be perceived to be as loud as at middle frequencies. For example, a sound at 10,000 Hz must have an intensity level of 30 dB to seem as loud as a 20 dB sound at 1000 Hz. Sounds above 120 phons are painful as well as damaging.

We do not often utilize our full range of hearing. This is particularly true for frequencies above 8000 Hz, which are rare in the environment and are unnecessary for understanding conversation or appreciating music. In fact, people who have lost the ability to hear such high frequencies are usually unaware of their loss until tested. The shaded region in Figure 17.35 is the frequency and intensity region where most conversational sounds fall. The curved lines indicate what effect hearing losses of 40 and 60 phons will have. A 40-phon hearing loss at all frequencies still allows a person to understand conversation, although it will seem very quiet. A person with a 60-phon loss at all frequencies will hear only the lowest frequencies and will not be able to understand speech unless it is much louder than normal. Even so, speech may seem indistinct, because higher frequencies are not as well perceived. The conversational speech region also has a gender component, in that female voices are usually characterized by higher frequencies. So the person with a 60-phon hearing impediment might have difficulty understanding the normal conversation of a woman.

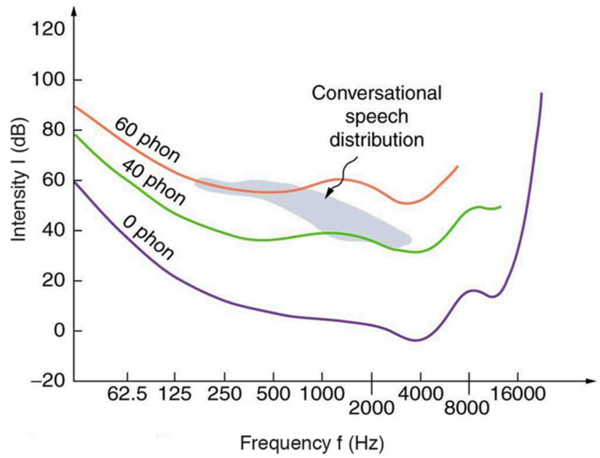


FIGURE 17.35 The shaded region represents frequencies and intensity levels found in normal conversational speech. The 0-phon line represents the normal hearing threshold, while those at 40 and 60 represent thresholds for people with 40- and 60-phon hearing losses, respectively.

Hearing tests are performed over a range of frequencies, usually from 250 to 8000 Hz, and can be displayed graphically in an audiogram like that in Figure 17.36. The hearing threshold is measured in dB *relative to the normal threshold,* so that normal hearing registers as 0 dB at all frequencies. Hearing loss caused by noise typically shows a dip near the 4000 Hz frequency, irrespective of the frequency that caused the loss and often affects both ears. The most common form of hearing loss comes with age and is called *presbycusis*—literally elder ear. Such loss is increasingly severe at higher frequencies, and interferes with music appreciation and speech recognition.

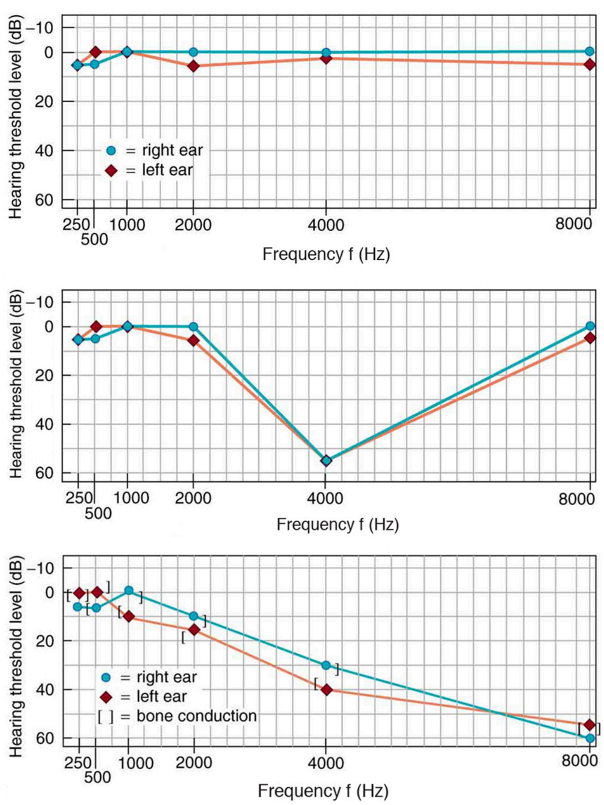


FIGURE 17.36 Audiograms showing the threshold in intensity level versus frequency for three different individuals. Intensity level is measured relative to the normal threshold. The top left graph is that of a person with normal hearing. The graph to its right has a dip at

4000 Hz and is that of a child who suffered hearing loss due to a cap gun. The third graph is typical of presbycusis, the progressive loss of higher frequency hearing with age. Tests performed by bone conduction (brackets) can distinguish nerve damage from middle ear damage.

The Hearing Mechanism

The hearing mechanism involves some interesting physics. The sound wave that impinges upon our ear is a pressure wave. The ear is a transducer that converts sound waves into electrical nerve impulses in a manner much more sophisticated than, but analogous to, a microphone. Figure 17.37 shows the gross anatomy of the ear with its division into three parts: the outer ear or ear canal; the middle ear, which runs from the eardrum to the cochlea; and the inner ear, which is the cochlea itself. The body part normally referred to as the ear is technically called the pinna.

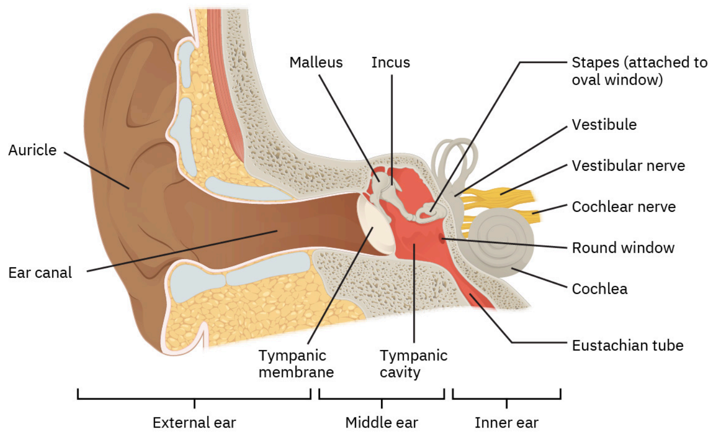


FIGURE 17.37 The illustration shows the gross anatomy of the human ear.

The outer ear, or ear canal, carries sound to the recessed protected eardrum. The air column in the ear canal resonates and is partially responsible for the sensitivity of the ear to sounds in the 2000 to 5000 Hz range. The middle ear converts sound into mechanical vibrations and applies these vibrations to the cochlea. The lever system of the middle ear takes the force exerted on the eardrum by sound pressure variations, amplifies it and transmits it to the inner ear via the oval window, creating pressure waves in the cochlea approximately 40 times greater than those impinging on the eardrum. (See Figure 17.38.) Two muscles in the middle ear (not shown) protect the inner ear from very intense sounds. They react to intense sound in a few milliseconds and reduce the force transmitted to the cochlea. This protective reaction can also be triggered by your own voice, so that humming while shooting a gun, for example, can reduce noise damage.

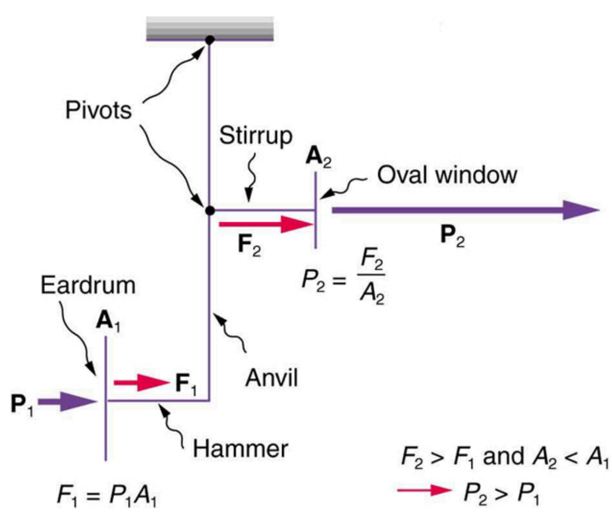


FIGURE 17.38 This schematic shows the middle ear's system for converting sound pressure into force, increasing that force through a lever system, and applying the increased force to a small area of the cochlea, thereby creating a pressure about 40 times that in the original sound wave. A protective muscle reaction to intense sounds greatly reduces the mechanical advantage of the lever system.

Figure 17.39 shows the middle and inner ear in greater detail. Pressure waves moving through the cochlea cause the tectorial membrane to vibrate, rubbing cilia (called hair cells), which stimulate nerves that send electrical signals to the brain. The membrane resonates at different positions for different frequencies, with high frequencies stimulating nerves at the near end and low frequencies at the far end. The complete operation of the cochlea is still not understood, but several mechanisms for sending information to the brain are known to be involved. For sounds below about 1000 Hz, the nerves send signals at the same frequency as the sound. For frequencies greater than about 1000 Hz, the nerves signal frequency by position. There is a structure to the cilia, and there are connections between nerve cells that perform signal processing before information is sent to the brain. Intensity information is partly indicated by the number of nerve signals and by volleys of signals. The brain processes the cochlear nerve signals to provide additional information such as source direction (based on time and intensity comparisons of sounds from both ears). Higher-level processing produces many nuances, such as music appreciation.

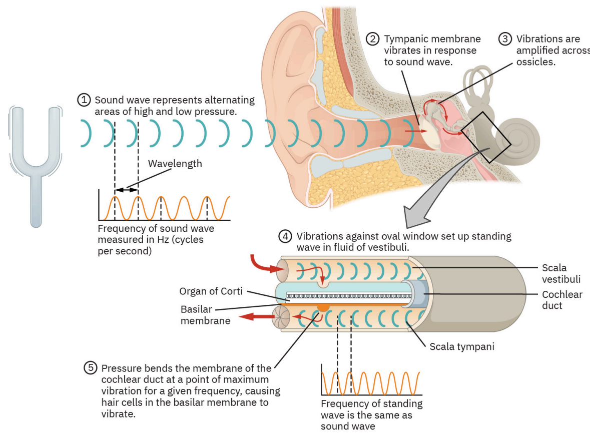


FIGURE 17.39 The inner ear, or cochlea, is a coiled tube about 3 mm in diameter and 3 cm in length if uncoiled. When the oval window is forced inward, as shown, a pressure wave travels through the perilymph in the direction of the arrows, stimulating nerves at the base of cilia in the organ of Corti.

Hearing losses can occur because of problems in the middle or inner ear. Conductive losses in the middle ear can be partially overcome by sending sound vibrations to the cochlea through the skull. Hearing aids for this purpose usually press against the bone behind the ear, rather than simply amplifying the sound sent into the ear canal as many hearing aids do. Damage to the nerves in the cochlea is not repairable, but amplification can partially compensate. There is a risk that amplification will produce further damage. Another common failure in the cochlea is damage or loss of the cilia but with nerves remaining functional. Cochlear implants that stimulate the nerves directly are now available and widely accepted. Over 100,000 implants are in use, in about equal numbers of adults and children.

