Survey of Physics

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

This text was compiled on 04/15/2025

TABLE OF CONTENTS

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

CHAPTER OVERVIEW

1: Mass and Inertia

It took just a moment for that head to fall, but a hundred years might not produce another like it. -- *Joseph-Louis Lagrange, referring to the execution of Lavoisier on May 8, 1794*

Contributors and Attributions

1.1: Mass

Change is impossible, claimed the ancient Greek philosopher Parmenides. His work was nonscientific, since he didn't state his ideas in a form that would allow them to be tested experimentally, but modern science nevertheless has a strong Parmenidean flavor. His main argument that change is an illusion was that something can't be turned into nothing, and likewise if you have nothing, you can't turn it into something. To make this into a scientific theory, we have to decide on a way to measure what “something” is, and we can then check by measurements whether the total amount of “something” in the universe really stays constant. How much “something” is there in a rock? Does a sunbeam count as “something?” Does heat count? Motion? Thoughts and feelings?

In physics, a conservation law is a statement that the total amount of a certain physical quantity always stays the same. This chapter deals with the conservation of mass. The metric system is designed around a unit of distance, the meter, a unit of mass, the kilogram, and a time unit, the second. Numerical measurement of distance and time probably date back almost as far into prehistory as counting money, but mass is a more modern concept. Until scientists figured out that mass was conserved, it wasn't obvious that there could be a single, consistent way of measuring an amount of matter, hence jiggers of whiskey and cords of wood. You may wonder why conservation of mass wasn't discovered until relatively modern times, but it wasn't obvious, for example, that gases had mass, and that the apparent loss of mass when wood was burned was exactly matched by the mass of the escaping gases.

Once scientists were on the track of the conservation of mass concept, they began looking for a way to define mass in terms of a definite measuring procedure. If they tried such a procedure, and the result was that it led to nonconservation of mass, then they would throw it out and try a different procedure. For instance, we might be tempted to define mass using kitchen measuring cups, i.e., as a measure of volume. Mass would then be perfectly conserved for a process like mixing marbles with peanut butter, but there would be processes like freezing water that led to a net increase in mass, and others like soaking up water with a sponge that caused a decrease. If, with the benefit of hindsight, it seems like the measuring cup definition was just plain silly, then here's a more subtle example of a wrong definition of mass. Suppose we define it using a bathroom scale, or a more precise device such as a postal scale that works on the same principle of using gravity to compress or twist a spring. The trouble is that gravity is not equally strong all over the surface of the earth, so for instance there would be nonconservation of mass when you brought an object up to the top of a mountain, where gravity is a little weaker.

Although some of the obvious possibilities have problems, there do turn out to be at least two approaches to defining mass that lead to its being a conserved quantity, so we consider these definitions to be “right” in the pragmatic sense that what's correct is what's useful.

One definition that works is to use balances, but compensate for the local strength of gravity. This is the method that is used by scientists who actually specialize in ultraprecise measurements. A standard kilogram, in the form of a platinum-iridium cylinder, is kept in a special shrine in Paris. Copies are made that balance against the standard kilogram in Parisian gravity, and they are then transported to laboratories in other parts of the world, where they are compared with other masses in the local gravity. The quantity defined in this way is called *gravitational mass*.

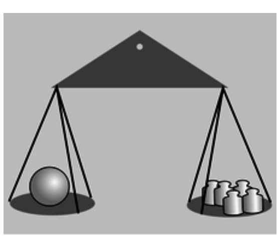


Figure b: A measurement of gravitational mass: the sphere has a gravitational mass of five kilograms.

A second and completely different approach is to measure how hard it is to change an object's state of motion. This tells us its *inertial mass*. For example, I'd be more willing to stand in the way of an oncoming poodle than in the path of a freight train, because my body will have a harder time convincing the freight train to stop. This is a dictionary-style conceptual definition, but in physics we need to back up a conceptual definition with an operational definition, which is one that spells out the operations required in order to measure the quantity being defined. We can operationalize our definition of inertial mass by throwing a standard kilogram at an object at a speed of 1 m/s (one meter per second) and measuring the recoiling object's velocity. Suppose we want to measure the mass of a particular block of cement. We put the block in a toy wagon on the sidewalk, and throw a standard kilogram at it. Suppose the standard kilogram hits the wagon, and then drops straight down to the sidewalk, having lost all its velocity, and the wagon and the block inside recoil at a velocity of 0.23 m/s. We then repeat the experiment with the block replaced by various numbers of standard kilograms, and find that we can reproduce the recoil velocity of 0.23 m/s with four standard kilograms in the wagon. We have determined the mass of the block to be four kilograms.[1] Although this definition of inertial mass has an appealing conceptual simplicity, it is obviously not very practical, at least in this crude form. Nevertheless, this method of collision is very much like the methods used for measuring the masses of subatomic particles, which, after all, can't be put on little postal scales!

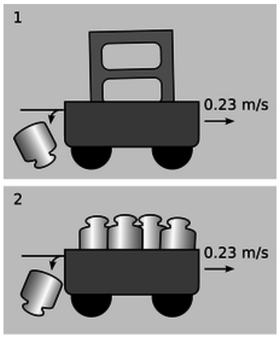


Figure c: A measurement of inertial mass: the wagon recoils with the same velocity in experiments 1 and 2, establishing that the inertial mass of the cement block is four kilograms.

Astronauts spending long periods of time in space need to monitor their loss of bone and muscle mass, and here as well, it's impossible to measure gravitational mass. Since they don't want to have standard kilograms thrown at them, they use a slightly different technique (figures d and e). They strap themselves to a chair which is attached to a large spring, and measure the time it takes for one cycle of vibration.

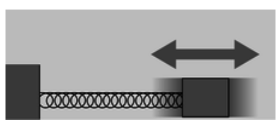
 

Figure d (left): The time for one cycle of vibration is related to the object's inertial mass.

Figure e (right): Astronaut Tamara Jernigan measures her inertial mass aboard the Space Shuttle.

1.1.1 Problem-solving techniques

How do we use a conservation law, such as conservation of mass, to solve problems? There are two basic techniques.

As an analogy, consider conservation of money, which makes it illegal for you to create dollar bills using your own inkjet printer. (Most people don't intentionally destroy their dollar bills, either!) Suppose the police notice that a particular store doesn't seem to have any customers, but the owner wears lots of gold jewelry and drives a BMW. They suspect that the store is a front for some kind of crime, perhaps counterfeiting. With intensive surveillance, there are two basic approaches they could use in their investigation. One method would be to have undercover agents try to find out how much money goes in the door, and how much money comes back out at the end of the day, perhaps by arranging through some trick to get access to the owner's briefcase in the morning and evening. If the amount of money that comes out every day is greater than the amount that went in, and if they're convinced there is no safe on the premises holding a large reservoir of money, then the owner must be counterfeiting. This inflow-equals-outflow technique is useful if we are sure that there is a region of space within which there is no supply of mass that is being built up or depleted.

Example 1: A stream of water

If you watch water flowing out of the end of a hose, you'll see that the stream of water is fatter near the mouth of the hose, and skinnier lower down. This is because the water speeds up as it falls. If the cross-sectional area of the stream was equal all along its length, then the rate of flow (kilograms per second) through a lower cross-section would be greater than the rate of flow through a cross-section higher up. Since the flow is steady, the amount of water between the two cross-sections stays constant. Conservation of mass therefore requires that the cross-sectional area of the stream shrink in inverse proportion to the increasing speed of the falling water.

f / Example 1.

self-check:

Suppose the you point the hose straight up, so that the water is rising rather than falling. What happens as the velocity gets smaller? What happens when the velocity becomes zero?

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

How can we apply a conservation law, such as conservation of mass, in a situation where mass might be stored up somewhere? To use a crime analogy again, a prison could contain a certain number of prisoners, who are not allowed to flow in or out at will. In physics, this is known as a *closed system*. A guard might notice that a certain prisoner's cell is empty, but that doesn't mean he's escaped. He could be sick in the infirmary, or hard at work in the shop earning cigarette money. What prisons actually do is to count all their prisoners every day, and make sure today's total is the same as yesterday's. One way of stating a conservation law is that for a closed system, the total amount of stuff (mass, in this chapter) stays constant.

Example 2: Lavoisier and chemical reactions in a closed system

The French chemist Antoine-Laurent Lavoisier is considered the inventor of the concept of conservation of mass. Before Lavoisier, chemists had never systematically weighed their chemicals to quantify the amount of each substance that was undergoing reactions. They also didn't completely understand that gases were just another state of matter, and hadn't tried performing reactions in sealed chambers to determine whether gases were being consumed from or released into the air. For this they had at least one practical excuse, which is that if you perform a gas-releasing reaction in a sealed chamber with no room for expansion, you get an explosion! Lavoisier invented a balance that was capable of measuring milligram masses, and figured out how to do reactions in an upside-down bowl in a basin of water, so that the gases could expand by pushing out some of the water. In a crucial experiment, Lavoisier heated a red mercury compound, which we would now describe as mercury oxide (HgO), in such a sealed chamber. A gas was produced (Lavoisier later named it “oxygen”), driving out some of the water, and the red compound was transformed into silvery liquid mercury metal. The crucial point was that the total mass of the entire apparatus was exactly the same before and after the reaction. Based on many observations of this type, Lavoisier proposed a general law of nature, that mass is always conserved. (In earlier experiments, in which closed systems were not used, chemists had become convinced that there was a mysterious substance, phlogiston, involved in combustion and oxidation reactions, and that phlogiston's mass could be positive, negative, or zero depending on the situation!)

a / Portrait of Monsieur Lavoisier and His Wife, by Jacques-Louis David, 1788. Lavoisier invented the concept of conservation of mass. The husband is depicted with his scientific apparatus, while in the background on the left is the portfolio belonging to Madame Lavoisier, who is thought to have been a student of David's.

1.1.2 Delta notation

A convenient notation used throughout physics is Δ, the uppercase Greek letter delta, which indicates “change in” or “after minus before.” For example, if b represents how much money you have in the bank, then a deposit of \$100 could be represented as $\Delta b = \$100$. That is, the change in your balance was \$100, or the balance after the transaction minus the balance before the transaction equals \$100. A withdrawal would be indicated by $\Delta b < 0$. We represent “before” and “after” using the subscripts i (initial) and f (final), e.g., $\Delta b = b_f - b_i$. Often the delta notation allows more precision than English words. For instance, “time” can be used to mean a point in time (“now's the time”), t, or it could mean a period of time (“the whole time, he had spit on his chin”), Δt.

This notation is particularly convenient for discussing conserved quantities. The law of conservation of mass can be stated simply as $\Delta m = 0$, where m is the total mass of any closed system.

self-check:

If x represents the location of an object moving in one dimension, then how would positive and negative signs of Δx be interpreted?

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

Discussion Questions

If an object had a straight-line $x - t$ graph with $\Delta x = 0$ and $\Delta t \neq 0$, what would be true about its velocity? What would this look like on a graph? What about $\Delta t = 0$ and $\Delta x \neq 0$?

Contributors and Attributions

1.2: Equivalence of Gravitational and Inertial Mass

We find experimentally that both gravitational and inertial mass are conserved to a high degree of precision for a great number of processes, including chemical reactions, melting, boiling, soaking up water with a sponge, and rotting of meat and vegetables. Now it's logically possible that both gravitational and inertial mass are conserved, but that there is no particular relationship between them, in which case we would say that they are separately conserved. On the other hand, the two conservation laws may be redundant, like having one law against murder and another law against killing people!

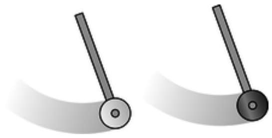


Figure a: The two pendulum bobs are constructed with equal gravitational masses. If their inertial masses are also equal, then each pendulum should take exactly the same amount of time per swing.

Here's an experiment that gets at the issue: stand up now and drop a coin and one of your shoes side by side. I used a 400-gram shoe and a 2-gram penny, and they hit the floor at the same time as far as I could tell by eye. This is an interesting result, but a physicist and an ordinary person will find it interesting for different reasons.

The layperson is surprised, since it would seem logical that heaver objects would always fall faster than light ones. However, it's fairly easy to prove that if air friction is negligible, any two objects made of the same substance must have identical motion when they fall. For instance, a 2-kg copper mass must exhibit the same falling motion as a 1-kg copper mass, because nothing would be changed by physically joining together two 1-kg copper masses to make a single 2-kg copper mass. Suppose, for example, that they are joined with a dab of glue; the glue isn't under any strain, because the two masses are doing the same thing side by side. Since the glue isn't really doing anything, it makes no difference whether the masses fall separately or side by side.[2]

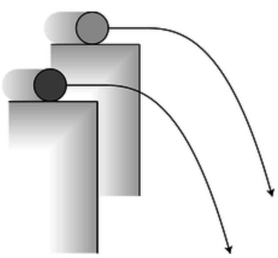
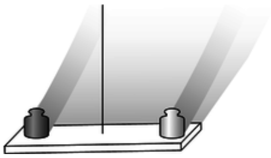


Figure b: If the cylinders have slightly unequal ratios of inertial to gravitational mass, their trajectories will be a little different.

What a physicist finds remarkable about the shoe-and-penny experiment is that it came out the way it did even though the shoe and the penny are made of *different* substances. There is absolutely no theoretical reason why this should be true. We could say that it happens because the greater gravitational mass of the shoe is exactly counteracted by its greater inertial mass, which makes it harder for gravity to get it moving, but that just leaves us wondering why inertial mass and gravitational mass are always in proportion to each other. It's possible that they are only approximately equivalent. Most of the mass of ordinary matter comes from neutrons and protons, and we could imagine, for instance, that neutrons and protons do not have exactly the same ratio of gravitational to inertial mass. This would show up as a different ratio of gravitational to inertial mass for substances containing different proportions of neutrons and protons.

Galileo did the first numerical experiments on this issue in the seventeenth century by rolling balls down inclined planes, although he didn't think about his results in these terms. A fairly easy way to improve on Galileo's accuracy is to use pendulums with bobs made of different materials. Suppose, for example, that we construct an aluminum bob and a brass bob, and use a double-pan balance to verify to good precision that their gravitational masses are equal. If we then measure the time required for each pendulum to perform a hundred cycles, we can check whether the results are the same. If their inertial masses are unequal, then the one with a smaller inertial mass will go through each cycle faster, since gravity has an easier time accelerating and decelerating it. With this type of experiment, one can easily verify that gravitational and inertial mass are proportional to each other to an accuracy of 10^{-3} or 10^{-4}.

In 1889, the Hungarian physicist Roland Eötvös used a slightly different approach to verify the equivalence of gravitational and inertial mass for various substances to an accuracy of about 10^{-8}, and the best such experiment, figure d, improved on even this phenomenal accuracy, bringing it to the 10^{-12} level.[3] In all the experiments described so far, the two objects move along similar trajectories: straight lines in the penny-and-shoe and inclined plane experiments, and circular arcs in the pendulum version. The Eötvös-style experiment looks for differences in the objects' trajectories. The concept can be understood by imagining the following simplified version. Suppose, as in figure b, we roll a brass cylinder off of a tabletop and measure where it hits the floor, and then do the same with an aluminum cylinder, making sure that both of them go over the edge with precisely the same velocity. An object with zero gravitational mass would fly off straight and hit the wall, while an object with zero inertial mass would make a sudden 90-degree turn and drop straight to the floor. If the aluminum and brass cylinders have ordinary, but slightly unequal, ratios of gravitational to inertial mass, then they will follow trajectories that are just slightly different. In other words, if inertial and gravitational mass are not exactly proportional to each other for all substances, then objects made of different substances will have different trajectories in the presence of gravity.



c / A simplified drawing of an Eötvös-style experiment. If the two masses, made out of two different substances, have slightly different ratios of inertial to gravitational mass, then the apparatus will twist slightly as the earth spins.

A simplified drawing of a practical, high-precision experiment is shown in figure c. Two objects made of different substances are balanced on the ends of a bar, which is suspended at the center from a thin fiber. The whole apparatus moves through space on a complicated, looping trajectory arising from the rotation of the earth superimposed on the earth's orbital motion around the sun. Both the earth's gravity and the sun's gravity act on the two objects. If their inertial masses are not exactly in proportion to their gravitational masses, then they will follow slightly different trajectories through space, which will result in a very slight twisting of the fiber between the daytime, when the sun's gravity is pulling upward, and the night, when the sun's gravity is downward. Figure d shows a more realistic picture of the apparatus.

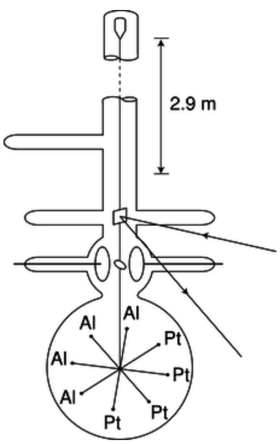


Figure d: A more realistic drawing of Braginskii and Panov's experiment. The whole thing was encased in a tall vacuum tube, which was placed in a sealed basement whose temperature was controlled to within 0.02°\textup{C}. The total mass of the platinum and aluminum test masses, plus the tungsten wire and the balance arms, was only 4.4 g. To detect tiny motions, a laser beam was bounced off of a mirror attached to the wire. There was so little friction that the balance would have taken on the order of several years to calm down completely after being put in place; to stop these vibrations, static electrical forces were applied through the two circular plates to provide very gentle twists on the ellipsoidal mass between them. After Braginskii and Panov.

This type of experiment, in which one expects a null result, is a tough way to make a career as a scientist. If your measurement comes out as expected, but with better accuracy than other people had previously achieved, your result is publishable, but won't be considered earth-shattering. On the other hand, if you build the most sensitive experiment ever, and the result comes out contrary to expectations, you're in a scary situation. You could be right, and earn a place in history, but if the result turns out to be due to a defect in your experiment, then you've made a fool of yourself.

Contributors and Attributions

1.3: 1.3 Galilean Relativity

I defined inertial mass conceptually as a measure of how hard it is to *change* an object's state of motion, the implication being that if you don't interfere, the object's motion won't change. Most people, however, believe that objects in motion have a natural tendency to slow down. Suppose I push my refrigerator to the west for a while at 0.1 m/s, and then stop pushing. The average person would say fridge just naturally stopped moving, but let's imagine how someone in China would describe the fridge experiment carried out in my house here in California. Due to the rotation of the earth, California is moving to the east at about 400 m/s. A point in China at the same latitude has the same speed, but since China is on the other side of the planet, China's east is my west. (If you're finding the three-dimensional visualization difficult, just think of China and California as two freight trains that go past each other, each traveling at 400 m/s.) If I insist on thinking of my dirt as being stationary, then China and its dirt are moving at 800 m/s to my west. From China's point of view, however, it's California that is moving 800 m/s in the opposite direction (my east). When I'm pushing the fridge to the west at 0.1 m/s, the observer in China describes its speed as 799.9 m/s. Once I stop pushing, the fridge speeds back up to 800 m/s. From my point of view, the fridge "naturally" slowed down when I stopped pushing, but according to the observer in China, it "naturally" sped up!

What's really happening here is that there's a tendency, due to friction, for the fridge to stop moving *relative to the floor*. In general, only relative motion has physical significance in physics, not absolute motion. It's not even possible to define absolute motion, since there is no special reference point in the universe that everyone can agree is at rest. Of course if we want to measure motion, we do have to pick some arbitrary reference point which we will say is standing still, and we can then define x, y, and z coordinates extending out from that point, which we can define as having $x=0$, $y=0$, $z=0$. Setting up such a system is known as choosing a *frame of reference*. The local dirt is a natural frame of reference for describing a game of basketball, but if the game was taking place on the deck of a moving ocean liner, we would probably pick a frame of reference in which the deck was at rest, and the land was moving.

1.4: A Preview of Some Modern Physics

"Mommy, why do you and Daddy have to go to work?" "To make money, sweetie-pie." "Why do we need money?" "To buy food." "Why does food cost money?" When small children ask a chain of "why" questions like this, it usually isn't too long before their parents end up saying something like, "Because that's just the way it is," or, more honestly, "I don't know the answer."

The same happens in physics. We may gradually learn to explain things more and more deeply, but there's always the possibility that a certain observed fact, such as conservation of mass, will never be understood on any deeper level. Science, after all, uses limited methods to achieve limited goals, so the ultimate reason for all existence will always be the province of religion. There is, however, an appealing explanation for conservation of mass, which is atomism, the theory that matter is made of tiny, unchanging particles. The atomic hypothesis dates back to ancient Greece, but the first solid evidence to support it didn't come until around the eighteenth century, and individual atoms were never detected until about 1900. The atomic theory implies not only conservation of mass, but a couple of other things as well.

First, it implies that the total mass of one particular element is conserved. For instance, lead and gold are both elements, and if we assume that lead atoms can't be turned into gold atoms, then the total mass of lead and the total mass of gold are separately conserved. It's as though there was not just a law against pickpocketing, but also a law against surreptitiously moving money from one of the victim's pockets to the other. It turns out, however, that although chemical reactions never change one type of atom into another, transmutation can happen in nuclear reactions, such as the ones that created most of the elements in your body out of the primordial hydrogen and helium that condensed out of the aftermath of the Big Bang.

Second, atomism implies that mass is *quantized*, meaning that only certain values of mass are possible and the ones in between can't exist. We can have three atoms of gold or four atoms of gold, but not three an a half. Although quantization of mass is a natural consequence of any theory in which matter is made up of tiny particles, it was discovered in the twentieth century that other quantities, such as energy, are quantized as well, which had previously not been suspected.

self-check:

Is money quantized?

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

If atomism is starting to make conservation of mass seem inevitable to you, then it may disturb you to know that Einstein discovered it isn't really conserved. If you put a 50-gram iron nail in some water, seal the whole thing up, and let it sit on a fantastically precise balance while the nail rusts, you'll find that the system loses about 6×10^{-12} kg of mass by the time the nail has turned completely to rust. This has to do with Einstein's famous equation $E = mc^2$. Rusting releases heat energy, which then escapes out into the room. Einstein's equation states that this amount of heat, E, is equivalent to a certain amount of mass, m. The c in the c^2 is the speed of light, which is a large number, and a large amount of energy is therefore equivalent to a very small amount of mass, so you don't notice nonconservation of mass under ordinary conditions. What is really conserved is not the mass, m, but the mass-plus-energy, $E + mc^2$. The point of this discussion is not to get you to do numerical exercises with $E = mc^2$ (at this point you don't even know what units are used to measure energy), but simply to point out to you the empirical nature of the laws of physics. If a previously accepted theory is contradicted by an experiment, then the theory needs to be changed. This is also a good example of something called the *correspondence principle*, which is a historical observation about how scientific theories change: when a new scientific theory replaces an old one, the old theory is always contained within the new one as an approximation that works within a certain restricted range of situations. Conservation of mass is an extremely good approximation for all chemical reactions, since chemical reactions never release or consume enough energy to change the total mass by a large percentage. Conservation of mass would not have been accepted for 110 years as a fundamental principle of physics if it hadn't been verified over and over again by a huge number of accurate experiments.

Contributors and Attributions

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

1.5: Footnotes

1. You might think intuitively that the recoil velocity should be exactly one fourth of a meter per second, and you'd be right except that the wagon has some mass as well. Our present approach, however, only requires that we give a way to test for equality of masses. To predict the recoil velocity from scratch, we'd need to use conservation of momentum, which is discussed in a later chapter.
2. The argument only fails for objects light enough to be affected appreciably by air friction: a bunch of feathers falls differently if you wad them up because the pattern of air flow is altered by putting them together.
3. V.B. Braginskii and V.I. Panov, Soviet Physics JETP 34, 463 (1972).
4. The principle of Galilean relativity is extended on page 190.

1.6: Problems

1. Thermometers normally use either mercury or alcohol as their working fluid. If the level of the fluid rises or falls, does this violate conservation of mass?

2. The ratios of the masses of different types of atoms were determined a century before anyone knew any actual atomic masses in units of kg. One finds, for example, that when ordinary table salt, NaCl, is melted, the chlorine atoms bubble off as a gas, leaving liquid sodium metal. Suppose the chlorine escapes, so that its mass cannot be directly determined by weighing. Experiments show that when 1.00000 kg of NaCl is treated in this way, the mass of the remaining sodium metal is 0.39337 kg. Based on this information, determine the ratio of the mass of a chlorine atom to that of a sodium atom (answer check available at lightandmatter.com)

3. An atom of the most common naturally occurring uranium isotope breaks up spontaneously into a thorium atom plus a helium atom. The masses are as follows:

uranium	3.95292849 X 10^{25} kg
thorium	3.88638748 X 10^{25} kg
helium	6.646481 X 10^{27} kg

Each of these experimentally determined masses is uncertain in its last decimal place. Is mass conserved in this process to within the accuracy of the experimental data? How would you interpret this?

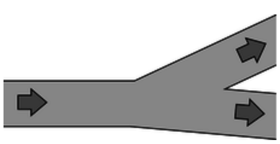
4. If two spherical water droplets of radius b combine to make a single droplet, what is its radius? (Assume that water has constant density.)

5. Make up an experiment that would test whether mass is conserved in an animal's metabolic processes.

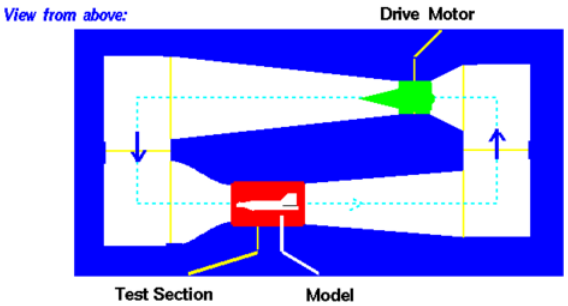
6. The figure shows a hydraulic jack. What is the relationship between the distance traveled by the plunger and the distance traveled by the object being lifted, in terms of the cross-sectional areas?

7. In an example in this chapter, I argued that a stream of water must change its cross-sectional area as it rises or falls. Suppose that the stream of water is confined to a constant-diameter pipe. Which assumption breaks down in this situation?

8. A river with a certain width and depth splits into two parts, each of which has the same width and depth as the original river. What can you say about the speed of the current after the split?



9. The diagram shows a cross-section of a wind tunnel of the kind used, for example, to test designs of airplanes. Under normal conditions of use, the density of the air remains nearly constant throughout the whole wind tunnel. How can the speed of the air be controlled and calculated? (Diagram by NASA, Glenn Research Center.)

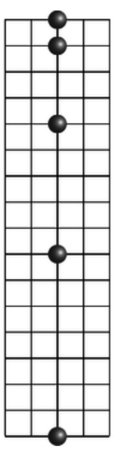


10. A water wave is in a tank that extends horizontally from $x=0$ to $x=a$, and from $z=0$ to $z=b$. We assume for simplicity that at a certain moment in time the height y of the water's surface only depends on x, not z, so that we can effectively ignore the z coordinate. Under these assumptions, the total volume of the water in the tank is

$$V=b\int_0^a y(x)\,dx \tag{1.6.1}$$

Since the density of the water is essentially constant, conservation of mass requires that V is always the same. When the water is calm, we have $y=h$, where $h=V/ab$. If two different wave patterns move into each other, we might imagine that they would add in the sense that $y_{total}-h=(y_1-h)+(y_2-h)$. Show that this type of addition is consistent with conservation of mass.

11. The figure shows the position of a falling ball at equal time intervals, depicted in a certain frame of reference. On a similar grid, show how the ball's motion would appear in a frame of reference that was moving horizontally at a speed of one box per unit time relative to the first frame.



12. (solution in the pdf version of the book) The figure shows the motion of a point on the rim of a rolling wheel. (The shape is called a cycloid.) Suppose bug A is riding on the rim of the wheel on a bicycle that is rolling, while bug B is on the spinning wheel of a bike that is sitting upside down on the floor. Bug A is moving along a cycloid, while bug B is moving in a circle. Both wheels are doing the same number of revolutions per minute. Which bug has a harder time holding on, or do they find it equally difficult?

Contributors and Attributions

CHAPTER OVERVIEW

2: Forces and Motion

2.1: Prelude to Dynamics- Newton’s Laws of Motion

Motion draws our attention. Motion itself can be beautiful, causing us to marvel at the forces needed to achieve spectacular motion, such as that of a dolphin jumping out of the water, or a pole vaulter, or the flight of a bird, or the orbit of a satellite. The study of motion is kinematics, but kinematics only *describes* the way objects move—their velocity and their acceleration. **Dynamics** considers the forces that affect the motion of moving objects and systems. Newton’s laws of motion are the foundation of dynamics. These laws provide an example of the breadth and simplicity of principles under which nature functions. They are also universal laws in that they apply to similar situations on Earth as well as in space.

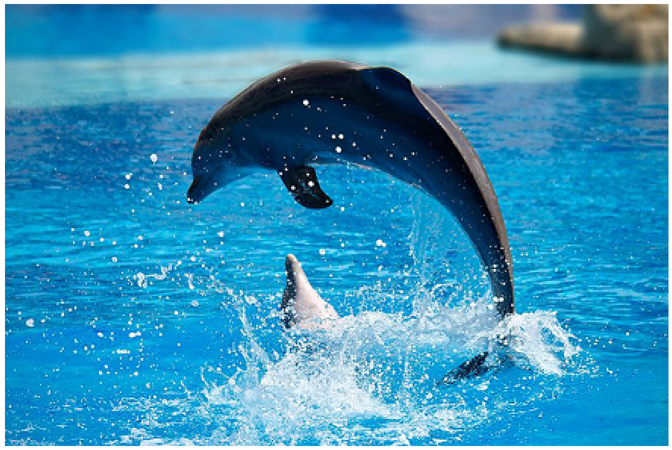


Figure 4.1.1. Newton’s laws of motion describe the motion of the dolphin’s path. (credit: Jin Jang)

Isaac Newton’s (1642–1727) laws of motion were just one part of the monumental work that has made him legendary. The development of Newton’s laws marks the transition from the Renaissance into the modern era. This transition was characterized by a revolutionary change in the way people thought about the physical universe. For many centuries natural philosophers had debated the nature of the universe based largely on certain rules of logic with great weight given to the thoughts of earlier classical philosophers such as Aristotle (384–322 BC). Among the many great thinkers who contributed to this change were Newton and Galileo.

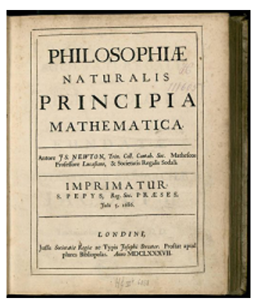


Figure 4.1.2. Issac Newton’s monumental work, Philosophiae Naturalis Principia Mathematica, was published in 1687. It proposed scientific laws that are still used today to describe the motion of objects. (credit: Service commun de la documentation de l'Université de Strasbourg)

Galileo was instrumental in establishing *observation* as the absolute determinant of truth, rather than “logical” argument. Galileo’s use of the telescope was his most notable achievement in demonstrating the importance of observation. He discovered moons orbiting Jupiter and made other observations that were inconsistent with certain ancient ideas and religious dogma. For this reason, and because of the manner in which he dealt with those in authority, Galileo was tried by the Inquisition and punished. He spent the final years of his life under a form of house arrest. Because others before Galileo had also made discoveries by *observing* the nature of the universe, and because repeated observations verified those of Galileo, his work could not be suppressed or denied. After his death, his work was verified by others, and his ideas were eventually accepted by the church and scientific communities.

Galileo also contributed to the formation of what is now called Newton’s first law of motion. Newton made use of the work of his predecessors, which enabled him to develop laws of motion, discover the law of gravity, invent calculus, and make great contributions to the theories of light and color. It is amazing that many of these developments were made with Newton working alone, without the benefit of the usual interactions that take place among scientists today.

It was not until the advent of modern physics early in the 20th century that it was discovered that Newton’s laws of motion produce a good approximation to motion only when the objects are moving at speeds much, much less than the speed of light and when those objects are larger than the size of most molecules (about m in diameter). These constraints define the realm of classical mechanics, as discussed in Introduction to the Nature of Science and Physics. At the beginning of the 20th century, Albert Einstein (1879–1955) developed the theory of relativity and, along with many other scientists, developed quantum theory. This theory does not have the constraints present in classical physics. All of the situations we consider in this chapter, and all those preceding the introduction of relativity in Special Relativity, are in the realm of classical physics.

MAKING CONNECTIONS: PAST AND PRESENT PHILOSOPHY

The importance of observation and the concept of *cause and effect* were not always so entrenched in human thinking. This realization was a part of the evolution of modern physics from natural philosophy. The achievements of Galileo, Newton, Einstein, and others were key milestones in the history of scientific thought. Most of the scientific theories that are described in this book descended from the work of these scientists.*The importance of observation* and the concept of *cause and effect* were not always so entrenched in human thinking. This realization was a part of the evolution of modern physics from natural philosophy. The achievements of Galileo, Newton, Einstein, and others were key milestones in the history of scientific thought. Most of the scientific theories that are described in this book descended from the work of these scientists.

2.2: Development of Force Concept

Learning Objectives

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

- Understand the definition of force.

Dynamics is the study of the forces that cause objects and systems to move. To understand this, we need a working definition of force. Our intuitive definition of **force**—that is, a push or a pull—is a good place to start. We know that a push or pull has both magnitude and direction (therefore, it is a vector quantity) and can vary considerably in each regard. For example, a cannon exerts a strong force on a cannonball that is launched into the air. In contrast, Earth exerts only a tiny downward pull on a flea.

A more quantitative definition of force can be based on some standard force, just as distance is measured in units relative to a standard distance. One possibility is to stretch a spring a certain fixed distance, as illustrated in Figure $2.2.2$, and use the force it exerts to pull itself back to its relaxed shape—called a *restoring force*—as a standard. The magnitude of all other forces can be stated as multiples of this standard unit of force. Many other possibilities exist for standard forces. Some alternative definitions of force will be given later in this chapter.

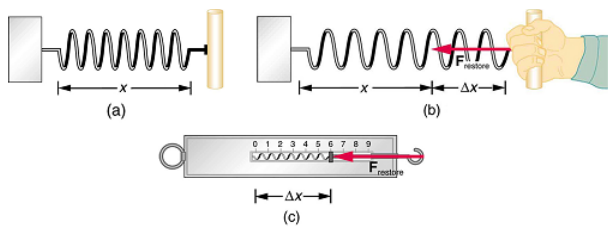


Figure $2.2.2$:The force exerted by a stretched spring can be used as a standard unit of force. (a) This spring has a length when undistorted. (b) When stretched a distance Δx the spring exerts a restoring force, $F_{restore}$ *which is reproducible. (c) A spring scale is one device that uses a spring to measure force. The force* $F_{restore}$ *is exerted on whatever is attached to the hook. Here* $F_{restore}$ *has a magnitude of 6 units in the force standard being employed.*

TAKE HOME EXPERIMENT: FORCE STANDARDS

To investigate force standards and cause and effect, get two identical rubber bands. Hang one rubber band vertically on a hook. Find a small household item that could be attached to the rubber band using a paper clip, and use this item as a weight to investigate the stretch of the rubber band. Measure the amount of stretch produced in the rubber band with one, two, and four of these (identical) items suspended from the rubber band. What is the relationship between the number of items and the amount of stretch? How large a stretch would you expect for the same number of items suspended from two rubber bands? What happens to the amount of stretch of the rubber band (with the weights attached) if the weights are also pushed to the side with a pencil?

Summary

- **Dynamics** is the study of how forces affect the motion of objects.
- **Force** is a push or pull that can be defined in terms of various standards, and it is a vector having both magnitude and direction.
- **External forces** are any outside forces that act on a body.

Glossary

dynamics

the study of how forces affect the motion of objects and systems

external force

a force acting on an object or system that originates outside of the object or system

free-body diagram

a sketch showing all of the external forces acting on an object or system; the system is represented by a dot, and the forces are represented by vectors extending outward from the dot

force

a push or pull on an object with a specific magnitude and direction; can be represented by vectors; can be expressed as a multiple of a standard force

Contributors and Attributions

2.3: Newton's First Law of Motion - Inertia

Learning Objectives

- Define mass and inertia.
- Understand Newton's first law of motion.

Experience suggests that an object at rest will remain at rest if left alone, and that an object in motion tends to slow down and stop unless some effort is made to keep it moving. What **Newton's first law of motion** states, however, is the following:

Newton's First Law of Motion

A body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force.

Note the repeated use of the verb "remains." We can think of this law as preserving the status quo of motion.

Rather than contradicting our experience, **Newton's first law of motion** states that there must be a *cause* (which is a net external force) *for there to be any change in velocity (either a change in magnitude or direction)*. We will define *net external force* in the next section. An object sliding across a table or floor slows down due to the net force of friction acting on the object. If friction disappeared, would the object still slow down?

The idea of cause and effect is crucial in accurately describing what happens in various situations. For example, consider what happens to an object sliding along a rough horizontal surface. The object quickly grinds to a halt. If we spray the surface with talcum powder to make the surface smoother, the object slides farther. If we make the surface even smoother by rubbing lubricating oil on it, the object slides farther yet. Extrapolating to a frictionless surface, we can imagine the object sliding in a straight line indefinitely. Friction is thus the *cause* of the slowing (consistent with Newton's first law). The object would not slow down at all if friction were completely eliminated. Consider an air hockey table. When the air is turned off, the puck slides only a short distance before friction slows it to a stop. However, when the air is turned on, it creates a nearly frictionless surface, and the puck glides long distances without slowing down. Additionally, if we know enough about the friction, we can accurately predict how quickly the object will slow down. Friction is an external force.

Newton's first law is completely general and can be applied to anything from an object sliding on a table to a satellite in orbit to blood pumped from the heart. Experiments have thoroughly verified that any change in velocity (speed or direction) must be caused by an external force. The idea of *generally applicable or universal laws* is important not only here—it is a basic feature of all laws of physics. Identifying these laws is like recognizing patterns in nature from which further patterns can be discovered. The genius of Galileo, who first developed the idea for the first law, and Newton, who clarified it, was to ask the fundamental question, "What is the cause?" Thinking in terms of cause and effect is a worldview fundamentally different from the typical ancient Greek approach when questions such as "Why does a tiger have stripes?" would have been answered in Aristotelian fashion, "That is the nature of the beast." True perhaps, but not a useful insight.

Mass

The property of a body to remain at rest or to remain in motion with constant velocity is called **inertia**. Newton's first law is often called the law of inertia. As we know from experience, some objects have more inertia than others. It is obviously more difficult to change the motion of a large boulder than that of a basketball, for example. The inertia of an object is measured by its **mass**. Roughly speaking, mass is a measure of the amount of "stuff" (or matter) in something. The quantity or amount of matter in an object is determined by the numbers of atoms and molecules of various types it contains. Unlike weight, mass does not vary with location. The mass of an object is the same on Earth, in orbit, or on the surface of the Moon. In practice, it is very difficult to count and identify all of the atoms and molecules in an object, so masses are not often determined in this manner. Operationally, the masses of objects are determined by comparison with the standard kilogram.

Exercise 2.3.1

Which has more mass: a kilogram of cotton balls or a kilogram of gold?

Answer

They are equal. A kilogram of one substance is equal in mass to a kilogram of another substance. The quantities that might differ between them are volume and density.

Summary

- **Newton's first law of motion** states that a body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force. This is also known as the **law of inertia**.
- **Inertia** is the tendency of an object to remain at rest or remain in motion. Inertia is related to an object's mass.
- **Mass** is the quantity of matter in a substance.

Glossary

inertia

the tendency of an object to remain at rest or remain in motion

law of inertia

see Newton's first law of motion

mass

the quantity of matter in a substance; measured in kilograms

Newton's first law of motion

a body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force; also known as the law of inertia

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

2.4: Newton's Second Law of Motion- Concept of a System

Learning Objectives

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

- Define net force, external force, and system.
- Understand Newton's second law of motion.
- Apply Newton's second law to determine the weight of an object.

Newton's second law of motion is closely related to Newton's first law of motion. It mathematically states the cause and effect relationship between force and changes in motion. Newton's second law of motion is more quantitative and is used extensively to calculate what happens in situations involving a force. Before we can write down Newton's second law as a simple equation giving the exact relationship of force, mass, and acceleration, we need to sharpen some ideas that have already been mentioned.

First, what do we mean by a change in motion? The answer is that a change in motion is equivalent to a change in velocity. A change in velocity means, by definition, that there is an **acceleration**. Newton's first law says that a net external force causes a change in motion; thus, we see that a *net external force causes acceleration*.

Another question immediately arises. What do we mean by an external force? An intuitive notion of external is correct — an **external force** acts from outside the **system** of interest. For example, in Figure $2.4.1a$ the system of interest is the wagon plus the child in it. The two forces exerted by the other children are external forces. An internal force acts between elements of the system. Again looking at Figure $2.4.1a$, the force the child in the wagon exerts to hang onto the wagon is an internal force between elements of the system of interest. Only external forces affect the motion of a system, according to Newton's first law. (The internal forces actually cancel, as we shall see in the next section.) *You must define the boundaries of the system before you can determine which forces are external*. Sometimes the system is obvious, whereas other times identifying the boundaries of a system is more subtle. The concept of a system is fundamental to many areas of physics, as is the correct application of Newton's laws. This concept will be revisited many times on our journey through physics.

Figure $2.4.1$ contains our first example of a **free-body diagram**, which is a technique used to illustrate all the external forces acting on a body. The body is represented by a single isolated point (or free body), and only those forces acting on the body from the outside (external forces) are shown. Free-body diagrams are very useful in analyzing forces acting on a system and are employed extensively in the study and application of Newton's laws of motion.