The cochlear implant was pioneered in Melbourne, Australia, by Graeme Clark in the 1970s for his deaf father. The implant consists of three external components and two internal components. The external components are a microphone for picking up sound and converting it into an electrical signal, a speech processor to select certain frequencies and a transmitter to transfer the signal to the internal components through electromagnetic induction. The internal components consist of a receiver/transmitter secured in the bone beneath the skin, which converts the signals into electric impulses and sends them through an internal cable to the cochlea and an array of about 24 electrodes wound through the cochlea. These electrodes in turn send the impulses directly into the brain. The electrodes basically emulate the cilia.

CHECK YOUR UNDERSTANDING

Are ultrasound and infrasound imperceptible to all hearing organisms? Explain your answer.

Solution

No, the range of perceptible sound is based in the range of human hearing. Many other organisms perceive either infrasound or ultrasound.

17.7 Ultrasound

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define acoustic impedance and intensity reflection coefficient.
- Describe medical and other uses of ultrasound technology.
- Calculate acoustic impedance using density values and the speed of ultrasound.
- Calculate the velocity of a moving object using Doppler-shifted ultrasound.

FIGURE 17.40 Ultrasound is used in medicine to painlessly and noninvasively monitor patient health and diagnose a wide range of disorders. (credit: abbybatchelder, Flickr)

Any sound with a frequency above 20,000 Hz (or 20 kHz)—that is, above the highest audible frequency—is defined to be ultrasound. In practice, it is possible to create ultrasound frequencies up to more than a gigahertz. (Higher frequencies are difficult to create; furthermore, they propagate poorly because they are very strongly absorbed.) Ultrasound has a tremendous number of applications, which range from burglar alarms to use in cleaning delicate objects to the guidance systems of bats. We begin our discussion of ultrasound with some of its applications in medicine, in which it is used extensively both for diagnosis and for therapy.

Characteristics of Ultrasound

The characteristics of ultrasound, such as frequency and intensity, are wave properties common to all types of waves. Ultrasound also has a wavelength that limits the fineness of detail it can detect. This characteristic is true of all waves. We can never observe details significantly smaller than the wavelength of our probe; for example, we will never see individual atoms with visible light, because the atoms are so small compared with the wavelength of light.

Ultrasound in Medical Therapy

Ultrasound, like any wave, carries energy that can be absorbed by the medium carrying it, producing effects that vary with intensity. When focused to intensities of 10^3 to $10^5\ \mathrm{W/m^2}$, ultrasound can be used to shatter gallstones or pulverize cancerous tissue in surgical procedures. (See Figure 17.41.) Intensities this great can damage individual cells, variously causing their protoplasm to stream inside them, altering their permeability, or rupturing their walls through *cavitation*. Cavitation is the creation of vapor cavities in a fluid—the longitudinal vibrations in ultrasound alternatively compress and expand the medium, and at sufficient amplitudes the expansion separates molecules. Most cavitation damage is done when the cavities collapse, producing even greater shock pressures.

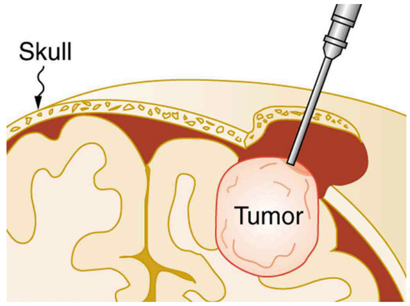


FIGURE 17.41 The tip of this small probe oscillates at 23 kHz with such a large amplitude that it pulverizes tissue on contact. The debris is then aspirated. The speed of the tip may exceed the speed of sound in tissue, thus creating shock waves and cavitation, rather than a smooth simple harmonic oscillator–type wave.

Most of the energy carried by high-intensity ultrasound in tissue is converted to thermal energy. In fact, intensities of 10^3 to $10^4\ \text{W/m}^2$ are commonly used for deep-heat treatments called ultrasound diathermy. Frequencies of 0.8 to 1 MHz are typical. In both athletics and physical therapy, ultrasound diathermy is most often applied to injured or overworked muscles to relieve pain and improve flexibility. Skill is needed by the therapist to avoid “bone burns” and other tissue damage caused by overheating and cavitation, sometimes made worse by reflection and focusing of the ultrasound by joint and bone tissue.

In some instances, you may encounter a different decibel scale, called the sound *pressure* level, when ultrasound travels in water or in human and other biological tissues. We shall not use the scale here, but it is notable that numbers for sound pressure levels range 60 to 70 dB higher than you would quote for β, the sound intensity level used in this text. Should you encounter a sound pressure level of 220 decibels, then, it is not an astronomically high intensity, but equivalent to about 155 dB—high enough to destroy tissue, but not as unreasonably high as it might seem at first.

Ultrasound in Medical Diagnostics

When used for imaging, ultrasonic waves are emitted from a transducer, a crystal exhibiting the piezoelectric effect (the expansion and contraction of a substance when a voltage is applied across it, causing a vibration of the crystal). These high-frequency vibrations are transmitted into any tissue in contact with the transducer. Similarly, if a pressure is applied to the crystal (in the form of a wave reflected off tissue layers), a voltage is produced which can be recorded. The crystal therefore acts as both a transmitter and a receiver of sound. Ultrasound is also partially absorbed by tissue on its path, both on its journey away from the transducer and on its return journey. From the time between when the original signal is sent and when the reflections from various boundaries between media are received, (as well as a measure of the intensity loss of the signal), the nature and position of each boundary between tissues and organs may be deduced.

Reflections at boundaries between two different media occur because of differences in a characteristic known as the **acoustic impedance** Z of each substance. Impedance is defined as

$$Z = \rho v, \tag{17.38}$$

where ρ is the density of the medium (in kg/m^3) and v is the speed of sound through the medium (in m/s). The units for Z are therefore $\text{kg/(m}^2 \cdot \text{s)}$.

Table 17.5 shows the density and speed of sound through various media (including various soft tissues) and the associated acoustic impedances. Note that the acoustic impedances for soft tissue do not vary much but that there is a big difference between the acoustic impedance of soft tissue and air and also between soft tissue and bone.

Medium	Density (kg/m^3)	Speed of Ultrasound (m/s)	Acoustic Impedance $\left(\text{kg}/\left(\text{m}^2 \cdot \text{s}\right)\right)$
Air	1.3	330	429
Water	1000	1500	1.5×10^6
Blood	1060	1570	1.66×10^6
Fat	925	1450	1.34×10^6
Muscle (average)	1075	1590	1.70×10^6
Bone (varies)	1400–1900	4080	5.7×10^6 to 7.8×10^6
Barium titanate (transducer material)	5600	5500	30.8×10^6

TABLE 17.5 The Ultrasound Properties of Various Media, Including Soft Tissue Found in the Body

At the boundary between media of different acoustic impedances, some of the wave energy is reflected and some is transmitted. The greater the *difference* in acoustic impedance between the two media, the greater the reflection and the smaller the transmission.

The **intensity reflection coefficient** a is defined as the ratio of the intensity of the reflected wave relative to the incident (transmitted) wave. This statement can be written mathematically as

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}, \tag{17.39}$$

where Z_1 and Z_2 are the acoustic impedances of the two media making up the boundary. A reflection coefficient of zero (corresponding to total transmission and no reflection) occurs when the acoustic impedances of the two media are the same. An impedance “match” (no reflection) provides an efficient coupling of sound energy from one medium to another. The image formed in an ultrasound is made by tracking reflections (as shown in Figure 17.42) and mapping the intensity of the reflected sound waves in a two-dimensional plane.

EXAMPLE 17.7

Calculate Acoustic Impedance and Intensity Reflection Coefficient: Ultrasound and Fat Tissue

(a) Using the values for density and the speed of ultrasound given in Table 17.5, show that the acoustic impedance of fat tissue is indeed $1.34 \times 10^6\ \text{kg/(m}^2\cdot\text{s)}$.

(b) Calculate the intensity reflection coefficient of ultrasound when going from fat to muscle tissue.

Strategy for (a)

The acoustic impedance can be calculated using $Z = \rho v$ and the values for ρ and v found in Table 17.5.

Solution for (a)

(1) Substitute known values from Table 17.5 into $Z = \rho v$.

$$Z = \rho v = \left(925\ \text{kg/m}^3\right)(1450\ \text{m/s}) \tag{17.40}$$

(2) Calculate to find the acoustic impedance of fat tissue.

$$1.34 \times 10^6 \text{ kg/(m}^2\cdot\text{s)} \quad (17.41)$$

This value is the same as the value given for the acoustic impedance of fat tissue.

Strategy for (b)

The intensity reflection coefficient for any boundary between two media is given by $a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}$, and the acoustic impedance of muscle is given in Table 17.5.

Solution for (b)

Substitute known values into $a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}$ to find the intensity reflection coefficient:

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2} = \frac{\left(1.34 \times 10^6 \text{ kg/(m}^2\cdot\text{ s)} - 1.70 \times 10^6 \text{ kg/(m}^2\cdot\text{ s)}\right)^2}{\left(1.70 \times 10^6 \text{ kg/(m}^2\cdot\text{ s)} + 1.34 \times 10^6 \text{ kg/(m}^2\cdot\text{ s)}\right)^2} = 0.014 \quad (17.42)$$

Discussion

This result means that only 1.4% of the incident intensity is reflected, with the remaining being transmitted.

The applications of ultrasound in medical diagnostics have produced untold benefits with no known risks. Diagnostic intensities are too low (about $10^{-2}\ \text{W/m}^2$) to cause thermal damage. More significantly, ultrasound has been in use for several decades and detailed follow-up studies do not show evidence of ill effects, quite unlike the case for x-rays.

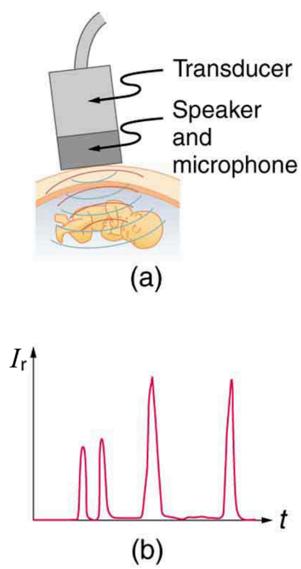
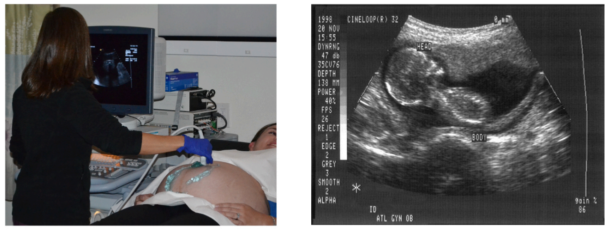


FIGURE 17.42 (a) An ultrasound speaker doubles as a microphone. Brief bleeps are broadcast, and echoes are recorded from various depths. (b) Graph of echo intensity versus time. The time for echoes to return is directly proportional to the distance of the reflector, yielding this information noninvasively.

The most common ultrasound applications produce an image like that shown in Figure 17.43. The speaker-microphone broadcasts a directional beam, sweeping the beam across the area of interest. This is accomplished by having multiple ultrasound sources in the probe's head, which are phased to interfere constructively in a given, adjustable direction. Echoes are measured as a function of position as well as depth. A computer constructs an image that reveals the shape and density of internal structures.



(a) (b)

FIGURE 17.43 (a) An ultrasonic image is produced by sweeping the ultrasonic beam across the area of interest, in this case the person's abdomen. Data are recorded and analyzed in a computer, providing a two-dimensional image (credit: COD Newsroom, Flickr). (b) Ultrasound image of 12-week-old fetus. (credit: Public Health Image Library, CDC)

How much detail can ultrasound reveal? The image in Figure 17.43 is typical of low-cost systems, but that in Figure 17.44 shows the remarkable detail possible with more advanced systems, including 3D imaging. Ultrasound today is commonly used in prenatal care. Such imaging can be used to see if the fetus is developing at a normal rate, and help in the determination of serious problems early in the pregnancy. Ultrasound is also in wide use to image the chambers of the heart and the flow of blood within the beating heart, using the Doppler effect (echocardiology).

Whenever a wave is used as a probe, it is very difficult to detect details smaller than its wavelength λ. Indeed, current technology cannot do quite this well. Abdominal scans may use a 7-MHz frequency, and the speed of sound in tissue is about 1540 m/s—so the wavelength limit to detail would be $\lambda = \frac{v_{\text{w}}}{f} = \frac{1540\text{ m/s}}{7{\times}10^6\text{ Hz}} = 0.22\text{ mm}$. In practice, 1-mm detail is attainable, which is sufficient for many purposes. Higher-frequency ultrasound would allow greater detail, but it does not penetrate as well as lower frequencies do. The accepted rule of thumb is that you can effectively scan to a depth of about 500λ into tissue. For 7 MHz, this penetration limit is $500 \times 0.22\text{ mm}$, which is 0.11 m. Higher frequencies may be employed in smaller organs, such as the eye, but are not practical for looking deep into the body.

FIGURE 17.44 A 3D ultrasound image of a fetus. (credit: Jennie Cu, Wikimedia Commons)

In addition to shape information, ultrasonic scans can produce density information superior to that found in X-rays, because the intensity of a reflected sound is related to changes in density. Sound is most strongly reflected at places where density changes are greatest.

Another major use of ultrasound in medical diagnostics is to detect motion and determine velocity through the Doppler shift of an echo, known as **Doppler-shifted ultrasound**. This technique is used to monitor fetal heartbeat, measure blood velocity, and detect occlusions in blood vessels, for example. (See Figure 17.45.) The magnitude of the Doppler shift in an echo is directly proportional to the velocity of whatever reflects the sound. Because an echo is involved, there is actually a double shift. The first occurs because the reflector (say a fetal heart) is a moving observer and receives a Doppler-shifted frequency. The reflector then acts as a moving source, producing a second Doppler shift.

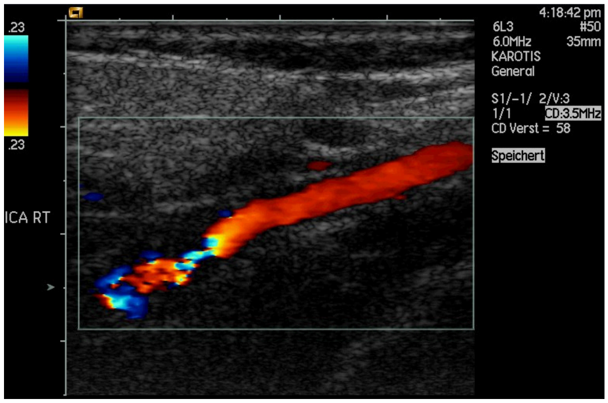


FIGURE 17.45 This Doppler-shifted ultrasonic image of a partially occluded artery uses color to indicate velocity. The highest velocities are in red, while the lowest are blue. The blood must move faster through the constriction to carry the same flow. (credit: Arning C, Grzyska U, Wikimedia Commons)

A clever technique is used to measure the Doppler shift in an echo. The frequency of the echoed sound is superimposed on the broadcast frequency, producing beats. The beat frequency is $F_B = |f_1 - f_2|$, and so it is directly proportional to the Doppler shift $(f_1 - f_2)$ and hence, the reflector's velocity. The advantage in this technique is that the Doppler shift is small (because the reflector's velocity is small), so that great accuracy would be needed to measure the shift directly. But measuring the beat frequency is easy, and it is not affected if the broadcast frequency varies somewhat. Furthermore, the beat frequency is in the audible range and can be amplified for audio feedback to the medical observer.

Uses for Doppler-Shifted Radar

Doppler-shifted radar echoes are used to measure wind velocities in storms as well as aircraft and automobile speeds. The principle is the same as for Doppler-shifted ultrasound. There is evidence that bats and dolphins may also sense the velocity of an object (such as prey) reflecting their ultrasound signals by observing its Doppler shift.

EXAMPLE 17.8

Calculate Velocity of Blood: Doppler-Shifted Ultrasound

Ultrasound that has a frequency of 2.50 MHz is sent toward blood in an artery that is moving toward the source at 20.0 cm/s, as illustrated in Figure 17.46. Use the speed of sound in human tissue as 1540 m/s. (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

a. What frequency does the blood receive?
b. What frequency returns to the source?
c. What beat frequency is produced if the source and returning frequencies are mixed?

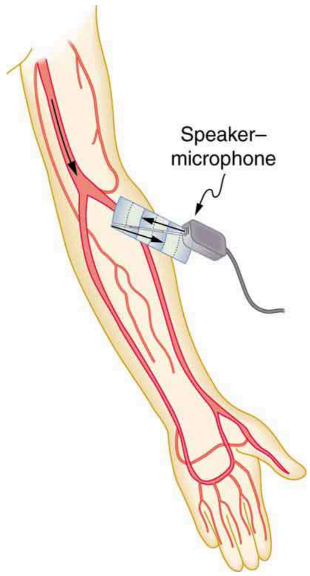


FIGURE 17.46 Ultrasound is partly reflected by blood cells and plasma back toward the speaker-microphone. Because the cells are moving, two Doppler shifts are produced—one for blood as a moving observer, and the other for the reflected sound coming from a moving source. The magnitude of the shift is directly proportional to blood velocity.

Strategy

The first two questions can be answered using $f_{obs} = f_s \left(\frac{v_w}{v_w \pm v_s} \right)$ and $f_{obs} = f_s \left(\frac{v_w \pm v_{obs}}{v_w} \right)$ for the Doppler shift. The last question asks for beat frequency, which is the difference between the original and returning frequencies.

Solution for (a)

(1) Identify knowns:

- The blood is a moving observer, and so the frequency it receives is given by

$$f_{obs} = f_s \left(\frac{v_w \pm v_{obs}}{v_w} \right). \quad (17.43)$$

- v_b is the blood velocity (v_{obs} here) and the plus sign is chosen because the motion is toward the source.

(2) Enter the given values into the equation.

$$f_{obs} = (2{,}500{,}000 \text{ Hz}) \left(\frac{1540 \text{ m/s} + 0.2 \text{ m/s}}{1540 \text{ m/s}} \right) \quad (17.44)$$

(3) Calculate to find the frequency: 2,500,325 Hz.

Solution for (b)

(1) Identify knowns:

- The blood acts as a moving source.
- The microphone acts as a stationary observer.
- The frequency leaving the blood is 2,500,325 Hz, but it is shifted upward as given by

$$f_{obs} = f_s \left(\frac{v_w}{v_w - v_b} \right). \quad (17.45)$$

f_{obs} is the frequency received by the speaker-microphone.

- The source velocity is v_b.
- The minus sign is used because the motion is toward the observer.

The minus sign is used because the motion is toward the observer.

(2) Enter the given values into the equation:

$$f_{\text{obs}} = (2{,}500{,}325\text{ Hz})\left(\frac{1540\text{ m/s}}{1540\text{ m/s} - 0.200\text{ m/s}}\right) \quad (17.46)$$

(3) Calculate to find the frequency returning to the source: 2,500,649 Hz.

Solution for (c)

(1) Identify knowns:

- The beat frequency is simply the absolute value of the difference between f_s and f_{obs}, as stated in:
$$f_B = |f_{\text{obs}} - f_s|. \quad (17.47)$$

(2) Substitute known values:

$$|2{,}500{,}649\text{ Hz} - 2{,}500{,}000\text{ Hz}| \quad (17.48)$$

(3) Calculate to find the beat frequency: 649 Hz.

Discussion

The Doppler shifts are quite small compared with the original frequency of 2.50 MHz. It is far easier to measure the beat frequency than it is to measure the echo frequency with an accuracy great enough to see shifts of a few hundred hertz out of a couple of megahertz. Furthermore, variations in the source frequency do not greatly affect the beat frequency, because both f_s and f_{obs} would increase or decrease. Those changes subtract out in $f_B = |f_{\text{obs}} - f_s|$.

Industrial and Other Applications of Ultrasound

Industrial, retail, and research applications of ultrasound are common. A few are discussed here. Ultrasonic cleaners have many uses. Jewelry, machined parts, and other objects that have odd shapes and crevices are immersed in a cleaning fluid that is agitated with ultrasound typically about 40 kHz in frequency. The intensity is great enough to cause cavitation, which is responsible for most of the cleansing action. Because cavitation-produced shock pressures are large and well transmitted in a fluid, they reach into small crevices where even a low-surface-tension cleaning fluid might not penetrate.

Sonar is a familiar application of ultrasound. Sonar typically employs ultrasonic frequencies in the range from 30.0 to 100 kHz. Bats, dolphins, submarines, and even some birds use ultrasonic sonar. Echoes are analyzed to give distance and size information both for guidance and finding prey. In most sonar applications, the sound reflects quite well because the objects of interest have significantly different density than the medium in which they travel. When the Doppler shift is observed, velocity information can also be obtained. Submarine sonar can be used to obtain such information, and there is evidence that some bats also sense velocity from their echoes.

Similarly, there are a range of relatively inexpensive devices that measure distance by timing ultrasonic echoes. Many cameras, for example, use such information to focus automatically. Some doors open when their ultrasonic ranging devices detect a nearby object, and certain home security lights turn on when their ultrasonic rangers observe motion. Ultrasonic “measuring tapes” also exist to measure such things as room dimensions. Sinks in public restrooms are sometimes automated with ultrasound devices to turn faucets on and off when people wash their hands. These devices reduce the spread of germs and can conserve water.