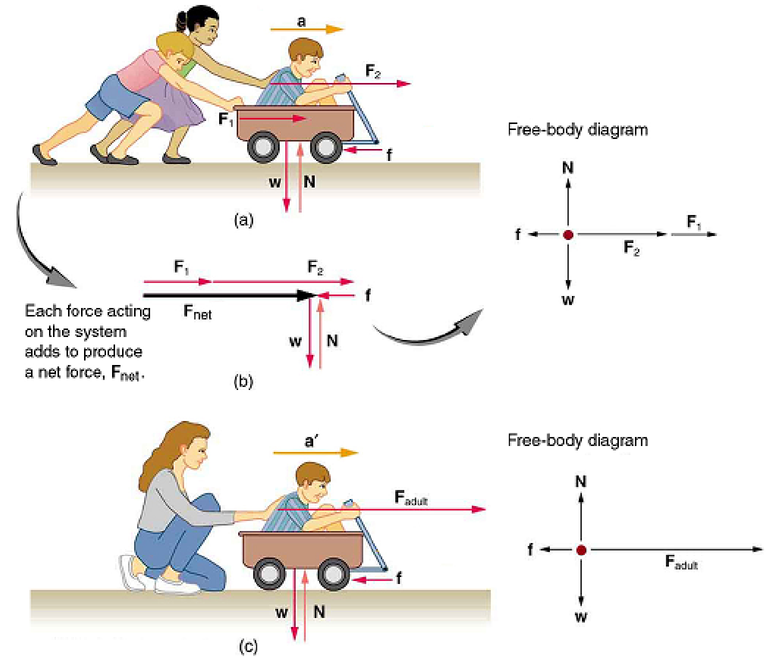


Figure 2.4.1: Different forces exerted on the same mass produce different accelerations. (a) Two children push a wagon with a child in it. Arrows representing all external forces are shown. The system of interest is the wagon and its rider. The weight w of the system and the support of the ground N are also shown for completeness and are assumed to cancel. The vector f represents the friction acting on the wagon, and it acts to the left, opposing the motion of the wagon. (b) All of the external forces acting on the system add together to produce a net force, F_{net}. The free-body diagram shows all of the forces acting on the system of interest. The dot represents the center of mass of the system. Each force vector extends from this dot. Because there are two forces acting to the right, we draw the vectors collinearly. (c) A larger net external force produces a larger acceleration $(al > a)$ when an adult pushes the child.

Now, it seems reasonable that acceleration should be directly proportional to and in the same direction as the net (total) external force acting on a system. This assumption has been verified experimentally and is illustrated in Figure. In part (a), a smaller force causes a smaller acceleration than the larger force illustrated in part (c). For completeness, the vertical forces are also shown; they are assumed to cancel since there is no acceleration in the vertical direction. The vertical forces are the weight w and the support of the ground N, and the horizontal force f represents the force of friction. These will be discussed in more detail in later sections. For now, we will define **friction** as a force that opposes the motion past each other of objects that are touching. Figure $2.4.1b$ shows how vectors representing the external forces add together to produce a net force, F_{net}.

To obtain an equation for Newton’s second law, we first write the relationship of acceleration and net external force as the proportionality

$$a \propto F_{net} \tag{2.4.1}$$

where the symbol $\propto$ means “proportional to,” and F_{net} is the **net external force**. (The net external force is the vector sum of all external forces and can be determined graphically, using the head-to-tail method, or analytically, using components. The techniques are the same as for the addition of other vectors, and are covered in the chapter section on Two-Dimensional Kinematics.) This proportionality states what we have said in words—*acceleration is directly proportional to the net external force*. Once the system of interest is chosen, it is important to identify the external forces and ignore the internal ones. It is a tremendous simplification not to have to consider the numerous internal forces acting between objects within the system, such as muscular forces within the child’s body, let alone the myriad of forces between atoms in the objects, but by doing so, we can easily solve some very complex problems with only minimal error due to our simplification

Now, it also seems reasonable that acceleration should be inversely proportional to the mass of the system. In other words, the larger the mass (the inertia), the smaller the acceleration produced by a given force. And indeed, as illustrated in Figure, the same net external force applied to a car produces a much smaller acceleration than when applied to a basketball. The proportionality is written as

$$a \propto \frac{1}{m}, \tag{2.4.2}$$

where m is the mass of the system. Experiments have shown that acceleration is exactly inversely proportional to mass, just as it is exactly linearly proportional to the net external force.

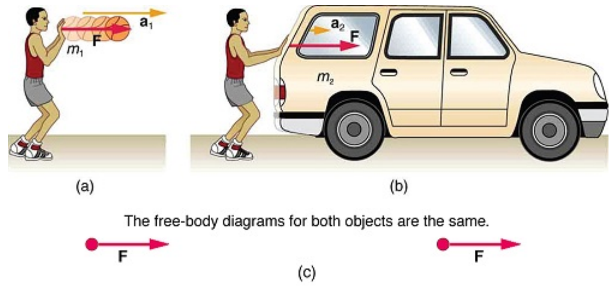


Figure 2.4.2:The same force exerted on systems of different masses produces different accelerations. (a) A basketball player pushes on a basketball to make a pass. (The effect of gravity on the ball is ignored.) (b) The same player exerts an identical force on a stalled SUV and produces a far smaller acceleration (even if friction is negligible). (c) The free-body diagrams are identical, permitting direct comparison of the two situations. A series of patterns for the free-body diagram will emerge as you do more problems.

It has been found that the acceleration of an object depends *only* on the net external force and the mass of the object. Combining the two proportionalities just given yields Newton's second law of motion.

Newton's Second Law of Motion

The acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass. In equation form, Newton's second law of motion is

$$a=\frac{F_{net}}{m} \tag{2.4.3}$$

This is often written in the more familiar form

$$F_{net}=ma. \tag{2.4.4}$$

When only the magnitude of force and acceleration are considered, this equation is simply

$$F_{net}=ma. \tag{2.4.5}$$

Although these last two equations are really the same, the first gives more insight into what Newton's second law means. The law is a *cause and effect relationship* among three quantities that is not simply based on their definitions. The validity of the second law is completely based on experimental verification.

Units of Force

$F_{net}=ma$ is used to define the units of force in terms of the three basic units for mass, length, and time. The SI unit of force is called the newton (abbreviated N) and is the force needed to accelerate a 1-kg system at the rate of $1m/s^2$ That is, since $F_{net}=ma$,

$$1N=1kg\cdot s^2 \tag{2.4.6}$$

While almost the entire world uses the newton for the unit of force, in the United States the most familiar unit of force is the pound (lb), where 1 N = 0.225 lb.

Weight and the Gravitational Force

When an object is dropped, it accelerates toward the center of Earth. Newton's second law states that a net force on an object is responsible for its acceleration. If air resistance is negligible, the net force on a falling object is the gravitational force, commonly called its **weight** w. Weight can be denoted as a vector w because it has a direction; down is, by definition, the direction of gravity, and hence weight is a downward force. The magnitude of weight is denoted as w. Galileo was instrumental in showing that, in the absence of air resistance, all objects fall with the same acceleration w. Using Galileo's result and Newton's second law, we can derive an equation for weight.

Consider an object with mass m falling downward toward Earth. It experiences only the downward force of gravity, which has magnitude w. Newton's second law states that the magnitude of the net external force on an object is $F_{net} = ma$.

Since the object experiences only the downward force of gravity, $F_{net} = w$. We know that the acceleration of an object due to gravity is g, or $a = g$. Substituting these into Newton's second law gives

WEIGHT

This is the equation for weight - the gravitational force on mass m:

$$w = mg \tag{2.4.7}$$

Since weight $g = 9.80 m/s^2$ on Earth, the weight of a 1.0 kg object on Earth is 9.8 N, as we see:

$$w = mg = (1.0 kg)(9.8 m/s^2) = 9.8 N. \tag{2.4.8}$$

Recall that g can take a positive or negative value, depending on the positive direction in the coordinate system. Be sure to take this into consideration when solving problems with weight.

When the net external force on an object is its weight, we say that it is in **free-fall**. That is, the only force acting on the object is the force of gravity. In the real world, when objects fall downward toward Earth, they are never truly in free-fall because there is always some upward force from the air acting on the object.

The acceleration due to gravity g varies slightly over the surface of Earth, so that the weight of an object depends on location and is not an intrinsic property of the object. Weight varies dramatically if one leaves Earth's surface. On the Moon, for example, the acceleration due to gravity is only $1.67 m/s^2$. A 1.0-kg mass thus has a weight of 9.8 N on Earth and only about 1.7 N on the Moon.

The broadest definition of weight in this sense is that *the weight of an object is the gravitational force on it from the nearest large body*, such as Earth, the Moon, the Sun, and so on. This is the most common and useful definition of weight in physics. It differs dramatically, however, from the definition of weight used by NASA and the popular media in relation to space travel and exploration. When they speak of "weightlessness" and "microgravity," they are really referring to the phenomenon we call "free-fall" in physics. We shall use the above definition of weight, and we will make careful distinctions between free-fall and actual weightlessness.

It is important to be aware that weight and mass are very different physical quantities, although they are closely related. Mass is the quantity of matter (how much "stuff") and does not vary in classical physics, whereas weight is the gravitational force and does vary depending on gravity. It is tempting to equate the two, since most of our examples take place on Earth, where the weight of an object only varies a little with the location of the object. Furthermore, the terms *mass* and *weight* are used interchangeably in everyday language; for example, our medical records often show our "weight" in kilograms, but never in the correct units of newtons.

COMMON MISCONCEPTIONS: MASS VS. WEIGHT

Mass and weight are often used interchangeably in everyday language. However, in science, these terms are distinctly different from one another. Mass is a measure of how much matter is in an object. The typical measure of mass is the kilogram (or the "slug" in English units). Weight, on the other hand, is a measure of the force of gravity acting on an object. Weight is equal to the mass of an object (m) multiplied by the acceleration due to gravity (g). Like any other force, weight is measured in terms of newtons (or pounds in English units).

Assuming the mass of an object is kept intact, it will remain the same, regardless of its location. However, because weight depends on the acceleration due to gravity, the weight of an object can change when the object enters into a region with stronger or weaker gravity. For example, the acceleration due to gravity on the Moon is $1.67 m/s^2$(which is much less than the acceleration due to gravity on Earth, $9.80 m/s^2$). If you measured your weight on Earth and then measured your weight on the Moon, you would find that you "weigh" much less, even though you do not look any skinnier. This is because the force of gravity is weaker on the Moon. In fact, when people say that they are "losing weight," they really mean that they are losing "mass" (which in turn causes them to weigh less)

TAKE-HOME EXPERIMENT: MASS AND WEIGHT

What do bathroom scales measure? When you stand on a bathroom scale, what happens to the scale? It depresses slightly. The scale contains springs that compress in proportion to your weight—similar to rubber bands expanding when pulled. The springs provide a measure of your weight (for an object which is not accelerating). This is a force in newtons (or pounds). In most countries, the measurement is divided by 9.80 to give a reading in mass units of kilograms. The scale measures weight but is calibrated to provide information about mass. While standing on a bathroom scale, push down on a table next to you. What happens to the reading? Why? Would your scale measure the same "mass" on Earth as on the Moon?

Example 2.4.1:What Acceleration Can a Person Produce when pushing a Lawn Mower?

Suppose that the net external force (push minus friction) exerted on a lawn mower is 51 N (about 11 lb) parallel to the ground. The mass of the mower is 24 kg. What is its acceleration?

Figure 2.4.3: The net force on a lawn mower is 51 N to the right. At what rate does the lawn mower accelerate to the right?

Strategy

Since F_{net} and m are given, the acceleration can be calculated directly from Newton's second law as stated in $F_{net} = ma$.

Solution

The magnitude of the acceleration a is $a = \frac{F_{net}}{m}$. Entering known values gives

$$a = \frac{51\ N}{24\ kg} \tag{2.4.9}$$

Substituting the units $kg \cdot m/s^2$ for N yields

$$a = \frac{51\ kg/s^2}{24\ kg} = 2.1\ m/s^2 \tag{2.4.10}$$

Discussion

The direction of the acceleration is the same direction as that of the net force, which is parallel to the ground. There is no information given in this example about the individual external forces acting on the system, but we can say something about their relative magnitudes. For example, the force exerted by the person pushing the mower must be greater than the friction opposing the motion (since we know the mower moves forward), and the vertical forces must cancel if there is to be no acceleration in the vertical direction (the mower is moving only horizontally). The acceleration found is small enough to be reasonable for a person pushing a mower. Such an effort would not last too long because the person's top speed would soon be reached.

Example 2.4.2:What Rocket Thrust Accelerates This Sled?

Prior to manned space flights, rocket sleds were used to test aircraft, missile equipment, and physiological effects on human subjects at high speeds. They consisted of a platform that was mounted on one or two rails and propelled by several rockets. Calculate the magnitude of force exerted by each rocket, called its thrust T for the four-rocket propulsion system shown in Figure. The sled's initial acceleration is $49 m/s^2$ the mass of the system is 2100 kg, and the force of friction opposing the motion is known to be 650 N.

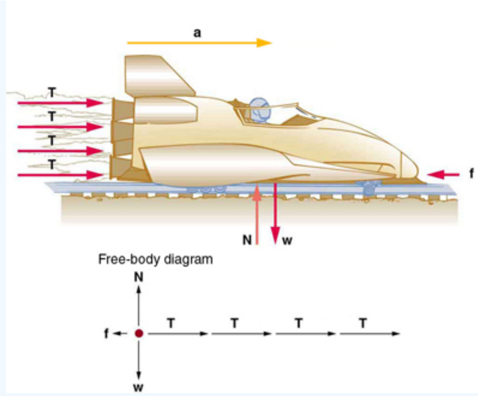


Figure 4.4.4. A sled experiences a rocket thrust that accelerates it to the right. Each rocket creates an identical thrust T. As in other situations where there is only horizontal acceleration, the vertical forces cancel. The ground exerts an upward force N on the system that is equal in magnitude and opposite in direction to its weight, w. The system here is the sled, its rockets, and rider, so none of the forces between these objects are considered. The arrow representing friction (f) is drawn larger than scale.

Strategy

Although there are forces acting vertically and horizontally, we assume the vertical forces cancel since there is no vertical acceleration. This leaves us with only horizontal forces and a simpler one-dimensional problem. Directions are indicated with plus or minus signs, with right taken as the positive direction. See the free-body diagram in the figure.

Solution

Since acceleration, mass, and the force of friction are given, we start with Newton's second law and look for ways to find the thrust of the engines. Since we have defined the direction of the force and acceleration as acting "to the right," we need to consider only the magnitudes of these quantities in the calculations. Hence we begin with

$$F_{net} = ma. \tag{2.4.11}$$

, where F_{net} is the net force along the horizontal direction. We can see from Figure that the engine thrusts add, while friction opposes the thrust. In equation form, the net external force is

$$F_{net} = 4T - f. \tag{2.4.12}$$

Substituting this into Newton's second law gives

$$F_{net} = ma = 4T - f. \tag{2.4.13}$$

Using a little algebra, we solve for the total thrust 4T:

$$4T = ma + f. \tag{2.4.14}$$

Substituting known values yields

$$4T = ma + f = (2100\, kg)(49\, m/s^2) + 650\, N \tag{2.4.15}$$

So the total thrust is

$$1 \times 10^5 N, \tag{2.4.16}$$

and the individual thrusts are

$$T = \frac{1 \times 10^5}{4} = 2.6 \times 10^4\, N \tag{2.4.17}$$

Discussion

The numbers are quite large, so the result might surprise you. Experiments such as this were performed in the early 1960s to test the limits of human endurance and the setup designed to protect human subjects in jet fighter emergency ejections. Speeds of 1000 km/h were obtained, with accelerations of 45 g-s. (Recall that g, the acceleration due gravity is $9.80\, m/s^2$. When we say that an acceleration is 45 g-s, it is $45 \times 9.80 m/s^2$, which is approximately $440 m/s^2$). While living subjects are not used any more, land speeds of 10,000 km/h have been obtained with rocket sleds. In this example, as in the preceding one, the system of interest is obvious. We will see in later examples that choosing the system of interest is crucial—and the choice is not always obvious.

Newton's second law of motion is more than a definition; it is a relationship among acceleration, force, and mass. It can help us make predictions. Each of those physical quantities can be defined independently, so the second law tells us something basic and universal about nature. The next section introduces the third and final law of motion.

Summary

- Acceleration, a, is defined as a change in velocity, meaning a change in its magnitude or direction, or both.
- An external force is one acting on a system from outside the system, as opposed to internal forces, which act between components within the system.
- Newton's second law of motion states that the acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass.
- In equation form, Newton's second law of motion is $a = \frac{F_{net}}{m}$
- This is often written in the more familiar form: $F_{net} = ma$.
- The weight w of an object is defined as the force of gravity acting on an object of mass m. The object experiences an acceleration due to gravity g:

$$w = mg. \tag{2.4.18}$$

- If the only force acting on an object is due to gravity, the object is in free fall.
- Friction is a force that opposes the motion past each other of objects that are touching.

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

2.5: Newton’s Third Law of Motion- Symmetry in Forces

Learning Objectives

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

- Understand Newton’s third law of motion.
- Apply Newton’s third law to define systems and solve problems of motion.

There is a passage in the musical *Man of la Mancha* that relates to Newton’s third law of motion. Sancho, in describing a fight with his wife to Don Quixote, says, “Of course I hit her back, Your Grace, but she’s a lot harder than me and you know what they say, ‘Whether the stone hits the pitcher or the pitcher hits the stone, it’s going to be bad for the pitcher.’” This is exactly what happens whenever one body exerts a force on another—the first also experiences a force (equal in magnitude and opposite in direction). Numerous common experiences, such as stubbing a toe or throwing a ball, confirm this. It is precisely stated in **Newton’s third law of motion.**

NEWTON’S THIRD LAW OF MOTION

Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts.

This law represents a certain *symmetry in nature*: Forces always occur in pairs, and one body cannot exert a force on another without experiencing a force itself. We sometimes refer to this law loosely as “action-reaction,” where the force exerted is the action and the force experienced as a consequence is the reaction. Newton’s third law has practical uses in analyzing the origin of forces and understanding which forces are external to a system.

We can readily see Newton’s third law at work by taking a look at how people move about. Consider a swimmer pushing off from the side of a pool, as illustrated in Figure. She pushes against the pool wall with her feet and accelerates in the direction *opposite* to that of her push. The wall has exerted an equal and opposite force back on the swimmer. You might think that two equal and opposite forces would cancel, but they do not *because they act on different systems*. In this case, there are two systems that we could investigate: the swimmer or the wall. If we select the swimmer to be the system of interest, as in the figure, then $F_{wall\ on\ feet}$ is an external force on this system and affects its motion. The swimmer moves in the direction of $F_{wall\ on\ feet}$. In contrast, the force $F_{feet\ on\ wall}$ acts on the wall and not on our system of interest. Thus $F_{feet\ on\ wall}$ does not directly affect the motion of the system and does not cancel $F_{wall\ on\ feet}$. Note that the swimmer pushes in the direction opposite to that in which she wishes to move. The reaction to her push is thus in the desired direction.

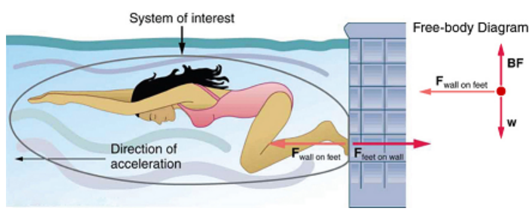


Figure 2.5.1: When the swimmer exerts a force $F_{feet\ on\ wall}$ on the wall, she accelerates in the direction opposite to that of her push. This means the net external force on her is in the direction opposite to $F_{feet\ on\ wall}$ This opposition occurs because, in accordance with Newton’s third law of motion, the wall exerts a force $F_{wall\ on\ feet}$ on her, equal in magnitude but in the direction opposite to the one she exerts on it. The line around the swimmer indicates the system of interest. Note that $F_{feet\ on\ wall}$ does not act on this system (the swimmer) and, thus, does not cancel $F_{wall\ on\ feet}$. Thus the free-body diagram shows only $F_{wall\ on\ feet}$, w, the gravitational force, and BF, the buoyant force of the water supporting the swimmer’s weight. The vertical forces w and BF cancel since there is no vertical motion.

Other examples of Newton’s third law are easy to find. As a professor paces in front of a whiteboard, she exerts a force backward on the floor. The floor exerts a reaction force forward on the professor that causes her to accelerate forward. Similarly, a car accelerates because the ground pushes forward on the drive wheels in reaction to the drive wheels pushing backward on the ground. You can see evidence of the wheels pushing backward when tires spin on a gravel road and throw rocks backward. In another example, rockets move forward by expelling gas backward at high velocity. This means the rocket exerts a large backward force on the gas in the rocket combustion chamber, and the gas therefore exerts a large reaction force forward on the rocket. This reaction force is called thrust. It is a common misconception that rockets propel themselves by pushing on the ground or on the air behind them. They actually work better in a vacuum, where they can more readily expel the exhaust gases. Helicopters similarly create lift by pushing air down, thereby experiencing an upward reaction force. Birds and airplanes also fly by exerting force on air in a direction opposite to that of whatever force they need. For example, the wings of a bird force air downward and backward in order to get lift and move forward. An octopus propels itself in the water by ejecting water through a funnel from its body, similar to a jet ski. In a situation similar to Sancho's, professional cage fighters experience reaction forces when they punch, sometimes breaking their hand by hitting an opponent's body.

Exercise 2.5.1: Getting up to speed: Choosing the Correct System

A physics professor pushes a cart of demonstration equipment to a lecture hall, as seen in Figure. Her mass is 65.0 kg, the cart's is 12.0 kg, and the equipment's is 7.0 kg. Calculate the acceleration produced when the professor exerts a backward force of 150 N on the floor. All forces opposing the motion, such as friction on the cart's wheels and air resistance, total 24.0 N.

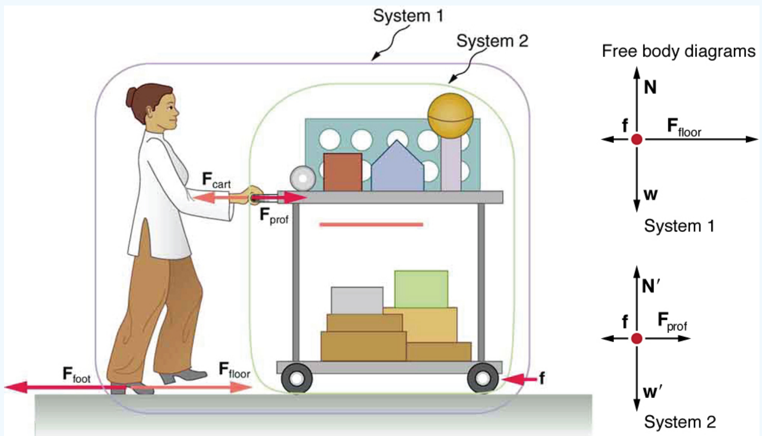


Figure 2.5.2: A professor pushes a cart of demonstration equipment. The lengths of the arrows are proportional to the magnitudes of the forces (except for f, since it is too small to draw to scale). Different questions are asked in each example; thus, the system of interest must be defined differently for each. System 1 is appropriate for Example, since it asks for the acceleration of the entire group of objects. Only F_{floor} and f are external forces acting on System 1 along the line of motion. All other forces either cancel or act on the outside world. System 2 is chosen for this example so that F_{prof} will be an external force and enter into Newton's second law. Note that the free-body diagrams, which allow us to apply Newton's second law, vary with the system chosen.

Strategy

Since they accelerate as a unit, we define the system to be the professor, cart, and equipment. This is System 1 in Figure. The professor pushes backward with a force F_{foot} of 150 N. According to Newton's third law, the floor exerts a forward reaction force F_{floor} of 150 N on System 1. Because all motion is horizontal, we can assume there is no net force in the vertical direction. The problem is therefore one-dimensional along the horizontal direction. As noted, f opposes the motion and is thus in the opposite direction of F_{floor}. Note that we do not include the forces F_{prof} or F_{cart} because these are internal forces, and we do not include F_{foot} because it acts on the floor, not on the system. There are no other significant forces acting on System 1. If the net external force can be found from all this information, we can use Newton's second law to find the acceleration as requested. See the free-body diagram in the figure.

Solution

Newton's second law is given by

$$a=\frac{F_{net}}{m}.$$

The net external force on System 1 is deduced from Figure and the discussion above to be

$$F_{net}=F_{floor}-f=150\,N-24.0\,N=126\,N.$$

The mass of System 1 is

$$m=(65.0+12.0+7.0)=84\,kg.$$

These values of F_{net} and m produce an acceleration of

$$a = \frac{F_{net}}{m}$$

$$a = \frac{126\,N}{84\,kg} = 1.5\,m/s^2.$$

Discussion

None of the forces between components of System 1, such as between the professor's hands and the cart, contribute to the net external force because they are internal to System 1. Another way to look at this is to note that forces between components of a system cancel because they are equal in magnitude and opposite in direction. For example, the force exerted by the professor on the cart results in an equal and opposite force back on her. In this case both forces act on the same system and, therefore, cancel. Thus internal forces (between components of a system) cancel. Choosing System 1 was crucial to solving this problem.

Example 2.5.2: Force on the Cart: Choosing a New System

Calculate the force the professor exerts on the cart in Figure using data from the previous example if needed.

Strategy

If we now define the system of interest to be the cart plus equipment (System 2 in Figure), then the net external force on System 2 is the force the professor exerts on the cart minus friction. The force she exerts on the cart, F_{prof} is an external force acting on System 2. F_{prof} was internal to System 1, but it is external to System 2 and will enter Newton's second law for System 2.

Solution

Newton's second law can be used to find F_{prof} Starting with

$$a = \frac{F_{net}}{m}$$

and noting that the magnitude of the net external force on System 2 is

$$F_{net} = F_{prof} - f,$$

we solve for F_{prof}, the desired quantity

$$F_{net} + f.$$

The value of f is given, so we must calculate net F_{net}. That can be done since both the acceleration and mass of System 2 are known. Using Newton's second law we see that

$$F_{net} = ma,$$

where the mass of System 2 is 19.0 kg (m = 12.0 kg + 7.0 kg) and its acceleration was found to be $a = 1.5\,m/s^2$ in the previous example. Thus,

$$F_{net} = ma$$

$$F_{net} = (19.0\,kg)(1.5\,m/s^2) = 29\,N.$$

Now we can find the desired force:

$$F_{prof} = F_{net} + f,$$

$$F_{prof} = 29\,N + 24.0\,N = 53\,N.$$

Discussion

It is interesting that this force is significantly less than the 150-N force the professor exerted backward on the floor. Not all of that 150-N force is transmitted to the cart; some of it accelerates the professor.

The choice of a system is an important analytical step both in solving problems and in thoroughly understanding the physics of the situation (which is not necessarily the same thing).

PHET EXPLORATIONS: GRAVITY FORCE LAB

Visualize the gravitational force that two objects exert on each other. Change properties of the objects in order to see how it changes the gravity force.

Section Summary

- **Newton's third law of motion** represents a basic symmetry in nature. It states: Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that the first body exerts.
- A **thrust** is a reaction force that pushes a body forward in response to a backward force. Rockets, airplanes, and cars are pushed forward by a thrust reaction force.

Glossary

Newton's third law of motion

whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that the first body exerts

thrust

a reaction force that pushes a body forward in response to a backward force; rockets, airplanes, and cars are pushed forward by a thrust reaction force

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

3: Work and Energy

3.1: Prelude to Work, Energy, and Energy Resources

Energy plays an essential role both in everyday events and in scientific phenomena. You can no doubt name many forms of energy, from that provided by our foods, to the energy we use to run our cars, to the sunlight that warms us on the beach. You can also cite examples of what people call energy that may not be scientific, such as someone having an energetic personality. Not only does energy have many interesting forms, it is involved in almost all phenomena, and is one of the most important concepts of physics. What makes it even more important is that the total amount of energy in the universe is constant. Energy can change forms, but it cannot appear from nothing or disappear without a trace. Energy is thus one of a handful of physical quantities that we say is *conserved*.

Figure 3.1.1: How many forms of energy can you identify in this photograph of a wind farm in Iowa? (credit: Jürgen from Sandesneben, Germany, Wikimedia Commons)

Conservation of energy (as physicists like to call the principle that energy can neither be created nor destroyed) is based on experiment. Even as scientists discovered new forms of energy, conservation of energy has always been found to apply. Perhaps the most dramatic example of this was supplied by Einstein when he suggested that mass is equivalent to energy (his famous equation $E = mc^2$.

From a societal viewpoint, energy is one of the major building blocks of modern civilization. Energy resources are key limiting factors to economic growth. The world use of energy resources, especially oil, continues to grow, with ominous consequences economically, socially, politically, and environmentally. We will briefly examine the world's energy use patterns at the end of this chapter.

There is no simple, yet accurate, scientific definition for energy. Energy is characterized by its many forms and the fact that it is conserved. We can loosely define energy as the ability to do work, admitting that in some circumstances not all **energy** is available to do work. Because of the association of energy with work, we begin the chapter with a discussion of work. Work is intimately related to energy and how energy moves from one system to another or changes form.

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

3.2: Work- The Scientific Definition

Learning Objectives

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

- Explain how an object must be displaced for a force on it to do work.
- Explain how relative directions of force and displacement determine whether the work done is positive, negative, or zero.

What It Means to Do Work

The scientific definition of work differs in some ways from its everyday meaning. Certain things we think of as hard work, such as writing an exam or carrying a heavy load on level ground, are not work as defined by a scientist. The scientific definition of work reveals its relationship to energy—whenever work is done, energy is transferred. For work, in the scientific sense, to be done, a force must be exerted and there must be motion or displacement in the direction of the force.

Formally, the work done on a system by a constant force is defined to be *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

$$W = F\,d \tag{3.2.1}$$

where W is work, d is the displacement of the system, and F is the magnitude of the applied force. Note that when the displacement and the applied force point in opposite directions, the work done is negative, which means that energy is being transferred out of the object as it is being displaced.

To find the work done on a system that undergoes motion that is not one-way or that is in two or three dimensions, we divide the motion into one-way one-dimensional segments and add up the work done over each segment.

What is Work?

The work done on a system by a constant force is *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

$$W = F\,d \tag{3.2.2}$$

where W is work, F is the magnitude of the force on the system, d is the magnitude of the displacement of the system.

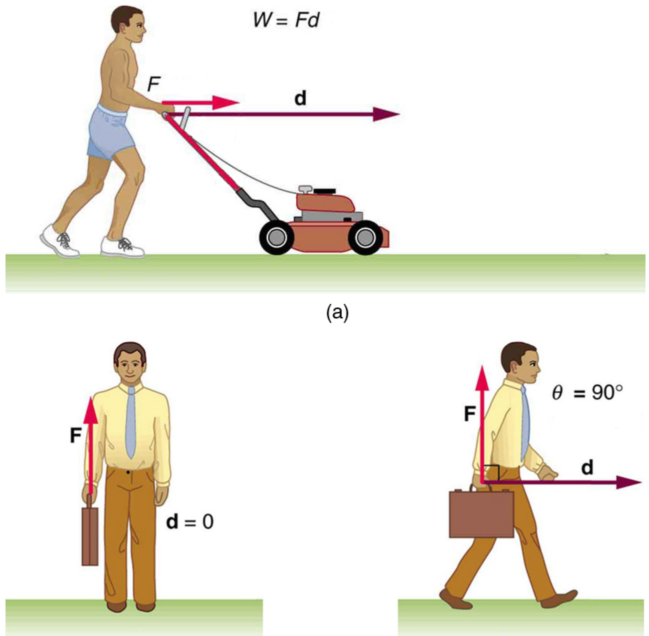


Figure $3.2.1$: Examples of work. (a) The work done by the force F on this lawn mower is $Fd\,cos\,\theta$. Note that $F\,cos\,\theta$ is the component of the force in the direction of motion. (b) A person holding a briefcase does no work on it, because there is no motion. No energy is transferred to or from the briefcase. (c) The person moving the briefcase horizontally at a constant speed does no work on it, and transfers no energy to it. (d) Work is done on the briefcase by carrying it upstairs at constant speed, because there is necessarily a component of force F in the direction of the motion. Energy is transferred to the briefcase and could in turn be used to do work. (e) When the briefcase is lowered, energy is transferred out of the briefcase and into an electric generator. Here the work done on the briefcase by the generator is negative, removing energy from the briefcase, because F and d are in opposite directions.

To examine what the definition of work means, let us consider the other situations shown in Figure. The person holding the briefcase in Figure $3.2.1b$ does no work, for example. Here $d = 0$, so $W = 0$. Why is it you get tired just holding a load? The answer is that your muscles are doing work against one another, *but they are doing no work on the system of interest* (the "briefcase-Earth system" - see Gravitational Potential Energy for more details). There must be motion for work to be done, and there must be a component of the force in the direction of the motion. For example, the person carrying the briefcase on level ground in Figure $3.2.1c$ does no work on it, because the force is perpendicular to the motion.

Summary

- Work is the transfer of energy by a force acting on an object as it is displaced.
- The work W that a force F does on an object is the product of the magnitude F of the force, times the magnitude d of the displacement. In symbols,

$$W = Fd \tag{3.2.3}$$

- The SI unit for work and energy is the joule (J), where $1\,J = 1\,N \cdot m = 1\,kg\,m^2/s^2$.
- The work done by a force is zero if the displacement is either zero or perpendicular to the force.
- The work done is positive if the force and displacement have the same direction, and negative if they have opposite direction.

Glossary

energy
the ability to do work

work
the transfer of energy by a force that causes an object to be displaced; the product of the component of the force in the direction of the displacement and the magnitude of the displacement

joule

SI unit of work and energy, equal to one newton-meter

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

3.3: Kinetic Energy and the Work-Energy Theorem

Learning Objectives

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

- Explain work as a transfer of energy and net work as the work done by the net force.
- Explain and apply the work-energy theorem.

Work Transfers Energy

What happens to the work done on a system? Energy is transferred into the system, but in what form? Does it remain in the system or move on? The answers depend on the situation. For example, if a lawn mower is pushed just hard enough to keep it going at a constant speed, then energy put into the mower by the person is removed continuously by friction, and eventually leaves the system in the form of heat transfer. In contrast, work done on the briefcase by the person carrying it up stairs in is stored in the briefcase-Earth system and can be recovered at any time by allowing the briefcase to fall back down to the ground floor. In fact, the building of the pyramids in ancient Egypt is an example of storing energy in a system by doing work on the system. Some of the energy imparted to the stone blocks in lifting them during construction of the pyramids remains in the stone-Earth system and has the potential to do work.

In this section we begin the study of various types of work and forms of energy. We will find that some types of work leave the energy of a system constant, for example, whereas others change the system in some way, such as making it move. We will also develop definitions of important forms of energy, such as the energy of motion.

Net Work and the Work-Energy Theorem

We know from the study of Newton's laws in Dynamics: Force and Newton's Laws of Motion that net force causes acceleration. We will see in this section that work done by the net force gives a system energy of motion, and in the process we will also find an expression for the energy of motion.

To begin with, let us consider a situation where a force is used to accelerate an object in a direction parallel to its initial velocity. Such a situation occurs for the package on the roller belt conveyor system shown below.

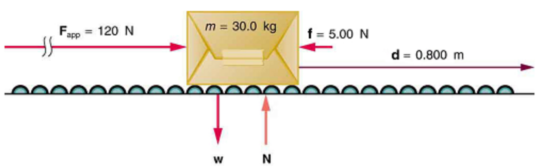


Figure 3.3.2: A package on a roller belt is pushed horizontally through a distance d.

The force of gravity and the normal force acting on the package are perpendicular to the displacement and do no work. Moreover, they are also equal in magnitude and opposite in direction so they cancel in calculating the net force. The net force arises solely from the horizontal applied force F_{app} and the horizontal friction force f. Thus, as expected, the net force is parallel to the displacement, and the net work is given by

$$W_{net} = F_{net} d. \tag{3.3.1}$$

The effect of the net force F_{net} is to accelerate the package from v_0 to v The kinetic energy of the package increases, indicating that the net work done on the system is positive. (See Example.) By using Newton's second law, and doing some algebra, we can reach an interesting conclusion. Substituting $F = ma$ from Newton's second law gives

$$W_{net} = mad. \tag{3.3.2}$$

To get a relationship between net work and the speed given to a system by the net force acting on it, we take $d = x - x_0$. It is possible to show that if the acceleration has a constant value a for an object traveling a distance d, then the final velocity v and initial velocity v_0 of the object are related by the equation $v^2 = v_0^2 + 2ad$. Solving for acceleration gives $a = \frac{v^2 - v_0^2}{2d}$. When a is substituted into the preceding expression for W_{net} we obtain

$$W_{net} = m\left(\frac{v^2 - v_0^2}{2d}\right)d. \tag{3.3.3}$$

The d cancels, and we rearrange this to obtain

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{3.3.4}$$

This expression is called the **work-energy theorem,** and it actually applies *in general* (even for forces that vary in direction and magnitude), although we have derived it for the special case of a constant force parallel to the displacement. The theorem implies that the net work on a system equals the change in the quantity $\frac{1}{2}mv^2$. This quantity is our first example of a form of energy.

Work-Energy Theorem

The net work on a system equals the change in the quantity $\frac{1}{2}mv^2$.

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{3.3.5}$$

The quantity $\frac{1}{2}mv^2$ in the work-energy theorem is defined to be the translational **kinetic energy** (KE) of a mass m moving at a speed v. (*Translational* kinetic energy is distinct from *rotational* kinetic energy, which is considered later.) In equation form, the translational kinetic energy,

$$KE = \frac{1}{2}mv^2, \tag{3.3.6}$$

is the energy associated with translational motion. Kinetic energy is a form of energy associated with the motion of a particle, single body, or system of objects moving together.

We are aware that it takes energy to get an object, like a car or the package in Figure, up to speed, but it may be a bit surprising that kinetic energy is proportional to speed squared. This proportionality means, for example, that a car traveling at 100 km/h has four times the kinetic energy it has at 50 km/h, helping to explain why high-speed collisions are so devastating. We will now consider a series of examples to illustrate various aspects of work and energy.

Example 3.3.1: Calculating the Kinetic Energy of a Package

Suppose a 30.0-kg package on the roller belt conveyor system in Figure 7.03.2 is moving at 0.500 m/s. What is its kinetic energy?

Strategy

Because the mass m and the speed v are given, the kinetic energy can be calculated from its definition as given in the equation $KE = \frac{1}{2}mv^2$.

Solution

The kinetic energy is given by

$$KE = \frac{1}{2}mv^2. \tag{3.3.7}$$

Entering known values gives

$$KE = 0.5(30.0\,kg)(0.500\,m/s)^2, \tag{3.3.8}$$

which yields

$$KE = 3.75\,kg \cdot m^2/s^2 = 3.75\,J \tag{3.3.9}$$

Discussion

Note that the unit of kinetic energy is the joule, the same as the unit of work, as mentioned when work was first defined. It is also interesting that, although this is a fairly massive package, its kinetic energy is not large at this relatively low speed. This fact is consistent with the observation that people can move packages like this without exhausting themselves.

Example $3.3.3$: Determining Speed from Work and Energy

Find the speed of the package in Figure 7.03.2. at the end of the push, using work and energy concepts.

Strategy

Here the work-energy theorem can be used, because we have just calculated the net work W_{net} and the initial kinetic energy, $\frac{1}{2}mv_0^2$ These calculations allow us to find the final kinetic energy, $\frac{1}{2}mv^2$ and thus the final speed v.

Solution

The work-energy theorem in equation form is

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{3.3.10}$$

Solving for $\frac{1}{2}mv^2$ gives

$$\frac{1}{2}mv^2 = W_{net} + \frac{1}{2}mv_0^2 \tag{3.3.11}$$

Thus,

$$\frac{1}{2}mv^2 = 92.0\,J + 3.75\,J = 95.75\,J. \tag{3.3.12}$$

Solving for the final speed as requested and entering known values gives

$$v = \sqrt{\frac{2(95.75\,J)}{m}} = \sqrt{\frac{191.5\,kg \cdot m^2/s^2}{30.0\,kg}} \tag{3.3.13}$$

$$= 2.53\,m/s \tag{3.3.14}$$

Discussion

Using work and energy, we not only arrive at an answer, we see that the final kinetic energy is the sum of the initial kinetic energy and the net work done on the package. This means that the work indeed adds to the energy of the package.

Summary

- The net work W_{net} is the work done by the net force acting on an object.
- Work done on an object transfers energy to the object.
- The translational kinetic energy of an object of mass m moving at speed v is $KE = \frac{1}{2}mv^2$.
- The work-energy theorem states that the net work W_{net} on a system changes its kinetic energy, $W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$.

Glossary

net work

work done by the net force, or vector sum of all the forces, acting on an object

work-energy theorem

the result, based on Newton’s laws, that the net work done on an object is equal to its change in kinetic energy

kinetic energy

the energy an object has by reason of its motion, equal to $\frac{1}{2}mv^2$ for the translational (i.e., non-rotational) motion of an object of mass m moving at speed v

Contributors and Attributions

3.4: Gravitational Potential Energy

Learning Objectives

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

- Explain gravitational potential energy in terms of work done against gravity.
- Show that the gravitational potential energy of an object of mass *m* at height *h* on Earth is given by $PE_g = mgh$
- Show how knowledge of the potential energy as a function of position can be used to simplify calculations and explain physical phenomena.

Work Done Against Gravity

Climbing stairs and lifting objects is work in both the scientific and everyday sense—it is work done against the gravitational force. When there is work, there is a transformation of energy. The work done against the gravitational force goes into an important form of stored energy that we will explore in this section.

Let us calculate the work done in lifting an object of mass m through a height h such as in Figure. If the object is lifted straight up at constant speed, then the force needed to lift it is equal to its weight mg. The work done on the mass is then $W = Fd = mgh$. We define this to be the **gravitational potential energy** (PE_g) put into (or gained by) the object-Earth system. This energy is associated with the state of separation between two objects that attract each other by the gravitational force. For convenience, we refer to this as the PE_g gained by the object, recognizing that this is energy stored in the gravitational field of Earth. Why do we use the word “system”? Potential energy is a property of a system rather than of a single object—due to its physical position. An object’s gravitational potential is due to its position relative to the surroundings within the Earth-object system. The force applied to the object is an external force, from outside the system. When it does positive work it increases the gravitational potential energy of the system. Because gravitational potential energy depends on relative position, we need a reference level at which to set the potential energy equal to 0. We usually choose this point to be Earth’s surface, but this point is arbitrary; what is important is the *difference* in gravitational potential energy, because this difference is what relates to the work done. The difference in gravitational potential energy of an object (in the Earth-object system) between two rungs of a ladder will be the same for the first two rungs as for the last two rungs.

Converting Between Potential Energy and Kinetic Energy

Gravitational potential energy may be converted to other forms of energy, such as kinetic energy. If we release the mass, gravitational force will do an amount of work equal to mgh on it, thereby increasing its kinetic energy by that same amount (by the work-energy theorem). We will find it more useful to consider just the conversion of PE_g to KE without explicitly considering the intermediate step of work. (See Example 3.4.2.) This shortcut makes it is easier to solve problems using energy (if possible) rather than explicitly using forces.

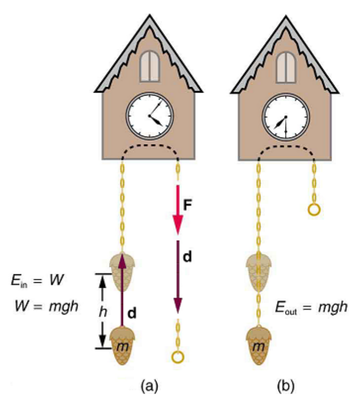


Figure 3.4.1: (a) The work done to lift the weight is stored in the mass-Earth system as gravitational potential energy. (b) As the weight moves downward, this gravitational potential energy is transferred to the cuckoo clock.

More precisely, we define the *change* in gravitational potential energy ΔPE_g to be

$$\Delta PE_g = mgh, \tag{3.4.1}$$

where, for simplicity, we denote the change in height by h rather than the usual Δh. Note that h is positive when the final height is greater than the initial height, and vice versa. For example, if a 0.500-kg mass hung from a cuckoo clock is raised 1.00 m, then its change in gravitational potential energy is

$$mgh = (0.500\,kg)(9.80\,m/s^2)(1.00\,m) \tag{3.4.2}$$

$$= 4.90\,kg \cdot m^2/s^2 = 4.90\,J. \tag{3.4.3}$$

Note that the units of gravitational potential energy turn out to be joules, the same as for work and other forms of energy. As the clock runs, the mass is lowered. We can think of the mass as gradually giving up its 4.90 J of gravitational potential energy, *without directly considering the force of gravity that does the work*.

Using Potential Energy to Simplify Calculations

The equation $\Delta PE_g = mgh$ applies for any path that has a change in height of h, not just when the mass is lifted straight up. (See Figure.) It is much easier to calculate mgh (a simple multiplication) than it is to calculate the work done along a complicated path. The idea of gravitational potential energy has the double advantage that it is very broadly applicable and it makes calculations easier. From now on, we will consider that any change in vertical position h of a mass m is accompanied by a change in gravitational potential energy mgh, and we will avoid the equivalent but more difficult task of calculating work done by or against the gravitational force.

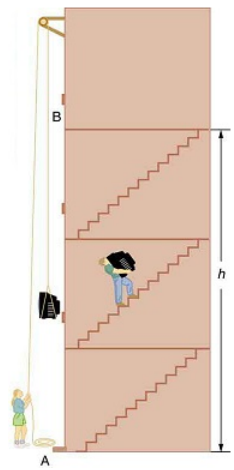


Figure $3.4.2$: The change in gravitational potential energy (ΔPE_g) between points A and B is independent of the path $\Delta PE_g = mgh$ for any path between the two points. Gravity is one of a small class of forces where the work done by or against the force depends only on the starting and ending points, not on the path between them.

Figure $3.4.3$: The work done by the ground upon the kangaroo reduces its kinetic energy to zero as it lands. However, by applying the force of the ground on the hind legs over a longer distance, the impact on the bones is reduced. (credit: Chris Samuel, Flickr)

Example 3.4.2: Finding the Speed of a Roller Coaster from its Height

(a) What is the final speed of the roller coaster shown in Figure, if it starts from rest at the top of the 20.0 m hill and work done by frictional forces is negligible? (b) What is its final speed (again assuming negligible friction) if its initial speed is 5.00 m/s?

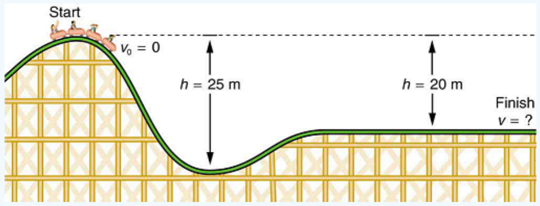


Figure 3.4.4: The speed of a roller coaster increases as gravity pulls it downhill and is greatest at its lowest point. Viewed in terms of energy, the roller-coaster-Earth system's gravitational potential energy is converted to kinetic energy. If work done by friction is negligible, all ΔPE_g is converted to KE.

Strategy

The roller coaster loses potential energy as it goes downhill. We neglect friction, so that the remaining force exerted by the track is the normal force, which is perpendicular to the direction of motion and does no work. The net work on the roller coaster is then done by gravity alone. The *loss* of gravitational potential energy from moving *downward* through a distance h equals the *gain* in kinetic energy. This can be written in equation form as $-\Delta PE = \Delta KE$. Using the equations for PE_g and KE we can solve for the final speed v, which is the desired quantity.

Solution for (a)

Here the initial kinetic energy is zero, so that $\Delta KE = \frac{1}{2}mv^2$. The equation for change in potential energy states that $\Delta PE_g = mgh$. Since h is negative in this case, we will rewrite this as $\Delta PE_g = -mg|h|$ to show the minus sign clearly. Thus,

$$-\Delta PE_g = \Delta KE \tag{3.4.4}$$

becomes

$$mg|h| = \frac{1}{2}mv^2 \tag{3.4.5}$$

Solving for v we find that mass cancels and that

$$v = \sqrt{2g|h|}. \tag{3.4.6}$$

Substituting known values,

$$v = \sqrt{2(9.80\,m/s^2)(20.0\,m)} \tag{3.4.7}$$

$$= 19.8\,m/s \tag{3.4.8}$$

Solution for (b)

Again $-\Delta PE_g = \Delta KE$. In this case there is initial kinetic energy, so $\Delta KE = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$. Thus,

$$mgh = \frac{1}{2}mv^2 - \frac{1}{2}m_0^2. \tag{3.4.9}$$

Rearranging gives

$$\frac{1}{2}mv^2 = mg|h| + \frac{1}{2}mv_0^2. \tag{3.4.10}$$

This means that the final kinetic energy is the sum of the initial kinetic energy and the gravitational potential energy. Mass again cancels, and

$$v = \sqrt{2g|h| + v_0^2}. \tag{3.4.11}$$

This equation is very similar to the kinematics equation $v=\sqrt{v_0^2+2ad}$, but it is more general—the kinematics equation is valid only for constant acceleration, whereas our equation above is valid for any path regardless of whether the object moves with a constant acceleration. Now, substituting known values gives

$$v=\sqrt{2(9.80\,m/s^2)(20.0\,m)+(5.00)^2} \tag{3.4.12}$$

$$=20.4\,m/s. \tag{3.4.13}$$

Discussion and Implications

First, note that mass cancels. This is quite consistent with observations made in Falling Objects that all objects fall at the same rate if friction is negligible. Second, only the speed of the roller coaster is considered; there is no information about its direction at any point. This reveals another general truth. When friction is negligible, the speed of a falling body depends only on its initial speed and height, and not on its mass or the path taken. For example, the roller coaster will have the same final speed whether it falls 20.0 m straight down or takes a more complicated path like the one in the figure. Third, and perhaps unexpectedly, the final speed in part (b) is greater than in part (a), but by far less than 5.00 m/s. Finally, note that speed can be found at *any* height along the way by simply using the appropriate value of h at the point of interest.

We have seen that work done by or against the gravitational force depends only on the starting and ending points, and not on the path between, allowing us to define the simplifying concept of gravitational potential energy. We can do the same thing for a few other forces, and we will see that this leads to a formal definition of the law of conservation of energy.

Making Connections: Take-Home Investigation—Converting Potential to Kinetic Energy

One can study the conversion of gravitational potential energy into kinetic energy in this experiment. On a smooth, level surface, use a ruler of the kind that has a groove running along its length and a book to make an incline (see Figure). Place a marble at the 10-cm position on the ruler and let it roll down the ruler. When it hits the level surface, measure the time it takes to roll one meter. Now place the marble at the 20-cm and the 30-cm positions and again measure the times it takes to roll 1 m on the level surface. Find the velocity of the marble on the level surface for all three positions. Plot velocity squared versus the distance traveled by the marble. What is the shape of each plot? If the shape is a straight line, the plot shows that the marble's kinetic energy at the bottom is proportional to its potential energy at the release point.

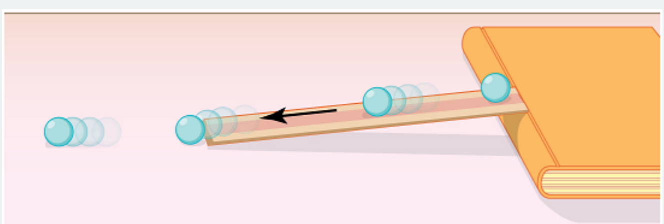


Figure 3.4.5: A marble rolls down a ruler, and its speed on the level surface is measured.

Summary

- Work done against gravity in lifting an object becomes potential energy of the object-Earth system.
- The change in gravitational potential energy ΔPE_g, is $\Delta PE_g=mgh$, with h being the increase in height and g the acceleration due to gravity.
- The gravitational potential energy of an object near Earth's surface is due to its position in the mass-Earth system. Only differences in gravitational potential energy, ΔPE_g, have physical significance.
- As an object descends without friction, its gravitational potential energy changes into kinetic energy corresponding to increasing speed, so that $\Delta KE=-\Delta PE_g$.

Glossary

gravitational potential energy
the energy an object has due to its position in a gravitational field

Contributors and Attributions

3.5: Conservation of Energy

Learning Objectives

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

- Explain the law of the conservation of energy.
- Describe some of the many forms of energy.
- Define efficiency of an energy conversion process as the fraction left as useful energy or work, rather than being transformed, for example, into thermal energy.

Energy, as we have noted, is conserved, making it one of the most important physical quantities in nature. The law of conservation of energy can be stated as follows:

We have explored some forms of energy and some ways it can be transferred from one system to another. This exploration led to the definition of two major types of energy—mechanical energy $(KE + PE)$ and energy transferred via work done by nonconservative forces (W_{nc}) But energy takes *many* other forms, manifesting itself in *many* different ways, and we need to be able to deal with all of these before we can write an equation for the above general statement of the conservation of energy.

Other Forms of Energy than Mechanical Energy

At this point, we deal with all other forms of energy by lumping them into a single group called other energy (OE). Then we can state the conservation of energy in equation form as

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f. \tag{3.5.1}$$

All types of energy and work can be included in this very general statement of conservation of energy. Kinetic energy is KE, work done by a conservative force is represented by PE, work done by nonconservative forces is W_{nc} and all other energies are included as OE. This equation applies to all previous examples; in those situations OE was constant, and so it subtracted out and was not directly considered.

Usefulness of the Energy Conservation Principle

The fact that energy is conserved and has many forms makes it very important. You will find that energy is discussed in many contexts, because it is involved in all processes. It will also become apparent that many situations are best understood in terms of energy and that problems are often most easily conceptualized and solved by considering energy.

When does OE play a role? One example occurs when a person eats. Food is oxidized with the release of carbon dioxide, water, and energy. Some of this chemical energy is converted to kinetic energy when the person moves, to potential energy when the person changes altitude, and to thermal energy (another form of OE).

Some of the Many Forms of Energy

What are some other forms of energy? You can probably name a number of forms of energy not yet discussed. Many of these will be covered in later chapters, but let us detail a few here. **Electrical energy** is a common form that is converted to many other forms and does work in a wide range of practical situations. Fuels, such as gasoline and food, carry **chemical energy** that can be transferred to a system through oxidation. Chemical fuel can also produce electrical energy, such as in batteries. Batteries can in turn produce light, which is a very pure form of energy. Most energy sources on Earth are in fact stored energy from the energy we receive from the Sun. We sometimes refer to this as **radiant energy**, or electromagnetic radiation, which includes visible light, infrared, and ultraviolet radiation. **Nuclear energy** comes from processes that convert measurable amounts of mass into energy. Nuclear energy is transformed into the energy of sunlight, into electrical energy in power plants, and into the energy of the heat transfer and blast in weapons. Atoms and molecules inside all objects are in random motion. This internal mechanical energy from the random motions is called **thermal energy**, because it is related to the temperature of the object. These and all other forms of energy can be converted into one another and can do work.

Table gives the amount of energy stored, used, or released from various objects and in various phenomena. The range of energies and the variety of types and situations is impressive.

Problem-Solving Strategies for Energy

You will find the following problem-solving strategies useful whenever you deal with energy. The strategies help in organizing and reinforcing energy concepts. In fact, they are used in the examples presented in this chapter. The familiar general problem-solving strategies presented earlier—involving identifying physical principles, knowns, and unknowns, checking units, and so on —continue to be relevant here.

Step 1. Determine the system of interest and identify what information is given and what quantity is to be calculated. A sketch will help.

Step 2. Examine all the forces involved and determine whether you know or are given the potential energy from the work done by the forces. Then use step 3 or step 4.

Step 3. If you know the potential energies for the forces that enter into the problem, then forces are all conservative, and you can apply conservation of mechanical energy simply in terms of potential and kinetic energy. The equation expressing conservation of energy is

$$KE_i + PE_i = KE_f + PE_f. \tag{3.5.2}$$

Step 4. If you know the potential energy for only some of the forces, possibly because some of them are nonconservative and do not have a potential energy, or if there are other energies that are not easily treated in terms of force and work, then the conservation of energy law in its most general form must be used.

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f. \tag{3.5.3}$$

In most problems, one or more of the terms is zero, simplifying its solution. Do not calculate W_c, the work done by conservative forces; it is already incorporated in the PE terms.

Step 5. You have already identified the types of work and energy involved (in step 2). Before solving for the unknown, *eliminate terms wherever possible* to simplify the algebra. For example, choose $h = 0$ at either the initial or final point, so that PE_g is zero there. Then solve for the unknown in the customary manner.

Step 6. *Check the answer to see if it is reasonable*. Once you have solved a problem, reexamine the forms of work and energy to see if you have set up the conservation of energy equation correctly. For example, work done against friction should be negative, potential energy at the bottom of a hill should be less than that at the top, and so on. Also check to see that the numerical value obtained is reasonable. For example, the final speed of a skateboarder who coasts down a 3-m-high ramp could reasonably be 20 km/h, but *not* 80 km/h.

Transformation of Energy

The transformation of energy from one form into others is happening all the time. The chemical energy in food is converted into thermal energy through metabolism; light energy is converted into chemical energy through photosynthesis. In a larger example, the chemical energy contained in coal is converted into thermal energy as it burns to turn water into steam in a boiler. This thermal energy in the steam in turn is converted to mechanical energy as it spins a turbine, which is connected to a generator to produce electrical energy. (In all of these examples, not all of the initial energy is converted into the forms mentioned. This important point is discussed later in this section.)

Another example of energy conversion occurs in a solar cell. Sunlight impinging on a solar cell (Figure 7.7.1) produces electricity, which in turn can be used to run an electric motor. Energy is converted from the primary source of solar energy into electrical energy and then into mechanical energy.

Figure 3.5.1: Solar energy is converted into electrical energy by solar cells, which is used to run a motor in this solar-power aircraft. (credit: NASA)

Object/phenomenon	Energy in joules
Big Bang	10^{68}
Energy released in a supernova	10^{44}
Fusion of all the hydrogen in Earth’s oceans	10^{34}
Annual world energy use	4×10^{20}
Large fusion bomb (9 megaton)	3.8×10^{16}
1 kg hydrogen (fusion to helium)	6.4×10^{14}
1 kg uranium (nuclear fission)	8.0×10^{13}
Hiroshima-size fission bomb (10 kiloton)	4.2×10^{13}
90,000-ton aircraft carrier at 30 knots	1.1×10^{10}
1 barrel crude oil	5.9×10^{9}
1 ton TNT	4.2×10^{9}
1 gallon of gasoline	1.2×10^{8}
Daily home electricity use (developed countries)	$7n \times 10^{7}$
Daily adult food intake (recommended)	1.2×10^{7}
1000-kg car at 90 km/h	3.1×10^{5}
1 g fat (9.3 kcal)	3.9×10^{4}
ATP hydrolysis reaction	3.2×10^{4}
1 g carbohydrate (4.1 kcal)	1.7×10^{4}
1 g protein (4.1 kcal)	1.7×10^{4}
Tennis ball at 100 km/h	22
Mosquito $^{-2}$ g at $0.5\ \mathrm{m/s})$	1.3×10^{-6}
Single electron in a TV tube beam	4.0×10^{-15}
Energy to break one DNA strand	4.0×10^{-19}

Efficiency

Even though energy is conserved in an energy conversion process, the output of *useful energy* or work will be less than the energy input. The efficiency E_{ff} of an energy conversion process is defined as

$$Efficiency\,(E_{ff}) = \frac{useful\,energy\,or\,work\,output}{total\,energy\,input} = \frac{W_{out}}{E_{in}}. \tag{3.5.4}$$

Table lists some efficiencies of mechanical devices and human activities. In a coal-fired power plant, for example, about 40% of the chemical energy in the coal becomes useful electrical energy. The other 60% transforms into other (perhaps less useful) energy forms, such as thermal energy, which is then released to the environment through combustion gases and cooling towers.

Activity/device	Efficiency (%)
Cycling and climbing	20
Swimming, surface	2
Swimming, submerged	4
Shoveling	3
Weightlifting	9
Steam engine	17
Gasoline engine	30
Diesel engine	35
Nuclear power plant	35
Coal power plant	42
Electric motor	98
Compact fluorescent light	20
Gas heater (residential)	90
Solar cell	10

Efficiency of the Human Body and Mechanical Devices

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 energies for each spring.

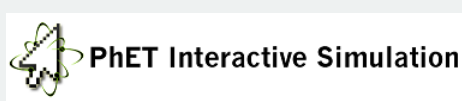


Figure 3.5.2: Masses and Springs

Summary

- The law of conservation of energy states that the total energy is constant in any process. Energy may change in form or be transferred from one system to another, but the total remains the same.
- When all forms of energy are considered, conservation of energy is written in equation form as

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f, \tag{3.5.5}$$

  where OE is all **other forms of energy** besides mechanical energy.
- Commonly encountered forms of energy include electric energy, chemical energy, radiant energy, nuclear energy, and thermal energy.

- Energy is often utilized to do work, but it is not possible to convert all the energy of a system to work.

The efficiency E_{ff} of a machine or human is defined to be $E_{ff} = \frac{W_{out}}{E_{in}}$, where W_{out} is useful work output and E_{in} s the energy consumed.

Glossary

law of conservation of energy
: the general law that total energy is constant in any process; energy may change in form or be transferred from one system to another, but the total remains the same

electrical energy
: the energy carried by a flow of charge

chemical energy
: the energy in a substance stored in the bonds between atoms and molecules that can be released in a chemical reaction

radiant energy
: the energy carried by electromagnetic waves

nuclear energy
: energy released by changes within atomic nuclei, such as the fusion of two light nuclei or the fission of a heavy nucleus

thermal energy
: the energy within an object due to the random motion of its atoms and molecules that accounts for the object's temperature

efficiency
: a measure of the effectiveness of the input of energy to do work; useful energy or work divided by the total input of energy

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

3.6: Work, Energy, and Power in Humans

Learning Objectives

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

- Explain the human body's consumption of energy when at rest vs. when engaged in activities that do useful work.
- Calculate the conversion of chemical energy in food into useful work.

Energy Conversion in Humans

Our own bodies, like all living organisms, are energy conversion machines. Conservation of energy implies that the chemical energy stored in food is converted into work, thermal energy, and/or stored as chemical energy in fatty tissue. (Figure 7.09.1.) The fraction going into each form depends both on how much we eat and on our level of physical activity. If we eat more than is needed to do work and stay warm, the remainder goes into body fat.

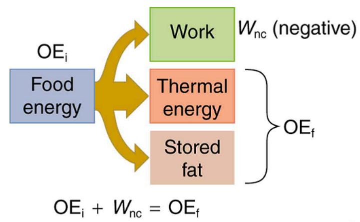


Figure 3.6.1: Energy consumed by humans is converted to work, thermal energy, and stored fat. By far the largest fraction goes to thermal energy, although the fraction varies depending on the type of physical activity.

Power Consumed at Rest

The *rate* at which the body uses food energy to sustain life and to do different activities is called the metabolic rate. The total energy conversion rate of a person *at rest* is called the basal metabolic rate (BMR) and is divided among various systems in the body, as shown in Table. The largest fraction goes to the liver and spleen, with the brain coming next. Of course, during vigorous exercise, the energy consumption of the skeletal muscles and heart increase markedly. About 75% of the calories burned in a day go into these basic functions. The BMR is a function of age, gender, total body weight, and amount of muscle mass (which burns more calories than body fat). Athletes have a greater BMR due to this last factor.

Basal Metabolic Rates (BMR):

Organ	Power consumed at rest (W)	Oxygen consumption (mL/min)	Percent of BMR
Liver & spleen	23	67	27
Brain	16	47	19
Skeletal muscle	15	45	18
Kidney	9	26	10
Heart	6	17	7
Other	16	48	19
Totals	85 W	250 mL/min	100%

Energy consumption is directly proportional to oxygen consumption because the digestive process is basically one of oxidizing food. We can measure the energy people use during various activities by measuring their oxygen use. (See Figure7.09.1.) Approximately 20 kJ of energy are produced for each liter of oxygen consumed, independent of the type of food. Table shows energy and oxygen consumption rates (power expended) for a variety of activities.

Power of Doing Useful Work

Work done by a person is sometimes called useful work, which is *work done on the outside world*, such as lifting weights. Useful work requires a force exerted through a distance on the outside world, and so it excludes internal work, such as that done by the heart when pumping blood. Useful work does include that done in climbing stairs or accelerating to a full run, because these are accomplished by exerting forces on the outside world. Forces exerted by the body are non-conservative, so that they can change the mechanical energy $(KE+PE)$ of the system worked upon, and this is often the goal. A baseball player throwing a ball, for example, increases both the ball's kinetic and potential energy.

If a person needs more energy than they consume, such as when doing vigorous work, the body must draw upon the chemical energy stored in fat. So exercise can be helpful in losing fat. However, the amount of exercise needed to produce a loss in fat, or to burn off extra calories consumed that day, can be large, as Example 7.09.1 illustrates.

Example 3.6.1: Calculating Weight Loss from Exercising

If a person who normally requires an average of 12,000 kJ (3000 kcal) of food energy per day consumes 13,000 kJ per day, he will steadily gain weight. How much bicycling per day is required to work off this extra 1000 kJ?

Solution

Table states that 400 W are used when cycling at a moderate speed. The time required to work off 1000 kJ at this rate is then

$$Time=\frac{energy}{\left(\frac{energy}{time}\right)}=\frac{1000\,kJ}{400\,W}=2500\,s=42\,min. \tag{3.6.1}$$

Discussion

If this person uses more energy than he or she consumes, the person's body will obtain the needed energy by metabolizing body fat. If the person uses 13,000 kJ but consumes only 12,000 kJ, then the amount of fat loss will be

$$Fat\,loss=(1000\,kJ)\left(\frac{1\,g\,fat}{30\,kJ}\right)=26\,g, \tag{3.6.2}$$

assuming the energy content of fat to be 39 kJ/g.

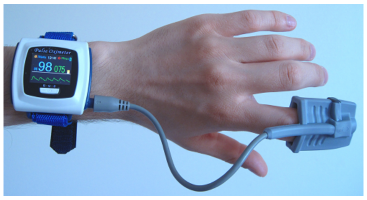


Figure 3.6.2: A pulse oxymeter is an apparatus that measures the amount of oxygen in blood. A knowledge of oxygen and carbon dioxide levels indicates a person's metabolic rate, which is the rate at which food energy is converted to another form. a person's metabolic rate, which is the rate at which food energy is converted to another form. Such measurements can indicate the level of athletic conditioning as well as certain medical problems. (credit: UusiAjaja, Wikimedia Commons)

Energy and Oxygen Consumption Rates:

Activity	Energy consumption in watts	Oxygen consumption in liters O_2/min
Sleeping	83	0.24
Sitting at rest	120	0.34
Standing relaxed	125	0.36
Sitting in class	210	0.60
Walking (5 km/h)	280	0.80
Cycling (13–18 km/h)	400	1.14
Shivering	425	1.21
Playing tennis	440	1.26
Swimming breaststroke	475	1.36
Ice skating (14.5 km/h)	545	1.56

Activity	Energy consumption in watts	Oxygen consumption in liters O_2/min
Climbing stairs (116/min)	685	1.96
Cycling (21 km/h)	700	2.00
Running cross-country	740	2.12
Playing basketball	800	2.28
Cycling, professional racer	1855	5.30
Sprinting	2415	6.90

All bodily functions, from thinking to lifting weights, require energy. (See Figure 7.09.3.) The many small muscle actions accompanying all quiet activity, from sleeping to head scratching, ultimately become thermal energy, as do less visible muscle actions by the heart, lungs, and digestive tract. Shivering, in fact, is an involuntary response to low body temperature that pits muscles against one another to produce thermal energy in the body (and do no work). The kidneys and liver consume a surprising amount of energy, but the biggest surprise of all it that a full 25% of all energy consumed by the body is used to maintain electrical potentials in all living cells. (Nerve cells use this electrical potential in nerve impulses.) This bioelectrical energy ultimately becomes mostly thermal energy, but some is utilized to power chemical processes such as in the kidneys and liver, and in fat production.

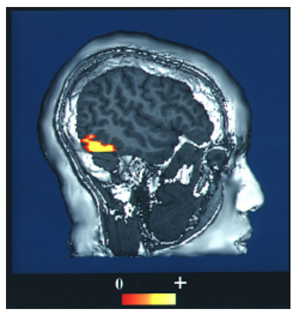


Figure 3.6.3: This MRI scan shows an increased level of energy consumption in the vision center of the brain. Here, the patient was being asked to recognize faces. (credit: NIH via Wikimedia Commons)

Summary

- The human body converts energy stored in food into work, thermal energy, and/or chemical energy that is stored in fatty tissue.
- The *rate* at which the body uses food energy to sustain life and to do different activities is called the metabolic rate, and the corresponding rate when at rest is called the basal metabolic rate (BMR)
- The energy included in the basal metabolic rate is divided among various systems in the body, with the largest fraction going to the liver and spleen, and the brain coming next.
- About 75% of food calories are used to sustain basic body functions included in the basal metabolic rate.
- The energy consumption of people during various activities can be determined by measuring their oxygen use, because the digestive process is basically one of oxidizing food.

Glossary

metabolic rate
the rate at which the body uses food energy to sustain life and to do different activities

basal metabolic rate
the total energy conversion rate of a person at rest

useful work
work done on an external system

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

3.7: World Energy Use

Learning Objectives

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

- Describe the distinction between renewable and nonrenewable energy sources.
- Explain why the inevitable conversion of energy to less useful forms makes it necessary to conserve energy resources.

Energy is an important ingredient in all phases of society. We live in a very interdependent world, and access to adequate and reliable energy resources is crucial for economic growth and for maintaining the quality of our lives. But current levels of energy consumption and production are not sustainable. About 40% of the world's energy comes from oil, and much of that goes to transportation uses. Oil prices are dependent as much upon new (or foreseen) discoveries as they are upon political events and situations around the world. The U.S., with 4.5% of the world's population, consumes 24% of the world's oil production per year; 66% of that oil is imported!

Renewable and Nonrenewable Energy Sources

The principal energy resources used in the world are shown in Figure 7.10.1. The fuel mix has changed over the years but now is dominated by oil, although natural gas and solar contributions are increasing. Renewable forms of energy are those sources that cannot be used up, such as water, wind, solar, and biomass. About 85% of our energy comes from nonrenewable fossil fuels—oil, natural gas, coal. The likelihood of a link between global warming and fossil fuel use, with its production of carbon dioxide through combustion, has made, in the eyes of many scientists, a shift to non-fossil fuels of utmost importance—but it will not be easy.

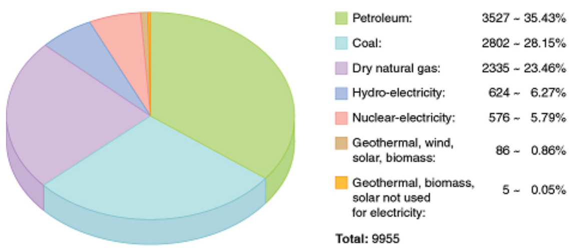


Figure 3.7.1: World energy consumption by source, in billions of kilowatt-hours: 2006. (credit: KVDP)

The World's Growing Energy Needs

World energy consumption continues to rise, especially in the developing countries. (See Figure 7.10.1.) Global demand for energy has tripled in the past 50 years and might triple again in the next 30 years. While much of this growth will come from the rapidly booming economies of China and India, many of the developed countries, especially those in Europe, are hoping to meet their energy needs by expanding the use of renewable sources. Although presently only a small percentage, renewable energy is growing very fast, especially wind energy. For example, Germany plans to meet 20% of its electricity and 10% of its overall energy needs with renewable resources by the year 2020. (See 7.10.2.) Energy is a key constraint in the rapid economic growth of China and India. In 2003, China surpassed Japan as the world's second largest consumer of oil. However, over 1/3 of this is imported. Unlike most Western countries, coal dominates the commercial energy resources of China, accounting for 2/3 of its energy consumption. In 2009 China surpassed the United States as the largest generator of CO_2 In India, the main energy resources are biomass (wood and dung) and coal. Half of India's oil is imported. About 70% of India's electricity is generated by highly polluting coal. Yet there are sizeable strides being made in renewable energy. India has a rapidly growing wind energy base, and it has the largest solar cooking program in the world.

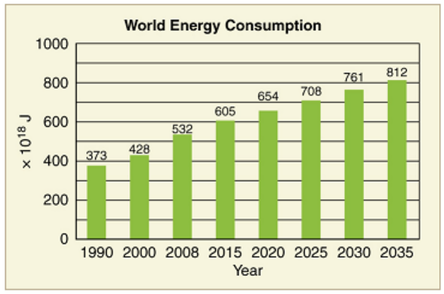


Figure 3.7.2: Past and projected world energy use (source: Based on data from U.S. Energy Information Administration, 2011)

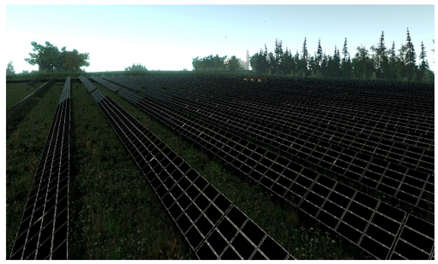


Figure 3.7.3: Solar cell arrays at a power plant in Steindorf, Germany (credit: Michael Betke, Flickr)

Table displays the 2006 commercial energy mix by country for some of the prime energy users in the world. While non-renewable sources dominate, some countries get a sizeable percentage of their electricity from renewable resources. For example, about 67% of New Zealand's electricity demand is met by hydroelectric. Only 10% of the U.S. electricity is generated by renewable resources, primarily hydroelectric. It is difficult to determine total contributihttp://physwiki.ucdavis.ed..._Energy_Useons of renewable energy in some countries with a large rural population, so these percentages in this table are left blank.

Energy Consumption—Selected Countries (2006):

Country	Consumption, in EJ (10^{18} J)	Oil	Natural Gas	Coal	Nuclear	Hydro	Other Renewables	Electricity Use per capita (kWh/yr)	Energy Use per capita (GJ/yr)
Australia	5.4	34%	17%	44%	0%	3%	1%	10000	260
Brazil	9.6	48%	7%	5%	1%	35%	2%	2000	50
China	63	22%	3%	69%	1%	6%		1500	35
Egypt	2.4	50%	41%	1%	0%	6%		990	32
Germany	16	37%	24%	24%	11%	1%	3%	6400	173
India	15	34%	7%	52%	1%	5%		470	13
Indonesia	4.9	51%	26%	16%	0%	2%	3%	420	22
Japan	24	48%	14%	21%	12%	4%	1%	7100	176
New Zealand	0.44	32%	26%	6%	0%	11%	19%	8500	102
Russia	31	19%	53%	16%	5%	6%		5700	202
U.S.	105	40%	23%	22%	8%	3%	1%	12500	340
World	**432**	**39%**	**23%**	**24%**	**6%**	**6%**	**2%**	**2600**	**71**

Energy and Economic Well-being

The last two columns in this table examine the energy and electricity use per capita. Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (gross domestic product) per capita, are matched by higher levels of energy consumption per capita. This is borne out in Figure 7.10.4. Increased efficiency of energy use will change this dependency. A global problem is balancing energy resource development against the harmful effects upon the environment in its extraction and use.

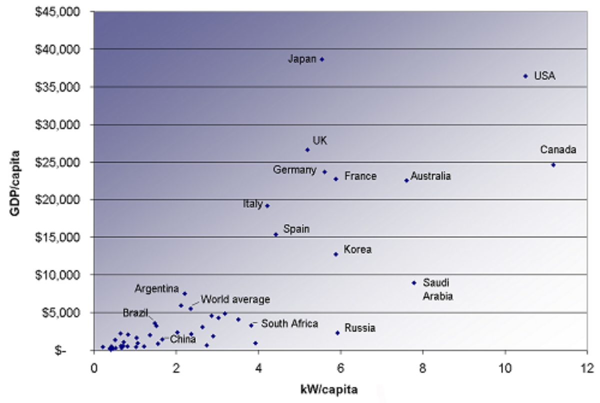


Figure 3.7.4: Power consumption per capita versus GDP per capita for various countries. Note the increase in energy usage with increasing GDP. (2007, credit: Frank van Mierlo, Wikimedia Commons)

Conserving Energy

As we finish this chapter on energy and work, it is relevant to draw some distinctions between two sometimes misunderstood terms in the area of energy use. As has been mentioned elsewhere, the “law of the conservation of energy” is a very useful principle in analyzing physical processes. It is a statement that cannot be proven from basic principles, but is a very good bookkeeping device, and no exceptions have ever been found. It states that the total amount of energy in an isolated system will always remain constant. Related to this principle, but remarkably different from it, is the important philosophy of energy conservation. This concept has to do with seeking to decrease the amount of energy used by an individual or group through (1) reduced activities (e.g., turning down thermostats, driving fewer kilometers) and/or (2) increasing conversion efficiencies in the performance of a particular task—such as developing and using more efficient room heaters, cars that have greater miles-per-gallon ratings, energy-efficient compact fluorescent lights, etc.

Since energy in an isolated system is not destroyed or created or generated, one might wonder why we need to be concerned about our energy resources, since energy is a conserved quantity. The problem is that the final result of most energy transformations is waste heat transfer to the environment and conversion to energy forms no longer useful for doing work. To state it in another way, the potential for energy to produce useful work has been “degraded” in the energy transformation. (This will be discussed in more detail in Thermodynamics.)

Summary

- The relative use of different fuels to provide energy has changed over the years, but fuel use is currently dominated by oil, although natural gas and solar contributions are increasing.
- Although non-renewable sources dominate, some countries meet a sizeable percentage of their electricity needs from renewable resources.
- The United States obtains only about 10% of its energy from renewable sources, mostly hydroelectric power.
- Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (Gross Domestic Product) per capita, are matched by higher levels of energy consumption per capita.
- Even though, in accordance with the law of conservation of energy, energy can never be created or destroyed, energy that can be used to do work is always partly converted to less useful forms, such as waste heat to the environment, in all of our uses of energy for practical purposes.

Glossary

renewable forms of energy

those sources that cannot be used up, such as water, wind, solar, and biomass

fossil fuels

oil, natural gas, and coal

Contributors and Attributions

CHAPTER OVERVIEW

4: Momentum

4.1: Prelude

We use the term momentum in various ways in everyday language, and most of these ways are consistent with its precise scientific definition. We speak of sports teams or politicians gaining and maintaining the momentum to win. We also recognize that momentum has something to do with collisions. For example, looking at the rugby players in the photograph colliding and falling to the ground, we expect their momenta to have great effects in the resulting collisions. Generally, momentum implies a tendency to continue on course—to move in the same direction—and is associated with great mass and speed.

Figure 4.1.1: Each rugby player has great momentum, which will affect the outcome of their collisions with each other and the ground. (credit: ozzzie, Flickr)

Momentum, like energy, is important because it is conserved. Only a few physical quantities are conserved in nature, and studying them yields fundamental insight into how nature works, as we shall see in our study of momentum.

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

4.2: Linear Momentum and Force

Learning Objectives

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

- Define linear momentum.
- Explain the relationship between momentum and force.
- State Newton's second law of motion in terms of momentum.
- Calculate momentum given mass and velocity.

The scientific definition of linear momentum is consistent with most people's intuitive understanding of momentum: a large, fast-moving object has greater momentum than a smaller, slower object. Linear momentum is defined as the product of a system's mass multiplied by its velocity.

Linear Momentum

Linear momentum is defined as the product of a system's mass multiplied by its velocity:

$$p = mv \tag{4.2.1}$$

Momentum is directly proportional to the object's mass and also its velocity. Thus the greater an object's mass or the greater its velocity, the greater its momentum. Momentum p is a vector having the same direction as the velocity v. The SI unit for momentum is $kg \cdot m/s$.

Example 4.2.1: Calculating Momentum: A Football Player and a Football

a. Calculate the momentum of a 110-kg football player running at 8.00 m/s.
b. Compare the player's momentum with the momentum of a hard-thrown 0.410-kg football that has a speed of 25.0 m/s.

Strategy

No information is given regarding direction, and so we can calculate only the magnitude of the momentum, p (As usual, a symbol that is in italics is a magnitude, whereas one that is italicized, boldfaced, and has an arrow is a vector.) In both parts of this example, the magnitude of momentum can be calculated directly from the definition of momentum given in Equation 4.2.1, which becomes

$$p = mv$$

when only magnitudes are considered.

Solution for (a)

To determine the momentum of the player, substitute the known values for the player's mass and speed into the equation.

$$\begin{aligned} p_{player} &= (110\, kg)(8.00\, m/s) \\ &= 880\, kg \cdot m/s \end{aligned}$$

Solution for (b)

To determine the momentum of the ball, substitute the known values for the ball's mass and speed into the equation.

$$\begin{aligned} p_{ball} &= (0.410\, kg)(25.0\, m/s) \\ &= 10.3\, kg \cdot m/s \end{aligned}$$

The ratio of the player's momentum to that of the ball is

$$\frac{p_{player}}{p_{ball}} = \frac{880}{10.3} = 85.0$$

Discussion

Although the ball has greater velocity, the player has a much greater mass. Thus the momentum of the player is much greater than the momentum of the football, as you might guess. As a result, the player's motion is only slightly affected if he catches the ball. We shall quantify what happens in such collisions in terms of momentum in later sections.

Momentum and Newton's Second Law

The importance of momentum, unlike the importance of energy, was recognized early in the development of classical physics. Momentum was deemed so important that it was called the "quantity of motion." Newton actually stated his second law of motion in terms of momentum: The net external force equals the change in momentum of a system divided by the time over which it changes.

Newton's Second Law of Motion in Terms of Momentum

The net external force equals the change in momentum of a system divided by the time over which it changes.

$$F_{net} = \frac{\Delta p}{\Delta t} \tag{4.2.2}$$

where F_{net} is the net external force, Δp is the change in momentum, and Δt is the change in time.

Making Connections: Force and Momentum

Force and momentum are intimately related. Force acting over time can change momentum, and Newton's second law of motion, can be stated in its most broadly applicable form in terms of momentum. Momentum continues to be a key concept in the study of atomic and subatomic particles in quantum mechanics.

This statement of Newton's second law of motion includes the more familiar $F_{net} = ma$ as a special case. We can derive this form as follows. First, note that the change in momentum Δp is given by

$$\Delta p = \Delta(mv) \tag{4.2.3}$$

If the mass of the system is constant, then

$$\Delta(mv) = m\Delta v. \tag{4.2.4}$$

So that for constant mass, Newton's second law of motion becomes

$$F_{net} = \frac{\Delta p}{\Delta t} = \frac{m\Delta v}{\Delta t}. \tag{4.2.5}$$

Because $\frac{\Delta v}{\Delta t} = a,$ we get the familiar equation

$$F_{net} = ma \tag{4.2.6}$$

when the mass of the system is *constant*.

Newton's second law of motion stated in terms of momentum is more generally applicable because it can be applied to systems where the mass is changing, such as rockets, as well as to systems of constant mass. We will consider systems with varying mass in some detail**;** however, the relationship between momentum and force remains useful when mass is constant, such as in the following example.

Example 4.2.2: Calculating Force: Venus Williams' Racquet

During the 2007 French Open, Venus Williams hit the fastest recorded serve in a premier women's match, reaching a speed of 58 m/s (209 km/h). What is the average force exerted on the 0.057-kg tennis ball by Venus Williams' racquet, assuming that the ball's speed just after impact is 58 m/s, that the initial horizontal component of the velocity before impact is negligible, and that the ball remained in contact with the racquet for 5.0 ms (milliseconds)?

Strategy

This problem involves only one dimension because the ball starts from having no horizontal velocity component before impact. Newton's second law stated in terms of momentum is then written as

$$F_{net} = \frac{\Delta p}{\Delta t}$$

As noted above, when mass is constant, the change in momentum is given by

$$\Delta p = m\Delta v = m(v_f - v_i).$$

In this example, the velocity just after impact and the change in time are given; thus, once Δp s calculated, $F_{net} = \frac{\Delta p}{\Delta t}$ can be used to find the force.

Solution

To determine the change in momentum, substitute the values for the initial and final velocities into the equation above.

$$\begin{aligned} \Delta p &= m(v_f - v_i) \\ &= (0.057\,kg)(58\,m/s - 0\,m/s) \\ &= 3.306\,kg\cdot m/s = 3.3\,kg\cdot m/s \end{aligned}$$

Now the magnitude of the net external force can determined by using $F_{net} = \frac{\Delta p}{\Delta t}$

$$\begin{aligned} F_{net} = \frac{\Delta p}{\Delta t} &= \frac{3.306\,kg}{5.0\times 10^{-3}} \\ &= 661\,N, \end{aligned}$$

where we have retained only two significant figures in the final step.