Ultrasound is used for nondestructive testing in industry and by the military. Because ultrasound reflects well from any large change in density, it can reveal cracks and voids in solids, such as aircraft wings, that are too small to be seen with x-rays. For similar reasons, ultrasound is also good for measuring the thickness of coatings, particularly where there are several layers involved.

Basic research in solid state physics employs ultrasound. Its attenuation is related to a number of physical characteristics, making it a useful probe. Among these characteristics are structural changes such as those found in liquid crystals, the transition of a material to a superconducting phase, as well as density and other properties.

These examples of the uses of ultrasound are meant to whet the appetites of the curious, as well as to illustrate the underlying physics of ultrasound. There are many more applications, as you can easily discover for yourself.

CHECK YOUR UNDERSTANDING

Why is it possible to use ultrasound both to observe a fetus in the womb and also to destroy cancerous tumors in the body?

Solution

Ultrasound can be used medically at different intensities. Lower intensities do not cause damage and are used for medical imaging. Higher intensities can pulverize and destroy targeted substances in the body, such as tumors.

Glossary

acoustic impedance property of medium that makes the propagation of sound waves more difficult
antinode point of maximum displacement
bow wake V-shaped disturbance created when the wave source moves faster than the wave propagation speed
Doppler effect an alteration in the observed frequency of a sound due to motion of either the source or the observer
Doppler shift the actual change in frequency due to relative motion of source and observer
Doppler-shifted ultrasound a medical technique to detect motion and determine velocity through the Doppler shift of an echo
fundamental the lowest-frequency resonance
harmonics the term used to refer collectively to the fundamental and its overtones
hearing the perception of sound
infrasound sounds below 20 Hz
intensity the power per unit area carried by a wave
intensity reflection coefficient a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave
loudness the perception of sound intensity
node point of zero displacement
note basic unit of music with specific names, combined to generate tunes
overtones all resonant frequencies higher than the fundamental
phon the numerical unit of loudness
pitch the perception of the frequency of a sound
sonic boom a constructive interference of sound created by an object moving faster than sound
sound a disturbance of matter that is transmitted from its source outward
sound intensity level a unitless quantity telling you the level of the sound relative to a fixed standard
sound pressure level the ratio of the pressure amplitude to a reference pressure
timbre number and relative intensity of multiple sound frequencies
tone number and relative intensity of multiple sound frequencies
ultrasound sounds above 20,000 Hz

Section Summary

17.1 Sound

- Sound is a disturbance of matter that is transmitted from its source outward.
- Sound is one type of wave.
- Hearing is the perception of sound.

17.2 Speed of Sound, Frequency, and Wavelength

The relationship of the speed of sound v_w, its frequency f, and its wavelength λ is given by

$$v_\mathrm{w} = f\lambda,$$

which is the same relationship given for all waves.

In air, the speed of sound is related to air temperature T by

$$v_\mathrm{w} = (331\ \mathrm{m/s})\sqrt{\frac{T}{273\ \mathrm{K}}}.$$

v_w is the same for all frequencies and wavelengths.

17.3 Sound Intensity and Sound Level

- Intensity is the same for a sound wave as was defined for all waves; it is

  $$I = \frac{P}{A},$$

  where P is the power crossing area A. The SI unit for I is watts per meter squared. The intensity of a sound wave is also related to the pressure amplitude Δp

  $$I = \frac{(\Delta p)^2}{2\rho v_\mathrm{w}},$$

  where ρ is the density of the medium in which the sound wave travels and v_w is the speed of sound in the medium.
- Sound intensity level in units of decibels (dB) is

  $$\beta\ (\mathrm{dB}) = 10\log_{10}\left(\frac{I}{I_0}\right),$$

  where $I_0 = 10^{-12}\ \mathrm{W/m^2}$ is the threshold intensity of hearing.

17.4 Doppler Effect and Sonic Booms

- The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer.
- The actual change in frequency is called the Doppler shift.
- A sonic boom is constructive interference of sound created by an object moving faster than sound.
- A sonic boom is a type of bow wake created when any wave source moves faster than the wave propagation speed.

- For a stationary observer and a moving source, the observed frequency f_{obs} is:
  $f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right)$,
  where f_s is the frequency of the source, v_s is the speed of the source, and v_w is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away.
- For a stationary source and moving observer, the observed frequency is:
  $f_{obs} = f_s\left(\frac{v_w \pm v_{obs}}{v_w}\right)$,
  where v_{obs} is the speed of the observer.

17.5 Sound Interference and Resonance: Standing Waves in Air Columns

- Sound interference and resonance have the same properties as defined for all waves.
- In air columns, the lowest-frequency resonance is called the fundamental, whereas all higher resonant frequencies are called overtones. Collectively, they are called harmonics.
- The resonant frequencies of a tube closed at one end are:
  $f_n = n\frac{v_w}{4L}, n = 1, 3, 5...,$
  f_1 is the fundamental and L is the length of the tube.
- The resonant frequencies of a tube open at both ends are:
  $f_n = n\frac{v_w}{2L}, n = 1, 2, 3...$

17.6 Hearing

- The range of audible frequencies is 20 to 20,000 Hz.
- Those sounds above 20,000 Hz are ultrasound, whereas those below 20 Hz are infrasound.
- The perception of frequency is pitch.
- The perception of intensity is loudness.
- Loudness has units of phons.

17.7 Ultrasound

- The acoustic impedance is defined as:
  $Z = \rho v,$
  ρ is the density of a medium through which the sound travels and v is the speed of sound through that medium.
- The intensity reflection coefficient a, a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave, is given by
  $a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}.$
- The intensity reflection coefficient is a unitless quantity.

Conceptual Questions

17.2 Speed of Sound, Frequency, and Wavelength

1. How do sound vibrations of atoms differ from thermal motion?

2. When sound passes from one medium to another where its propagation speed is different, does its frequency or wavelength change? Explain your answer briefly.

17.3 Sound Intensity and Sound Level

3. Six members of a synchronized swim team wear earplugs to protect themselves against water pressure at depths, but they can still hear the music and perform the combinations in the water perfectly. One day, they were asked to leave the pool so the dive team could practice a few dives, and they tried to practice on a mat, but seemed to have a lot more difficulty. Why might this be?

4. A community is concerned about a plan to bring train service to their downtown from the town's outskirts. The current sound intensity level, even though the rail yard is blocks away, is 70 dB downtown. The mayor assures the public that there will be a difference of only 30 dB in sound in the downtown area. Should the townspeople be concerned? Why?

17.4 Doppler Effect and Sonic Booms

5. Is the Doppler shift real or just a sensory illusion?

6. Due to efficiency considerations related to its bow wake, the supersonic transport aircraft must maintain a cruising speed that is a constant ratio to the speed of sound (a constant Mach number). If the aircraft flies from warm air into colder air, should it increase or decrease its speed? Explain your answer.

7. When you hear a sonic boom, you often cannot see the plane that made it. Why is that?

17.5 Sound Interference and Resonance: Standing Waves in Air Columns

8. How does an unamplified guitar produce sounds so much more intense than those of a plucked string held taut by a simple stick?
9. You are given two wind instruments of identical length. One is open at both ends, whereas the other is closed at one end. Which is able to produce the lowest frequency?
10. What is the difference between an overtone and a harmonic? Are all harmonics overtones? Are all overtones harmonics?

17.6 Hearing

11. Why can a hearing test show that your threshold of hearing is 0 dB at 250 Hz, when Figure 17.35 implies that no one can hear such a frequency at less than 20 dB?

17.7 Ultrasound

12. If audible sound follows a rule of thumb similar to that for ultrasound, in terms of its absorption, would you expect the high or low frequencies from your neighbor's stereo to penetrate into your house? How does this expectation compare with your experience?
13. Elephants and whales are known to use infrasound to communicate over very large distances. What are the advantages of infrasound for long distance communication?
14. It is more difficult to obtain a high-resolution ultrasound image in the abdominal region of someone who is overweight than for someone who has a slight build. Explain why this statement is accurate.
15. Suppose you read that 210-dB ultrasound is being used to pulverize cancerous tumors. You calculate the intensity in watts per centimeter squared and find it is unreasonably high ($10^5\ \text{W/cm}^2$). What is a possible explanation?

Problems & Exercises

17.2 Speed of Sound, Frequency, and Wavelength

1. When poked by a spear, an operatic soprano lets out a 1200-Hz shriek. What is its wavelength if the speed of sound is 345 m/s?
2. What frequency sound has a 0.10-m wavelength when the speed of sound is 340 m/s?
3. Calculate the speed of sound on a day when a 1500 Hz frequency has a wavelength of 0.221 m.
4. (a) What is the speed of sound in a medium where a 100-kHz frequency produces a 5.96-cm wavelength? (b) Which substance in Table 17.1 is this likely to be?
5. Show that the speed of sound in 20.0°C air is 343 m/s, as claimed in the text.
6. Air temperature in the Sahara Desert can reach 56.0°C (about 134°F). What is the speed of sound in air at that temperature?
7. Dolphins make sounds in air and water. What is the ratio of the wavelength of a sound in air to its wavelength in seawater? Assume air temperature is 20.0°C.
8. A sonar echo returns to a submarine 1.20 s after being emitted. What is the distance to the object creating the echo? (Assume that the submarine is in the ocean, not in fresh water.)
9. (a) If a submarine's sonar can measure echo times with a precision of 0.0100 s, what is the smallest difference in distances it can detect? (Assume that the submarine is in the ocean, not in fresh water.) (b) Discuss the limits this time resolution imposes on the ability of the sonar system to detect the size and shape of the object creating the echo.
10. A physicist at a fireworks display times the lag between seeing an explosion and hearing its sound, and finds it to be 0.400 s. (a) How far away is the explosion if air temperature is 24.0°C and if you neglect the time taken for light to reach the physicist? (b) Calculate the distance to the explosion taking the speed of light into account. Note that this distance is negligibly greater.
11. Suppose a bat uses sound echoes to locate its insect prey, 3.00 m away. (See Figure 17.9.) (a) Calculate the echo times for temperatures of 5.00°C and 35.0°C. (b) What percent uncertainty does this cause for the bat in locating the insect? (c) Discuss the significance of this uncertainty and whether it could cause difficulties for the bat. (In practice, the bat continues to use sound as it closes in, eliminating most of any difficulties imposed by this and other effects, such as motion of the prey.)

17.3 Sound Intensity and Sound Level

12. What is the intensity in watts per meter squared of 85.0-dB sound?

13. The warning tag on a lawn mower states that it produces noise at a level of 91.0 dB. What is this in watts per meter squared?

14. A sound wave traveling in 20°C air has a pressure amplitude of 0.5 Pa. What is the intensity of the wave?

15. What intensity level does the sound in the preceding problem correspond to?

16. What sound intensity level in dB is produced by earphones that create an intensity of $4.00 \times 10^{-2}\ \text{W/m}^2$?

17. Show that an intensity of $10^{-12}\ \text{W/m}^2$ is the same as $10^{-16}\ \text{W/cm}^2$.

18. (a) What is the decibel level of a sound that is twice as intense as a 90.0-dB sound? (b) What is the decibel level of a sound that is one-fifth as intense as a 90.0-dB sound?