Discussion

This quantity was the average force exerted by Venus Williams’ racquet on the tennis ball during its brief impact (note that the ball also experienced the 0.56-N force of gravity, but that force was not due to the racquet). This problem could also be solved by first finding the acceleration and then using $F = ma$ but one additional step would be required compared with the strategy used in this example.

Summary

- Linear momentum (*momentum* for brevity) is defined as the product of a system’s mass multiplied by its velocity.
- In symbols, linear momentum p is defined to be

  $$p = mv$$

  where m is the mass of the system and v is its velocity.
- The SI unit for momentum is $kg\cdot m/s$.
- Newton’s second law of motion in terms of momentum states that the net external force equals the change in momentum of a system divided by the time over which it changes.
- In symbols, Newton’s second law of motion is defined to be

  $$F_{net} = \frac{\Delta p}{\Delta t}$$

  where F_{net} is the net external force, Δp s the change in momentum, and Δt is change in time.

Glossary

linear momentum

the product of mass and velocity

second law of motion

physical law that states that the net external force equals the change in momentum of a system divided by the time over which it changes

4.3: Impulse

Learning Objectives

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

- Define impulse.
- Describe effects of impulses in everyday life.
- Determine the average effective force using graphical representation.
- Calculate average force and impulse given mass, velocity, and time.

The effect of a force on an object depends on how long it acts, as well as how great the force is. In [link], a very large force acting for a short time had a great effect on the momentum of the tennis ball. A small force could cause the same change in momentum, but it would have to act for a much longer time. For example, if the ball were thrown upward, the gravitational force (which is much smaller than the tennis racquet's force) would eventually reverse the momentum of the ball. Quantitatively, the effect we are talking about is the change in momentum Δp.

By rearranging the equation $\Delta F_{net} = \frac{\Delta p}{\Delta t}$ to be

$$\Delta p = F_{net}\Delta t, \tag{4.3.1}$$

we can see how the change in momentum equals the average net external force multiplied by the time this force acts. The quantity $F_{net}\Delta t$ is given the name impulse. Impulse is the same as the change in momentum.

Impulse: Change in Momentum

Change in momentum equals the average net external force multiplied by the time this force acts.

$$\Delta p = F_{net}\Delta t \tag{4.3.2}$$

The quantity $F_{net}\Delta t$ is given the name impulse.

There are many ways in which an understanding of impulse can save lives, or at least limbs. The dashboard padding in a car, and certainly the airbags, allow the net force on the occupants in the car to act over a much longer time when there is a sudden stop. The momentum change is the same for an occupant, whether an air bag is deployed or not, but the force (to bring the occupant to a stop) will be much less if it acts over a larger time. Cars today have many plastic components. One advantage of plastics is their lighter weight, which results in better gas mileage. Another advantage is that a car will crumple in a collision, especially in the event of a head-on collision. A longer collision time means the force on the car will be less. Deaths during car races decreased dramatically when the rigid frames of racing cars were replaced with parts that could crumple or collapse in the event of an accident.

Bones in a body will fracture if the force on them is too large. If you jump onto the floor from a table, the force on your legs can be immense if you land stiff-legged on a hard surface. Rolling on the ground after jumping from the table, or landing with a parachute, extends the time over which the force (on you from the ground) acts.

MAKING CONNECTIONS: Take-Home Investigation—Hand Movement and Impulse

Try catching a ball while "giving" with the ball, pulling your hands toward your body. Then, try catching a ball while keeping your hands still. Hit water in a tub with your full palm. After the water has settled, hit the water again by diving your hand with your fingers first into the water. (Your full palm represents a swimmer doing a belly flop and your diving hand represents a swimmer doing a dive.) Explain what happens in each case and why. Which orientations would you advise people to avoid and why?

Summary

- Impulse, or change in momentum, equals the average net external force multiplied by the time this force acts:

$$\Delta p = F_{net}\Delta t. \tag{4.3.3}$$

- Forces are usually not constant over a period of time.

Glossary

change in momentum

the difference between the final and initial momentum; the mass times the change in velocity

impulse

the average net external force times the time it acts; equal to the change in momentum

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

4.4: Conservation of Momentum

Learning Objectives

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

- Describe the principle of conservation of momentum.
- Derive an expression for the conservation of momentum.
- Explain conservation of momentum with examples.
- Explain the principle of conservation of momentum as it relates to atomic and subatomic particles.

Momentum is an important quantity because it is conserved. Yet it was not conserved in the examples in Impulse and Linear Momentum and Force, where large changes in momentum were produced by forces acting on the system of interest. Under what circumstances is momentum conserved?

The answer to this question entails considering a sufficiently large system. It is always possible to find a larger system in which total momentum is constant, even if momentum changes for components of the system. If a football player runs into the goalpost in the end zone, there will be a force on him that causes him to bounce backward. However, the Earth also recoils —conserving momentum—because of the force applied to it through the goalpost. Because Earth is many orders of magnitude more massive than the player, its recoil is immeasurably small and can be neglected in any practical sense, but it is real nevertheless.

Consider what happens if the masses of two colliding objects are more similar than the masses of a football player and Earth—for example, one car bumping into another, as shown in Figure $4.4.1$. Both cars are coasting in the same direction when the lead car (labeled m_2 is bumped by the trailing car (labeled m_1. The only unbalanced force on each car is the force of the collision. (Assume that the effects due to friction are negligible.) Car 1 slows down as a result of the collision, losing some momentum, while car 2 speeds up and gains some momentum. We shall now show that the total momentum of the two-car system remains constant.

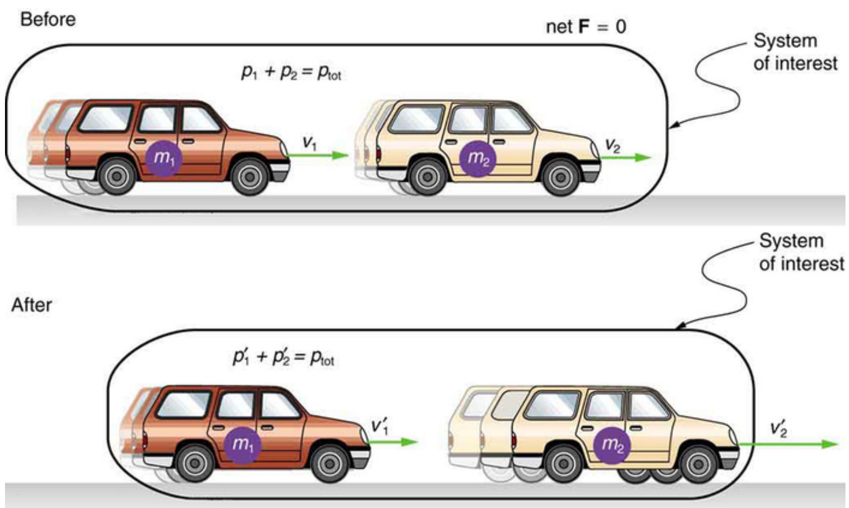


Figure $4.4.1$: A car of mass m_1 moving with a velocity of v_1 bumps into another car of mass m_2 and velocity v_2 that it is following. As a result, the first car slows down to a velocity of v'_1 and the second speeds up to a velocity of v'_2 The momentum of each car is changed, but the total momentum p_{tot} of the two cars is the same before and after the collision (if you assume friction is negligible).

Using the definition of impulse, the change in momentum of car 1 is given by

$$\Delta p_1 = F_1 \Delta t, \tag{4.4.1}$$

is the force on car 1 due to car 2, and Δt

where F_1 is the time the force acts (the duration of the collision). Intuitively, it seems obvious that the collision time is the same for both cars, but it is only true for objects traveling at ordinary speeds. This assumption must be modified for objects travelling near the speed of light, without affecting the result that momentum is conserved.

Similarly, the change in momentum of car 2 is

$$\Delta p_2 = F_2 \Delta t, \tag{4.4.2}$$

where F_2 is the force on car 2 due to car 1, and we assume the duration of the collision Δt is the same for both cars. We know from Newton's third law that $F_2 = -F_1$, and so

$$\Delta p_2 = -F_1 \Delta t = -\Delta p_1. \tag{4.4.3}$$

Thus, the changes in momentum are equal and opposite, and

$$\Delta p_1 + \Delta p_2 = 0. \tag{4.4.4}$$

Because the changes in momentum add to zero, the total momentum of the two-car system is constant. That is,

$$p_1 + p_2 = constant \tag{4.4.5}$$

$$p_1 + p_2 = p'_1 + p'_2, \tag{4.4.6}$$

where p'_1 and p'_2 are the momenta of cars 1 and 2 after the collision. (We often use primes to denote the final state.)

This result—that momentum is conserved—has validity far beyond the preceding one-dimensional case. It can be similarly shown that total momentum is conserved for any isolated system, with any number of objects in it. In equation form, the conservation of momentum principle for an isolated system is written

$$p_{tot} = constant, \tag{4.4.7}$$

or

$$p_{tot} = p_{tot}, \tag{4.4.8}$$

where p_{tot} is the total momentum (the sum of the momenta of the individual objects in the system) and p_{tot}, is the total momentum some time later. (The total momentum can be shown to be the momentum of the center of mass of the system.) An isolated system is defined to be one for which the net external force is zero $(F_{net} = 0)$.

Conservation of Momentum Principle

$$p_{tot} = constant \tag{4.4.9}$$

$$p_{tot} = p'_{tot} \; (isolated\ system) \tag{4.4.10}$$

Isolated System

An isolated system is defined to be one for which the net external force is zero $(F_{net} = 0)$.

Perhaps an easier way to see that momentum is conserved for an isolated system is to consider Newton's second law in terms of momentum, $F_{net} = \frac{\Delta p_{tot}}{\Delta t}$. For an isolated system, $(F_{net} = 0)$; thus $\Delta p_{tot} = 0$ and Δp is constant.

The conservation of momentum principle can be applied to systems as different as a comet striking Earth and a gas containing huge numbers of atoms and molecules. Conservation of momentum is violated only when the net external force is not zero. But another larger system can always be considered in which momentum is conserved by simply including the source of the external force. For example, in the collision of two cars considered above, the two-car system conserves momentum while each one-car system does not.

MAKING CONNECTIONS: TAKE-HOME Investigation—Drop of Tennis Ball and a Basketball

Hold a tennis ball side by side and in contact with a basketball. Drop the balls together. (Be careful!) What happens? Explain your observations. Now hold the tennis ball above and in contact with the basketball. What happened? Explain your observations. What do you think will happen if the basketball ball is held above and in contact with the tennis ball?

MAKING CONNECTIONS: TAKE-HOME Investigation—Two Tennis Balls in a Ballistic Trajectory

Tie two tennis balls together with a string about a foot long. Hold one ball and let the other hang down and throw it in a ballistic trajectory. Explain your observations. Now mark the center of the string with bright ink or attach a brightly colored sticker to it and throw again. What happened? Explain your observations.

Some aquatic animals such as jellyfish move around based on the principles of conservation of momentum. A jellyfish fills its umbrella section with water and then pushes the water out resulting in motion in the opposite direction to that of the jet of water. Squids propel themselves in a similar manner but, in contrast with jellyfish, are able to control the direction in which they move by aiming their nozzle forward or backward. Typical squids can move at speeds of 8 to 12 km/h.

The ballistocardiograph (BCG) was a diagnostic tool used in the second half of the 20th century to study the strength of the heart. About once a second, your heart beats, forcing blood into the aorta. A force in the opposite direction is exerted on the rest of your body (recall Newton's third law). A ballistocardiograph is a device that can measure this reaction force. This measurement is done by using a sensor (resting on the person) or by using a moving table suspended from the ceiling. This technique can gather information on the strength of the heart beat and the volume of blood passing from the heart. However, the electrocardiogram (ECG or EKG) and the echocardiogram (cardiac ECHO or ECHO; a technique that uses ultrasound to see an image of the heart) are more widely used in the practice of cardiology.

Making Connections: Conservation of Momentum and Collision

Conservation of momentum is quite useful in describing collisions. Momentum is crucial to our understanding of atomic and subatomic particles because much of what we know about these particles comes from collision experiments.

Subatomic Collisions and Momentum

The conservation of momentum principle not only applies to the macroscopic objects, it is also essential to our explorations of atomic and subatomic particles. Giant machines hurl subatomic particles at one another, and researchers evaluate the results by assuming conservation of momentum (among other things).

On the small scale, we find that particles and their properties are invisible to the naked eye but can be measured with our instruments, and models of these subatomic particles can be constructed to describe the results. Momentum is found to be a property of all subatomic particles including massless particles such as photons that compose light. Momentum being a property of particles hints that momentum may have an identity beyond the description of an object's mass multiplied by the object's velocity. Indeed, momentum relates to wave properties and plays a fundamental role in what measurements are taken and how we take these measurements. Furthermore, we find that the conservation of momentum principle is valid when considering systems of particles. We use this principle to analyze the masses and other properties of previously undetected particles, such as the nucleus of an atom and the existence of quarks that make up particles of nuclei. Figure $4.4.3$ below illustrates how a particle scattering backward from another implies that its target is massive and dense. Experiments seeking evidence that quarks make up protons (one type of particle that makes up nuclei) scattered high-energy electrons off of protons (nuclei of hydrogen atoms). Electrons occasionally scattered straight backward in a manner that implied a very small and very dense particle makes up the proton—this observation is considered nearly direct evidence of quarks. The analysis was based partly on the same conservation of momentum principle that works so well on the large scale.

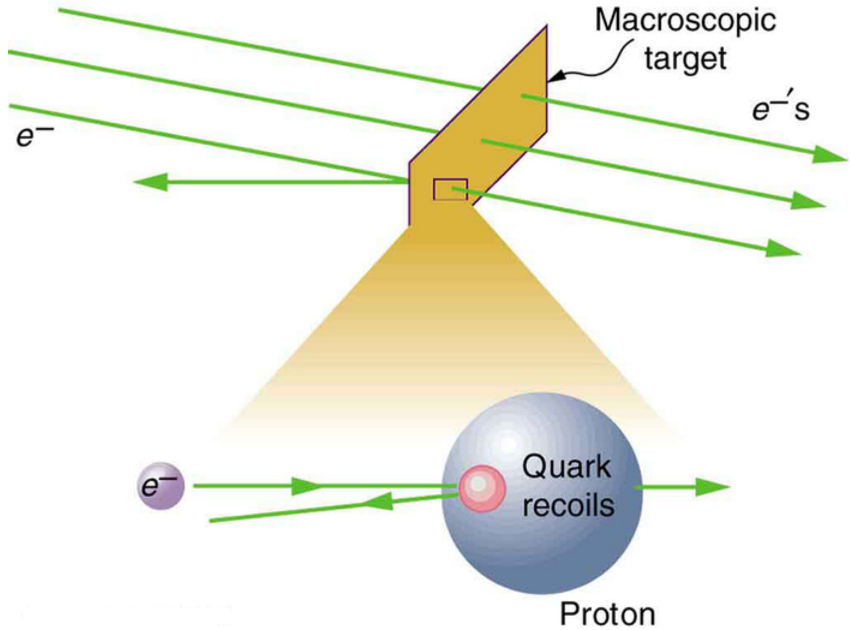


Figure 4.4.3: A subatomic particle scatters straight backward from a target particle. In experiments seeking evidence for quarks, electrons were observed to occasionally scatter straight backward from a proton.

Summary

- The conservation of momentum principle is written

$$p_{tot} = constant \tag{4.4.11}$$

or

$$p_{tot} = p'_{tot} \; (isolated\; system), \tag{4.4.12}$$

- p_{tot} is the initial total momentum and p'_{tot} is the total momentum some time later. An isolated system is defined to be one for which the net external force is zero $(F_{net} = 0$
- During projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero.
- Conservation of momentum applies only when the net external force is zero.
- The conservation of momentum principle is valid when considering systems of particles.

Glossary

conservation of momentum principle

when the net external force is zero, the total momentum of the system is conserved or constant

isolated system

a system in which the net external force is zero

quark

fundamental constituent of matter and an elementary particle

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

- **8.3: Conservation of Momentum** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

4.5: Elastic Collisions in One Dimension

Learning Objectives

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

- Describe an elastic collision of two objects in one dimension.
- Define internal kinetic energy.
- Derive an expression for conservation of internal kinetic energy in a one dimensional collision.
- Determine the final velocities in an elastic collision given masses and initial velocities.

Let us consider various types of two-object collisions. These collisions are the easiest to analyze, and they illustrate many of the physical principles involved in collisions. The conservation of momentum principle is very useful here, and it can be used whenever the net external force on a system is zero.

We start with the elastic collision of two objects moving along the same line—a one-dimensional problem. An elastic collision is one that also conserves internal kinetic energy. Internal kinetic energy is the sum of the kinetic energies of the objects in the system. Figure illustrates an elastic collision in which internal kinetic energy and momentum are conserved.

Truly elastic collisions can only be achieved with subatomic particles, such as electrons striking nuclei. Macroscopic collisions can be very nearly, but not quite, elastic—some kinetic energy is always converted into other forms of energy such as heat transfer due to friction and sound. One macroscopic collision that is nearly elastic is that of two steel blocks on ice. Another nearly elastic collision is that between two carts with spring bumpers on an air track. Icy surfaces and air tracks are nearly frictionless, more readily allowing nearly elastic collisions on them.

Elastic Collision

An elastic collision is one that conserves internal kinetic energy.

Internal Kinetic Energy

Internal kinetic energy is the sum of the kinetic energies of the objects in the system.

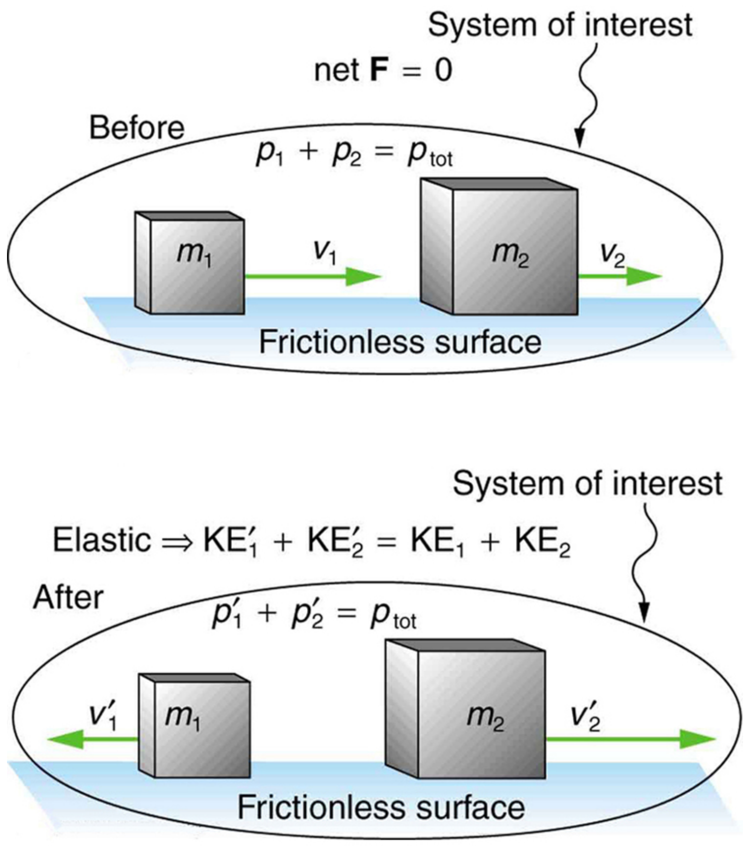


Figure 4.5.1: An elastic one-dimensional two-object collision. Momentum and internal kinetic energy are conserved.

Now, to solve problems involving one-dimensional elastic collisions between two objects we can use the equations for conservation of momentum and conservation of internal kinetic energy. First, the equation for conservation of momentum for two objects in a one-dimensional collision is

$$p_1 + p_2 = p'_1 + p'_2 \; (F_{net} = 0) \tag{4.5.1}$$

or

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m'_2 v'_2 \; (F_{net} = 0), \tag{4.5.2}$$

where the primes (') indicate values after the collision. By definition, an elastic collision conserves internal kinetic energy, and so the sum of kinetic energies before the collision equals the sum after the collision. Thus,

$$\frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 = \frac{1}{2} m_1 v_1'^2 + \frac{1}{2} m_2 v_2'^2 \tag{4.5.3}$$

expresses the equation for conservation of internal kinetic energy in a one-dimensional collision.

Example 4.5.1: Calculating Velocities Following an Elastic Collision

Calculate the velocities of two objects following an elastic collision, given that

$$m_1 = 0.500\,kg,\; m_2 = 3.50\,kg,\; v_1 = 4.00\,m/s,\; and\, v_2 = 0, \tag{4.5.4}$$

Strategy and Concept

First, visualize what the initial conditions mean—a small object strikes a larger object that is initially at rest. This situation is slightly simpler than the situation shown in Figure where both objects are initially moving. We are asked to find two unknowns (the final velocities v'_1 and v'_2). To find two unknowns, we must use two independent equations. Because this collision is elastic, we can use the above two equations. Both can be simplified by the fact that object 2 is initially at rest, and thus $v_2 = 0$. Once we simplify these equations, we combine them algebraically to solve for the unknowns.

Solution

For this problem, note that $v_2 = 0$ and use conservation of momentum. Thus,

$$p_1 = p'_1 + p'_2 \tag{4.5.5}$$

or

$$m_1 v_1 = m_1 v'_1 + m_2 v'_2. \tag{4.5.6}$$

Using conservation of internal kinetic energy and that $v_2 = 0$,

$$\frac{1}{2} m_1 v_1^2 = \frac{1}{2} m v_1 v_1'^2 + \frac{1}{2} m v_2 v_2'^2. \tag{4.5.7}$$

Solving the first equation (momentum equation) for v'_2, we obtain

$$v'_2 = \frac{m_1}{m_2}(v_1 - v'_1). \tag{4.5.8}$$

Substituting this expression into the second equation (internal kinetic energy equation) eliminates the variable v'_2, leaving only v'_1 as an unknown (the algebra is left as an exercise for the reader). There are two solutions to any quadratic equation; in this example, they are as an unknown (the algebra is left as an exercise for the reader). There are two solutions to any quadratic equation; in this example, they are

$$v'_1 = 4.00\, m/s \tag{4.5.9}$$

and

$$v'_1 = -3.00\, m/s. \tag{4.5.10}$$

As noted when quadratic equations were encountered in earlier chapters, both solutions may or may not be meaningful. In this case, the first solution is the same as the initial condition. The first solution thus represents the situation before the collision and is discarded. The second solution $(v'_1 = -3.00\, m/s)$ is negative, meaning that the first object bounces backward. When this negative value of v'_1 is used to find the velocity of the second object after the collision, we get

$$v'_2 = \frac{m_1}{m_2}(v_1 - v'_1) = \frac{0.500\, kg}{3.50\, kg}[4.00 - (-3.00)]\, m/s \tag{4.5.11}$$

or

$$v'_2 = 1.00\, m/s. \tag{4.5.12}$$

Discussion

The result of this example is intuitively reasonable. A small object strikes a larger one at rest and bounces backward. The larger one is knocked forward, but with a low speed. (This is like a compact car bouncing backward off a full-size SUV that is initially at rest.) As a check, try calculating the internal kinetic energy before and after the collision. You will see that the internal kinetic energy is unchanged at 4.00 J. Also check the total momentum before and after the collision; you will find it, too, is unchanged.

The equations for conservation of momentum and internal kinetic energy as written above can be used to describe any one-dimensional elastic collision of two objects. These equations can be extended to more objects if needed.′ $4.00\, \mathrm{m/s}$

Making Connections: Take-Home Investigation—Ice Cubes and Elastic Collision

Find a few ice cubes which are about the same size and a smooth kitchen tabletop or a table with a glass top. Place the ice cubes on the surface several centimeters away from each other. Flick one ice cube toward a stationary ice cube and observe the path and velocities of the ice cubes after the collision. Try to avoid edge-on collisions and collisions with rotating ice cubes. Have you created approximately elastic collisions? Explain the speeds and directions of the ice cubes using momentum.

> **PHET EXPLORATIONS: COLLISIONS LAB**
>
> Investigate collisions on an air hockey table. Set up your own experiments: vary the number of discs, masses and initial conditions. Is momentum conserved? Is kinetic energy conserved? Vary the elasticity and see what happens.
>
> 
>
>
> Figure 4.5.2: Collision Lab

Summary

- An elastic collision is one that conserves internal kinetic energy.
- Conservation of kinetic energy and momentum together allow the final velocities to be calculated in terms of initial velocities and masses in one dimensional two-body collisions.

Glossary

elastic collision

a collision that also conserves internal kinetic energy

internal kinetic energy

the sum of the kinetic energies of the objects in a system

4.6: Inelastic Collisions in One Dimension

Learning Objectives

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

- Define inelastic collision.
- Explain perfectly inelastic collision.
- Apply an understanding of collisions to sports.
- Determine recoil velocity and loss in kinetic energy given mass and initial velocity.

We have seen that in an elastic collision, internal kinetic energy is conserved. An inelastic collision is one in which the internal kinetic energy changes (it is not conserved). This lack of conservation means that the forces between colliding objects may remove or add internal kinetic energy. Work done by internal forces may change the forms of energy within a system. For inelastic collisions, such as when colliding objects stick together, this internal work may transform some internal kinetic energy into heat transfer. Or it may convert stored energy into internal kinetic energy, such as when exploding bolts separate a satellite from its launch vehicle.

Definition: Inelastic Collisions

An inelastic collision is one in which the internal kinetic energy changes (it is not conserved).

Figure $4.6.1$ shows an example of an inelastic collision. Two objects that have equal masses head toward one another at equal speeds and then stick together. Their total internal kinetic energy is initially

$$\frac{1}{2}mv^2 + \frac{1}{2}mv^2 = mv^2. \tag{4.6.1}$$

The two objects come to rest after sticking together, conserving momentum. But the internal kinetic energy is zero after the collision. A collision in which the objects stick together is sometimes called a perfectly inelastic collision because it reduces internal kinetic energy more than does any other type of inelastic collision. In fact, such a collision reduces internal kinetic energy to the minimum it can have while still conserving momentum.

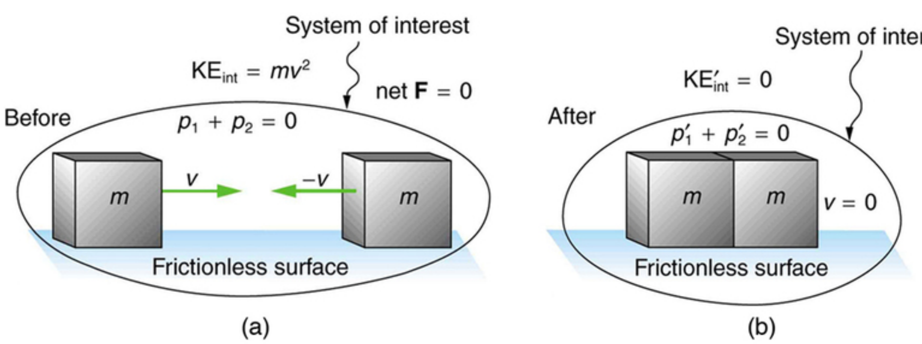


Figure $4.6.1$: An inelastic one-dimensional two-object collision. Momentum is conserved, but internal kinetic energy is not conserved. (a) Two objects of equal mass initially head directly toward one another at the same speed. (b) The objects stick together (a perfectly inelastic collision), and so their final velocity is zero. The internal kinetic energy of the system changes in any inelastic collision and is reduced to zero in this example.

Definition: Perfectly Inelastic Collisions

A collision in which the objects stick together is sometimes called “perfectly inelastic.”

Example $4.6.1$: Calculating Velocity and Change in Kinetic Energy - Inelastic Collision of a Puck and a Goalie

a. Find the recoil velocity of a 70.0-kg ice hockey goalie, originally at rest, who catches a 0.150-kg hockey puck slapped at him at a velocity of 35.0 m/s.
b. How much kinetic energy is lost during the collision? Assume friction between the ice and the puck-goalie system is negligible (Figure $4.6.2$)

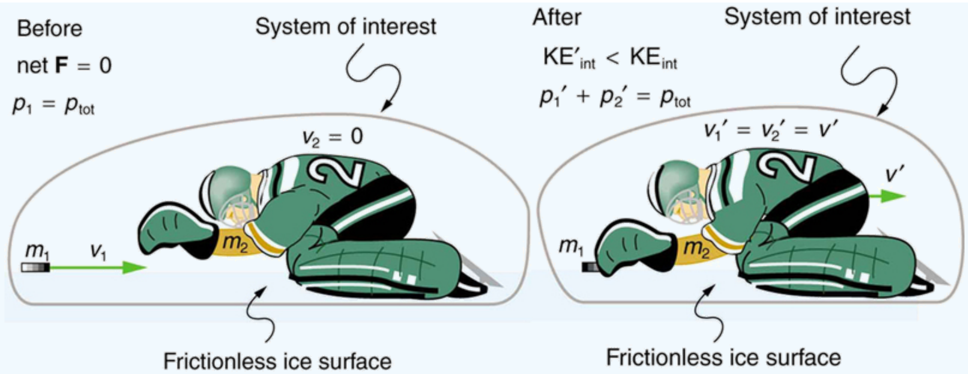


Figure 4.6.2: An ice hockey goalie catches a hockey puck and recoils backward. The initial kinetic energy of the puck is almost entirely converted to thermal energy and sound in this inelastic collision.

Strategy

Momentum is conserved because the net external force on the puck-goalie system is zero. We can thus use conservation of momentum to find the final velocity of the puck and goalie system. Note that the initial velocity of the goalie is zero and that the final velocity of the puck and goalie are the same. Once the final velocity is found, the kinetic energies can be calculated before and after the collision and compared as requested.

Solution for (a)

Momentum is conserved because the net external force on the puck-goalie system is zero.

Conservation of momentum is

$$p_1 + p_2 = p'_1 + p'_2$$

or

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m_2 v'_2.$$

Because the goalie is initially at rest, we know $v_2 = 0$. Because the goalie catches the puck, the final velocities are equal, or $v'_1 = v'_2 = v'$. Thus, the conservation of momentum equation simplifies to

$$m_1 v_1 = (m_1 + m_2) v'.$$

Solving for v' yields

$$v' = \frac{m_1}{m_1 + m_2} v_1.$$

Entering known values in this equation, we get

$$\begin{aligned} v' &= \left(\frac{0.150\, kg}{70.0\, kg + 0.150\, kg}\right)(35.0\, m/s) \\ &= 7.48 \times 10^{-2} m/s)^2 \\ &= 0.196\, J. \end{aligned}$$

The change in internal kinetic energy is thus

$$\begin{aligned} KE'_{int} - KE_{int} &= 0.196\, J - 91.9\, J \\ &= -91.7\, J \end{aligned}$$

where the minus sign indicates that the energy was lost.

Discussion for (b)

Nearly all of the initial internal kinetic energy is lost in this perfectly inelastic collision. KE_{int} is mostly converted to thermal energy and sound.

During some collisions, the objects do not stick together and less of the internal kinetic energy is removed—such as happens in most automobile accidents. Alternatively, stored energy may be converted into internal kinetic energy during a collision. Figure $4.6.3$ shows a one-dimensional example in which two carts on an air track collide, releasing potential energy from a compressed spring. Example $4.6.2$ deals with data from such a collision.

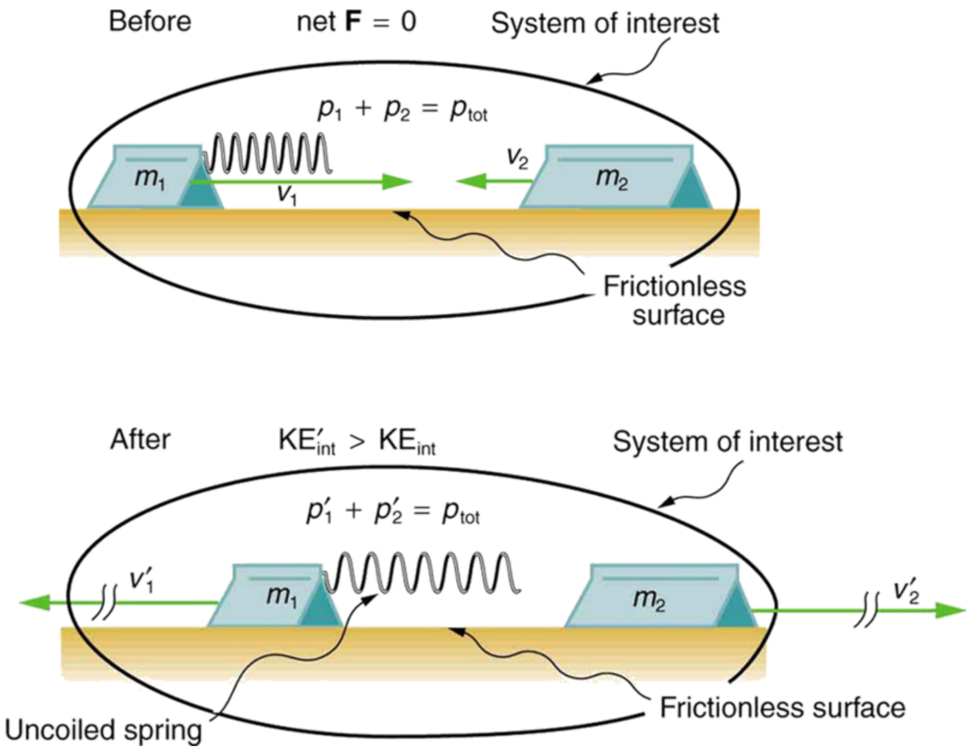


Figure $4.6.3$: An air track is nearly frictionless, so that momentum is conserved. Motion is one-dimensional. In this collision, examined in Example $4.6.2$, the potential energy of a compressed spring is released during the collision and is converted to internal kinetic energy.

Collisions are particularly important in sports and the sporting and leisure industry utilizes elastic and inelastic collisions. Let us look briefly at tennis. Recall that in a collision, it is momentum and not force that is important. So, a heavier tennis racquet will have the advantage over a lighter one. This conclusion also holds true for other sports—a lightweight bat (such as a softball bat) cannot hit a hardball very far.

The location of the impact of the tennis ball on the racquet is also important, as is the part of the stroke during which the impact occurs. A smooth motion results in the maximizing of the velocity of the ball after impact and reduces sports injuries such as tennis elbow. A tennis player tries to hit the ball on the "sweet spot" on the racquet, where the vibration and impact are minimized and the ball is able to be given more velocity. Sports science and technologies also use physics concepts such as momentum and rotational motion and vibrations.

Take-Home Experiment—Bouncing of Tennis Ball

1. Find a racquet (a tennis, badminton, or other racquet will do). Place the racquet on the floor and stand on the handle. Drop a tennis ball on the strings from a measured height. Measure how high the ball bounces. Now ask a friend to hold the racquet firmly by the handle and drop a tennis ball from the same measured height above the racquet. Measure how high the ball bounces and observe what happens to your friend's hand during the collision. Explain your observations and measurements.
2. The coefficient of restitution (c) is a measure of the elasticity of a collision between a ball and an object, and is defined as the ratio of the speeds after and before the collision. A perfectly elastic collision has a c of 1. For a ball bouncing off the floor (or a racquet on the floor), c can be shown to be $c=(h/H)^{1/2}$ where h is the height to which the ball bounces and H is the height from which the ball is dropped. Determine c for the cases in Part 1 and for the case of a tennis ball bouncing off a concrete or wooden floor ($c=0.85$ for new tennis balls used on a tennis court).

Example 4.6.2: Calculating Final Velocity and Energy Release - Two Carts Collide

In the collision pictured in Figure 4.6.3, two carts collide inelastically. Cart 1 (denoted m_1 carries a spring which is initially compressed. During the collision, the spring releases its potential energy and converts it to internal kinetic energy. The mass of cart 1 and the spring is 0.350 kg, and the cart and the spring together have an initial velocity of $-0.500\,m/s$. After the collision, cart 1 is observed to recoil with a velocity of $-4.00\,m/s$.

a. What is the final velocity of cart 2?
b. How much energy was released by the spring (assuming all of it was converted into internal kinetic energy)?

Strategy

We can use conservation of momentum to find the final velocity of cart 2, because $F_{net} = 0$ (the track is frictionless and the force of the spring is internal). Once this velocity is determined, we can compare the internal kinetic energy before and after the collision to see how much energy was released by the spring.

Solution for (a)

As before, the equation for conservation of momentum in a two-object system is

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m_2 v'_2.$$

The only unknown in this equation is v'_2. Solving for v'_2 and substituting known values into the previous equation fields

$$\begin{aligned} v'_2 &= \frac{m_1 v_1 + m_2 v_2 - m_1 v'_1}{m_2} \\ &= \frac{0.350\,kg)(2.00\,m/s) + (0.500\,kg)(-0.500\,m/s)}{0.500\,kg} - \frac{(0.350\,kg)(-4.00\,m/s)}{0.500\,kg} \\ &= 3.70\,m/s. \end{aligned}$$

Solution for (b)

The internal kinetic energy before the collision is

$$\begin{aligned} KE_{int} &= \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 \\ &= \frac{1}{2}(0.350\,kg)(2.00\,m/s)^2 + \frac{1}{2}(0.500\,kg)(-0.500\,m/s)^2 \\ &= 0.763\,J. \end{aligned}$$

After the collision, the internal kinetic energy is

$$\begin{aligned} KE'_{int} &= \frac{1}{2} m_1 v_1'^2 + \frac{1}{2} m_2 v_2'^2 \\ &= \frac{1}{2}(0.350\,kg)(-4.00\,m/s)^2 + \frac{1}{2}(0.500\,kg)(0.370\,m/s)^2 \\ &= 6.22\,J. \end{aligned}$$

The change in internal kinetic energy is thus

$$\begin{aligned} KE' - KE &= 6.22\,J - 0.763\,J \\ &= 5.46\,J. \end{aligned}$$

Discussion

The final velocity of cart 2 is large and positive, meaning that it is moving to the right after the collision. The internal kinetic energy in this collision increases by 5.46 J. That energy was released by the spring.

Summary

- An inelastic collision is one in which the internal kinetic energy changes (it is not conserved).
- A collision in which the objects stick together is sometimes called perfectly inelastic because it reduces internal kinetic energy more than does any other type of inelastic collision.
- Sports science and technologies also use physics concepts such as momentum and rotational motion and vibrations.

Glossary

inelastic collision
a collision in which internal kinetic energy is not conserved

perfectly inelastic collision
a collision in which the colliding objects stick together

CHAPTER OVERVIEW

5: Torque and Angular Momentum

How can we guarantee that a body is in equilibrium and what can we learn from systems that are in equilibrium? There are actually two conditions that must be satisfied to achieve equilibrium. These conditions are the topics of the first two sections of this chapter.

Thumbnails: Relationship between force (F), torque(τ), momentum (p), and angular momentum (L) vectors in a rotating system. (r) is the radius. (Public domain; Yawe).

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

5.1: Prelude to Statics and Torque

What might desks, bridges, buildings, trees, and mountains have in common—at least in the eyes of a physicist? The answer is that they are ordinarily motionless relative to the Earth. Furthermore, their acceleration is zero because they remain motionless. That means they also have something in common with a car moving at a constant velocity, because anything with a constant velocity also has an acceleration of zero. Now, the important part—Newton's second law states that net $F = ma$, and so the net external force is zero for all stationary objects and for all objects moving at constant velocity. There are forces acting, but they are balanced. That is, they are in *equilibrium*.

Figure 5.1.1: On a short time scale, rocks like these in Australia's Kings Canyon are static, or motionless relative to the Earth. (credit: freeaussiestock.com)

Statics

Statics is the study of forces in equilibrium, a large group of situations that makes up a special case of Newton's second law. We have already considered a few such situations; in this chapter, we cover the topic more thoroughly, including consideration of such possible effects as the rotation and deformation of an object by the forces acting on it.

How can we guarantee that a body is in equilibrium and what can we learn from systems that are in equilibrium? There are actually two conditions that must be satisfied to achieve equilibrium. These conditions are the topics of the first two sections of this chapter.

5.2: The First Condition for Equilibrium

Learning Objectives

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

- State the first condition of equilibrium.
- Explain static equilibrium.
- Explain dynamic equilibrium.

The first condition necessary to achieve equilibrium is the one already mentioned: the net external force on the system must be zero. Expressed as an equation, this is simply

$$net\ F=0 \tag{5.2.1}$$

Note that if net F is zero, then the net external force in *any* direction is zero. For example, the net external forces along the typical *x*- and *y*-axes are zero. This is written as

$$net\ F_x\ and\ F_y=0 \tag{5.2.2}$$

Figuress $5.2.1$ and $5.2.2$ illustrate situations where $net\ F=0$ for both static equilibrium (motionless), and dynamic equilibrium (constant velocity).

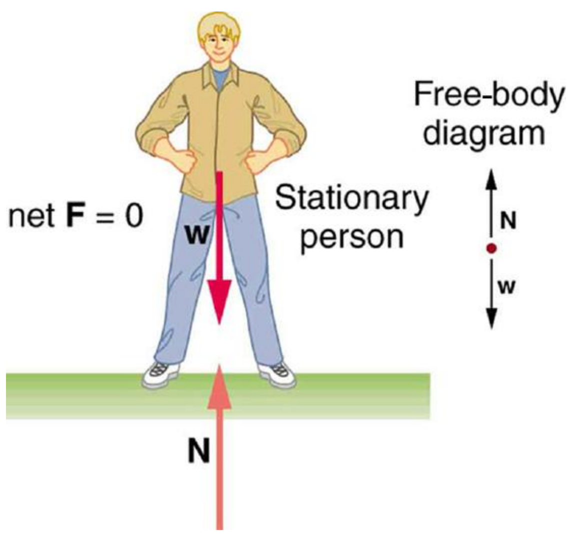


Figure $5.2.1$: This motionless person is in static equilibrium. The forces acting on him add up to zero. Both forces are vertical in this case.