19. (a) What is the intensity of a sound that has a level 7.00 dB lower than a $4.00 \times 10^{-9}\ \text{W/m}^2$ sound? (b) What is the intensity of a sound that is 3.00 dB higher than a $4.00 \times 10^{-9}\ \text{W/m}^2$ sound?

20. (a) How much more intense is a sound that has a level 17.0 dB higher than another? (b) If one sound has a level 23.0 dB less than another, what is the ratio of their intensities?

21. People with good hearing can perceive sounds as low in level as $-8.00\ \text{dB}$ at a frequency of 3000 Hz. What is the intensity of this sound in watts per meter squared?

22. If a large housefly 3.0 m away from you makes a noise of 40.0 dB, what is the noise level of 1000 flies at that distance, assuming interference has a negligible effect?

23. Ten cars in a circle at a boom box competition produce a 120-dB sound intensity level at the center of the circle. What is the average sound intensity level produced there by each stereo, assuming interference effects can be neglected?

24. The amplitude of a sound wave is measured in terms of its maximum gauge pressure. By what factor does the amplitude of a sound wave increase if the sound intensity level goes up by 40.0 dB?

25. If a sound intensity level of 0 dB at 1000 Hz corresponds to a maximum gauge pressure (sound amplitude) of $10^{-9}\ \text{atm}$, what is the maximum gauge pressure in a 60-dB sound? What is the maximum gauge pressure in a 120-dB sound?

26. An 8-hour exposure to a sound intensity level of 90.0 dB may cause hearing damage. What energy in joules falls on a 0.800-cm-diameter eardrum so exposed?

27. (a) Ear trumpets were never very common, but they did aid people with hearing losses by gathering sound over a large area and concentrating it on the smaller area of the eardrum. What decibel increase does an ear trumpet produce if its sound gathering area is $900\ \text{cm}^2$ and the area of the eardrum is $0.500\ \text{cm}^2$, but the trumpet only has an efficiency of 5.00% in transmitting the sound to the eardrum? (b) Comment on the usefulness of the decibel increase found in part (a).

28. Sound is more effectively transmitted into a stethoscope by direct contact than through the air, and it is further intensified by being concentrated on the smaller area of the eardrum. It is reasonable to assume that sound is transmitted into a stethoscope 100 times as effectively compared with transmission though the air. What, then, is the gain in decibels produced by a stethoscope that has a sound gathering area of $15.0\ \text{cm}^2$, and concentrates the sound onto two eardrums with a total area of $0.900\ \text{cm}^2$ with an efficiency of 40.0%?

29. Loudspeakers can produce intense sounds with surprisingly small energy input in spite of their low efficiencies. Calculate the power input needed to produce a 90.0-dB sound intensity level for a 12.0-cm-diameter speaker that has an efficiency of 1.00%. (This value is the sound intensity level right at the speaker.)

17.4 Doppler Effect and Sonic Booms

30. (a) What frequency is received by a person watching an oncoming ambulance moving at 110 km/h and emitting a steady 800-Hz sound from its siren? The speed of sound on this day is 345 m/s. (b) What frequency does she receive after the ambulance has passed?

31. (a) At an air show a jet flies directly toward the stands at a speed of 1200 km/h, emitting a frequency of 3500 Hz, on a day when the speed of sound is 342 m/s. What frequency is received by the observers? (b) What frequency do they receive as the plane flies directly away from them?

32. What frequency is received by a mouse just before being dispatched by a hawk flying at it at 25.0 m/s and emitting a screech of frequency 3500 Hz? Take the speed of sound to be 331 m/s.

33. A spectator at a parade receives an 888-Hz tone from an oncoming trumpeter who is playing an 880-Hz note. At what speed is the musician approaching if the speed of sound is 338 m/s?

34. A commuter train blows its 200-Hz horn as it approaches a crossing. The speed of sound is 335 m/s. (a) An observer waiting at the crossing receives a frequency of 208 Hz. What is the speed of the train? (b) What frequency does the observer receive as the train moves away?

35. Can you perceive the shift in frequency produced when you pull a tuning fork toward you at 10.0 m/s on a day when the speed of sound is 344 m/s? To answer this question, calculate the factor by which the frequency shifts and see if it is greater than 0.300%.

36. Two eagles fly directly toward one another, the first at 15.0 m/s and the second at 20.0 m/s. Both screech, the first one emitting a frequency of 3200 Hz and the second one emitting a frequency of 3800 Hz. What frequencies do they receive if the speed of sound is 330 m/s?

37. What is the minimum speed at which a source must travel toward you for you to be able to hear that its frequency is Doppler shifted? That is, what speed produces a shift of 0.300% on a day when the speed of sound is 331 m/s?

17.5 Sound Interference and Resonance: Standing Waves in Air Columns

38. A “showy” custom-built car has two brass horns that are supposed to produce the same frequency but actually emit 263.8 and 264.5 Hz. What beat frequency is produced?

39. What beat frequencies will be present: (a) If the musical notes A and C are played together (frequencies of 220 and 264 Hz)? (b) If D and F are played together (frequencies of 297 and 352 Hz)? (c) If all four are played together?

40. What beat frequencies result if a piano hammer hits three strings that emit frequencies of 127.8, 128.1, and 128.3 Hz?

41. A piano tuner hears a beat every 2.00 s when listening to a 264.0-Hz tuning fork and a single piano string. What are the two possible frequencies of the string?

42. (a) What is the fundamental frequency of a 0.672-m-long tube, open at both ends, on a day when the speed of sound is 344 m/s? (b) What is the frequency of its second harmonic?

43. If a wind instrument, such as a tuba, has a fundamental frequency of 32.0 Hz, what are its first three overtones? It is closed at one end. (The overtones of a real tuba are more complex than this example, because it is a tapered tube.)

44. What are the first three overtones of a bassoon that has a fundamental frequency of 90.0 Hz? It is open at both ends. (The overtones of a real bassoon are more complex than this example, because its double reed makes it act more like a tube closed at one end.)

45. How long must a flute be in order to have a fundamental frequency of 262 Hz (this frequency corresponds to middle C on the evenly tempered chromatic scale) on a day when air temperature is 20.0°C? It is open at both ends.

46. What length should an oboe have to produce a fundamental frequency of 110 Hz on a day when the speed of sound is 343 m/s? It is open at both ends.

47. What is the length of a tube that has a fundamental frequency of 176 Hz and a first overtone of 352 Hz if the speed of sound is 343 m/s?

48. (a) Find the length of an organ pipe closed at one end that produces a fundamental frequency of 256 Hz when air temperature is 18.0°C. (b) What is its fundamental frequency at 25.0°C?

49. By what fraction will the frequencies produced by a wind instrument change when air temperature goes from 10.0°C to 30.0°C? That is, find the ratio of the frequencies at those temperatures.

50. The ear canal resonates like a tube closed at one end. (See Figure 17.37.) If ear canals range in length from 1.80 to 2.60 cm in an average population, what is the range of fundamental resonant frequencies? Take air temperature to be 37.0°C, which is the same as body temperature. How does this result correlate with the intensity versus frequency graph (Figure 17.35 of the human ear?

51. Calculate the first overtone in an ear canal, which resonates like a 2.40-cm-long tube closed at one end, by taking air temperature to be 37.0°C. Is the ear particularly sensitive to such a frequency? (The resonances of the ear canal are complicated by its nonuniform shape, which we shall ignore.)

52. A crude approximation of voice production is to consider the breathing passages and mouth to be a resonating tube closed at one end. (See Figure 17.29.) (a) What is the fundamental frequency if the tube is 0.240-m long, by taking air temperature to be 37.0°C? (b) What would this frequency become if the person replaced the air with helium? Assume the same temperature dependence for helium as for air.
53. (a) Students in a physics lab are asked to find the length of an air column in a tube closed at one end that has a fundamental frequency of 256 Hz. They hold the tube vertically and fill it with water to the top, then lower the water while a 256-Hz tuning fork is rung and listen for the first resonance. What is the air temperature if the resonance occurs for a length of 0.336 m? (b) At what length will they observe the second resonance (first overtone)?
54. What frequencies will a 1.80-m-long tube produce in the audible range at 20.0°C if: (a) The tube is closed at one end? (b) It is open at both ends?

17.6 Hearing

55. The factor of 10^{-12} in the range of intensities to which the ear can respond, from threshold to that causing damage after brief exposure, is truly remarkable. If you could measure distances over the same range with a single instrument and the smallest distance you could measure was 1 mm, what would the largest be?
56. The frequencies to which the ear responds vary by a factor of 10^{3}. Suppose the speedometer on your car measured speeds differing by the same factor of 10^{3}, and the greatest speed it reads is 90.0 mi/h. What would be the slowest nonzero speed it could read?
57. What are the closest frequencies to 500 Hz that an average person can clearly distinguish as being different in frequency from 500 Hz? The sounds are not present simultaneously.
58. Can the average person tell that a 2002-Hz sound has a different frequency than a 1999-Hz sound without playing them simultaneously?
59. If your radio is producing an average sound intensity level of 85 dB, what is the next lowest sound intensity level that is clearly less intense?
60. Can you tell that your roommate turned up the sound on the TV if its average sound intensity level goes from 70 to 73 dB?
61. Based on the graph in Figure 17.34, what is the threshold of hearing in decibels for frequencies of 60, 400, 1000, 4000, and 15,000 Hz? Note that many AC electrical appliances produce 60 Hz, music is commonly 400 Hz, a reference frequency is 1000 Hz, your maximum sensitivity is near 4000 Hz, and many older TVs produce a 15,750 Hz whine.
62. What sound intensity levels must sounds of frequencies 60, 3000, and 8000 Hz have in order to have the same loudness as a 40-dB sound of frequency 1000 Hz (that is, to have a loudness of 40 phons)?
63. What is the approximate sound intensity level in decibels of a 600-Hz tone if it has a loudness of 20 phons? If it has a loudness of 70 phons?
64. (a) What are the loudnesses in phons of sounds having frequencies of 200, 1000, 5000, and 10,000 Hz, if they are all at the same 60.0-dB sound intensity level? (b) If they are all at 110 dB? (c) If they are all at 20.0 dB?
65. Suppose a person has a 50-dB hearing loss at all frequencies. By how many factors of 10 will low-intensity sounds need to be amplified to seem normal to this person? Note that smaller amplification is appropriate for more intense sounds to avoid further hearing damage.
66. If a woman needs an amplification of 5.0×10^{12} times the threshold intensity to enable her to hear at all frequencies, what is her overall hearing loss in dB? Note that smaller amplification is appropriate for more intense sounds to avoid further damage to her hearing from levels above 90 dB.
67. (a) What is the intensity in watts per meter squared of a just barely audible 200-Hz sound? (b) What is the intensity in watts per meter squared of a barely audible 4000-Hz sound?
68. (a) Find the intensity in watts per meter squared of a 60.0-Hz sound having a loudness of 60 phons. (b) Find the intensity in watts per meter squared of a 10,000-Hz sound having a loudness of 60 phons.
69. A person has a hearing threshold 10 dB above normal at 100 Hz and 50 dB above normal at 4000 Hz. How much more intense must a 100-Hz tone be than a 4000-Hz tone if they are both barely audible to this person?
70. A child has a hearing loss of 60 dB near 5000 Hz, due to noise exposure, and normal hearing elsewhere. How much more intense is a 5000-Hz tone than a 400-Hz tone if they are both barely audible to the child?