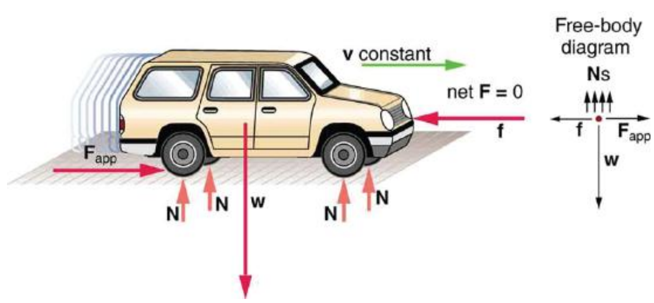


Figure $5.2.2$: This car is in dynamic equilibrium because it is moving at constant velocity. There are horizontal and vertical forces, but the net external force in any direction is zero. The applied force F_{app} between the tires and the road is balanced by air friction, and the weight of the car is supported by the normal forces, here shown to be equal for all four tires.

However, it is not sufficient for the net external force of a system to be zero for a system to be in equilibrium. Consider the two situations illustrated in Figures $5.2.3$ and $5.2.4$ where forces are applied to an ice hockey stick lying flat on ice. The net external force is zero in both situations shown in the figure; but in one case, equilibrium is achieved, whereas in the other, it is not. In Figure $5.2.3$, the ice hockey stick remains motionless. But in Figure $5.2.4$, with the same forces applied in different places, the stick experiences accelerated rotation. Therefore, we know that the point at which a force is applied is another factor in determining whether or not equilibrium is achieved. This will be explored further in the next section.

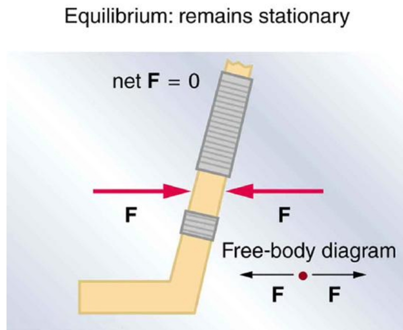


Figure 5.2.3: An ice hockey stick lying flat on ice with two equal and opposite horizontal forces applied to it. Friction is negligible, and the gravitational force is balanced by the support of the ice (a normal force). Thus, $net\, F = 0$. Equilibrium is achieved, which is static equilibrium in this case.

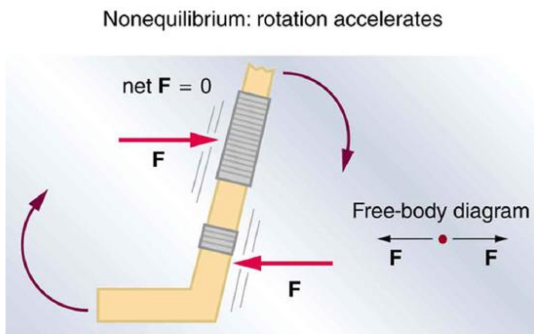


Figure 5.2.4:The same forces are applied at other points and the stick rotates—in fact, it experiences an accelerated rotation. Here $net\, F = 0$ but the system is not at equilibrium. Hence, the $net\, F = 0$ is a necessary—but not sufficient—condition for achieving equilibrium.

PhET Explorations: Torque

Investigate how torque causes an object to rotate. Discover the relationships between angular acceleration, moment of inertia, angular momentum and torque.

Summary

- Statics is the study of forces in equilibrium.
- Two conditions must be met to achieve equilibrium, which is defined to be motion without linear or rotational acceleration.
- The first condition necessary to achieve equilibrium is that the net external force on the system must be zero, so that $F = 0$.

Glossary

static equilibrium

a state of equilibrium in which the net external force and torque acting on a system is zero

dynamic equilibrium

a state of equilibrium in which the net external force and torque on a system moving with constant velocity are zero

5.3: The Second Condition for Equilibrium

Learning Objectives

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

- State the second condition that is necessary to achieve equilibrium.
- Explain torque and the factors on which it depends.
- Describe the role of torque in rotational mechanics.

Definition: Torque

The second condition necessary to achieve equilibrium involves avoiding accelerated rotation (maintaining a constant angular velocity. A rotating body or system can be in equilibrium if its rate of rotation is constant and remains unchanged by the forces acting on it. To understand what factors affect rotation, let us think about what happens when you open an ordinary door by rotating it on its hinges.

Several familiar factors determine how effective you are in opening the door (Figure $5.3.1$). First of all, the larger the force, the more effective it is in opening the door—obviously, the harder you push, the more rapidly the door opens. Also, the point at which you push is crucial. If you apply your force too close to the hinges, the door will open slowly, if at all. Most people have been embarrassed by making this mistake and bumping up against a door when it did not open as quickly as expected. Finally, the direction in which you push is also important. The most effective direction is perpendicular to the door—we push in this direction almost instinctively.

The magnitude, direction, and point of application of the force are incorporated into the definition of the physical quantity called torque. Torque is the rotational equivalent of a force. It is a measure of the effectiveness of a force in changing or accelerating a rotation (changing the angular velocity over a period of time). For forces applied strictly at right angles, we can express the magnitude of the torque in equation form as

$$\tau = r_{\perp} F \tag{5.3.1}$$

where τ (the Greek letter tau) is the symbol for torque, $r_{\perp}$ is the perpendicular lever arm, and F is the magnitude of the force.

The perpendicular lever arm $r_{\perp}$ is the shortest distance from the pivot point to the line along which F acts. Note that the line segment that defines the distance $r_{\perp}$ is perpendicular to F, as its name implies.

The SI unit of torque is newtons times meters, usually written as N·m. For example, if you push perpendicular to the door with a force of 40 N at a distance of 0.800 m from the hinges, you exert a torque of 32 N·m (0.800 m $\times$ 40 N) relative to the hinges. If you reduce the force to 20 N, the torque is reduced to 16 N·m, and so on.

The torque is always calculated with reference to some chosen pivot point. For the same applied force, a different choice for the location of the pivot will give you a different value for the torque. Any point in any object can be chosen to calculate the torque about that point. The object may not actually pivot about the chosen "pivot point."

Note that for rotation in a plane, torque has two possible directions. Torque is either clockwise or counterclockwise relative to the chosen pivot point.

Now, *the second condition necessary to achieve equilibrium* is that *the net external torque on a system must be zero*. An external torque is one that is created by an external force. You can choose the point around which the torque is calculated. The point can be the physical pivot point of a system or any other point in space—but it must be the same point for all torques. If the second condition (net external torque on a system is zero) is satisfied for one choice of pivot point, it will also hold true for any other choice of pivot point in or out of the system of interest. The second condition necessary to achieve equilibrium is stated in equation form as

$$net\ \tau = 0 \tag{5.3.2}$$

where net means total. Torques which are in opposite directions are assigned opposite signs. A common convention is to call counterclockwise (ccw) torques positive and clockwise (cw) torques negative.

When two children balance a seesaw as shown in Figure $5.3.3$, they satisfy the two conditions for equilibrium. Most people have perfect intuition about seesaws, knowing that the lighter child must sit farther from the pivot and that a heavier child can keep a lighter one off the ground indefinitely.

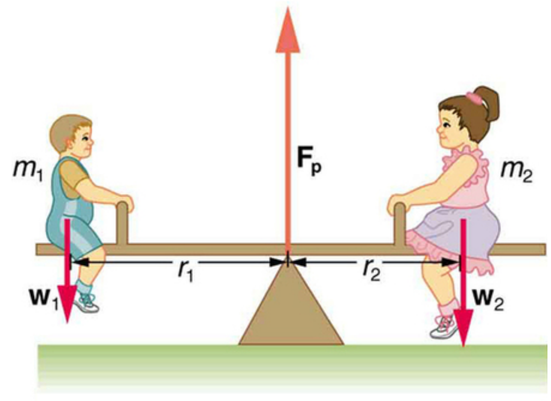


Figure $5.3.3$: Two children balancing a seesaw satisfy both conditions for equilibrium. The lighter child sits farther from the pivot to create a torque equal in magnitude to that of the heavier child.

Example $5.3.1$: She Saw Torques On A Seesaw

The two children shown in Figure $5.3.3$ are balanced on a seesaw of negligible mass. (This assumption is made to keep the example simple—more involved examples will follow.) The first child has a mass of 26.0 kg and sits 1.60 m from the pivot.

a. If the second child has a mass of 32.0 kg, how far is she from the pivot?
b. What is F_p, the supporting force exerted by the pivot?

Strategy

Both conditions for equilibrium must be satisfied. In part (a), we are asked for a distance; thus, the second condition (regarding torques) must be used, since the first (regarding only forces) has no distances in it. To apply the second condition for equilibrium, we first identify the system of interest to be the seesaw plus the two children. We take the supporting pivot to be the point about which the torques are calculated. We then identify all external forces acting on the system.

Solution (a)

The three external forces acting on the system are the weights of the two children and the supporting force of the pivot. Let us examine the torque produced by each. Torque is defined to be

$$\tau = r_\perp F$$

The torques exerted by the three forces are first,

$$\tau_1 = r_1 w_1$$

second,

$$\tau_2 = -r_2 w_2$$

and third,

$$\begin{aligned} \tau_p &= r_p F_p \\ &= 0 \cdot F_p \\ &= 0. \end{aligned}$$

Note that a minus sign has been inserted into the second equation because this torque is clockwise and is therefore negative by convention. Since F_p acts directly on the pivot point, the distance r_p is zero. A force acting on the pivot cannot cause a rotation, just as pushing directly on the hinges of a door will not cause it to rotate. Now, the second condition for equilibrium is that the sum of the torques on both children is zero. Therefore

$$\tau_2 = -\tau_1,$$

or

$$r_2 w_2 = r_1 w_1.$$

Weight is mass times the acceleration due to gravity. Entering mg for w, we get

$$r_2 m_2 g = r_1 w_1 g.$$

Solve this for the unknown r_2:

$$r_2 = r_1 \frac{m_1}{m_2}.$$

The quantities on the right side of the equation are known; thus, r_2 is

$$\begin{aligned} r_2 &= (1.60\, m) \frac{26.0\, kg}{32.0\, kg} \\ &= 1.30\, m \end{aligned}$$

As expected, the heavier child must sit closer to the pivot (1.30 m versus 1.60 m) to balance the seesaw.

Solution (b)

This part asks for a force F_p. The easiest way to find it is to use the first condition for equilibrium, which is

$$net\ F = 0.$$

The forces are all vertical, so that we are dealing with a one-dimensional problem along the vertical axis; hence, the condition can be written as

$$net\ F_y = 0$$

where we again call the vertical axis the *y*-axis. Choosing upward to be the positive direction, and using plus and minus signs to indicate the directions of the forces, we see that

$$F_p - w_1 - w_2 = 0.$$

This equation yields what might have been guessed at the beginning:

$$F_p = w_1 + w_2.$$

So, the pivot supplies a supporting force equal to the total weight of the system:

$$F_p = m_1 g + m_2 g.$$

Entering known values gives

$$\begin{aligned} F_p &= (26.0\, kg)(9.80\, m/s^2) + (32.0\, kg)(9.80\, m/s^2) \\ &= 568\, N. \end{aligned}$$

Discussion

The two results make intuitive sense. The heavier child sits closer to the pivot. The pivot supports the weight of the two children. Part (b) can also be solved using the second condition for equilibrium, since both distances are known, but only if the pivot point is chosen to be somewhere other than the location of the seesaw's actual pivot!

Several aspects of the preceding example have broad implications. First, the choice of the pivot as the point around which torques are calculated simplified the problem. Since F_p is exerted on the pivot point, its lever arm is zero. Hence, the torque exerted by the supporting force F_p is zero relative to that pivot point. The second condition for equilibrium holds for any choice of pivot point, and so we choose the pivot point to simplify the solution of the problem.

Second, the acceleration due to gravity canceled in this problem, and we were left with a ratio of masses. *This will not always be the case*. Always enter the correct forces—do not jump ahead to enter some ratio of masses.

Third, the weight of each child is distributed over an area of the seesaw, yet we treated the weights as if each force were exerted at a single point. This is not an approximation—the distances $r_{\perp}$ and r_2 are the distances to points directly below the center of gravity of each child. As we shall see in the next section, the mass and weight of a system can act as if they are located at a single point.

Finally, note that the concept of torque has an importance beyond static equilibrium. *Torque plays the same role in rotational motion that force plays in linear motion.* We will examine this in the next chapter.

Take-Home Experiment

- Take a piece of modeling clay and put it on a table, then mash a cylinder down into it so that a ruler can balance on the round side of the cylinder while everything remains still. Put a penny 8 cm away from the pivot. Where would you need to put two pennies to balance? Three pennies?

Summary

- The second condition assures those torques are also balanced. Torque is the rotational equivalent of a force in producing a rotation and is defined to be

$$\tau = r_{\perp} F$$

  where τ is torque, $r_{\perp}$ is the perpendicular distance from the pivot point to the line containing the point where the force is applied, and F is magnitude of the force.
- The perpendicular lever arm $r_{\perp}$ is the shortest distance from the pivot point to the line along which F acts. The SI unit for torque is the newton-meter N·m. The second condition necessary to achieve equilibrium is that the net external torque on a system must be zero:

$$net\,\tau = 0$$

  By convention, counterclockwise torques are positive, and clockwise torques are negative.

Glossary

torque
: turning or twisting effectiveness of a force

perpendicular lever arm
: the shortest distance from the pivot point to the line along which F lies

SI units of torque
: newton times meters, usually written as N·m

center of gravity
: the point where the total weight of the body is assumed to be concentrated

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

5.4: Simple Machines

Learning Objectives

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

- Describe different simple machines.
- Calculate the mechanical advantage.

Simple machines are devices that can be used to multiply or augment a force that we apply – often at the expense of a distance through which we apply the force. The word for "machine" comes from the Greek word meaning "to help make things easier." Levers, gears, pulleys, wedges, and screws are some examples of machines. Energy is still conserved for these devices because a machine cannot do more work than the energy put into it. However, machines can reduce the input force that is needed to perform the job. The ratio of output to input force magnitudes for any simple machine is called its mechanical advantage (MA).

$$MA = \frac{F_o}{F_i} \tag{5.4.1}$$

One of the simplest machines is the lever, which is a rigid bar pivoted at a fixed place called the fulcrum. Torques are involved in levers, since there is rotation about a pivot point. Distances from the physical pivot of the lever are crucial, and we can obtain a useful expression for the MA in terms of these distances.

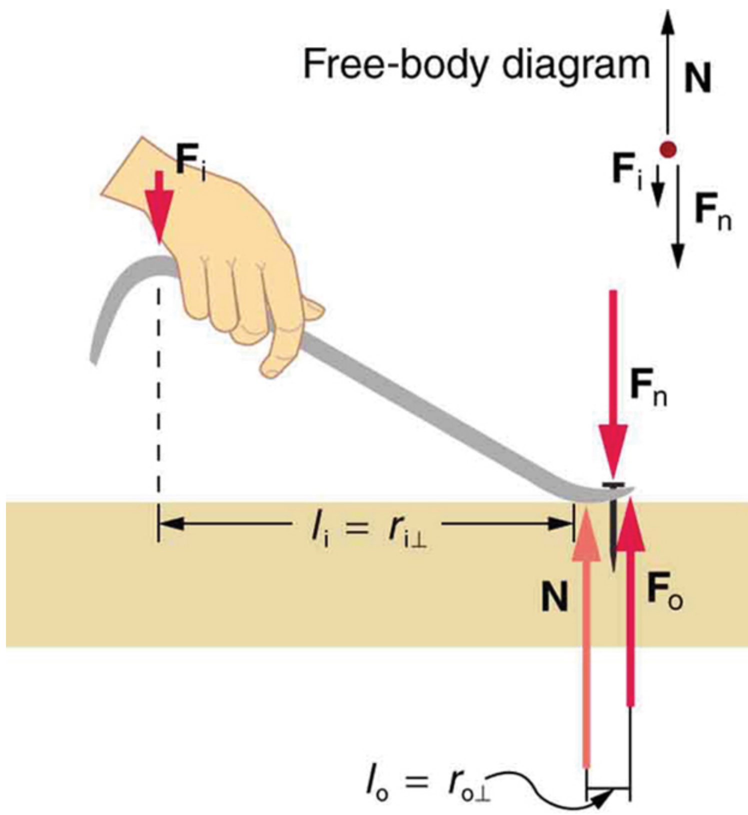


Figure 5.4.1: A nail puller is a lever with a large mechanical advantage. The external forces on the nail puller are represented by solid arrows. The force that the nail puller applies to the nail (F_o) is not a force on the nail puller. The reaction force the nail exerts back on the puller (F_n) is an external force and is equal and opposite to F_o. The perpendicular lever arms of the input and output forces are l_i and l_o.

Figure shows a lever type that is used as a nail puller. Crowbars, seesaws, and other such levers are all analogous to this one F_i is the input force and F_o is the output force. There are three vertical forces acting on the nail puller (the system of interest) – these are F_i, F_o, and N. F_n is the reaction force back on the system, equal and opposite to F_o. (note that F_o is not a force on the system.) N is the normal force upon the lever, and its torque is zero since it is exerted at the pivot. The torques due to F_i and F_n must be equal to each other if the nail is not moving, to satisfy the second condition for equilibrium $(net\,\tau = 0)$. (In order for the nail to actually move, the torque due to F_i must be ever-so-slightly greater than torque due to F_n.) Hence,

$$l_i F_i = l_o F_o \tag{5.4.2}$$

where l_i and l_o are the distances from where the input and output forces are applied to the pivot, as shown in the figure. Rearranging the last equation gives

$$\frac{F_o}{F_i} = \frac{l_i}{l_o}. \tag{5.4.3}$$

What interests us most here is that the magnitude of the force exerted by the nail puller,F_o, is much greater than the magnitude of the input force applied to the puller at the other end, F_i. For the nail puller,

$$MA = \frac{F_o}{F_i} = \frac{l_i}{l_o}. \tag{5.4.4}$$

This equation is true for levers in general. For the nail puller, the MA is certainly greater than one. The longer the handle on the nail puller, the greater the force you can exert with it.

Two other types of levers that differ slightly from the nail puller are a wheelbarrow and a shovel, shown in Figure. All these lever types are similar in that only three forces are involved – the input force, the output force, and the force on the pivot – and thus their MAs are given by $MA = \frac{F_o}{F_i}$ and $MA = \frac{d_1}{d_2}$, with distances being measured relative to the physical pivot. The wheelbarrow and shovel differ from the nail puller because both the input and output forces are on the same side of the pivot.

In the case of the wheelbarrow, the output force or load is between the pivot (the wheel's axle) and the input or applied force. In the case of the shovel, the input force is between the pivot (at the end of the handle) and the load, but the input lever arm is shorter than the output lever arm. In this case, the MA is less than one.

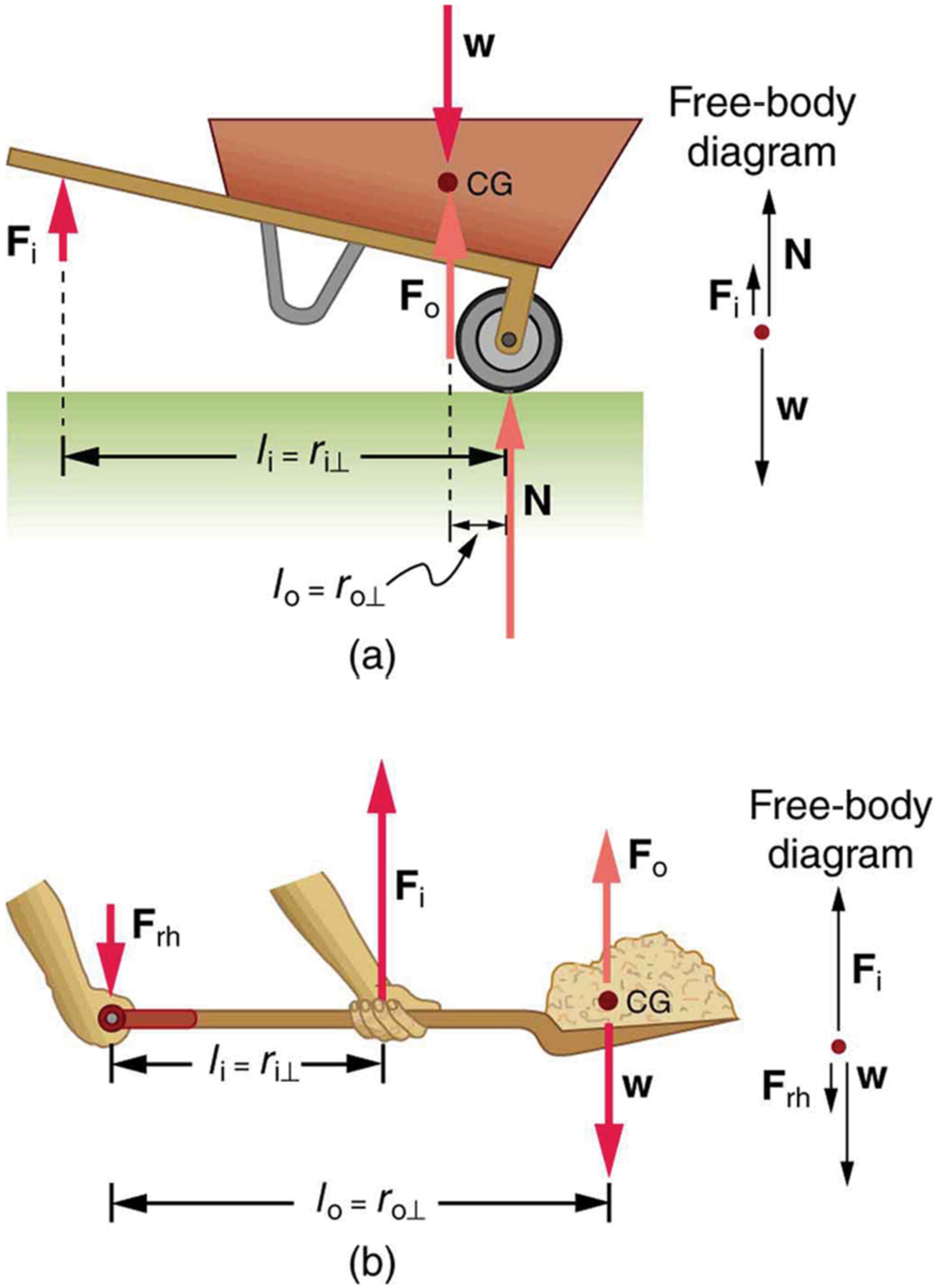


Figure 5.4.2: (a) In the case of the wheelbarrow, the output force or load is between the pivot and the input force. The pivot is the wheel's axle. Here, the output force is greater than the input force. Thus, a wheelbarrow enables you to lift much heavier loads than you could with your body alone. (b) In the case of the shovel, the input force is between the pivot and the load, but the input lever arm is shorter than the output lever arm. The pivot is at the handle held by the right hand. Here, the output force (supporting the shovel's load) is less than the input force (from the hand nearest the load), because the input is exerted closer to the pivot than is the output.

Example 5.4.1: What is the Advantage for the Wheelbarrow?

In the wheelbarrow of Figure, the load has a perpendicular lever arm of 7.50 cm, while the hands have a perpendicular lever arm of 1.02 m. (a) What upward force must you exert to support the wheelbarrow and its load if their combined mass is 45.0 kg? (b) What force does the wheelbarrow exert on the ground?

Strategy

Here, we use the concept of mechanical advantage.

Solution

(a) In this case, $\frac{F_o}{F_i} = \frac{l_i}{l_o}$ becomes

$$F_i = F_o \frac{l_o}{l_i}, \tag{5.4.5}$$

Adding values into this equation yields

$$F_i = (45.0\,kg)(9.80\,m/s^2)\frac{0.075\,m}{1.02\,m} = 32.4\,N. \tag{5.4.6}$$

The free-body diagram (see Figure) gives the following normal force:

$$F_i = +N = W. \tag{5.4.7}$$

Therefore,

$$N = (45.0\,kg)(9.80\,m/s^2) - 32.4\,N = 409\,N. \tag{5.4.8}$$

N is the normal force acting on the wheel; by Newton's third law, the force the wheel exerts on the ground is $409\,N$.

Discussion

An even longer handle would reduce the force needed to lift the load. The MA here is $MA = 1.01/0.0750 = 13.6$

Another very simple machine is the inclined plane. Pushing a cart up a plane is easier than lifting the same cart straight up to the top using a ladder, because the applied force is less. However, the work done in both cases (assuming the work done by friction is negligible) is the same. Inclined lanes or ramps were probably used during the construction of the Egyptian pyramids to move large blocks of stone to the top.

A crank is a lever that can be rotated 360^o about its pivot, as shown in Figure. Such a machine may not look like a lever, but the physics of its actions remain the same. The MA for a crank is simply the ratio of the radii r_i/r_o. Wheels and gears have this simple expression for their MAs too. The MA can be greater than 1, as it is for the crank, or less than 1, as it is for the simplified car axle driving the wheels, as shown. If the axle's radius is $2.0\,cm$ and the wheel's radius is $24.0\,cm$, then $MA = 2.0/24.0 = 0.083$ and the axle would have to exert a force of $12,000\,N$ on the wheel to enable it to exert a force of $1000\,N$ on the ground.

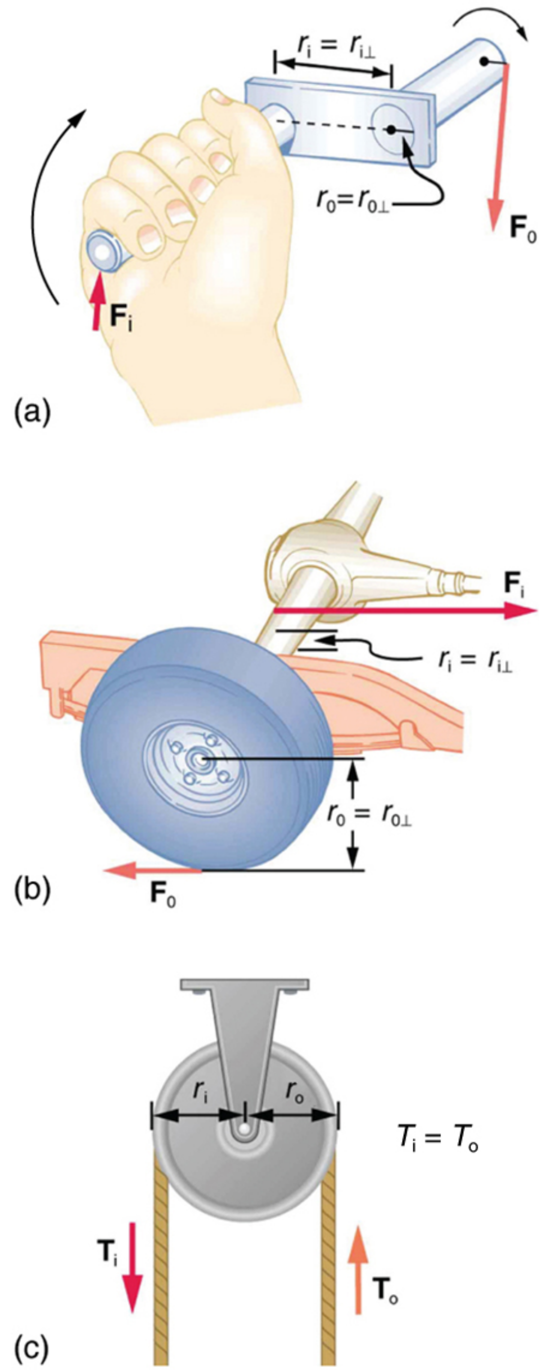


Figure $5.4.3$: (a) A crank is a type of lever that can be rotated 360^o about its pivot. Cranks are usually designed to have a large MA. (b) A simplified automobile axle drives a wheel, which has a much larger diameter than the axle. The MA is less than 1. (c) An ordinary pulley is used to lift a heavy load. The pulley changes the direction of the force T exerted by the cord without changing its magnitude. Hence, this machine has an MA of 1.

An ordinary pulley has an MA of 1; it only changes the direction of the force and not its magnitude. Combinations of pulleys, such as those illustrated in Figure, are used to multiply force. If the pulleys are friction-free, then the force output is approximately an integral multiple of the tension in the cable. The number of cables pulling directly upward on the system of interest, as illustrated in the figures given below, is approximately the MA of the pulley system. Since each attachment applies an external force in approximately the same direction as the others, they add, producing a total force that is nearly an integral multiple of the input force T.

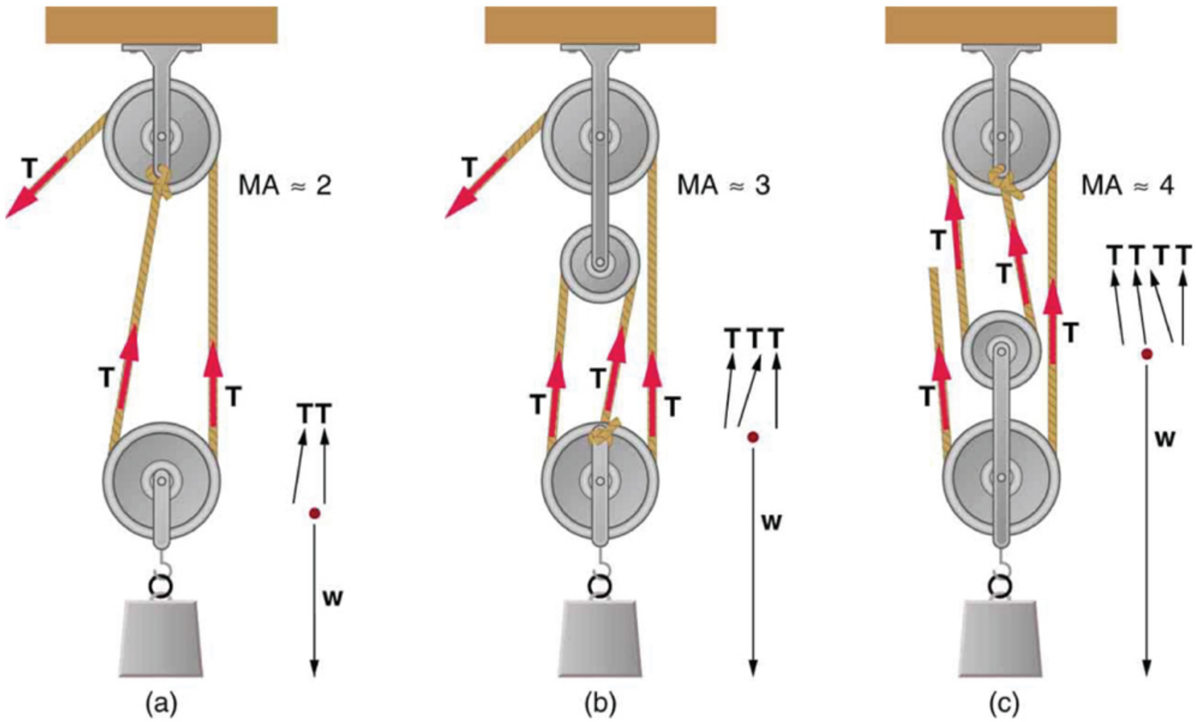


Figure 9.6.4. (a) The combination of pulleys is used to multiply force. The force is an integral multiple of tension if the pulleys are frictionless. This pulley system has two cables attached to its load, thus applying a force of approximately $2T$. This machine has $MA \approx 2$. (b) Three pulleys are used to lift a load in such a way that the mechanical advantage is about 3. Effectively, there are three cables attached to the load. (c) This pulley system applies a force of $4T$, so that it has $MA \approx 4$. Effectively, four cables are pulling on the system of interest.

Summary

- Simple machines are devices that can be used to multiply or augment a force that we apply – often at the expense of a distance through which we have to apply the force.
- The ratio of output to input forces for any simple machine is called its mechanical advantage
- A few simple machines are the lever, nail puller, wheelbarrow, crank, etc.

Glossary

mechanical advantage

the ratio of output to input forces for any simple machine

5.5: Forces and Torques in Muscles and Joints

Learning Objectives

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

- Explain the forces exerted by muscles.
- State how a bad posture causes back strain.
- Discuss the benefits of skeletal muscles attached close to joints.
- Discuss various complexities in the real system of muscles, bones, and joints.

Muscles, bones, and joints are some of the most interesting applications of statics. There are some surprises. Muscles, for example, exert far greater forces than we might think. Figure shows a forearm holding a book and a schematic diagram of an analogous lever system. The schematic is a good approximation for the forearm, which looks more complicated than it is, and we can get some insight into the way typical muscle systems function by analyzing it.

Muscles can only contract, so they occur in pairs. In the arm, the biceps muscle is a flexor—that is, it closes the limb. The triceps muscle is an extensor that opens the limb. This configuration is typical of skeletal muscles, bones, and joints in humans and other vertebrates. Most skeletal muscles exert much larger forces within the body than the limbs apply to the outside world. The reason is clear once we realize that most muscles are attached to bones via tendons close to joints, causing these systems to have mechanical advantages much less than one. Viewing them as simple machines, the input force is much greater than the output force, as seen in Figure.

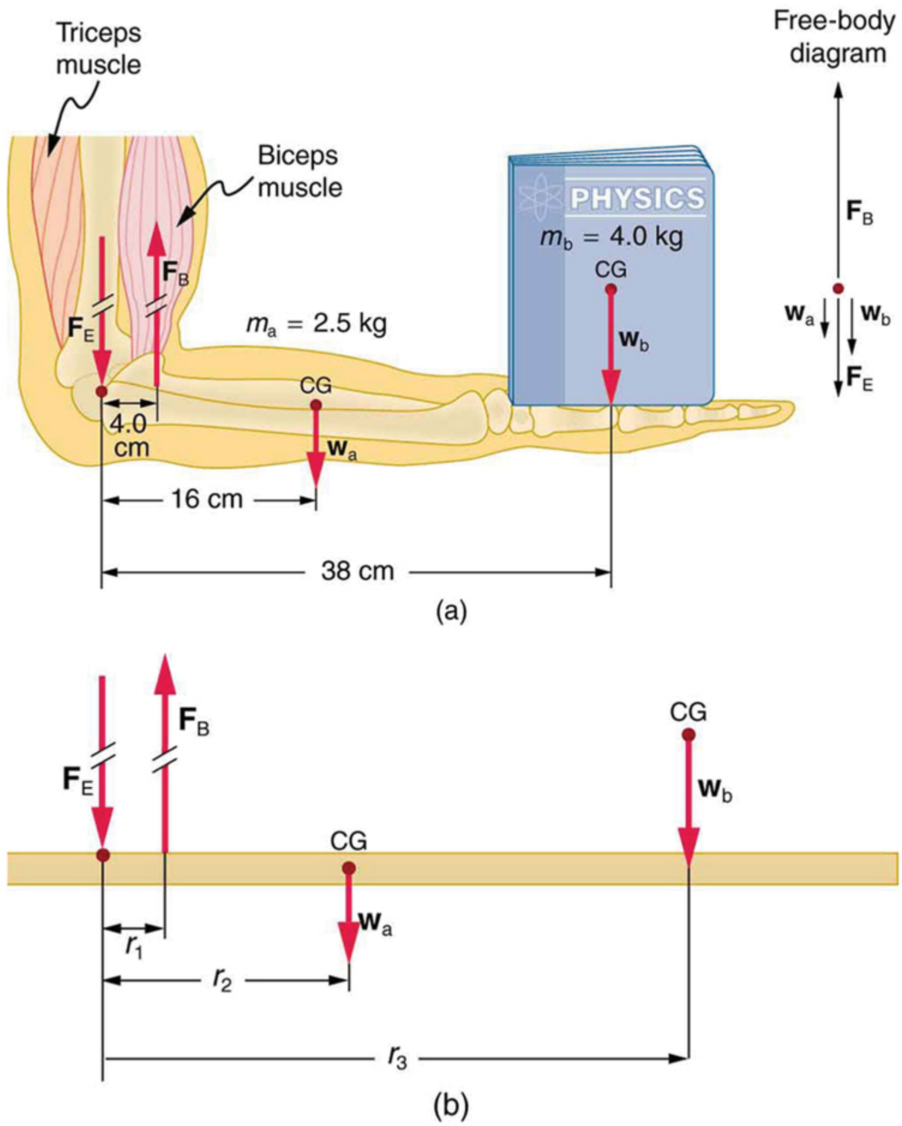


Figure 5.5.1: (a) The figure shows the forearm of a person holding a book. The biceps exert a force F_B to support the weight of the forearm and the book. The triceps are assumed to be relaxed. (b) Here, you can view an approximately equivalent mechanical system with the pivot at the elbow joint.

In the above example of the biceps muscle, the angle between the forearm and upper arm is 90°. If this angle changes, the force exerted by the biceps muscle also changes. In addition, the length of the biceps muscle changes. The force the biceps muscle can exert depends upon its length; it is smaller when it is shorter than when it is stretched.

Very large forces are also created in the joints. In the previous example, the downward force F_E exerted by the humerus at the elbow joint equals 407 N, or 6.38 times the total weight supported. (The calculation of F_E is straightforward and is left as an end-of-chapter problem.) Because muscles can contract, but not expand beyond their resting length, joints and muscles often exert forces that act in opposite directions and thus subtract. (In the above example, the upward force of the muscle minus the downward force of the joint equals the weight supported—that is, (470 N - 407 N) = 63 N, approximately equal to the weight supported.) Forces in muscles and joints are largest when their load is a long distance from the joint, as the book is in the previous example.

In racquet sports such as tennis the constant extension of the arm during game play creates large forces in this way. The mass times the lever arm of a tennis racquet is an important factor, and many players use the heaviest racquet they can handle. It is no wonder that joint deterioration and damage to the tendons in the elbow, such as "tennis elbow," can result from repetitive motion, undue torques, and possibly poor racquet selection in such sports. Various tried techniques for holding and using a racquet or bat or stick not only increases sporting prowess but can minimize fatigue and long-term damage to the body. For example, tennis balls correctly hit at the "sweet spot" on the racquet will result in little vibration or impact force being felt in the racquet and the body.

Twisting the hand to provide top spin on the ball or using an extended rigid elbow in a backhand stroke can also aggravate the tendons in the elbow.

Training coaches and physical therapists use the knowledge of relationships between forces and torques in the treatment of muscles and joints. In physical therapy, an exercise routine can apply a particular force and torque which can, over a period of time, revive muscles and joints. Some exercises are designed to be carried out under water, because this requires greater forces to be exerted, further strengthening muscles. However, connecting tissues in the limbs, such as tendons and cartilage as well as joints are sometimes damaged by the large forces they carry. Often, this is due to accidents, but heavily muscled athletes, such as weightlifters, can tear muscles and connecting tissue through effort alone.

The back is considerably more complicated than the arm or leg, with various muscles and joints between vertebrae, all having mechanical advantages less than 1. Back muscles must, therefore, exert very large forces, which are borne by the spinal column. Discs crushed by mere exertion are very common. The jaw is somewhat exceptional—the masseter muscles that close the jaw have a mechanical advantage greater than 1 for the back teeth, allowing us to exert very large forces with them. A cause of stress headaches is persistent clenching of teeth where the sustained large force translates into fatigue in muscles around the skull.

Figure shows how bad posture causes back strain. In part (a), we see a person with good posture. Note that her upper body's cg is directly above the pivot point in the hips, which in turn is directly above the base of support at her feet. Because of this, her upper body's weight exerts no torque about the hips. The only force needed is a vertical force at the hips equal to the weight supported. No muscle action is required, since the bones are rigid and transmit this force from the floor. This is a position of unstable equilibrium, but only small forces are needed to bring the upper body back to vertical if it is slightly displaced. Bad posture is shown in part (b); we see that the upper body's cg is in front of the pivot in the hips. This creates a clockwise torque around the hips that is counteracted by muscles in the lower back. These muscles must exert large forces, since they have typically small mechanical advantages. (In other words, the perpendicular lever arm for the muscles is much smaller than for the cg.) Poor posture can also cause muscle strain for people sitting at their desks using computers. Special chairs are available that allow the body's CG to be more easily situated above the seat, to reduce back pain. Prolonged muscle action produces muscle strain. Note that the cg of the entire body is still directly above the base of support in part (b) of Figure. This is compulsory; otherwise the person would not be in equilibrium. We lean forward for the same reason when carrying a load on our backs, to the side when carrying a load in one arm, and backward when carrying a load in front of us, as seen in Figure.

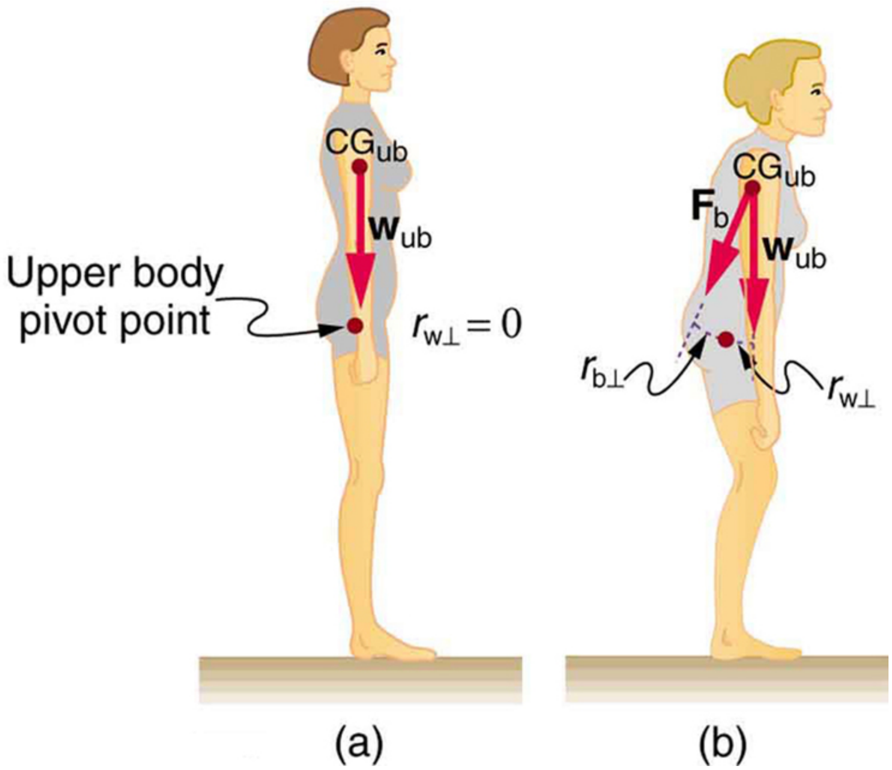


Figure 5.5.2: (a) Good posture places the upper body's cg over the pivots in the hips, eliminating the need for muscle action to balance the body. (b) Poor posture requires exertion by the back muscles to counteract the clockwise torque produced around the pivot by the upper body's weight. The back muscles have a small effective perpendicular lever arm, $r_{b\perp}$, and must therefore exert a large force F_b. Note that the legs lean backward to keep the cg of the entire body above the base of support in the feet.