71. What is the ratio of intensities of two sounds of identical frequency if the first is just barely discernible as louder to a person than the second?

17.7 Ultrasound

Unless otherwise indicated, for problems in this section, assume that the speed of sound through human tissues is 1540 m/s.

72. What is the sound intensity level in decibels of ultrasound of intensity $10^5\ \mathrm{W/m^2}$, used to pulverize tissue during surgery?

73. Is 155-dB ultrasound in the range of intensities used for deep heating? Calculate the intensity of this ultrasound and compare this intensity with values quoted in the text.

74. Find the sound intensity level in decibels of $2.00 \times 10^{-2}\ \mathrm{W/m^2}$ ultrasound used in medical diagnostics.

75. The time delay between transmission and the arrival of the reflected wave of a signal using ultrasound traveling through a piece of fat tissue was 0.13 ms. At what depth did this reflection occur?

76. In the clinical use of ultrasound, transducers are always coupled to the skin by a thin layer of gel or oil, replacing the air that would otherwise exist between the transducer and the skin. (a) Using the values of acoustic impedance given in Table 17.5 calculate the intensity reflection coefficient between transducer material and air. (b) Calculate the intensity reflection coefficient between transducer material and gel (assuming for this problem that its acoustic impedance is identical to that of water). (c) Based on the results of your calculations, explain why the gel is used.

77. (a) Calculate the minimum frequency of ultrasound that will allow you to see details as small as 0.250 mm in human tissue. (b) What is the effective depth to which this sound is effective as a diagnostic probe?

78. (a) Find the size of the smallest detail observable in human tissue with 20.0-MHz ultrasound. (b) Is its effective penetration depth great enough to examine the entire eye (about 3.00 cm is needed)? (c) What is the wavelength of such ultrasound in $0^\circ\mathrm{C}$ air?

79. (a) Echo times are measured by diagnostic ultrasound scanners to determine distances to reflecting surfaces in a patient. What is the difference in echo times for tissues that are 3.50 and 3.60 cm beneath the surface? (This difference is the minimum resolving time for the scanner to see details as small as 0.100 cm, or 1.00 mm. Discrimination of smaller time differences is needed to see smaller details.) (b) Discuss whether the period T of this ultrasound must be smaller than the minimum time resolution. If so, what is the minimum frequency of the ultrasound and is that out of the normal range for diagnostic ultrasound?

80. (a) How far apart are two layers of tissue that produce echoes having round-trip times (used to measure distances) that differ by $0.750\ \mu\mathrm{s}$? (b) What minimum frequency must the ultrasound have to see detail this small?

81. (a) A bat uses ultrasound to find its way among trees. If this bat can detect echoes 1.00 ms apart, what minimum distance between objects can it detect? (b) Could this distance explain the difficulty that bats have finding an open door when they accidentally get into a house?

82. A dolphin is able to tell in the dark that the ultrasound echoes received from two sharks come from two different objects only if the sharks are separated by 3.50 m, one being that much farther away than the other. (a) If the ultrasound has a frequency of 100 kHz, show this ability is not limited by its wavelength. (b) If this ability is due to the dolphin's ability to detect the arrival times of echoes, what is the minimum time difference the dolphin can perceive?

83. A diagnostic ultrasound echo is reflected from moving blood and returns with a frequency 500 Hz higher than its original 2.00 MHz. What is the velocity of the blood? (Assume that the frequency of 2.00 MHz is accurate to seven significant figures and 500 Hz is accurate to three significant figures.)

84. Ultrasound reflected from an oncoming bloodstream that is moving at 30.0 cm/s is mixed with the original frequency of 2.50 MHz to produce beats. What is the beat frequency? (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

CHAPTER 18

Electric Charge and Electric Field

FIGURE 18.1 Static electricity from this plastic slide causes the child's hair to stand on end. The sliding motion stripped electrons away from the child's body, leaving an excess of positive charges, which repel each other along each strand of hair. (credit: Ken Bosma/Wikimedia Commons)

CHAPTER OUTLINE

INTRODUCTION TO ELECTRIC CHARGE AND ELECTRIC FIELD The image of American politician and scientist Benjamin Franklin (1706–1790) flying a kite in a thunderstorm is familiar to every schoolchild. (See Figure 18.2.) In this experiment, Franklin demonstrated a connection between lightning and **static electricity**. Sparks were drawn from a key hung on a kite string during an electrical storm. These sparks were like those produced by static electricity, such as the spark that jumps from your finger to a metal doorknob after you walk across a wool carpet. What Franklin demonstrated in his dangerous experiment was a connection between phenomena on two different scales: one the grand power of an electrical storm, the other an effect of more human proportions. Connections like this one reveal the underlying unity of the laws of nature, an aspect we humans find particularly appealing.

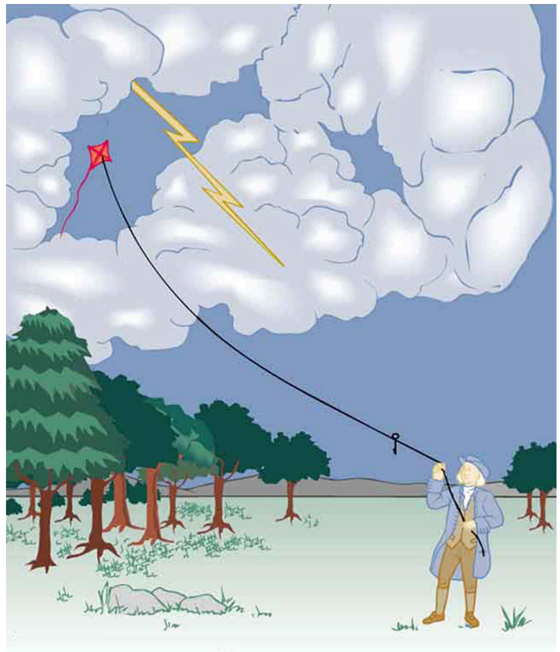


FIGURE 18.2 When Benjamin Franklin demonstrated that lightning was related to static electricity, he made a connection that is now part of the evidence that all directly experienced forces except the gravitational force are manifestations of the electromagnetic force.

Much has been written about Franklin. His experiments were only part of the life of a man who was a scientist, inventor, revolutionary, statesman, and writer. Franklin's experiments were not performed in isolation, nor were they the only ones to reveal connections.

For example, the Italian scientist Luigi Galvani (1737–1798) performed a series of experiments in which static electricity was used to stimulate contractions of leg muscles of dead frogs, an effect already known in humans subjected to static discharges. But Galvani also found that if he joined two metal wires (say copper and zinc) end to end and touched the other ends to muscles, he produced the same effect in frogs as static discharge. Alessandro Volta (1745–1827), partly inspired by Galvani's work, experimented with various combinations of metals and developed the battery.

During the same era, other scientists made progress in discovering fundamental connections. The periodic table was developed as the systematic properties of the elements were discovered. This influenced the development and refinement of the concept of atoms as the basis of matter. Such submicroscopic descriptions of matter also help explain a great deal more.

Atomic and molecular interactions, such as the forces of friction, cohesion, and adhesion, are now known to be manifestations of the **electromagnetic force**. Static electricity is just one aspect of the electromagnetic force, which also includes moving electricity and magnetism.

All the macroscopic forces that we experience directly, such as the sensations of touch and the tension in a rope, are due to the electromagnetic force, one of the four fundamental forces in nature. The gravitational force, another fundamental force, is actually sensed through the electromagnetic interaction of molecules, such as between those in our feet and those on the top of a bathroom scale. (The other two fundamental forces, the strong nuclear force and the weak nuclear force, cannot be sensed on the human scale.)

This chapter begins the study of electromagnetic phenomena at a fundamental level. The next several chapters will cover static electricity, moving electricity, and magnetism—collectively known as electromagnetism. In this chapter, we begin with the study of electric phenomena due to charges that are at least temporarily stationary, called electrostatics, or static electricity.

[Click to view content (https://openstax.org/books/college-physics-2e/pages/18-introduction-to-electric-charge-and-electric-field)](https://openstax.org/books/college-physics-2e/pages/18-introduction-to-electric-charge-and-electric-field)

18.1 Static Electricity and Charge: Conservation of Charge

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define electric charge, and describe how the two types of charge interact.
- Describe three common situations that generate static electricity.
- State the law of conservation of charge.

FIGURE 18.3 Borneo amber was mined in Sabah, Malaysia, from shale-sandstone-mudstone veins. When a piece of amber is rubbed with a piece of silk, the amber gains more electrons, giving it a net negative charge. At the same time, the silk, having lost electrons, becomes positively charged. (credit: Sebakoamber, Wikimedia Commons)

What makes plastic wrap cling? Static electricity. Not only are applications of static electricity common these days, its existence has been known since ancient times. The first record of its effects dates to ancient Greeks who noted more than 500 years B.C. that polishing amber temporarily enabled it to attract bits of straw (see Figure 18.3). The very word *electric* derives from the Greek word for amber (*electron*).

Many of the characteristics of static electricity can be explored by rubbing things together. Rubbing creates the spark you get from walking across a wool carpet, for example. Static cling generated in a clothes dryer and the attraction of straw to recently polished amber also result from rubbing. Similarly, lightning results from air movements under certain weather conditions. You can also rub a balloon on your hair, and the static electricity created can then make the balloon cling to a wall. We also have to be cautious of static electricity, especially in dry climates. When we pump gasoline, we are warned to discharge ourselves (after sliding across the seat) on a metal surface before grabbing the gas nozzle. Attendants in hospital operating rooms must wear booties with a conductive strip of aluminum foil on the bottoms to avoid creating sparks which may ignite flammable anesthesia gases combined with the oxygen being used.

Some of the most basic characteristics of static electricity include:

- The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge.
- There are only two types of charge, one called positive and the other called negative.
- Like charges repel, whereas unlike charges attract.
- The force between charges decreases with distance.

How do we know there are two types of **electric charge**? When various materials are rubbed together in controlled ways, certain combinations of materials always produce one type of charge on one material and the opposite type on the other. By convention, we call one type of charge “positive”, and the other type “negative.” For example, when glass is rubbed with silk, the glass becomes positively charged and the silk negatively charged. Since the glass and silk have opposite charges, they attract one another like clothes that have rubbed together in a dryer. Two glass rods rubbed with silk in this manner will repel one another, since each rod has positive charge on it. Similarly, two silk cloths so rubbed will repel, since both cloths have negative charge. Figure 18.4 shows how these simple materials can be used to explore the nature of the force between charges.

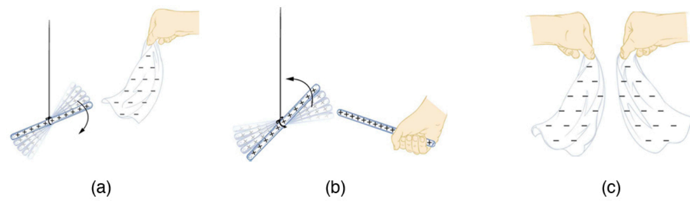


FIGURE 18.4 A glass rod becomes positively charged when rubbed with silk, while the silk becomes negatively charged. (a) The glass rod is attracted to the silk because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

More sophisticated questions arise. Where do these charges come from? Can you create or destroy charge? Is there a smallest unit of charge? Exactly how does the force depend on the amount of charge and the distance between charges? Such questions obviously occurred to Benjamin Franklin and other early researchers, and they interest us even today.

Charge Carried by Electrons and Protons

Franklin wrote in his letters and books that he could see the effects of electric charge but did not understand what caused the phenomenon. Today we have the advantage of knowing that normal matter is made of atoms, and that atoms contain positive and negative charges, usually in equal amounts.

Figure 18.5 shows a simple model of an atom with negative **electrons** orbiting its positive nucleus. The nucleus is positive due to the presence of positively charged **protons**. Nearly all charge in nature is due to electrons and protons, which are two of the three building blocks of most matter. (The third is the neutron, which is neutral, carrying no charge.) Other charge-carrying particles are observed in cosmic rays and nuclear decay, and are created in particle accelerators. All but the electron and proton survive only a short time and are quite rare by comparison.

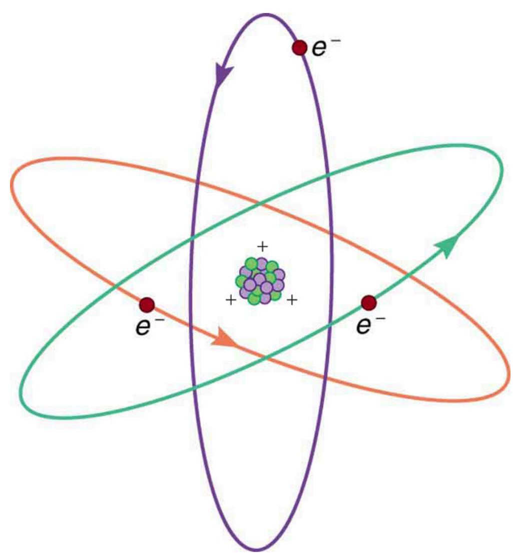


FIGURE 18.5 This simplified (and not to scale) view of an atom is called the planetary model of the atom. Negative electrons orbit a much heavier positive nucleus, as the planets orbit the much heavier sun. There the similarity ends, because forces in the atom are electromagnetic, whereas those in the planetary system are gravitational. Normal macroscopic amounts of matter contain immense numbers of atoms and molecules and, hence, even greater numbers of individual negative and positive charges.

The charges of electrons and protons are identical in magnitude but opposite in sign. Furthermore, all charged objects in nature are integral multiples of this basic quantity of charge, meaning that all charges are made of combinations of a basic unit of charge. Usually, charges are formed by combinations of electrons and protons. The magnitude of this basic charge is

$$|q_e| = 1.60 \times 10^{-19}\ \text{C}. \quad (18.1)$$

The symbol q is commonly used for charge and the subscript e indicates the charge of a single electron (or proton).

The SI unit of charge is the coulomb (C). The number of protons needed to make a charge of 1.00 C is

$$1.00 \text{ C} \times \frac{1 \text{ proton}}{1.60 \times 10^{-19} \text{ C}} = 6.25 \times 10^{18} \text{ protons}. \tag{18.2}$$

Similarly, 6.25×10^{18} electrons have a combined charge of −1.00 coulomb. Just as there is a smallest bit of an element (an atom), there is a smallest bit of charge. There is no directly observed charge smaller than $|q_e|$ (see Things Great and Small: The Submicroscopic Origin of Charge), and all observed charges are integral multiples of $|q_e|$.

Things Great and Small: The Submicroscopic Origin of Charge

With the exception of exotic, short-lived particles, all charge in nature is carried by electrons and protons. Electrons carry the charge we have named negative. Protons carry an equal-magnitude charge that we call positive. (See Figure 18.6.) Electron and proton charges are considered fundamental building blocks, since all other charges are integral multiples of those carried by electrons and protons. Electrons and protons are also two of the three fundamental building blocks of ordinary matter. The neutron is the third and has zero total charge.

Figure 18.6 shows a person touching a Van de Graaff generator and receiving excess positive charge. The expanded view of a hair shows the existence of both types of charges but an excess of positive. The repulsion of these positive like charges causes the strands of hair to repel other strands of hair and to stand up. The further blowup shows an artist's conception of an electron and a proton perhaps found in an atom in a strand of hair.

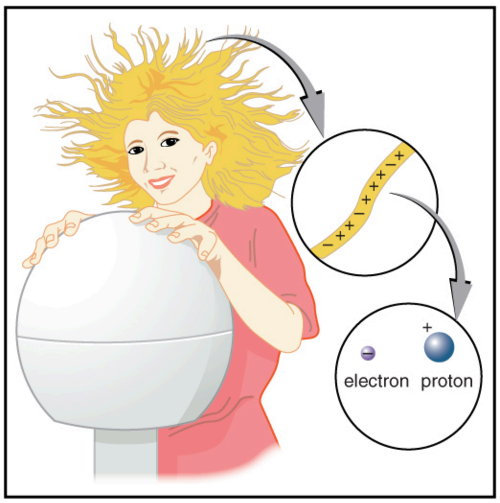


FIGURE 18.6 When this person touches a Van de Graaff generator, some electrons are attracted to the generator, resulting in an excess of positive charge, causing her hair to stand on end. The charges in one hair are shown. An artist's conception of an electron and a proton illustrate the particles carrying the negative and positive charges. We cannot really see these particles with visible light because they are so small (the electron seems to be an infinitesimal point), but we know a great deal about their measurable properties, such as the charges they carry.

The electron seems to have no substructure; in contrast, when the substructure of protons is explored by scattering extremely energetic electrons from them, it appears that there are point-like particles inside the proton. These sub-particles, named quarks, have never been directly observed, but they are believed to carry fractional charges as seen in Figure 18.7. Charges on electrons and protons and all other directly observable particles are unitary, but these quark substructures carry charges of either $-\frac{1}{3}$ or $+\frac{2}{3}$. There are continuing attempts to observe fractional charge directly and to learn of the properties of quarks, which are perhaps the ultimate substructure of matter.

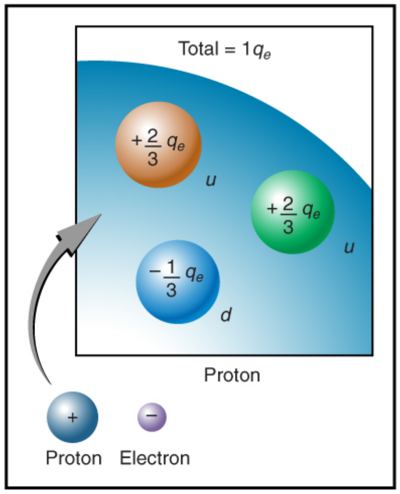


FIGURE 18.7 Artist's conception of fractional quark charges inside a proton. A group of three quark charges add up to the single positive charge on the proton: $-\frac{1}{3}q_e + \frac{2}{3}q_e + \frac{2}{3}q_e = +1q_e$.

Separation of Charge in Atoms

Charges in atoms and molecules can be separated—for example, by rubbing materials together. Some atoms and molecules have a greater affinity for electrons than others and will become negatively charged by close contact in rubbing, leaving the other material positively charged. (See Figure 18.8.) Positive charge can similarly be induced by rubbing. Methods other than rubbing can also separate charges. Batteries, for example, use combinations of substances that interact in such a way as to separate charges. Chemical interactions may transfer negative charge from one substance to the other, making one battery terminal negative and leaving the first one positive.

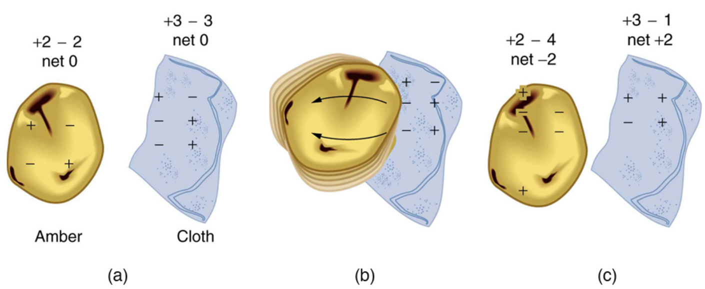


FIGURE 18.8 When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

No charge is actually created or destroyed when charges are separated as we have been discussing. Rather, existing charges are moved about. In fact, in all situations the total amount of charge is always constant. This universally obeyed law of nature is called the **law of conservation of charge**.

Law of Conservation of Charge

Total charge is constant in any process.

In more exotic situations, such as in particle accelerators, mass, Δm, can be created from energy in the amount $\Delta m = \frac{E}{c^2}$. Sometimes, the created mass is charged, such as when an electron is created. Whenever a charged particle is created, another having an opposite charge is always created along with it, so that the total charge created is zero. Usually, the two particles are "matter-antimatter" counterparts. For example, an antielectron would usually be created at the same time as an electron. The antielectron has a positive charge (it is called a positron), and so the total charge created is zero. (See Figure 18.9.) All particles have antimatter counterparts with opposite signs. When matter and antimatter counterparts are brought together, they completely annihilate one another. By annihilate, we mean that the mass of the two particles is converted to energy *E*, again obeying the relationship $\Delta m = \frac{E}{c^2}$. Since the two particles have equal and opposite charge, the total charge is zero before and after the annihilation; thus, total charge is conserved.

Making Connections: Conservation Laws

Only a limited number of physical quantities are universally conserved. Charge is one—energy, momentum, and angular momentum are others. Because they are conserved, these physical quantities are used to explain more phenomena and form more connections than other, less basic quantities. We find that conserved quantities give us great insight into the rules followed by nature and hints to the organization of nature. Discoveries of conservation laws have led to further discoveries, such as the weak nuclear force and the quark substructure of protons and other particles.