You have probably been warned against lifting objects with your back. This action, even more than bad posture, can cause muscle strain and damage discs and vertebrae, since abnormally large forces are created in the back muscles and spine.

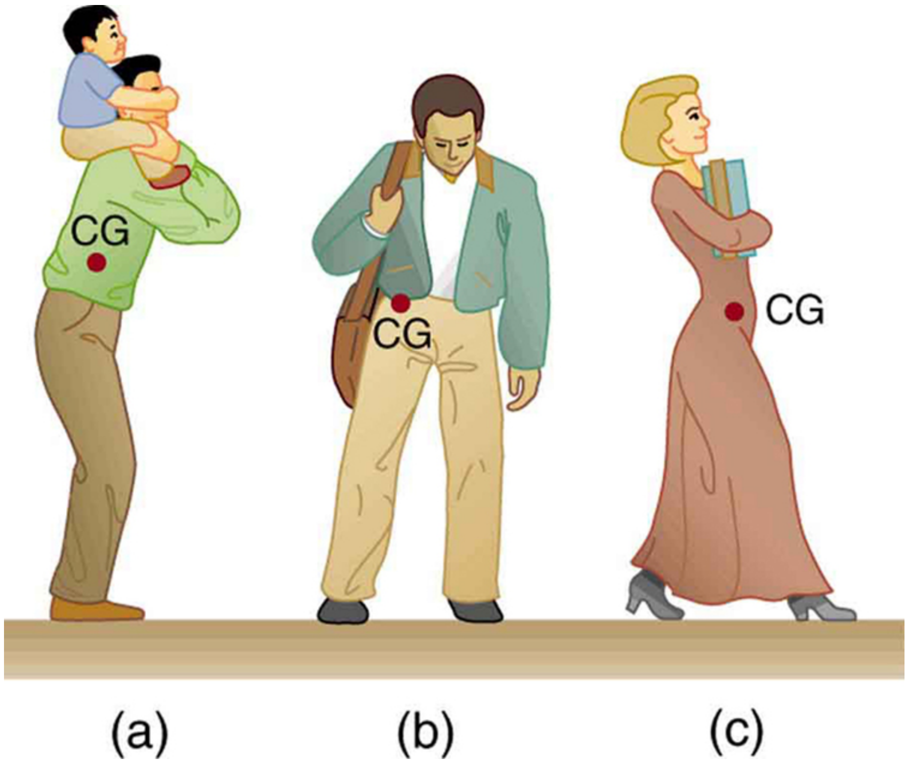


Figure 5.5.3: People adjust their stance to maintain balance. (a) A father carrying his son piggyback leans forward to position their overall cg above the base of support at his feet. (b) A student carrying a shoulder bag leans to the side to keep the overall cg over his feet. (c) Another student carrying a load of books in her arms leans backward for the same reason.

What are the benefits of having most skeletal muscles attached so close to joints? One advantage is speed because small muscle contractions can produce large movements of limbs in a short period of time. Other advantages are flexibility and agility, made possible by the large numbers of joints and the ranges over which they function. For example, it is difficult to imagine a system with biceps muscles attached at the wrist that would be capable of the broad range of movement we vertebrates possess.

There are some interesting complexities in real systems of muscles, bones, and joints. For instance, the pivot point in many joints changes location as the joint is flexed, so that the perpendicular lever arms and the mechanical advantage of the system change, too. Thus the force the biceps muscle must exert to hold up a book varies as the forearm is flexed. Similar mechanisms operate in the legs, which explain, for example, why there is less leg strain when a bicycle seat is set at the proper height. The methods employed in this section give a reasonable description of real systems provided enough is known about the dimensions of the system. There are many other interesting examples of force and torque in the body.

Summary

- Statics plays an important part in understanding everyday strains in our muscles and bones.
- Many lever systems in the body have a mechanical advantage of significantly less than one, as many of our muscles are attached close to joints.
- Someone with good posture stands or sits in such a way that their center of gravity lies directly above the pivot point in their hips, thereby avoiding back strain and damage to disks.

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney).

- **9.6: Forces and Torques in Muscles and Joints** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

5.6: Prelude to Rotational Motion and Angular Momentum

Why do tornadoes spin at all (Figure $5.6.1$)? And why do tornados spin so rapidly? The answer is that air masses that produce tornadoes are themselves rotating, and when the radii of the air masses decrease, their rate of rotation increases. An ice skater increases her spin in an exactly analogous manner as seen in Figure $5.6.2$. The skater starts her rotation with outstretched limbs and increases her spin by pulling them in toward her body. The same physics describes the exhilarating spin of a skater and the wrenching force of a tornado.

Figure $5.6.1$: The mention of a tornado conjures up images of raw destructive power. Tornadoes blow houses away as if they were made of paper and have been known to pierce tree trunks with pieces of straw. They descend from clouds in funnel-like shapes that spin violently, particularly at the bottom where they are most narrow, producing winds as high as 500 km/h. (credit: Daphne Zaras, U.S. National Oceanic and Atmospheric Administration)

Clearly, force, energy, and power are associated with rotational motion. These and other aspects of rotational motion are covered in this chapter. We shall see that all important aspects of rotational motion either have already been defined for linear motion or have exact analogs in linear motion. First, we look at angular acceleration—the rotational analog of linear acceleration.

Figure $5.6.2$: This figure skater increases her rate of spin by pulling her arms and her extended leg closer to her axis of rotation. (credit: Luu, Wikimedia Commons)

5.7: Angular Momentum and Its Conservation

Learning Objectives

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

- Understand the analogy between angular momentum and linear momentum.
- Observe the relationship between torque and angular momentum.
- Apply the law of conservation of angular momentum.

Why does Earth keep on spinning? What started it spinning to begin with? And how does an ice skater manage to spin faster and faster simply by pulling her arms in? Why does she not have to exert a torque to spin faster? Questions like these have answers based in angular momentum, the rotational analog to linear momentum.

By now the pattern is clear—every rotational phenomenon has a direct translational analog. It seems quite reasonable, then, to define **angular momentum L** as

$$L = I\omega \tag{5.7.1}$$

This equation is an analog to the definition of linear momentum as

$$p = mv \tag{5.7.2}$$

Units for linear momentum are $kg \cdot m/s$, while units for angular momentum are $kg \cdot m/s^2$. As we would expect, an object that has a large moment of inertia I, such as Earth, has a very large angular momentum. An object that has a large angular velocity ω, such as a centrifuge, also has a rather large angular momentum.

> Angular momentum is completely analogous to linear momentum, first presented in Uniform Circular Motion and Gravitation. It has the same implications in terms of carrying rotation forward, and it is conserved when the net external torque is zero. Angular momentum, like linear momentum, is also a property of the atoms and subatomic particles.

Example 5.7.1: Calculating Angular Momentum of the Earth

What is the angular momentum of the earth?

Strategy

No information is given in the statement of the problem; so we must look up pertinent data before we can calculate $L = I\omega$ First, the formula for the moment of inertia of a sphere is

$$I = \frac{2MR^2}{5} \tag{5.7.3}$$

so that

$$L = I\omega = \frac{2MR^2\omega}{5}. \tag{5.7.4}$$

Earth's mass M is $5.979 \times 10^{24}\ kg$ and its radius R is $6.376 \times 10^6\ m$. The Earth's angular velocity ω is, of course, exactly one revolution per day, but we must covert ω to radians per second to do the calculation in SI units.

Solution

Substituting known information into the expression for L and converting ω to radians per second gives

$$L = 0.4(5.979 \times 10^{24}\ kg)(6.376 \times 10^6\ m)\left(\frac{1\ rev}{d}\right) \tag{5.7.5}$$

$$= 9.72 \times 10^{37}\ kg \cdot m^2 \cdot rev/d \tag{5.7.6}$$

Substituting 2π for 1 rev and $8.64 \times 10^4\ s$ for 1 day gives

$$L=(9.72\times 10^{37}\,kg\cdot m^2)\left(\frac{2\pi\,rad/rev}{8.64\times 10^4\,s/d}\right)(1\,rev/d) \tag{5.7.7}$$

$$=7.07\times 10^{33}\,kg\cdot m^2/s. \tag{5.7.8}$$

Discussion

This number is large, demonstrating that Earth, as expected, has a tremendous angular momentum. The answer is approximate, because we have assumed a constant density for Earth in order to estimate its moment of inertia.

When you push a merry-go-round, spin a bike wheel, or open a door, you exert a torque. If the torque you exert is greater than opposing torques, then the rotation accelerates, and angular momentum increases. The greater the net torque, the more rapid the increase in L. The relationship between torque and angular momentum is

$$net\,\tau=\frac{\Delta L}{\Delta t}. \tag{5.7.9}$$

This expression is exactly analogous to the relationship between force and linear momentum, $F=\Delta p/\Delta t$. The equation $net\,\tau=\Delta L/\Delta t$ is very fundamental and broadly applicable. It is, in fact, the rotational form of Newton's second law.

Example 5.7.1: Calculating the Torque Putting Angular Momentum Into a Lazy Susan

Figure 5.7.1 shows a Lazy Susan food tray being rotated by a person in quest of sustenance. Suppose the person exerts a 2.50 N force perpendicular to the lazy Susan's 0.260-m radius for 0.150 s.

a. What is the final angular momentum of the lazy Susan if it starts from rest, assuming friction is negligible?
b. What is the final angular velocity of the lazy Susan, given that its mass is 4.00 kg and assuming its moment of inertia is that of a disk?

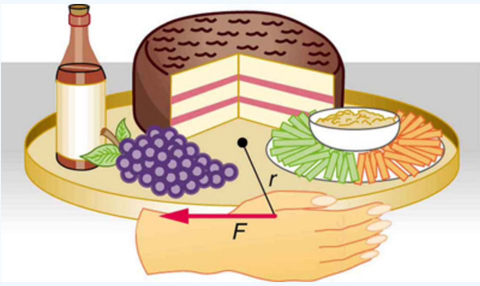


Figure 5.7.1: *A partygoer exerts a torque on a lazy Susan to make it rotate. The equation* $net\,\tau=\frac{\Delta L}{\Delta t}$ *gives the relationship between torque and the angular momentum produced.*

Strategy

We can find the angular momentum by solving $net\,\tau=\frac{\Delta L}{\Delta t}$ for ΔL, and using the given information to calculate the torque. The final angular momentum equals the change in angular momentum, because the lazy Susan starts from rest. That is, $\Delta L=L$ To find the final velocity, we must calculate ω from the definition of L in $L=I\omega$.

Solution for (a)

$$\text{Solving } net\,\tau=\frac{\Delta L}{\Delta t} \text{ for } \Delta L \text{ gives}$$

$$\Delta L=(net\,\tau)\Delta t. \tag{5.7.10}$$

Because the force is perpendicular to r, we see that $net\,\tau=rF$ so that

$$L=rF\Delta t=(0.260\,m)(2.50\,N)(0.150\,s) \tag{5.7.11}$$

$$=9.75\times 10^{-2}\,kg\cdot m^2/s. \tag{5.7.12}$$

Solution for (b)

The final angular velocity can be calculated from the definition of angular momentum,

$$L = I\omega. \tag{5.7.13}$$

Solving for ω and substituting the formula for the moment of inertia of a disk into the resulting equation gives

$$\omega = \frac{L}{I} = \frac{L}{\frac{1}{2}MR^2}. \tag{5.7.14}$$

And substituting known values into the preceding equation yields

$$\omega = \frac{9.75 \times 10^{-2}\, kg \cdot m^2/s}{(0.500)(4.00\, kg)(0.260\, m)} = 0.721\, rad/s. \tag{5.7.15}$$

Discussion

Note that the imparted angular momentum does not depend on any property of the object but only on torque and time. The final angular velocity is equivalent to one revolution in 8.71 s (determination of the time period is left as an exercise for the reader), which is about right for a lazy Susan.

Example 5.7.3: Calculating the Torque in a Kick

The person whose leg is shown in 5.7.1 kicks his leg by exerting a 2000-N force with his upper leg muscle. The effective perpendicular lever arm is 2.20 cm. Given the moment of inertia of the lower leg is $1.25\, kg \cdot m^2$.

a. find the angular acceleration of the leg.
b. Neglecting the gravitational force, what is the rotational kinetic energy of the leg after it has rotated through 57.3^0 (1.00 rad)?

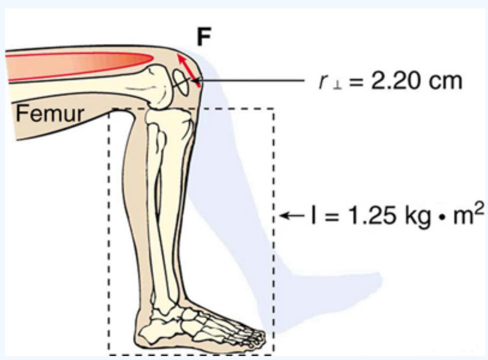


Figure 5.7.2: The muscle in the upper leg gives the lower leg an angular acceleration and imparts rotational kinetic energy to it by exerting a torque about the knee. F is a vector that is perpendicular to r. This example examines the situation.

Strategy

The angular acceleration can be found using the rotational analog to Newton's second law, or $\alpha = net\, \tau / I$. The moment of inertia I is given and the torque can be found easily from the given force and perpendicular lever arm. Once the angular acceleration α is known, the final angular velocity and rotational kinetic energy can be calculated.

Solution to (a)

From the rotational analog to Newton's second law, the angular acceleration α is

$$\alpha = \frac{net\, \tau}{I}. \tag{5.7.16}$$

Because the force and the perpendicular lever arm are given and the leg is vertical so that its weight does not create a torque, the net torque is thus

$$net\, \tau = r_\perp F \tag{5.7.17}$$

$$= (0.0220\, m)(2000\, N) \tag{5.7.18}$$

$$= 44.0\, N \cdot m \tag{5.7.19}$$

Substituting this value for the torque and the given value for the moment of inertia into the expression for α gives

$$\alpha = \frac{44.0\, N \cdot m}{1.25\, kg \cdot m^2} = 35.2\, rad/s^2 \tag{5.7.20}$$

Solution to (b)

The final angular velocity can be calculated from the kinematic expression

$$\omega^2 = \omega_0^2 + 2\alpha\theta \tag{5.7.21}$$

or

$$\omega^2 = 2\alpha\theta \tag{5.7.22}$$

because the initial angular velocity is zero. The kinetic energy of rotation is

$$KE_{rot} = \frac{1}{2} I\omega^2 \tag{5.7.23}$$

so it is most convenient to use the value of ω^2 just found and the given value for the moment of inertia. The kinetic energy is then

$$KE_{rot} = 0.5(1.25\, kg \cdot m^2)(70.4\, rad^2/s^2) \tag{5.7.24}$$

$$= 44\, J \tag{5.7.25}$$

Discussion

These values are reasonable for a person kicking his leg starting from the position shown. The weight of the leg can be neglected in part (a) because it exerts no torque when the center of gravity of the lower leg is directly beneath the pivot in the knee. In part (b), the force exerted by the upper leg is so large that its torque is much greater than that created by the weight of the lower leg as it rotates. The rotational kinetic energy given to the lower leg is enough that it could give a ball a significant velocity by transferring some of this energy in a kick.

Angular momentum, like energy and linear momentum, is conserved. This universally applicable law is another sign of underlying unity in physical laws. Angular momentum is conserved when net external torque is zero, just as linear momentum is conserved when the net external force is zero.

Conservation of Angular Momentum

We can now understand why Earth keeps on spinning. As we saw in the previous example, $\Delta L = (net\, \tau)\Delta t$. This equation means that, to change angular momentum, a torque must act over some period of time. Because Earth has a large angular momentum, a large torque acting over a long time is needed to change its rate of spin. So what external torques are there? Tidal friction exerts torque that is slowing Earth’s rotation, but tens of millions of years must pass before the change is very significant. Recent research indicates the length of the day was 18 h some 900 million years ago. Only the tides exert significant retarding torques on Earth, and so it will continue to spin, although ever more slowly, for many billions of years.

What we have here is, in fact, another conservation law. If the net torque is *zero*, then angular momentum is constant or *conserved*. We can see this rigorously by considering $net\, \tau = \frac{\Delta L}{\Delta t}$ for the situation in which the net torque is zero. In that case,

$$net\, \tau = 0 \tag{5.7.26}$$

implying that

$$\frac{\Delta L}{\Delta t} = 0. \tag{5.7.27}$$

If the change in angular momentum ΔL is zero, then the angular momentum is constant; thus,

$$L = constant\ (net\ \tau = 0) \tag{5.7.28}$$

or

$$L = L'(net\ \tau = 0). \tag{5.7.29}$$

These expressions are the law of conservation of angular momentum. Conservation laws are as scarce as they are important.

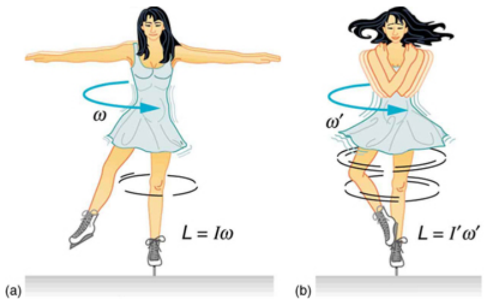


Figure $5.7.3$: (a) An ice skater is spinning on the tip of her skate with her arms extended. Her angular momentum is conserved because the net torque on her is negligibly small. In the next image, her rate of spin increases greatly when she pulls in her arms, decreasing her moment of inertia. The work she does to pull in her arms results in an increase in rotational kinetic energy.

An example of conservation of angular momentum is seen in Figure $5.7.3$, in which an ice skater is executing a spin. The net torque on her is very close to zero, because there is relatively little friction between her skates and the ice and because the friction is exerted very close to the pivot point. (Both F and τ are small, and so τ is negligibly small.) Consequently, she can spin for quite some time. She can do something else, too. She can increase her rate of spin by pulling her arms and legs in. Why does pulling her arms and legs in increase her rate of spin? The answer is that her angular momentum is constant, so that

$$L = L'. \tag{5.7.30}$$

Expressing this equation in terms of the moment of inertia,

$$I\omega = I'\omega' \tag{5.7.31}$$

where the primed quantities refer to conditions after she has pulled in her arms and reduced her moment of inertia. Because I' is smaller, the angular velocity ω' must increase to keep the angular momentum constant. The change can be dramatic, as the following example shows.

Example $5.7.4$: Calculating the Angular Momentum of a Spinning Skater

Suppose an ice skater, such as the one in Figure $5.7.3$, is spinning at 0.800 rev/ s with her arms extended. She has a moment of inertia of $2.34\ kg \cdot m^2$ with her arms extended and of $0.363\ kg \cdot m^2$ with her arms close to her body. (These moments of inertia are based on reasonable assumptions about a 60.0-kg skater.) (a) What is her angular velocity in revolutions per second after she pulls in her arms? (b) What is her rotational kinetic energy before and after she does this?

Strategy

In the first part of the problem, we are looking for the skater's angular velocity ω' after she has pulled in her arms. To find this quantity, we use the conservation of angular momentum and note that the moments of inertia and initial angular velocity are given. To find the initial and final kinetic energies, we use the definition of rotational kinetic energy given by

$$KE_{rot} = \frac{1}{2}I\omega^2 \tag{5.7.32}$$

Solution for (a)

Because torque is negligible (as discussed above), the conservation of angular momentum given in $I\omega = I'\omega'$ is applicable. Thus,

$$L = L' \tag{5.7.33}$$

or

$$I\omega^2 = I'\omega' \tag{5.7.34}$$

Solving for ω' and substituting known values into the resulting equation gives

$$\omega' = \frac{I}{I'}\omega = \left(\frac{2.34\,kg\cdot m^2}{0.363\,kg\cdot m^2}\right)(0.800\,rev/s) \tag{5.7.35}$$

$$= 5.16 rev/s. \tag{5.7.36}$$

Solution for (b)

Rotational kinetic energy is given by

$$KE_{rot} = \frac{1}{2}I\omega^2 \tag{5.7.37}$$

The initial value is found by substituting known values into the equation and converting the angular velocity to rad/s:

$$KE_{rot} = (0.5)(2.34\,kg\cdot m^2)((0.800\,rev/s)(2\pi\,rad/rev))^2 \tag{5.7.38}$$

$$= 29.6\,J. \tag{5.7.39}$$

The final rotational kinetic energy is

$$KE'_{rot} = \frac{1}{2}I'\omega'^2 \tag{5.7.40}$$

Substituting known values into this equation gives

$$KE'_{rot} = (0.5)(0.363\,kg\cdot m^2)[(5.16\,rev/s)((2\pi\,rad/rev)]^2 \tag{5.7.41}$$

$$= 191\,J. \tag{5.7.42}$$

Discussion

In both parts, there is an impressive increase. First, the final angular velocity is large, although most world-class skaters can achieve spin rates about this great. Second, the final kinetic energy is much greater than the initial kinetic energy. The increase in rotational kinetic energy comes from work done by the skater in pulling in her arms. This work is internal work that depletes some of the skater's food energy.

There are several other examples of objects that increase their rate of spin because something reduced their moment of inertia. Tornadoes are one example. Storm systems that create tornadoes are slowly rotating. When the radius of rotation narrows, even in a local region, angular velocity increases, sometimes to the furious level of a tornado. Earth is another example. Our planet was born from a huge cloud of gas and dust, the rotation of which came from turbulence in an even larger cloud. Gravitational forces caused the cloud to contract, and the rotation rate increased as a result (Figure 5.7.3).

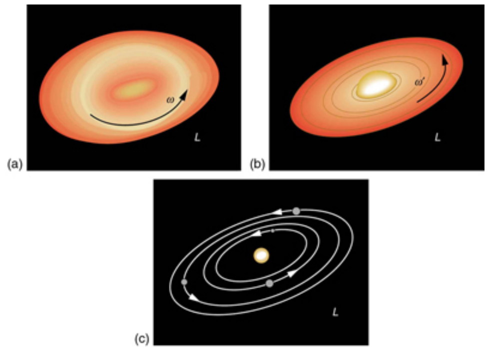


Figure 5.7.4: The Solar System coalesced from a cloud of gas and dust that was originally rotating. The orbital motions and spins of the planets are in the same direction as the original spin and conserve the angular momentum of the parent cloud.

Exercise 5.7.1: Check Your Undestanding

Is angular momentum completely analogous to linear momentum? What, if any, are their differences?

Solution

Yes, angular and linear momenta are completely analogous. While they are exact analogs they have different units and are not directly inter-convertible like forms of energy are.

Summary

- Every rotational phenomenon has a direct translational analog , likewise angular momentum L can be defined as $L = I\omega$.
- This equation is an analog to the definition of linear momentum as $p = mv$. The relationship between torque and angular momentum is $net\, \tau = \Delta L / \Delta t$.
- Angular momentum, like energy and linear momentum, is conserved. This universally applicable law is another sign of underlying unity in physical laws. Angular momentum is conserved when net external torque is zero, just as linear momentum is conserved when the net external force is zero.

Glossary

angular momentum
: the product of moment of inertia and angular velocity

law of conservation of angular momentum
: angular momentum is conserved, i.e., the initial angular momentum is equal to the final angular momentum when no external torque is applied to the system

---

CHAPTER OVERVIEW

6: Fluid Statics

6.1: Prelude to Fluid Statics

Much of what we value in life is fluid: a breath of fresh winter air; the hot blue flame in our gas cooker; the water we drink, swim in, and bathe in; the blood in our veins. What exactly is a fluid? Can we understand fluids with the laws already presented, or will new laws emerge from their study? The physical characteristics of static or stationary fluids and some of the laws that govern their behavior are the topics of this chapter. Fluid Dynamics and Its Biological and Medical Applications explores aspects of fluid flow.

Figure 6.1.1: The fluid essential to all life has a beauty of its own. It also helps support the weight of this swimmer. (credit: Terren, Wikimedia Commons)

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

6.2: What Is a Fluid?

Learning Objectives

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

- State the common phases of matter.
- Explain the physical characteristics of solids, liquids, and gases.
- Describe the arrangement of atoms in solids, liquids, and gases.

Matter most commonly exists as a solid, liquid, or gas; these states are known as the three common *phases of matter*. Solids have a definite shape and a specific volume, liquids have a definite volume but their shape changes depending on the container in which they are held, and gases have neither a definite shape nor a specific volume as their molecules move to fill the container in which they are held (Figure 6.2.1). Liquids and gases are considered to be fluids because they yield to shearing forces, whereas solids resist them. Note that the extent to which fluids yield to shearing forces (and hence flow easily and quickly) depends on a quantity called the viscosity which is discussed in detail in Viscosity and Laminar Flow; Poiseuille's Law. We can understand the phases of matter and what constitutes a fluid by considering the forces between atoms that make up matter in the three phases.

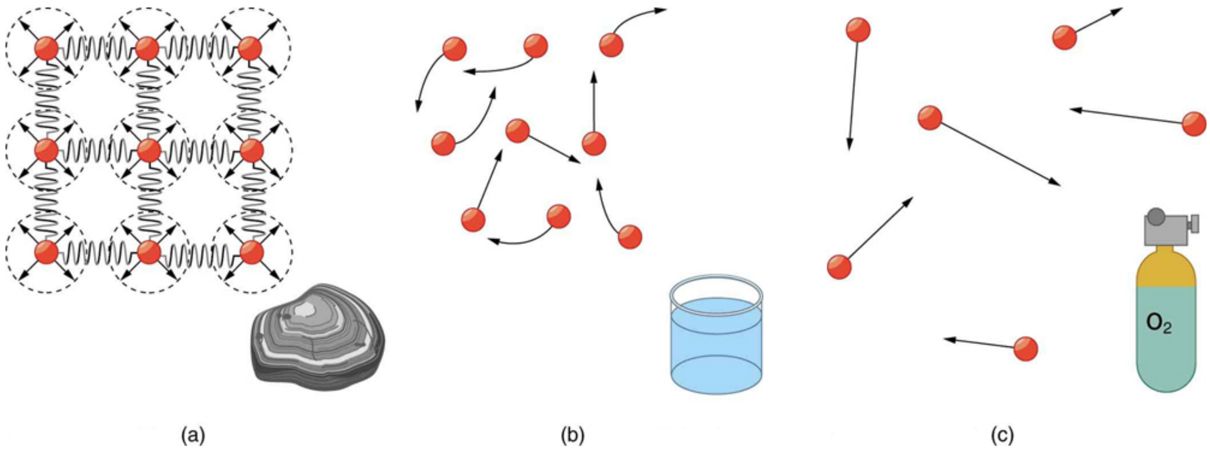


Figure 6.2.1: (a) Atoms in a solid always have the same neighbors, held near home by forces represented here by springs. These atoms are essentially in contact with one another. A rock is an example of a solid. This rock retains its shape because of the forces holding its atoms together. (b) Atoms in a liquid are also in close contact but can slide over one another. Forces between them strongly resist attempts to push them closer together and also hold them in close contact. Water is an example of a liquid. Water can flow, but it also remains in an open container because of the forces between its atoms. (c) Atoms in a gas are separated by distances that are considerably larger than the size of the atoms themselves, and they move about freely. A gas must be held in a closed container to prevent it from moving out freely.

Atoms in *solids* are in close contact, with forces between them that allow the atoms to vibrate but not to change positions with neighboring atoms. (These forces can be thought of as springs that can be stretched or compressed, but not easily broken.) Thus a solid *resists* all types of stress. A solid cannot be easily deformed because the atoms that make up the solid are not able to move about freely. Solids also resist compression, because their atoms form part of a lattice structure in which the atoms are a relatively fixed distance apart. Under compression, the atoms would be forced into one another. Most of the examples we have studied so far have involved solid objects which deform very little when stressed.

Submicroscopic Explanation of Solids and Liquids

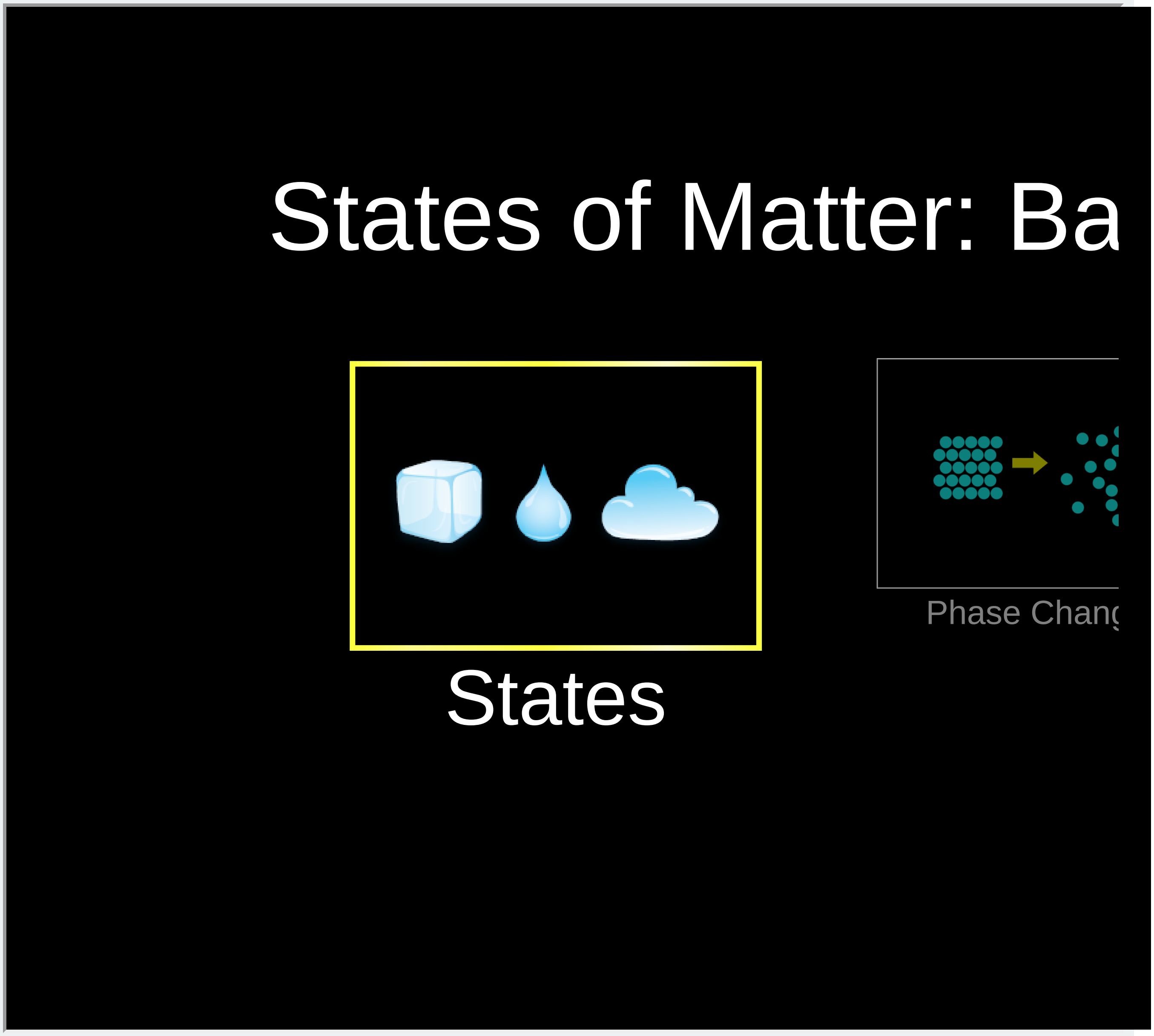
Atomic and molecular characteristics explain and underlie the macroscopic characteristics of solids and fluids. This submicroscopic explanation is one theme of this text and is highlighted in Conservation of Momentum. See, for example, microscopic description of collisions and momentum or microscopic description of pressure in a gas. This present section is devoted entirely to the submicroscopic explanation of solids and liquids.

- In contrast, *liquids* deform easily when stressed and do not spring back to their original shape once the force is removed because the atoms are free to slide about and change neighbors—that is, they *flow* (so they are a type of fluid), with the molecules held together by their mutual attraction. When a liquid is placed in a container with no lid on, it remains in the container (providing the container has no holes below the surface of the liquid!). Because the atoms are closely packed, liquids, like solids, resist compression.

Atoms in *gases* are separated by distances that are large compared with the size of the atoms. The forces between gas atoms are therefore very weak, except when the atoms collide with one another. Gases thus not only flow (and are therefore considered to be fluids) but they are relatively easy to compress because there is much space and little force between atoms. When placed in an open container gases, unlike liquids, will escape. The major distinction is that gases are easily compressed, whereas liquids are not. We shall generally refer to both gases and liquids simply as fluids, and make a distinction between them only when they behave differently.

PhET Explorations: States of Matter - Basics

Heat, cool, and compress atoms and molecules and watch as they change between solid, liquid, and gas phases.



Summary

- A fluid is a state of matter that yields to sideways or shearing forces. Liquids and gases are both fluids. Fluid statics is the physics of stationary fluids.

Glossary

fluids

liquids and gases; a fluid is a state of matter that yields to shearing forces

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

6.3: Density

Learning Objectives

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

- Define density.
- Calculate the mass of a reservoir from its density.
- Compare and contrast the densities of various substances.

Which weighs more, a ton of feathers or a ton of bricks? This old riddle plays with the distinction between mass and density. A ton is a ton, of course; but bricks have much greater density than feathers, and so we are tempted to think of them as heavier (Figure $6.3.1$).

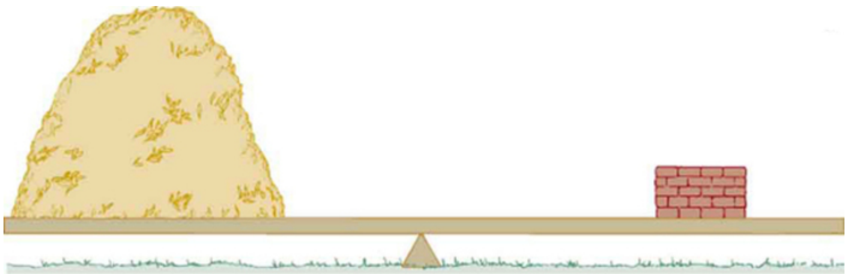


Figure $6.3.1$: A ton of feathers and a ton of bricks have the same mass, but the feathers make a much bigger pile because they have a much lower density.

Density, as you will see, is an important characteristic of substances. It is crucial, for example, in determining whether an object sinks or floats in a fluid.

Definition: Density

Density is mass per unit volume.

$$\rho = \frac{m}{V}, \tag{6.3.1}$$

where the Greek letter ρ (rho) is the symbol for density, m is the mass, and V is the volume occupied by the substance.

In the riddle regarding the feathers and bricks, the masses are the same, but the volume occupied by the feathers is much greater, since their density is much lower. The SI unit of density is kg/m^3, representative values are given in Table $6.3.1$. The metric system was originally devised so that water would have a density of $1\ g/cm^3$, equivalent to $10^3\ kg/m^3$. Thus the basic mass unit, the kilogram, was first devised to be the mass of 1000 mL of water, which has a volume of $1000\ cm^3$.

Table $6.3.1$: Densities of Various Substances

Substance	$\rho(10^3 \frac{kg}{m^3}\ or\ \frac{g}{mL})$	Substance	$\rho(10^3 \frac{kg}{m^3}\ or\ \frac{g}{mL})$	Substance	$\rho(10^3 \frac{kg}{m^3}\ or\ \frac{g}{mL})$
Solids		**Liquids**		**Gases**	
Aluminum	2.7	Water (4ºC)	1.000	Air	1.29×10^{-3}
Brass	8.44	Blood	1.05	Carbon dioxide	1.98×10^{-3}
Copper (average)	8.8	Sea water	1.025	Carbon monoxide	1.25×10^{-3}
Gold	19.32	Mercury	13.6	Hydrogen	0.090×10^{-3}
Iron or steel	7.8	Ethyl alcohol	0.79	Helium	0.18×10^{-3}
Lead	11.3	Petrol	0.68	Methane	0.72×10^{-3}
Polystyrene	0.10	Glycerin	1.26	Nitrogen	1.25×10^{-3}
Tungsten	19.30	Olive oil	0.92	Nitrous oxide	1.98×10^{-3}

Substance	$\rho(10^3 \frac{kg}{m^3} \text{ or } \frac{g}{mL})$	Substance	$\rho(10^3 \frac{kg}{m^3} \text{ or } \frac{g}{mL})$	Substance	$\rho(10^3 \frac{kg}{m^3} \text{ or } \frac{g}{mL})$
Uranium	18.70			Oxygen	1.43×10^{-3}
Concrete	2.30–3.0			Steam 100^o	0.60×10^{-3}
Cork	0.24				
Glass, common (average)	2.6				
Granite	2.7				
Earth's crust	3.3				
Wood	0.3–0.9				
Ice (0°C)	0.917				
Bone	1.7–2.0				

As you can see by examining Table $6.3.1$, the density of an object may help identify its composition. The density of gold, for example, is about 2.5 times the density of iron, which is about 2.5 times the density of aluminum. Density also reveals something about the phase of the matter and its substructure. Notice that the densities of liquids and solids are roughly comparable, consistent with the fact that their atoms are in close contact. The densities of gases are much less than those of liquids and solids, because the atoms in gases are separated by large amounts of empty space.

TAKE-HOME EXPERIMENT SUGAR AND SALT

A pile of sugar and a pile of salt look pretty similar, but which weighs more? If the volumes of both piles are the same, any difference in mass is due to their different densities (including the air space between crystals). Which do you think has the greater density? What values did you find? What method did you use to determine these values?

Example $6.3.1$: Calculating the Mass of a Reservoir From Its Volume

A reservoir has a surface area of $50\ km^2$ and an average depth of 40.0 m. What mass of water is held behind the dam? (See Figure $6.3.2$ for a view of a large reservoir—the Three Gorges Dam site on the Yangtze River in central China.)

Figure $6.3.2$: Three Gorges Dam in central China. When completed in 2008, this became the world's largest hydroelectric plant, generating power equivalent to that generated by 22 average-sized nuclear power plants. The concrete dam is 181 m high and 2.3 km across. The reservoir made by this dam is 660 km long. Over 1 million people were displaced by the creation of the reservoir. (credit: Le Grand Portage)

Strategy

We can calculate the volume V of the reservoir from its dimensions, and find the density of water ρ in Table 6.3.1. Then the mass m can be found from the definition of density (Equation 6.3.1).

Solution

Solving Equation 6.3.1 for m gives

$$m = \rho V.$$

The volume V of the reservoir is its surface area A times its average depth h:

$$\begin{aligned} V &= Ah \\ &= (50.0\,km^2)(40.0\,m) \\ &= \left[(50.0\,km^2)\left(\frac{10^3\,m}{1\,km}\right)\right](40.0\,m) \\ &= 2.00 \times 10^9\,m^3 \end{aligned}$$

The density of water ρ from Table 6.3.1 is $1.000 \times 10^3\,kg/m^3$. Substituting V and ρ into the expression for mass gives

$$\begin{aligned} m &= (1.00 \times 10^3\,kg/m^3)(2.00 \times 10^9\,m^3) \\ &= 2.00 \times 10^{12}\,kg. \end{aligned}$$

Discussion

A large reservoir contains a very large mass of water. In this example, the weight of the water in the reservoir is $mg = 1.96 \times 10^{13}\,N$, where g is the acceleration due to the Earth's gravity (about $9.80\,m/s^2$). It is reasonable to ask whether the dam must supply a force equal to this tremendous weight. The answer is no. As we shall see in the following sections, the force the dam must supply can be much smaller than the weight of the water it holds back.

Summary

- Density is the mass per unit volume of a substance or object. In equation form, density is defined as

$$\rho = \frac{m}{V}.n$$

- The SI unit of density is kg/m^3.

Glossary

density
the mass per unit volume of a substance or object

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

6.4: Pressure

Learning Objectives

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

- Define pressure.
- Explain the relationship between pressure and force.
- Calculate force given pressure and area.

You have no doubt heard the word pressure being used in relation to blood (high or low blood pressure) and in relation to the weather (high- and low-pressure weather systems). These are only two of many examples of pressures in fluids.

Definition: Pressure

Pressure is defined as the force divided by the area perpendicular to the force over which the force is applied, or

$$P=\frac{F}{A}. \tag{6.4.1}$$

where F is a force applied to an area A that is perpendicular to the force.

A given force can have a significantly different effect depending on the area over which the force is exerted, as shown in Figure Figure $6.4.1$. The SI unit for pressure is the *pascal*, where

$$1\,Pa=1\,N/m^2. \tag{6.4.2}$$

In addition to the pascal, there are many other units for pressure that are in common use. In meteorology, atmospheric pressure is often described in units of millibar (mb), where

$$100\,mb=1\times10^4\,Pa. \tag{6.4.3}$$

Pounds per square inch $(lb/in^2\,or\,psi)$ is still sometimes used as a measure of tire pressure, and millimeters of mercury (mm Hg) is still often used in the measurement of blood pressure. Pressure is defined for all states of matter but is particularly important when discussing fluids.