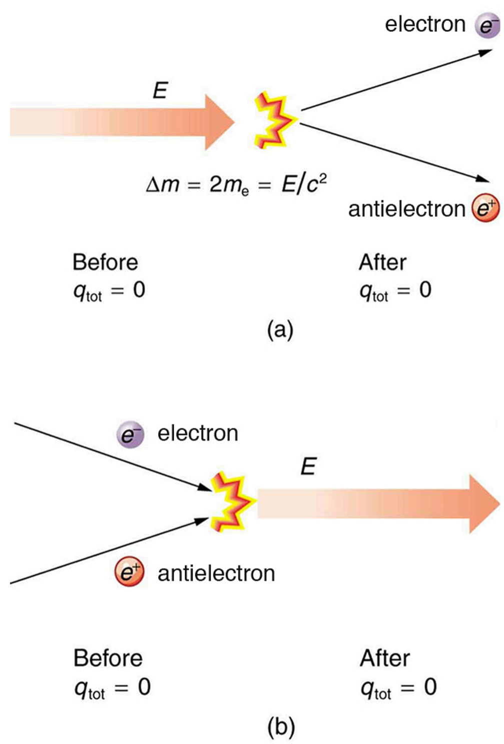


FIGURE 18.9 (a) When enough energy is present, it can be converted into matter. Here the matter created is an electron–antielectron pair. (m_e is the electron's mass.) The total charge before and after this event is zero. (b) When matter and antimatter collide, they annihilate each other; the total charge is conserved at zero before and after the annihilation.

The law of conservation of charge is absolute—it has never been observed to be violated. Charge, then, is a special physical quantity, joining a very short list of other quantities in nature that are always conserved. Other conserved quantities include energy, momentum, and angular momentum.

PHET EXPLORATIONS

Balloons and Static Electricity

Why does a balloon stick to your sweater? Rub a balloon on a sweater, then let go of the balloon and it flies over and sticks to the sweater. View the charges in the sweater, balloons, and the wall.

[Click to view content (https://openstax.org/books/college-physics-2e/pages/18-1-static-electricity-and-charge-conservation-of-charge)](https://openstax.org/books/college-physics-2e/pages/18-1-static-electricity-and-charge-conservation-of-charge)

18.2 Conductors and Insulators

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define conductor and insulator, explain the difference, and give examples of each.
- Describe three methods for charging an object.
- Explain what happens to an electric force as you move farther from the source.
- Define polarization.

FIGURE 18.10 This power adapter uses metal wires and connectors to conduct electricity from the wall socket to a laptop computer. The conducting wires allow electrons to move freely through the cables, which are shielded by rubber and plastic. These materials act as insulators that don't allow electric charge to escape outward. (credit: Evan-Amos, Wikimedia Commons)

Some substances, such as metals and salty water, allow charges to move through them with relative ease. Some of the electrons in metals and similar conductors are not bound to individual atoms or sites in the material. These **free electrons** can move through the material much as air moves through loose sand. Any substance that has free electrons and allows charge to move relatively freely through it is called a **conductor**. The moving electrons may collide with fixed atoms and molecules, losing some energy, but they can move in a conductor. Superconductors allow the movement of charge without any loss of energy. Salty water and other similar conducting materials contain free ions that can move through them. An ion is an atom or molecule having a positive or negative (nonzero) total charge. In other words, the total number of electrons is not equal to the total number of protons.

Other substances, such as glass, do not allow charges to move through them. These are called **insulators**. Electrons and ions in insulators are bound in the structure and cannot move easily—as much as 10^{23} times more slowly than in conductors. Pure water and dry table salt are insulators, for example, whereas molten salt and salty water are conductors.

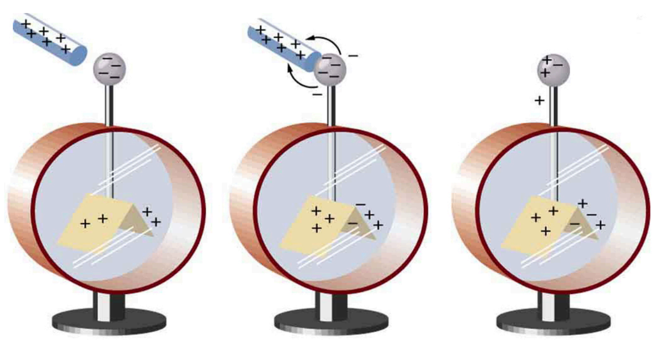


FIGURE 18.11 An electroscope is a favorite instrument in physics demonstrations and student laboratories. It is typically made with gold foil leaves hung from a (conducting) metal stem and is insulated from the room air in a glass-walled container. (a) A positively charged glass rod is brought near the tip of the electroscope, attracting electrons to the top and leaving a net positive charge on the leaves. Like charges in the light flexible gold leaves repel, separating them. (b) When the rod is touched against the ball, electrons are attracted and transferred, reducing the net charge on the glass rod but leaving the electroscope positively charged. (c) The excess charges are evenly distributed in the stem and leaves of the electroscope once the glass rod is removed.

Charging by Contact

Figure 18.11 shows an electroscope being charged by touching it with a positively charged glass rod. Because the glass rod is an insulator, it must actually touch the electroscope to transfer charge to or from it. (Note that the extra positive charges reside on the surface of the glass rod as a result of rubbing it with silk before starting the experiment.) Since only electrons move in metals, we see that they are attracted to the top of the electroscope. There, some are transferred to the positive rod by touch, leaving the electroscope with a net positive charge.

Electrostatic repulsion in the leaves of the charged electroscope separates them. The electrostatic force has a horizontal component that results in the leaves moving apart as well as a vertical component that is balanced by the gravitational force. Similarly, the electroscope can be negatively charged by contact with a negatively charged object.

Charging by Induction

It is not necessary to transfer excess charge directly to an object in order to charge it. Figure 18.12 shows a method of **induction** wherein a charge is created in a nearby object, without direct contact. Here we see two neutral metal spheres in contact with one another but insulated from the rest of the world. A positively charged rod is brought near one of them, attracting negative charge to that side, leaving the other sphere positively charged.

This is an example of induced **polarization** of neutral objects. Polarization is the separation of charges in an object that remains neutral. If the spheres are now separated (before the rod is pulled away), each sphere will have a net charge. Note that the object closest to the charged rod receives an opposite charge when charged by induction. Note also that no charge is removed from the charged rod, so that this process can be repeated without depleting the supply of excess charge.

Another method of charging by induction is shown in Figure 18.13. The neutral metal sphere is polarized when a charged rod is brought near it. The sphere is then grounded, meaning that a conducting wire is run from the sphere to the ground. Since the earth is large and most ground is a good conductor, it can supply or accept excess charge easily. In this case, electrons are attracted to the sphere through a wire called the ground wire, because it supplies a conducting path to the ground. The ground connection is broken before the charged rod is removed, leaving the sphere with an excess charge opposite to that of the rod. Again, an opposite charge is achieved when charging by induction and the charged rod loses none of its excess charge.

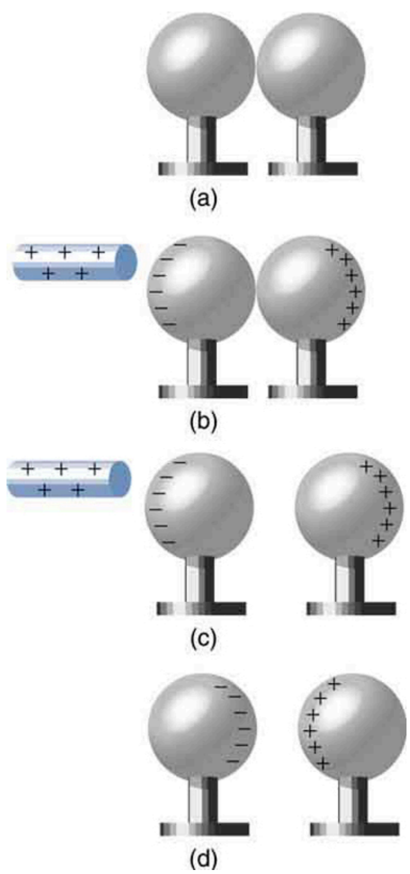


FIGURE 18.12 Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other

sphere positively charged. (c) The spheres are separated before the rod is removed, thus separating negative and positive charge. (d) The spheres retain net charges after the inducing rod is removed—without ever having been touched by a charged object.

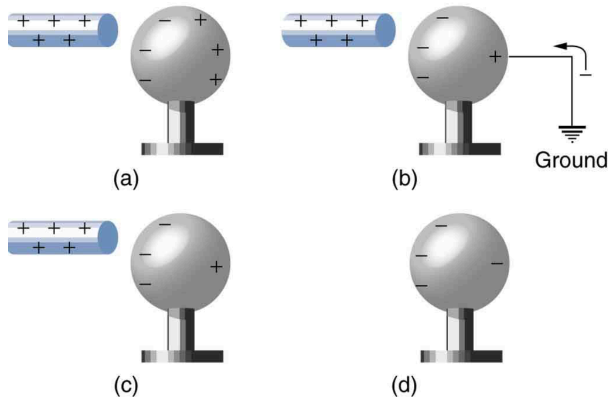


FIGURE 18.13 Charging by induction, using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from the earth's ample supply. (c) The ground connection is broken. (d) The positive rod is removed, leaving the sphere with an induced negative charge.

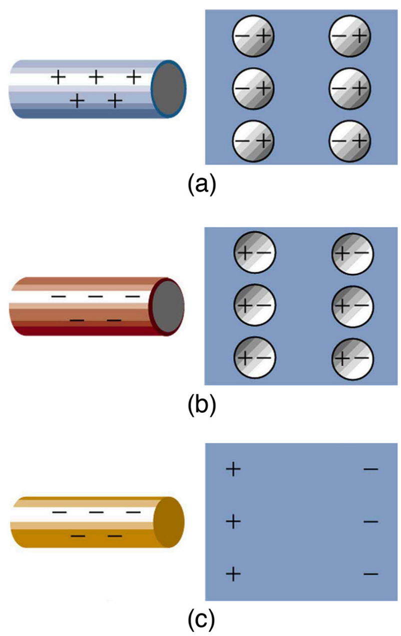


FIGURE 18.14 Both positive and negative objects attract a neutral object by polarizing its molecules. (a) A positive object brought near a neutral insulator polarizes its molecules. There is a slight shift in the distribution of the electrons orbiting the molecule, with unlike charges being brought nearer and like charges moved away. Since the electrostatic force decreases with distance, there is a net attraction. (b) A negative object produces the opposite polarization, but again attracts the neutral object. (c) The same effect occurs for a conductor; since the unlike charges are closer, there is a net attraction.

Neutral objects can be attracted to any charged object. The pieces of straw attracted to polished amber are neutral, for example. If you run a plastic comb through your hair, the charged comb can pick up neutral pieces of paper. Figure 18.14 shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

When a charged rod is brought near a neutral substance, an insulator in this case, the distribution of charge in atoms and molecules is shifted slightly. Opposite charge is attracted nearer the external charged rod, while like charge is repelled. Since the electrostatic force decreases with distance, the repulsion of like charges is weaker than the attraction of unlike charges, and so there is a net attraction. Thus a positively charged glass rod attracts neutral pieces of paper, as will a negatively charged rubber rod. Some molecules, like water, are polar molecules. Polar molecules have a natural or inherent separation of charge, although they are neutral overall. Polar molecules are particularly affected by other charged objects and show greater polarization effects than molecules with naturally uniform charge distributions.