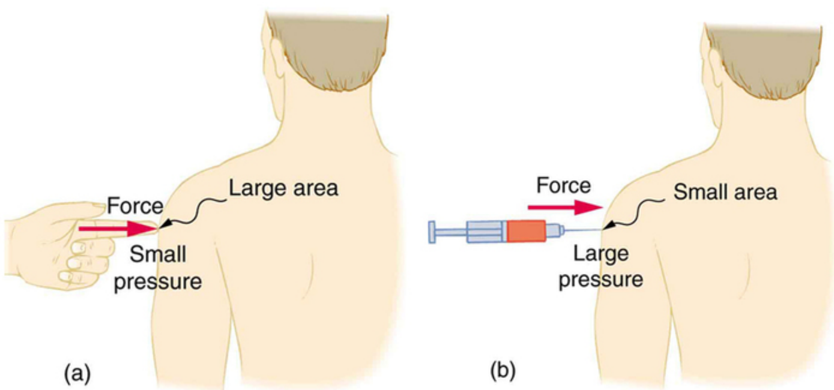


Figure $6.4.1$: (a) While the person being poked with the finger might be irritated, the force has little lasting effect. (b) In contrast, the same force applied to an area the size of the sharp end of a needle is great enough to break the skin.

Example $6.4.1$: Calculating Force Exerted by the Air - What Force Does a Pressure Exert?

An astronaut is working outside the International Space Station where the atmospheric pressure is essentially zero. The pressure gauge on her air tank reads $6.9\times10^6\,Pa$. What force does the air inside the tank exert on the flat end of the cylindrical tank, a disk 0.150 m in diameter?

Strategy

We can find the force exerted from the definition of pressure (Equation \red{pressure}) provided we can find the area A acted upon.

Solution

By rearranging the definition of pressure (Equation \red{pressure}) to solve for force, we see that

$$F = PA.$$

Here, the pressure P is given, as is the area of the end of the cylinder A, given by $A = \pi r^2$. Thus

$$\begin{aligned} F &= (6.90 \times 10^6\, Pa)(3.14)(0.0750\, m)^2 \\ &= 1.22 \times 10^5\, N. \end{aligned}$$

Discussion

Wow! No wonder the tank must be strong. Since we found $F = PA$, we see that the force exerted by a pressure is directly proportional to the area acted upon as well as the pressure itself.

The force exerted on the end of the tank is perpendicular to its inside surface. This direction is because the force is exerted by a static or stationary fluid. We have already seen that fluids cannot *withstand* shearing (sideways) forces; they cannot *exert* shearing forces, either. Fluid pressure has no direction, being a scalar quantity. The forces due to pressure have well-defined directions: they are always exerted perpendicular to any surface. (See the tire in Figure $6.4.2$, for example.)

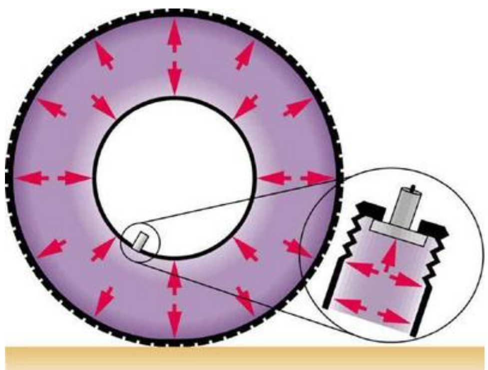


Figure $6.4.2$: Pressure inside this tire exerts forces perpendicular to all surfaces it contacts. The arrows give representative directions and magnitudes of the forces exerted at various points. Note that static fluids do not exert shearing forces.

Finally, note that pressure is exerted on all surfaces. Swimmers, as well as the tire, feel pressure on all sides (Figure $6.4.3$).

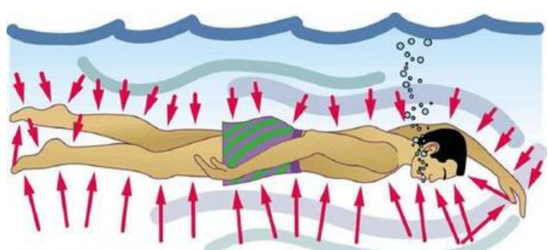


Figure $6.4.3$: Pressure is exerted on all sides of this swimmer, since the water would flow into the space he occupies if he were not there. The arrows represent the directions and magnitudes of the forces exerted at various points on the swimmer. Note that the forces are larger underneath, due to greater depth, giving a net upward or buoyant force that is balanced by the weight of the swimmer.

PHET EXPLORATIONS: GAS PROPERTIES

Pump gas molecules in this simulation to a box and see what happens as you change the volume, add or remove heat, change gravity, and more. Measure the temperature and pressure, and discover how the properties of the gas vary in relation to each other.

Summary

- Pressure is the force per unit perpendicular area over which the force is applied. In equation form, pressure is defined as

$$F = PA.$$

- The SI unit of pressure is pascal and 1 $Pa = 1\ N/m^2$.

Glossary

pressure

the force per unit area perpendicular to the force, over which the force acts

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

6.5: Variation of Pressure with Depth in a Fluid

Learning Objectives

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

- Define pressure in terms of weight.
- Explain the variation of pressure with depth in a fluid.
- Calculate density given pressure and altitude.

If your ears have ever popped on a plane flight or ached during a deep dive in a swimming pool, you have experienced the effect of depth on pressure in a fluid. At the Earth's surface, the air pressure exerted on you is a result of the weight of air above you. This pressure is reduced as you climb up in altitude and the weight of air above you decreases. Under water, the pressure exerted on you increases with increasing depth. In this case, the pressure being exerted upon you is a result of both the weight of water above you *and* that of the atmosphere above you. You may notice an air pressure change on an elevator ride that transports you many stories, but you need only dive a meter or so below the surface of a pool to feel a pressure increase. The difference is that water is much denser than air, about 775 times as dense.

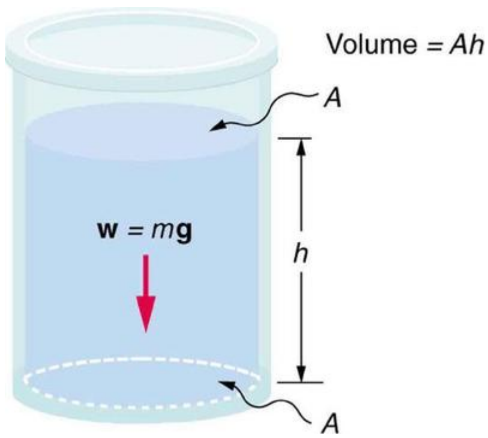


Figure 6.5.1: The bottom of this container supports the entire weight of the fluid in it. The vertical sides cannot exert an upward force on the fluid (since it cannot withstand a shearing force), and so the bottom must support it all.

Consider the container in Figure 6.5.1. Its bottom supports the weight of the fluid in it. Let us calculate the pressure exerted on the bottom by the weight of the fluid. That pressure is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):

$$P = \frac{mg}{A}. \tag{6.5.1}$$

We can find the mass of the fluid from its volume and density:

$$m = \rho V. \tag{6.5.2}$$

The volume of the fluid V is related to the dimensions of the container. It is

$$V = Ah, \tag{6.5.3}$$

where A is the cross-sectional area and h is the depth. Combining the last two equations gives

$$m = \rho A h. \tag{6.5.4}$$

If we enter this into the expression for pressure, we obtain

$$P = \frac{(\rho A h) g}{A}. \tag{6.5.5}$$

The area cancels, and rearranging the variables yields

$$P = h\rho g. \tag{6.5.6}$$

This value is the *pressure due to the weight of a fluid*. Equation 6.5.6 has general validity beyond the special conditions under which it is derived here. Even if the container were not there, the surrounding fluid would still exert this pressure, keeping the fluid static. Thus Equation 6.5.6 represents the pressure due to the weight of any fluid of *average density ρ at any depth h* below its surface. For liquids, which are nearly incompressible, this equation holds to great depths. For gases, which are quite compressible, one can apply this equation as long as the density changes are small over the depth considered. Example 6.5.1 illustrates this situation.

Example 6.5.1: Calculating the Average Pressure and Force Exerted: What Force Must a Dam Withstand?

In [link], we calculated the mass of water in a large reservoir. We will now consider the pressure and force acting on the dam retaining water (Figure 6.5.2). The dam is 500 m wide, and the water is 80.0 m deep at the dam.

a. What is the average pressure on the dam due to the water?
b. Calculate the force exerted against the dam and compare it with the weight of water in the dam (previously found to be $1.96 \times 10^{13}\,N$.

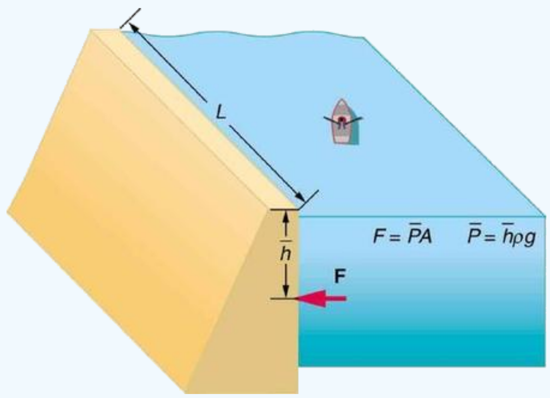


Figure 6.5.2: The dam must withstand the force exerted against it by the water it retains. This force is small compared with the weight of the water behind the dam.

Strategy for (a)

The average pressure $\bar{P}$ due to the weight of the water is the pressure at the average depth $\bar{h}$ of 40.0 m, since pressure increases linearly with depth.

Solution for (a)

The average pressure due to the weight of a fluid (Equation 6.5.6) is

$$\bar{P} = \bar{h}\rho g.$$

Entering the density of water from [link] and taking $\bar{h}$ to be the average depth of 40.0 m, we obtain

$$\bar{P} = (40.0\,m)\left(10^3\,\frac{kg}{m^3}\right)\left(9.80\,\frac{m}{s^2}\right)$$

$$= 3.92 \times 10^5\,\frac{N}{m^2} = 392\,kPa.$$

Strategy for (b)

The force exerted on the dam by the water is the average pressure times the area of contact:

$$F = \bar{P}A.$$

Solution for (b)

We have already found the value for $\bar{P}$. The area of the dam is

$$A = 80.0\,m \times 500\,m = 4.00 \times 10^4\,m^2,$$

so that

$$\begin{aligned} F &= (3.92 \times 10^5\ N/m^2)(4.00 \times 10^4\ m^2) \\ &= 1.57 \times 10^{10}\ N. \end{aligned}$$

Discussion

Although this force seems large, it is small compared with the $1.96 \times 10^{13}\ N$ weight of the water in the reservoir—in fact, it is only $0.0800\ \%$ of the weight. Note that the pressure found in part (a) is completely independent of the width and length of the lake—it depends only on its average depth at the dam. Thus the force depends only on the water's average depth and the dimensions of the dam, *not* on the horizontal extent of the reservoir. In the diagram, the thickness of the dam increases with depth to balance the increasing force due to the increasing pressure.epth to balance the increasing force due to the increasing pressure.

Atmospheric pressure is another example of pressure due to the weight of a fluid, in this case due to the weight of *air* above a given height. The atmospheric pressure at the Earth's surface varies a little due to the large-scale flow of the atmosphere induced by the Earth's rotation (this creates weather "highs" and "lows"). However, the average pressure at sea level is given by the *standard atmospheric pressure* P_{atm}, measured to be

$$1\ atmosphere\ (atm) = P_{atm} = 1.01 \times 10^5\ N/m^2 = 101\ kPa. \tag{6.5.7}$$

This relationship means that, on average, at sea level, a column of air above $1.00\ m^2$ of the Earth's surface has a weight of $1.01 \times 10^5\ N$, equivalent to 1 atm (Figure $6.5.3$).

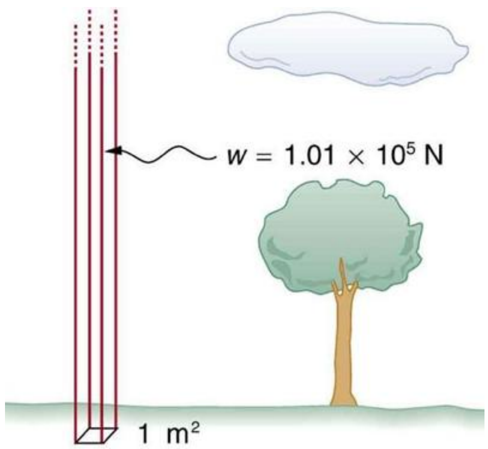


Figure $6.5.3$: Atmospheric pressure at sea level averages $1.01 \times 10^5\ Pa$ (equivalent to 1 atm), since the column of air over this $1\ m^2$, extending to the top of the atmosphere, weighs $1.01 \times 10^5\ N$.

Example $6.5.2$: Calculating Average Density: How Dense Is the Air?

Calculate the average density of the atmosphere, given that it extends to an altitude of 120 km. Compare this density with that of air listed in [link].

Strategy

If we solve $P = h\rho g$ for density, we see that

$$\bar{\rho} = \frac{P}{hg}.$$

We then take P to be atmospheric pressure, h is given, and g is known, and so we can use this to calculate $\bar{\rho}$.

Solution

Entering known values into the expression for $\bar{\rho}$ yields

$$\begin{aligned} \bar{\rho} &= \frac{1.01 \times 10^5\ N/m^2}{(120 \times 10^3\ m)(9.80\ m/s^2)} \\ &= 8.59 \times 10^{-2}\ kg/m^3. \end{aligned}$$

Discussion

This result is the average density of air between the Earth's surface and the top of the Earth's atmosphere, which essentially ends at 120 km. The density of air at sea level is given in [link] as $1.29\,kg/m^3$ - about 15 times its average value. Because air is so compressible, its density has its highest value near the Earth's surface and declines rapidly with altitude.

Example 6.5.3: Calculating Depth Below the Surface of Water: What Depth of Water Creates the Same Pressure as the Entire Atmosphere?

Calculate the depth below the surface of water at which the pressure due to the weight of the water equals 1.00 atm.

Strategy

We begin by solving the equation $P = h\rho g$ for depth h:

$$h = \frac{P}{\rho g}.$$

Then we take P to be 1.00 atm and ρ

to be the density of the water that creates the pressure.

Solution

Entering the known values into the expression for h gives

$$\begin{aligned} h &= \frac{1.01 \times 10^5\,N/m^2}{(1.00 \times 10^3\,m)(9.80\,m/s^2)} \\ &= 10.3\,m. \end{aligned}$$

Discussion

Just 10.3 m of water creates the same pressure as 120 km of air. Since water is nearly incompressible, we can neglect any change in its density over this depth.

What do you suppose is the *total* pressure at a depth of 10.3 m in a swimming pool? Does the atmospheric pressure on the water's surface affect the pressure below? The answer is yes. This seems only logical, since both the water's weight and the atmosphere's weight must be supported. So the *total* pressure at a depth of 10.3 m is 2 atm—half from the water above and half from the air above. Fluid pressures always add in this way.

Summary

- Pressure is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):

$$P = \frac{mg}{A}.$$

- Pressure due to the weight of a liquid is given by

$$P = h\rho g,$$

  where P is the pressure, h is the height of the liquid, ρ is the density of the liquid, and g is the acceleration due to the gravity.

Glossary

pressure
: the weight of the fluid divided by the area supporting it

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

6.6: Archimedes’ Principle

Learning Objectives

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

- Define buoyant force.
- State Archimedes’ principle.
- Understand why objects float or sink.
- Understand the relationship between density and Archimedes’ principle.

When you rise from lounging in a warm bath, your arms feel strangely heavy. This is because you no longer have the buoyant support of the water. Where does this buoyant force come from? Why is it that some things float and others do not? Do objects that sink get any support at all from the fluid? Is your body buoyed by the atmosphere, or are only helium balloons affected (Figure 6.6.1)?

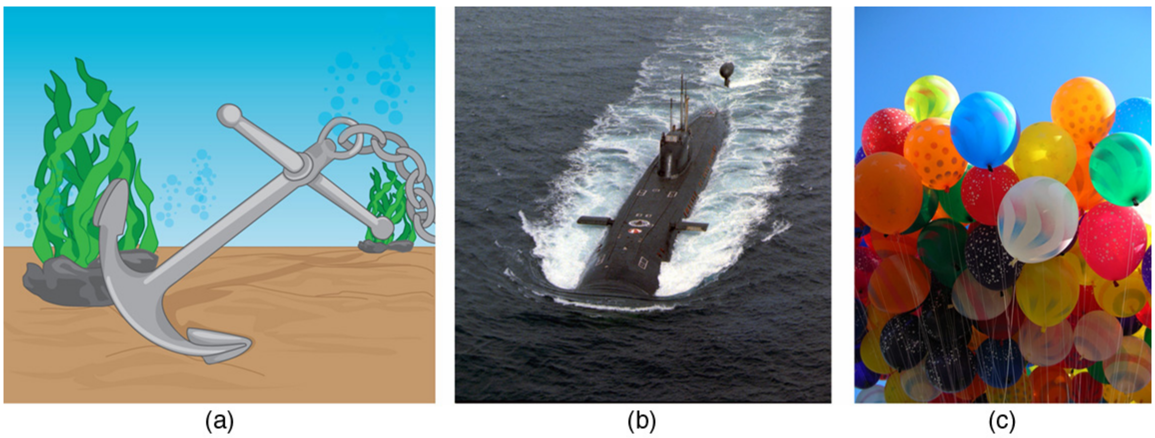


Figure 6.6.1: (a) Even objects that sink, like this anchor, are partly supported by water when submerged. (b) Submarines have adjustable density (ballast tanks) so that they may float or sink as desired. (credit: Allied Navy) (c) Helium-filled balloons tug upward on their strings, demonstrating air’s buoyant effect. (credit: Crystl)

Answers to all these questions, and many others, are based on the fact that pressure increases with depth in a fluid. This means that the upward force on the bottom of an object in a fluid is greater than the downward force on the top of the object. There is a net upward, or buoyant force on any object in any fluid (Figure 6.6.2). If the buoyant force is greater than the object’s weight, the object will rise to the surface and float. If the buoyant force is less than the object’s weight, the object will sink. If the buoyant force equals the object’s weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.

Defintion: Buoyant Force

The buoyant force is the net upward force on any object in any fluid.

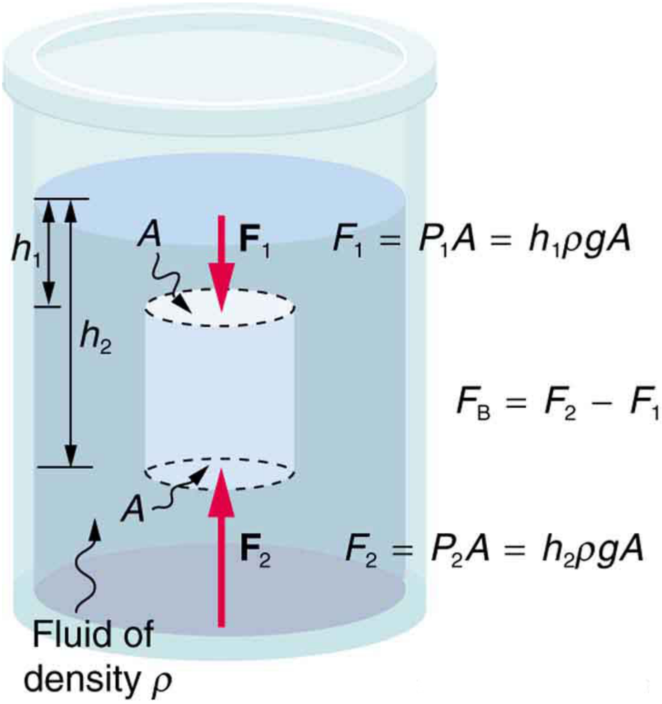


Figure 6.6.2: Pressure due to the weight of a fluid increases with depth since $P = h\rho g$. This pressure and associated upward force on the bottom of the cylinder are greater than the downward force on the top of the cylinder. Their difference is the buoyant force F_B. (Horizontal forces cancel.)

Just how great is this buoyant force? To answer this question, think about what happens when a submerged object is removed from a fluid, as in Figure 6.6.3.

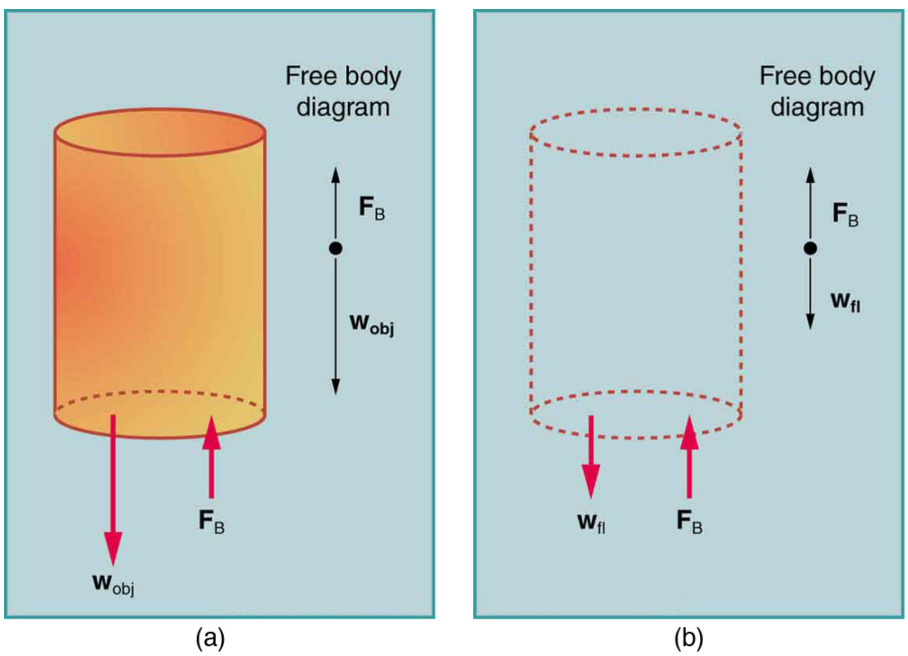


Figure 6.6.3: (a) An object submerged in a fluid experiences a buoyant force F_B. If F_B is greater than the weight of the object, the object will rise. If F_B is less than the weight of the object, the object will sink. (b) If the object is removed, it is replaced by fluid having weight w_{fl}. Since this weight is supported by surrounding fluid, the buoyant force must equal the weight of the fluid displaced. That is, $F_B = w_{fl}$, a statement of Archimedes'principle.

The space it occupied is filled by fluid having a weight w_{fl}. This weight is supported by the surrounding fluid, and so the buoyant force must equal w_{fl}, the weight of the fluid displaced by the object. It is a tribute to the genius of the Greek mathematician and inventor Archimedes (ca. 287–212 B.C.) that he stated this principle long before concepts of force were well established. Stated in words, Archimedes' principle is as follows: The buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

$$F_B = w_{fl}, \tag{6.6.1}$$

where F_B is the buoyant force and w_{fl} is the weight of the fluid displaced by the object. Archimedes' principle is valid in general, for any object in any fluid, whether partially or totally submerged.

Archimedes' Principle

According to this principle the buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

$$F_B = w_{fl}, \tag{6.6.2}$$

where F_B is the buoyant force and w_{fl} is the weight of the fluid displaced by the object.

Humm … High-tech body swimsuits were introduced in 2008 in preparation for the Beijing Olympics. One concern (and international rule) was that these suits should not provide any buoyancy advantage. How do you think that this rule could be verified?

Making Connections: Take-Home Investigation

The density of aluminum foil is 2.7 times the density of water. Take a piece of foil, roll it up into a ball and drop it into water. Does it sink? Why or why not? Can you make it sink?

Floating and Sinking

Drop a lump of clay in water. It will sink. Then mold the lump of clay into the shape of a boat, and it will float. Because of its shape, the boat displaces more water than the lump and experiences a greater buoyant force. The same is true of steel ships.

Example 6.6.1: Calculating buoyant force: dependency on shape

a. Calculate the buoyant force on 10,000 metric tons $(1.00 \times 10^7\ kg)$ of solid steel completely submerged in water, and compare this with the steel's weight.
b. What is the maximum buoyant force that water could exert on this same steel if it were shaped into a boat that could displace $1.00 \times 10^5\ m^3$ of water?

Strategy for (a)

To find the buoyant force, we must find the weight of water displaced. We can do this by using the densities of water and steel given in [link]. We note that, since the steel is completely submerged, its volume and the water's volume are the same. Once we know the volume of water, we can find its mass and weight.

Solution for (a)

First, we use the definition of density $\rho = \frac{m}{V}$ to find the steel's volume, and then we substitute values for mass and density. This gives

$$V_{st} = \frac{m_{st}}{\rho_{st}} = \frac{1.00 \times 10^7\ kg}{7.8 \times 10^3\ kg/m^3} = 1.28 \times 10^3\ m^3. \tag{6.6.3}$$

Because the steel is completely submerged, this is also the volume of water displaced, V_W. We can now find the mass of water displaced from the relationship between its volume and density, both of which are known. This gives

$$m_W = \rho_W V_W = (1.000 \times 10^3\ kg/m^3)(1.28 \times 10^3\ m^3) \tag{6.6.4}$$

$$= 1.3 \times 10^6\ kg. \tag{6.6.5}$$

By Archimedes' principle, the weight of water displaced is $m_W g$, so the buoyance force is

$$F_B = w_W = m_W g = (1.28 \times 10^6\ kg)(9.80\ m/s^2) \tag{6.6.6}$$

$$= 1.3 \times 10^7 \; N. \tag{6.6.7}$$

The steel's weight is $m_W g = 9.80 \times 10^7 \; N$,

which is much greater than the buoyant force, so the steel will remain submerged. Note that the buoyant force is rounded to two digits because the density of steel is given to only two digits.

Strategy for (b)

Here we are given the maximum volume of water the steel boat can displace. The buoyant force is the weight of this volume of water.

Solution for (b)

The mass of water displaced is found from its relationship to density and volume, both of which are known. That is,

$$m_W = \rho_W V_W = (1.000 \times 10^3 \; kg/m^3)(1.00 \times 10^5 \; m^3) \tag{6.6.8}$$

$$= 9.80 \times 10^8 \; kg. \tag{6.6.9}$$

The maximum buoyant force is the weight of this much water, or

$$F_B = w_W = m_W g = (1.00 \times 10^8 \; kg)(9.80 \, m/s^2) \tag{6.6.10}$$

$$= \times 10^8 \; N. \tag{6.6.11}$$

Discussion

The maximum buoyant force is ten times the weight of the steel, meaning the ship can carry a load nine times its own weight without sinking.

Making Connections: Take-Home Investigation

- A piece of household aluminum foil is 0.016 mm thick. Use a piece of foil that measures 10 cm by 15 cm. (a) What is the mass of this amount of foil? (b) If the foil is folded to give it four sides, and paper clips or washers are added to this "boat," what shape of the boat would allow it to hold the most "cargo" when placed in water? Test your prediction.

Density and Archimedes' Principle

Density plays a crucial role in Archimedes' principle. The average density of an object is what ultimately determines whether it floats. If its average density is less than that of the surrounding fluid, it will float. This is because the fluid, having a higher density, contains more mass and hence more weight in the same volume. The buoyant force, which equals the weight of the fluid displaced, is thus greater than the weight of the object. Likewise, an object denser than the fluid will sink.

The extent to which a floating object is submerged depends on how the object's density is related to that of the fluid. In Figure $6.6.4$, for example, the unloaded ship has a lower density and less of it is submerged compared with the same ship loaded. We can derive a quantitative expression for the fraction submerged by considering density. The fraction submerged is the ratio of the volume submerged to the volume of the object, or

$$fraction\ submerged = \frac{V_{sub}}{V_{obj}} = \frac{V_{fl}}{V_{obj}}. \tag{6.6.12}$$

The volume submerged equals the volume of fluid displaced, which we call V_{fl}. Now we can obtain the relationship between the densities by substituting $\rho = \frac{m}{V}$ into the expression. This gives

$$\frac{V_{fl}}{V_{obj}} = \frac{m_{fl}/\rho_{fl}}{m_{obj}/\bar{\rho}_{obj}}, \tag{6.6.13}$$

where $\bar{\rho}_{obj}$ is the average density of the object and ρ_{fl} is the density of the fluid. Since the object floats, its mass and that of the displaced fluid are equal, and so they cancel from the equation, leaving

$$fraction\ submerged = \frac{\bar{\rho}_{obj}}{\rho_{fl}}. \tag{6.6.14}$$

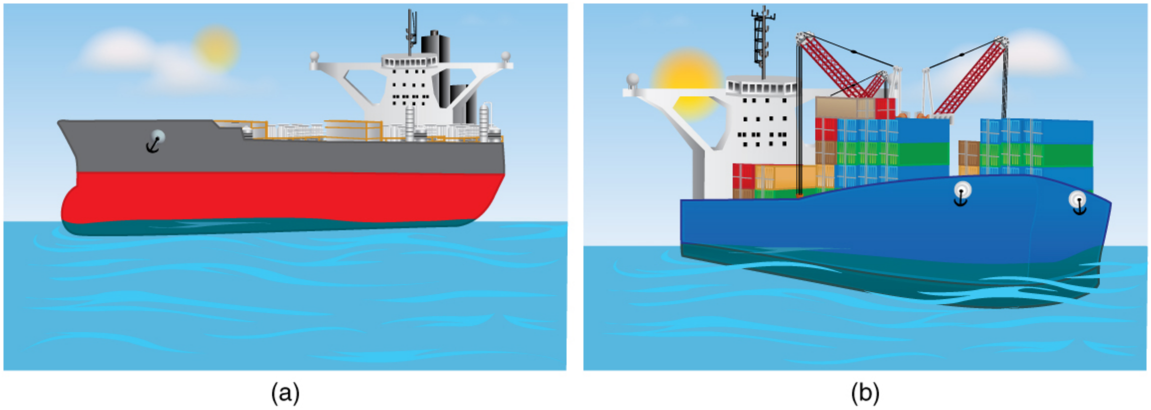


Figure 6.6.4: An unloaded ship (a) floats higher in the water than a loaded ship (b).

We use this last relationship to measure densities. This is done by measuring the fraction of a floating object that is submerged—for example, with a hydrometer. It is useful to define the ratio of the density of an object to a fluid (usually water) as specific gravity:

$$specific\,gravity = \frac{\bar{\rho}}{\rho_W}, \tag{6.6.15}$$

where $\bar{\rho}$ is the average density of the object or substance and ρ_W is the density of water at 4.00°C. Specific gravity is dimensionless, independent of whatever units are used for ρ. If an object floats, its specific gravity is less than one. If it sinks, its specific gravity is greater than one. Moreover, the fraction of a floating object that is submerged equals its specific gravity. If an object's specific gravity is exactly 1, then it will remain suspended in the fluid, neither sinking nor floating. Scuba divers try to obtain this state so that they can hover in the water. We measure the specific gravity of fluids, such as battery acid, radiator fluid, and urine, as an indicator of their condition. One device for measuring specific gravity is shown in Figure 6.6.5.

Definition: Specific Gravity

Specific gravity is the ratio of the density of an object to a fluid (usually water).

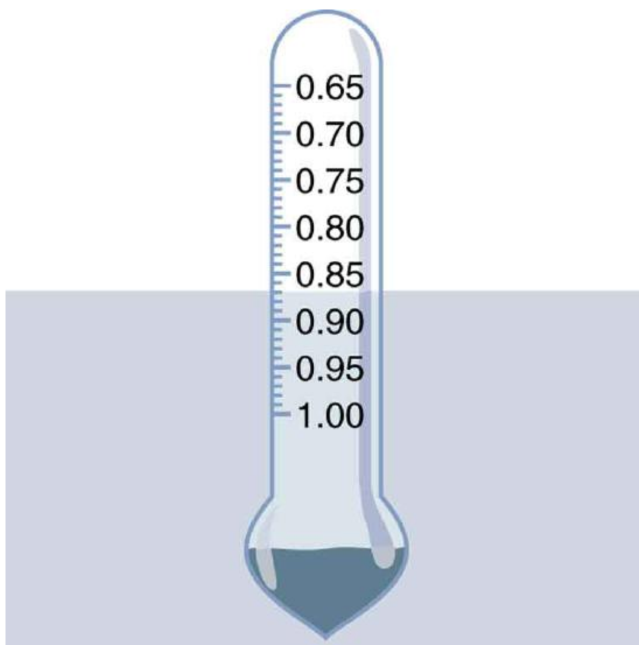


Figure 6.6.5: This hydrometer is floating in a fluid of specific gravity 0.87. The glass hydrometer is filled with air and weighted with lead at the bottom. It floats highest in the densest fluids and has been calibrated and labeled so that specific gravity can be read from it directly.

Example 6.6.2: Calculating Average Density: Floating Woman

Suppose a 60.0-kg woman floats in freshwater with 97.0% of her volume submerged when her lungs are full of air. What is her average density?

Strategy

We can find the woman's density by solving the equation

$$fraction\ submerged = \frac{\bar{\rho}_{obj}}{\rho_{fl}} \tag{6.6.16}$$

for the density of the object. This yields

$$\bar{\rho}_{obj} = \bar{\rho}_{person} = (fraction\ submerged) \cdot \rho_{fl}. \tag{6.6.17}$$

We know both the fraction submerged and the density of water, and so we can calculate the woman's density.

Solution

Entering the known values into the expression for her density, we obtain

$$\bar{\rho}_{person} = 0.970 \cdot \left(10^3 \frac{kg}{m^3}\right) = 970 \frac{kg}{m^3}. \tag{6.6.18}$$

Discussion

Her density is less than the fluid density. We expect this because she floats. Body density is one indicator of a person's percent body fat, of interest in medical diagnostics and athletic training. (See Figure $6.6.7$)

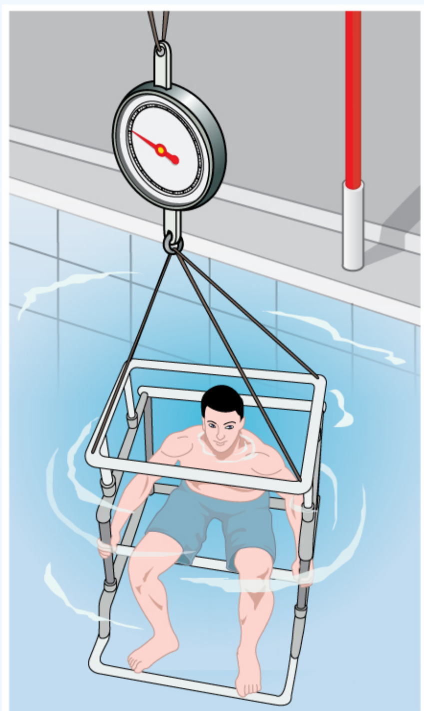


Figure $6.6.6$: Subject in a "fat tank," where he is weighed while completely submerged as part of a body density determination. The subject must completely empty his lungs and hold a metal weight in order to sink. Corrections are made for the residual air in his lungs (measured separately) and the metal weight. His corrected submerged weight, his weight in air, and pinch tests of strategic fatty areas are used to calculate his percent body fat.

There are many obvious examples of lower-density objects or substances floating in higher-density fluids—oil on water, a hot-air balloon, a bit of cork in wine, an iceberg, and hot wax in a "lava lamp," to name a few. Less obvious examples include lava rising in a volcano and mountain ranges floating on the higher-density crust and mantle beneath them. Even seemingly solid Earth has fluid characteristics.

More Density Measurements

One of the most common techniques for determining density is shown in Figure $6.6.7$. An object, here a coin, is weighed in air and then weighed again while submerged in a liquid. The density of the coin, an indication of its authenticity, can be calculated if the fluid density is known. This same technique can also be used to determine the density of the fluid if the density of the coin is known. All of these calculations are based on Archimedes' principle.

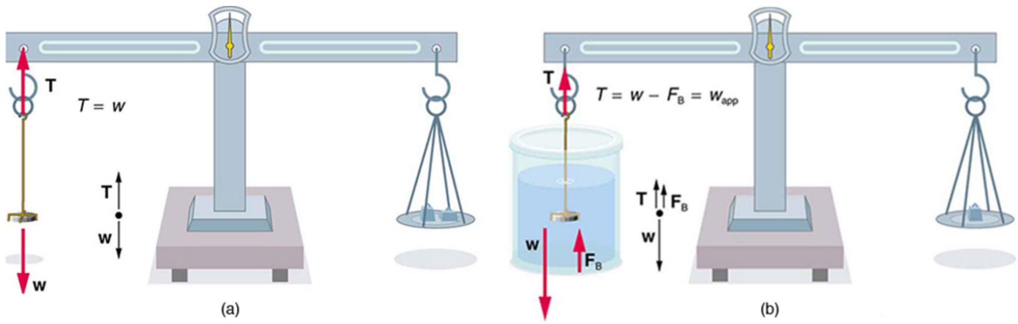


Figure $6.6.7$: (a) A coin is weighed in air. (b) The apparent weight of the coin is determined while it is completely submerged in a fluid of known density. These two measurements are used to calculate the density of the coin.

Archimedes' principle states that the buoyant force on the object equals the weight of the fluid displaced. This, in turn, means that the object *appears* to weigh less when submerged; we call this measurement the object's *apparent weight*. The object suffers an *apparent weight loss* equal to the weight of the fluid displaced. Alternatively, on balances that measure mass, the object suffers an *apparent mass loss* equal to the mass of fluid displaced. That is

$$apparent\, weight\, loss = weight\, of\, fluid\, displaced \tag{6.6.19}$$

or

$$apparent\, mass\, loss = mass\, of\, fluid\, displaced. \tag{6.6.20}$$

The next example illustrates the use of this technique.

Example $6.6.3$: Calculating Density: Is the Coin Authentic?

The mass of an ancient Greek coin is determined in air to be 8.630 g. When the coin is submerged in water as shown in Figure $6.6.7$, its apparent mass is 7.800 g. Calculate its density, given that water has a density of $1.000\, g/m^3$

and that effects caused by the wire suspending the coin are negligible.

Strategy

To calculate the coin's density, we need its mass (which is given) and its volume. The volume of the coin equals the volume of water displaced. The volume of water displaced $\rho = \frac{m}{V}$ for V.

Solution

The volume of water is $V_W = \frac{m_W}{\rho_W}$ where m_W is the mass of water displaced. As noted, the mass of the water displaced equals the apparent mass loss, which is $m_W = 8.630\, g - 7.800\, g = 0.830\, g$. Thus the volume of water is $V_W = \frac{0.830\, g}{1.000\, g/cm^3} = 0.830\, cm^3$. This is also the volume of the coin, since it is completely submerged. We can now find the density of the coin using the definition of density:

$$\rho_c = \frac{m_c}{V_c} = \frac{8.630\, g}{0.830\, cm^3} = 10.4\, g/cm^3. \tag{6.6.21}$$

Discussion

You can see from [link] that this density is very close to that of pure silver, appropriate for this type of ancient coin. Most modern counterfeits are not pure silver.

This brings us back to Archimedes' principle and how it came into being. As the story goes, the king of Syracuse gave Archimedes the task of determining whether the royal crown maker was supplying a crown of pure gold. The purity of gold is difficult to determine by color (it can be diluted with other metals and still look as yellow as pure gold), and other analytical techniques had not yet been conceived. Even ancient peoples, however, realized that the density of gold was greater than that of any other then-known substance. Archimedes purportedly agonized over his task and had his inspiration one day while at the public baths, pondering the support the water gave his body. He came up with his now-famous principle, saw how to apply it to determine density, and ran naked down the streets of Syracuse crying "Eureka!" (Greek for "I have found it"). Similar behavior can be observed in contemporary physicists from time to time!

PhET Explorations: Buoyancy

When will objects float and when will they sink? Learn how buoyancy works with blocks. Arrows show the applied forces, and you can modify the properties of the blocks and the fluid.

Summary

- Buoyant force is the net upward force on any object in any fluid. If the buoyant force is greater than the object's weight, the object will rise to the surface and float. If the buoyant force is less than the object's weight, the object will sink. If the buoyant force equals the object's weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.
- Archimedes' principle states that the buoyant force on an object equals the weight of the fluid it displaces.
- Specific gravity is the ratio of the density of an object to a fluid (usually water).

Glossary

Archimedes' principle
the buoyant force on an object equals the weight of the fluid it displaces

buoyant force
the net upward force on any object in any fluid

specific gravity
the ratio of the density of an object to a fluid (usually water)

Contributors and Attributions

6.7: Cohesion and Adhesion in Liquids - Surface Tension and Capillary Action

Learning Objectives

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

- Understand cohesive and adhesive forces.
- Define surface tension.
- Understand capillary action.

Children blow soap bubbles and play in the spray of a sprinkler on a hot summer day (Figure $6.7.1$). An underwater spider keeps his air supply in a shiny bubble he carries wrapped around him. A technician draws blood into a small-diameter tube just by touching it to a drop on a pricked finger. A premature infant struggles to inflate her lungs. What is the common thread? All these activities are dominated by the attractive forces between atoms and molecules in liquids—both within a liquid and between the liquid and its surroundings.

Figure $6.7.1$:The soap bubbles in this photograph are caused by cohesive forces among molecules in liquids. (credit: Steve Ford Elliott)

Attractive forces between molecules of the same type are called cohesive forces. Liquids can, for example, be held in open containers because cohesive forces hold the molecules together. Attractive forces between molecules of different types are called adhesive forces. Such forces cause liquid drops to cling to window panes, for example. In this section we examine effects directly attributable to cohesive and adhesive forces in liquids.

Definition: Cohesive Forces

Attractive forces between molecules of the same type are called cohesive forces.

Definition: Adhesive Forces

Attractive forces between molecules of different types are called adhesive forces.

Surface Tension

Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension. Molecules on the surface are pulled inward by cohesive forces, reducing the surface area. Molecules inside the liquid experience zero net force, since they have neighbors on all sides.

Definition: Surface Tension

Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension.

Surface Tension

Forces between atoms and molecules underlie the macroscopic effect called surface tension. These attractive forces pull the molecules closer together and tend to minimize the surface area. This is another example of a submicroscopic explanation for a macroscopic phenomenon.

The model of a liquid surface acting like a stretched elastic sheet can effectively explain surface tension effects. For example, some insects can walk on water (as opposed to floating in it) as we would walk on a trampoline—they dent the surface as shown in Figure $6.7.2a$. Figure $6.7.2b$ shows another example, where a needle rests on a water surface. The iron needle cannot, and does not, float, because its density is greater than that of water. Rather, its weight is supported by forces in the stretched surface that try to make the surface smaller or flatter. If the needle were placed point down on the surface, its weight acting on a smaller area would break the surface, and it would sink.

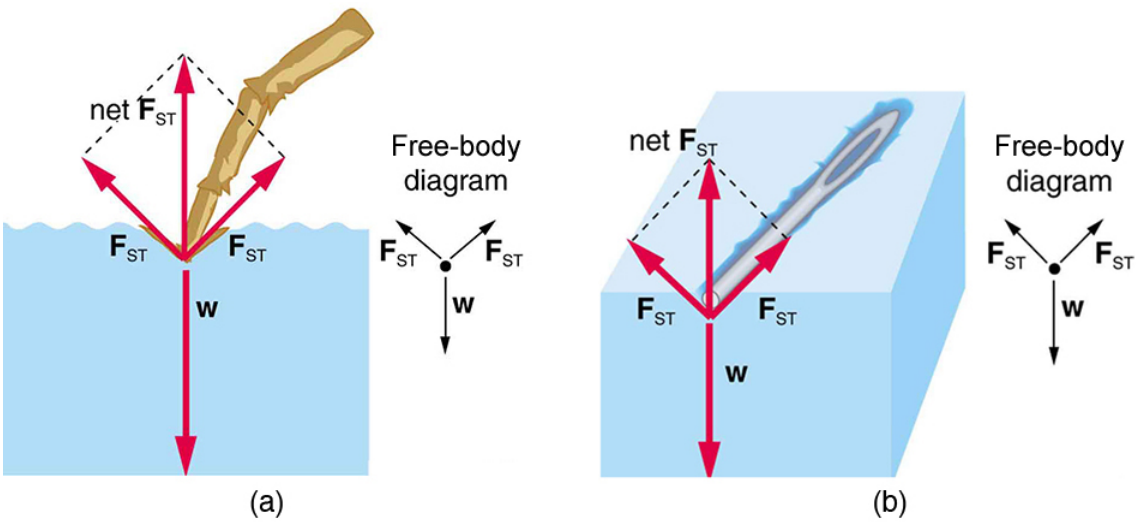


Figure $6.7.2$: Surface tension supporting the weight of an insect and an iron needle, both of which rest on the surface without penetrating it. They are not floating; rather, they are supported by the surface of the liquid. (a) An insect leg dents the water surface. F_{ST} is a restoring force (surface tension) parallel to the surface. (b) An iron needle similarly dents a water surface until the restoring force (surface tension) grows to equal its weight.

Surface tension is proportional to the strength of the cohesive force, which varies with the type of liquid. Surface tension $\overline{\gamma}$ is defined to be the force F per unit length L exerted by a stretched liquid membrane:

$$\gamma = \frac{F}{L}. \tag{6.7.1}$$

Table $6.7.1$ lists values of $\overline{\gamma}$ for some liquids.

Table 6.7.1

Liquid	Surface tension γ(N/m)
Water at 0^oC	0.0756
Water at 20^oC	0.0728
Water at 100^OC	0.0589
Soapy water (typical)	0.0370
Ethyl alcohol	0.0223
Glycerin	0.0631

Liquid	Surface tension γ(N/m)
Mercury	0.465
Olive Oil	0.032
Tissue fluids (typical)	0.050
Blood, whole at 37^oC	0.058
Blood plasma at 37^oC	0.073
Gold at 1070^oC	1.000
Oxygen at -193^oC	0.0157
Helium at -269^oC	0.00012

For the insect of Figure Figure $6.7.1a$, its weight w is supported by the upward components of the surface tension force: $w = \gamma L \sin\theta$, where L is the circumference of the insect's foot in contact with the water. Figure $6.7.3$ shows one way to measure surface tension. The liquid film exerts a force on the movable wire in an attempt to reduce its surface area. The magnitude of this force depends on the surface tension of the liquid and can be measured accurately.

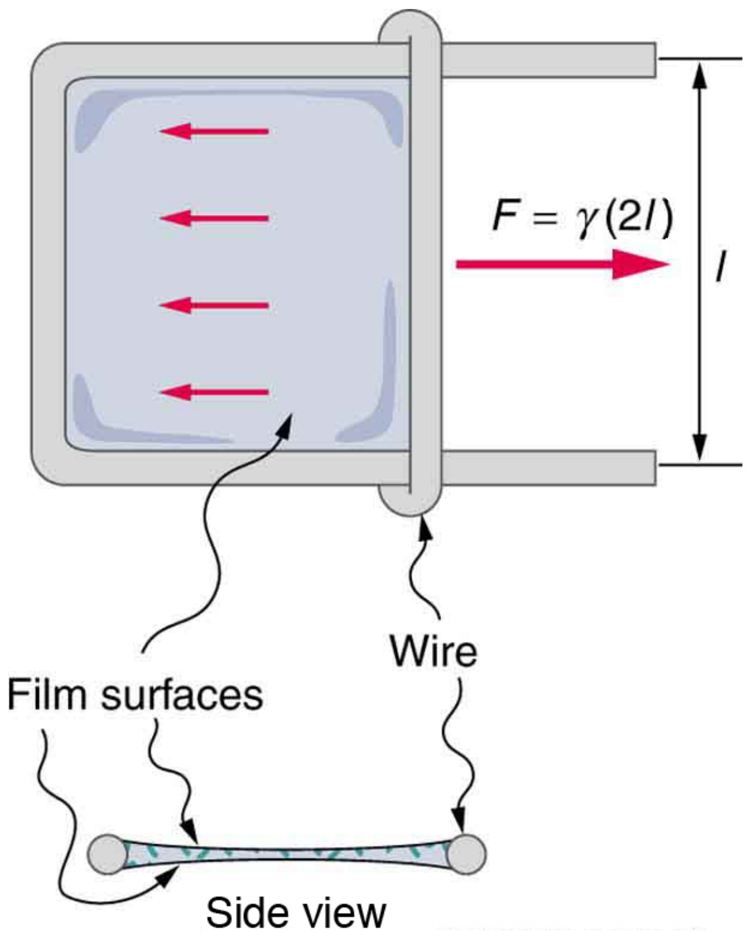


Figure $6.7.3$: Sliding wire device used for measuring surface tension; the device exerts a force to reduce the film's surface area. The force needed to hold the wire in place is $F = \gamma L = \gamma(2l)$, since there are two liquid surfaces attached to the wire. This force remains nearly constant as the film is stretched, until the film approaches its breaking point.

Surface tension is the reason why liquids form bubbles and droplets. The inward surface tension force causes bubbles to be approximately spherical and raises the pressure of the gas trapped inside relative to atmospheric pressure outside. It can be shown that the gauge pressure P inside a spherical bubble is given by

$$P=\frac{4\gamma}{r}, \tag{6.7.2}$$

where r is the radius of the bubble. Thus the pressure inside a bubble is greatest when the bubble is the smallest. Another bit of evidence for this is illustrated in Figure $6.7.4$. When air is allowed to flow between two balloons of unequal size, the smaller balloon tends to collapse, filling the larger balloon.

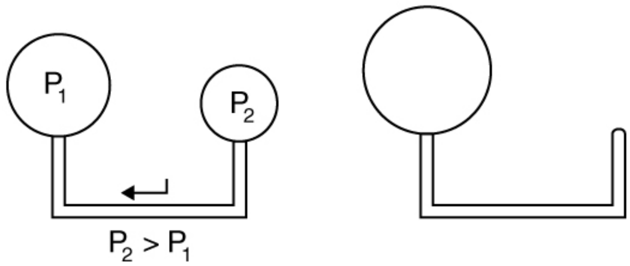


Figure $6.7.4$: With the valve closed, two balloons of different sizes are attached to each end of a tube. Upon opening the valve, the smaller balloon decreases in size with the air moving to fill the larger balloon. The pressure in a spherical balloon is inversely proportional to its radius, so that the smaller balloon has a greater internal pressure than the larger balloon, resulting in this flow.

Example 6.7.1: Surface Tension: Pressure Inside a Bubble

Calculate the gauge pressure inside a soap bubble $2.00\times 10^{-4} m$ in radius using the surface tension for soapy water in Table. Convert this pressure to mm Hg.

Strategy

The radius is given and the surface tension can be found in Table, and so P can be found directly from the equation $P=\frac{4\gamma}{r}$.

Solution

Substituting r and γ into this equation $P=\frac{4\gamma}{r}$, we obtaiin

$$P=\frac{4\gamma}{r}=\frac{4(0.037\,N/m)}{2.00\times 10^{-4} m}=740\,N/m^2=740\,Pa. \tag{6.7.3}$$

We use a conversion factor to get this into units of mm Hg:

$$P=(740\,N/m^2)\frac{1.00\,mm\,Hg}{133\,N/m^2}=5.56\,mm\,Hg. \tag{6.7.4}$$

Discussion

Note that if a hole were to be made in the bubble, the air would be forced out, the bubble would decrease in radius, and the gauge pressure would reduce to zero, and the absolute pressure inside would *decrease* to atmospheric pressure (760 mm Hg).

Our lungs contain hundreds of millions of mucus-lined sacs called *alveoli*, which are very similar in size, and about 0.1 mm in diameter. (See Figure.) You can exhale without muscle action by allowing surface tension to contract these sacs. Medical patients whose breathing is aided by a positive pressure respirator have air blown into the lungs, but are generally allowed to exhale on their own. Even if there is paralysis, surface tension in the alveoli will expel air from the lungs. Since pressure increases as the radii of the alveoli decrease, an occasional deep cleansing breath is needed to fully reinflate the alveoli. Respirators are programmed to do this and we find it natural, as do our companion dogs and cats, to take a cleansing breath before settling into a nap.

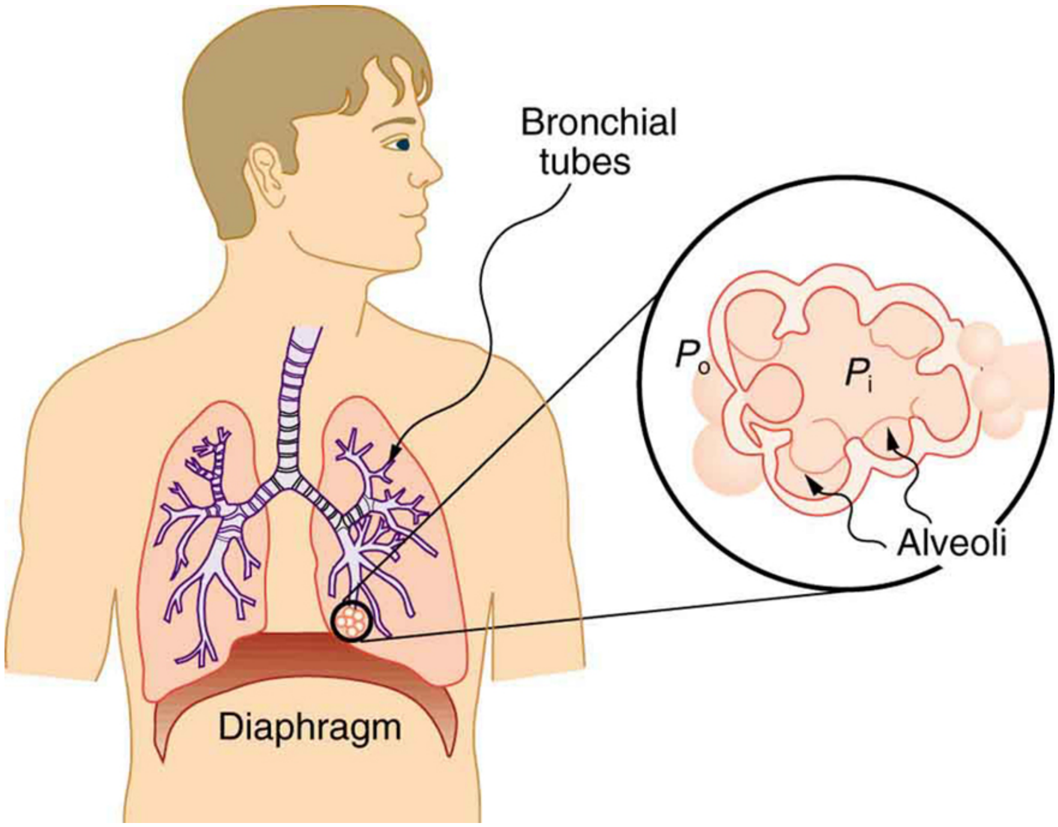


Figure 6.7.5: Bronchial tubes in the lungs branch into ever-smaller structures, finally ending in alveoli. The alveoli act like tiny bubbles. The surface tension of their mucous lining aids in exhalation and can prevent inhalation if too great.

The tension in the walls of the alveoli results from the membrane tissue and a liquid on the walls of the alveoli containing a long lipoprotein that acts as a surfactant (a surface-tension reducing substance). The need for the surfactant results from the tendency of small alveoli to collapse and the air to fill into the larger alveoli making them even larger (as demonstrated in Figure). During inhalation, the lipoprotein molecules are pulled apart and the wall tension increases as the radius increases (increased surface tension). During exhalation, the molecules slide back together and the surface tension decreases, helping to prevent a collapse of the alveoli. The surfactant therefore serves to change the wall tension so that small alveoli don't collapse and large alveoli are prevented from expanding too much. This tension change is a unique property of these surfactants, and is not shared by detergents (which simply lower surface tension). (See Figure.)

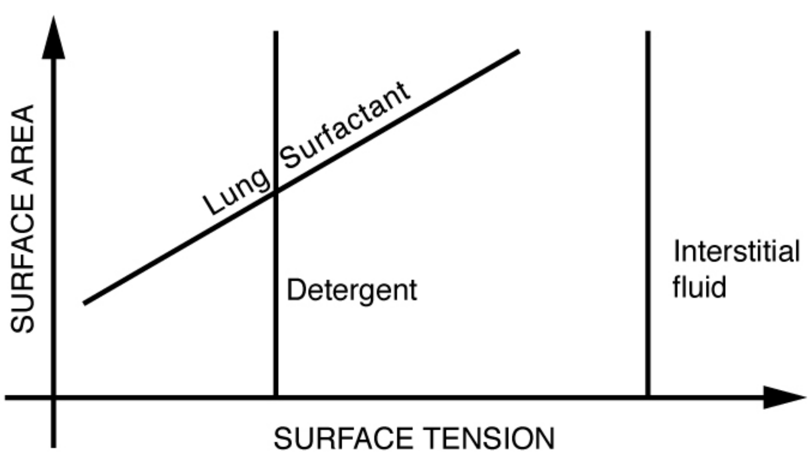


Figure 6.7.6: Surface tension as a function of surface area. The surface tension for lung surfactant decreases with decreasing area. This ensures that small alveoli don't collapse and large alveoli are not able to over expand.

If water gets into the lungs, the surface tension is too great and you cannot inhale. This is a severe problem in resuscitating drowning victims. A similar problem occurs in newborn infants who are born without this surfactant—their lungs are very difficult to inflate. This condition is known as *hyaline membrane disease* and is a leading cause of death for infants, particularly in premature births. Some success has been achieved in treating hyaline membrane disease by spraying a surfactant into the infant's breathing passages. Emphysema produces the opposite problem with alveoli. Alveolar walls of emphysema victims deteriorate, and the sacs combine to form larger sacs. Because pressure produced by surface tension decreases with increasing radius, these larger sacs produce smaller pressure, reducing the ability of emphysema victims to exhale. A common test for emphysema is to measure the pressure and volume of air that can be exhaled.

Making Connections: Take-Home Investigation

1. Try floating a sewing needle on water. In order for this activity to work, the needle needs to be very clean as even the oil from your fingers can be sufficient to affect the surface properties of the needle.
2. Place the bristles of a paint brush into water. Pull the brush out and notice that for a short while, the bristles will stick together. The surface tension of the water surrounding the bristles is sufficient to hold the bristles together. As the bristles dry out, the surface tension effect dissipates.
3. Place a loop of thread on the surface of still water in such a way that all of the thread is in contact with the water. Note the shape of the loop. Now place a drop of detergent into the middle of the loop. What happens to the shape of the loop? Why?
4. Sprinkle pepper onto the surface of water. Add a drop of detergent. What happens? Why?
5. Float two matches parallel to each other and add a drop of detergent between them. What happens? Note: For each new experiment, the water needs to be replaced and the bowl washed to free it of any residual detergent.

Summary

- Attractive forces between molecules of the same type are called cohesive forces.
- Attractive forces between molecules of different types are called adhesive forces.
- Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension.

Glossary

adhesive forces
the attractive forces between molecules of different type

cohesive forces
the attractive forces between molecules of the same type

surface tension
the cohesive forces between molecules which cause the surface of a liquid to contract to the smallest possible surface area

Contributors and Attributions

CHAPTER OVERVIEW

7: Electricity

Contributors and Attributions

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

7.1: Prelude 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. (Figure 7.1.1) 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 7.1.1: 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)

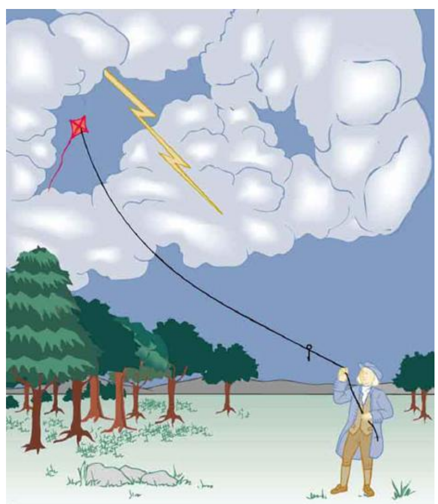


Figure 7.1.2: When Benjamin Franklin demonstrated that lightning was related to static electricity, he made a connection that is now part of the evidence that all directly experienced forces except the gravitational force are manifestations of the electromagnetic force.

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.

Glossary

static electricity

a buildup of electric charge on the surface of an object

electromagnetic force

one of the four fundamental forces of nature; the electromagnetic force consists of static electricity, moving electricity and magnetism

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

7.2: 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.

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 (Figure $7.2.1$). The very word *electric* derives from the Greek word for amber (*electron*).

Figure $7.2.1$: 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).

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 $7.2.2$ shows how these simple materials can be used to explore the nature of the force between charges.

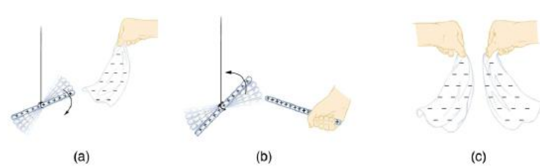


Figure $7.2.2$: A glass rod becomes positively charged when rubbed with silk, while the silk becomes negatively charged. (a) The glass rod is attracted to the silk because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

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 $7.2.3$ 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.

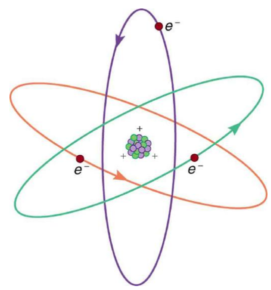


Figure $7.2.3$: This simplified (and not to scale) view of an atom is called the planetary model of the atom. Negative electrons orbit a much heavier positive nucleus, as the planets orbit the much heavier sun. There the similarity ends, because forces in the atom are electromagnetic, whereas those in 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

$$|q_e| = 1.60 \times 10^{-19} C$$

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

$$1.00C \times \frac{1 proton}{1.60 \times 10^{-19} C} = 6.25 \times 10^{18} protons$$

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. (Figure $7.2.4$) 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 $7.2.4$ 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.

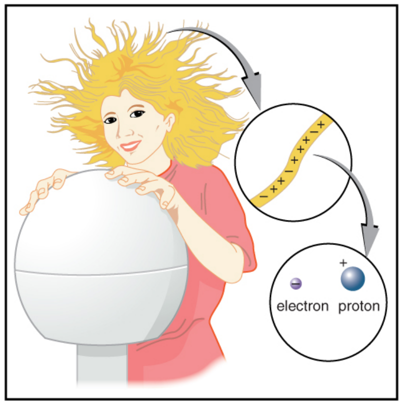


Figure $7.2.4$: When this person touches a Van de Graaff generator, some electrons are attracted to the generator, resulting in an excess of positive charge, causing her hair to stand on end. The charges in one hair are shown. An artist's conception of an electron and a proton illustrate the particles 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 $7.2.5$. 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.

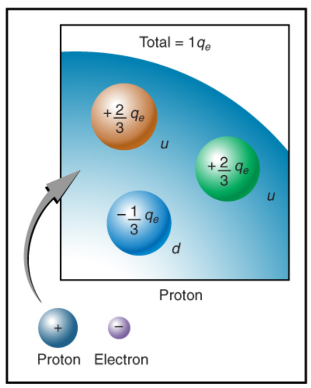


Figure $7.2.5$: Artist's conception of fractional quark charges inside a proton. A group of three quark charges add up to the single positive charge on the proton: $-\frac{1}{3}q_e + \frac{2}{3}q_e + \frac{2}{3}q_e = +1q_e$.

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. (Figure $7.2.6$) 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.

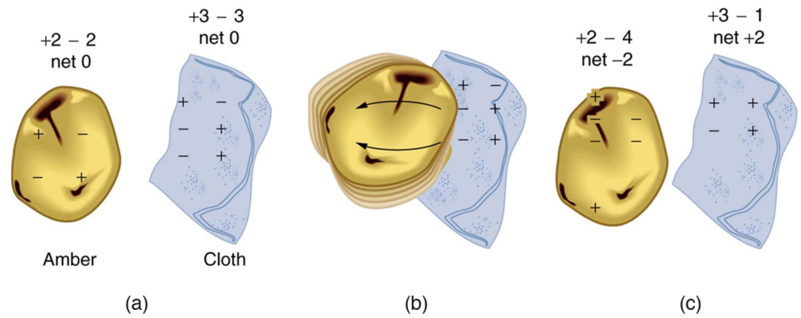


Figure 7.2.6: When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, 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. (Figure 7.2.7) 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.

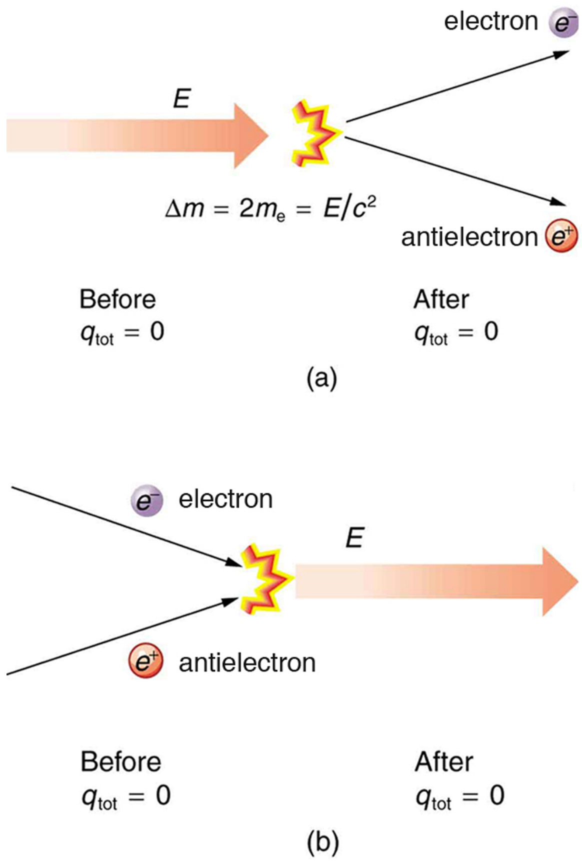


Figure 7.2.7: (a) When enough energy is present, it can be converted into matter. Here the matter created is an electron–antielectron pair. (m_e is the electron's mass.) The total charge before and after this event is zero. (b) When matter and antimatter collide, they annihilate each other; the total 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.

Summary

- There are only two types of charge, which we call positive and negative.
- Like charges repel, unlike charges attract, and the force between charges decreases with the square of the distance.
- The vast majority of positive charge in nature is carried by protons, while the vast majority of negative charge is carried by electrons.
- The electric charge of one electron is equal in magnitude and opposite in sign to the charge of one proton.
- An ion is an atom or molecule that has nonzero total charge due to having unequal numbers of electrons and protons.
- The SI unit for charge is the coulomb (C), with protons and electrons having charges of opposite sign but equal magnitude; the magnitude of this basic charge $|q_e|$ is
- $|q_e| = 1.60 \times 10^{-19} C$
- Whenever charge is created or destroyed, equal amounts of positive and negative are involved.
- Most often, existing charges are separated from neutral objects to obtain some net charge.
- Both positive and negative charges exist in neutral objects and can be separated by rubbing one object with another. For macroscopic objects, negatively charged means an excess of electrons and positively charged means a depletion of electrons.
- The law of conservation of charge ensures that whenever a charge is created, an equal charge of the opposite sign is created at the same time.

Glossary

electric charge

a physical property of an object that causes it to be attracted toward or repelled from another charged object; each charged object generates and is influenced by a force called an electromagnetic force

law of conservation of charge

states that whenever a charge is created, an equal amount of charge with the opposite sign is created simultaneously

electron

a particle orbiting the nucleus of an atom and carrying the smallest unit of negative charge

proton

a particle in the nucleus of an atom and carrying a positive charge equal in magnitude and opposite in sign to the amount of negative charge carried by an electron

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

7.3: 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.

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.

Figure $7.3.1$: 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)

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.

Charging by Contact

Figure $7.3.2$ 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.

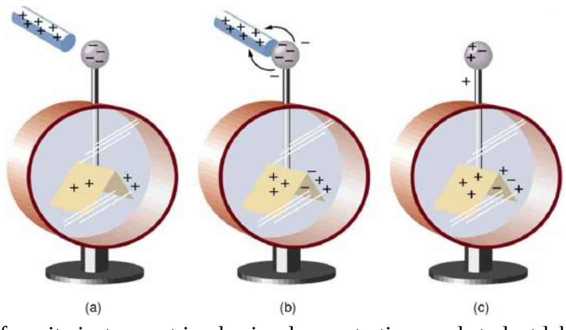


Figure $7.3.2$: An electroscope is a favorite instrument in physics demonstrations and student laboratories. It is typically made with gold foil leaves hung from a (conducting) metal stem and is insulated from the room air in a glass-walled container. (a) A positively charged glass rod is brought near 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.

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 $7.3.3$ 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.

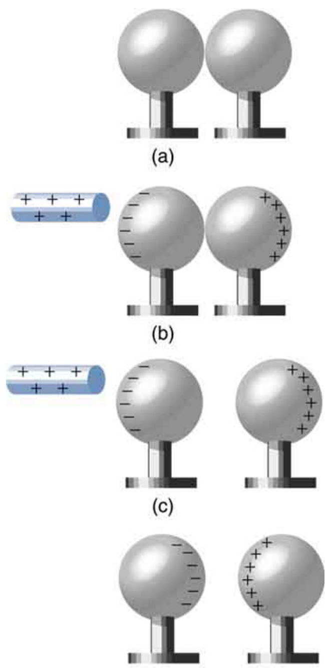


Figure $7.3.3$: Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other sphere positively charged. (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.

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.

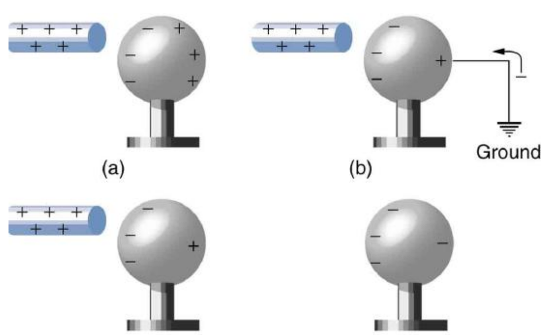


Figure $7.3.4$: Charging by induction, using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from the earth's ample supply. (c) The ground connection is broken. (d) The positive rod is removed, leaving the sphere with an induced negative charge.

Another method of charging by induction is shown in Figure 7.3.4. 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.

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 7.3.5 shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

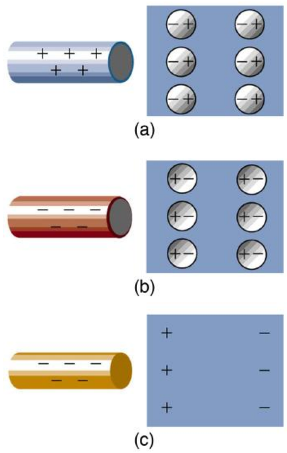


Figure 7.3.5: Both positive and negative objects attract a neutral object by polarizing its molecules. (a) A positive object brought near a neutral insulator polarizes its molecules. There is a slight shift in the distribution of the electrons orbiting the molecule, with unlike charges being brought 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.

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.

Check Your Understanding

Can you explain the attraction of water to the charged rod in the figure below?

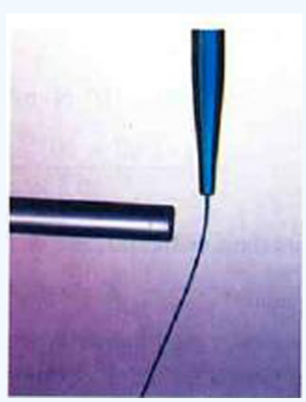


Figure 7.3.6.

Solution

Water molecules are polarized, giving them slightly positive and slightly negative sides. This makes water even more susceptible to a charged rod's attraction. As the water flows downward, due to the force of gravity, the charged conductor exerts a net attraction to the opposite charges in the stream of water, pulling it closer.

PHET EXPLORATIONS: JOHN TRAVOLTAGE

Make sparks fly with John Travoltage. Wiggle Johnnie's foot and he picks up charges from the carpet. Bring his hand close to the door knob and get rid of the excess charge.

Summary

- Polarization is the separation of positive and negative charges in a neutral object.
- A conductor is a substance that allows charge to flow freely through its atomic structure.
- An insulator holds charge within its atomic structure.
- Objects with like charges repel each other, while those with unlike charges attract each other.
- A conducting object is said to be grounded if it is connected to the Earth through a conductor. Grounding allows transfer of charge to and from the earth's large reservoir.
- Objects can be charged by contact with another charged object and obtain the same sign charge.
- If an object is temporarily grounded, it can be charged by induction, and obtains the opposite sign charge.
- Polarized objects have their positive and negative charges concentrated in different areas, giving them a non-symmetrical charge.
- Polar molecules have an inherent separation of charge.

Glossary

free electron
an electron that is free to move away from its atomic orbit

conductor
a material that allows electrons to move separately from their atomic orbits

insulator
a material that holds electrons securely within their atomic orbits

grounded
when a conductor is connected to the Earth, allowing charge to freely flow to and from Earth's unlimited reservoir

induction
the process by which an electrically charged object brought near a neutral object creates a charge in that object

polarization
slight shifting of positive and negative charges to opposite sides of an atom or molecule

electrostatic repulsion
the phenomenon of two objects with like charges repelling each other

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney).

7.4: Coulomb's Law

Learning Objectives

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

- State Coulomb's law in terms of how the electrostatic force changes with the distance between two objects.
- Calculate the electrostatic force between two charged point forces, such as electrons or protons.
- Compare the electrostatic force to the gravitational attraction for a proton and an electron; for a human and the Earth.

Through the work of scientists in the late 18th century, the main features of the **electrostatic force**—the existence of two types of charge, the observation that like charges repel, unlike charges attract, and the decrease of force with distance—were eventually refined, and expressed as a mathematical formula. The mathematical formula for the electrostatic force is called **Coulomb's law** after the French physicist Charles Coulomb (1736–1806), who performed experiments and first proposed a formula to calculate it.

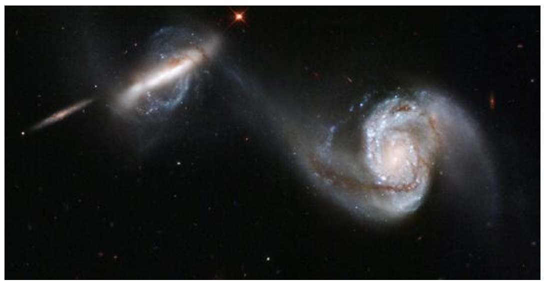


Figure $7.4.1$: This NASA image of Arp 87 shows the result of a strong gravitational attraction between two galaxies. In contrast, at the subatomic level, the electrostatic attraction between two objects, such as an electron and a proton, is far greater than their mutual attraction due to gravity. (credit: NASA/HST)

Definition: Coulomb's Law

Coulomb's law calculates the magnitude of the force F between two point charges, q_1 and q_2, separated by a distance r.

$$F = k\frac{|q_1 q_2|}{r^2}. \tag{7.4.1}$$

In SI units, the constant$k$ is equal to

$$k = 8.988 \times 10^9 \frac{N \cdot m^2}{C^2} \approx 8.99 \times 10^9 \frac{N \cdot m^2}{C^2}. \tag{7.4.2}$$

The electrostatic force is a vector quantity and is expressed in units of newtons. The force is understood to be along the line joining the two charges. (Figure $7.4.2$)

Although the formula for Coulomb's law is simple, it was no mean task to prove it. The experiments Coulomb did, with the primitive equipment then available, were difficult. Modern experiments have verified Coulomb's law to great precision. For example, it has been shown that the force is inversely proportional to distance between two objects squared $(F \propto 1/r^2)$ to an accuracy of 1 part in 10^{16}. No exceptions have ever been found, even at the small distances within the atom.

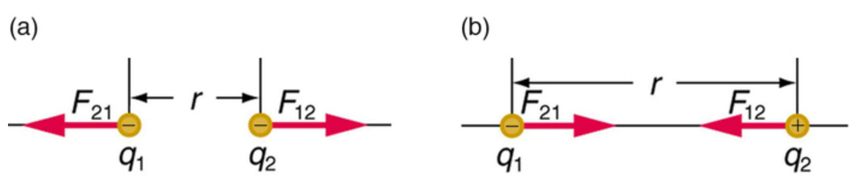


Figure $7.4.2$: The magnitude of the electrostatic force$F$ between point charges q_1 and q_2 separated by a distance r is given by Coulomb's law. Note that Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2. (a) Like charges. (b) Unlike charges.

Example 7.4.1: How Strong is the Coulomb Force Relative to the Gravitational Force?

Compare the electrostatic force between an electron and proton separated by $0.530 \times 10^{-10} m$ with the gravitational force between them. This distance is their average separation in a hydrogen atom.

Strategy

To compare the two forces, we first compute the electrostatic force using Coulomb's law, $F = k\frac{|q_1 q_2}{r^2}$. We then calculate the gravitational force using Newton's universal law of gravitation. Finally, we take a ratio to see how the forces compare in magnitude.

Solution

Entering the given and known information about the charges and separation of the electron and proton into the expression of Coulomb's law yields

$$F = k\frac{|q_1 q_2|}{r^2}$$

$$= (8.99 \times 10^9 N \cdot m^2/C^2) \times \frac{(1.60 \times 10^{-19} C)(1.60 \times 10^{-19} C)}{(0.530 \times 10^{-10} m)^2}$$

Thus the Coulomb force is

$$F = 8.19 \times 10^{-8} N.$$

The charges are opposite in sign, so this is an attractive force. This is a very large force for an electron—it would cause an acceleration of $8.99 \times 10^{22} m/s^2$(verification is left as an end-of-section problem).The gravitational force is given by Newton's law of gravitation as:

$$F_G = G\frac{mM}{r^2},$$

where $G = 6.67 \times 10^{-11} N \cdot m^2/kg^2$. Here m and M represent the electron and proton masses, which can be found in the appendices. Entering values for the knowns yields

$$F_G = (6.67 \times 10^{-11} N \cdot m^2/kg^2) \times \frac{(9.11 \times 10^{-31} kg)(1.67 \times 10^{-27} kg)}{(0.530 \times 10^{-10} m)^2} = 3.61 \times 10^{-47} N$$

This is also an attractive force, although it is traditionally shown as positive since gravitational force is always attractive. The ratio of the magnitude of the electrostatic force to gravitational force in this case is, thus,

$$\frac{F}{F_G} = 2.27 \times 10^{39}.$$

Discussion

This is a remarkably large ratio! Note that this will be the ratio of electrostatic force to gravitational force for an electron and a proton at any distance (taking the ratio before entering numerical values shows that the distance cancels). This ratio gives some indication of just how much larger the Coulomb force is than the gravitational force between two of the most common particles in nature.

As the example implies, gravitational force is completely negligible on a small scale, where the interactions of individual charged particles are important. On a large scale, such as between the Earth and a person, the reverse is true. Most objects are nearly electrically neutral, and so attractive and repulsive **Coulomb forces** nearly cancel. Gravitational force on a large scale dominates interactions between large objects because it is always attractive, while Coulomb forces tend to cancel.

Summary

- Frenchman Charles Coulomb was the first to publish the mathematical equation that describes the electrostatic force between two objects.

- Coulomb's law gives the magnitude of the force between point charges. It is $F = k\frac{|q_1 q_2|}{r^2}$, where q_1 and q_2 are two point charges separated by a distance r, and $k \approx 8.99 \times 10^9 N \cdot m^2/C^2$
- This Coulomb force is extremely basic, since most charges are due to point-like particles. It is responsible for all electrostatic effects and underlies most macroscopic forces.
- The Coulomb force is extraordinarily strong compared with the gravitational force, another basic force—but unlike gravitational force it can cancel, since it can be either attractive or repulsive.
- The electrostatic force between two subatomic particles is far greater than the gravitational force between the same two particles.

Glossary

Coulomb's law
the mathematical equation calculating the electrostatic force vector between two charged particles

Coulomb force
another term for the electrostatic force

electrostatic force
the amount and direction of attraction or repulsion between two charged bodies

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

7.5: Introduction to Electric Potential and Electric Energy

In Electric Charge and Electric Field, we just scratched the surface (or at least rubbed it) of electrical phenomena. Two of the most familiar aspects of electricity are its energy and *voltage*. We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at molecular levels, *ions* cross cell membranes and transfer information. We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. A motorcycle battery, for example, is small and would not be very successful in replacing the much larger car battery, yet each has the same voltage. In this chapter, we shall examine the relationship between voltage and electrical energy and begin to explore some of the many applications of electricity.

Figure 7.5.1: Automated external defibrillator unit (AED) (credit: U.S. Defense Department photo/Tech. Sgt. Suzanne M. Day)

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

7.6: Electric Potential Energy- Potential Difference

Learning Objectives

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

- Define electric potential and electric potential energy.
- Describe the relationship between potential difference and electrical potential energy.
- Explain electron volt and its usage in submicroscopic process.
- Determine electric potential energy given potential difference and amount of charge.

When a free positive charge q is accelerated by an electric field, such as shown in Figure $7.6.1$, it is given kinetic energy. The process is analogous to an object being accelerated by a gravitational field. It is as if the charge is going down an electrical hill where its electric potential energy is converted to kinetic energy. Let us explore the work done on a charge q by the electric field in this process, so that we may develop a definition of electric potential energy.

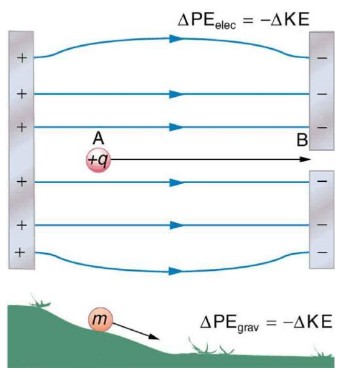


Figure $7.6.1$: A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases potential energy is converted to another form. Work is done by a force, but since this force is conservative, we can write $W = -\Delta\text{PE}$.

The electrostatic or Coulomb force is conservative, which means that the work done on q is independent of the path taken. This is exactly analogous to the gravitational force in the absence of dissipative forces such as friction. When a force is conservative, it is possible to define a potential energy associated with the force, and it is usually easier to deal with the potential energy (because it depends only on position) than to calculate the work directly.

We use the letters PE to denote electric potential energy, which has units of joules (J). The change in potential energy, ΔPE, is crucial, since the work done by a conservative force is the negative of the change in potential energy; that is, $W = -\Delta\text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

POTENTIAL ENERGY

$W = -\Delta\text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of voltage (related to electric potential energy) than to deal with the Coulomb force directly.

Calculating the work directly is generally difficult, since $W = Fd\cos\theta$ and the direction and magnitude of F can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that, since $F = qE$, the work, and hence ΔPE, is proportional to the test charge q To have a physical quantity that is independent of test charge, we define **electric potential** V (or simply potential, since electric is understood) to be the potential energy per unit charge:

$$V = \frac{\text{PE}}{q}. \tag{7.6.1}$$

ELECTRIC POTENTIAL

This is the electric potential energy per unit charge.

$$V = \frac{\text{PE}}{q} \tag{7.6.2}$$

Since PE is proportional to q, the dependence on q cancels. Thus V does not depend on q. The change in potential energy ΔPE is crucial, and so we are concerned with the difference in potential or potential difference ΔV between two points, where

$$\Delta V = V_B - V_A = \frac{\Delta\text{PE}}{q}. \tag{7.6.3}$$

The **potential difference** between points A and B, $V_B - V_A$, is thus defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

$$1\text{V} = 1\frac{\text{J}}{\text{C}} \tag{7.6.4}$$

POTENTIAL DIFFERENCE

The potential difference between points A and B, $V_B - V_A$, is defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

$$1\text{V} = 1\frac{\text{J}}{\text{C}} \tag{7.6.5}$$

The familiar term **voltage** is the common name for potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor.

In summary, the relationship between potential difference (or voltage) and electrical potential energy is given by

$$\Delta V = \frac{\Delta\text{PE}}{q} \text{ and } \Delta\text{PE} = q\Delta V. \tag{7.6.6}$$

POTENTIAL DIFFERENCE AND ELECTRICAL POTENTIAL ENERGY

The relationship between potential difference (or voltage) and electrical potential energy is given by

$$\Delta = \frac{\Delta\text{PE}}{q} \text{ and } \Delta\text{PE} = q\Delta V. \tag{7.6.7}$$

The second equation is equivalent to the first.

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other since $\Delta PE = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12 V batteries.

Example 7.6.1:Calculating Energy

Suppose you have a 12.0 V motorcycle battery that can move 5000 C of charge, and a 12.0 V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)

Strategy

To say we have a 12.0 V battery means that its terminals have a 12.0 V potential difference. When such a battery moves charge, it puts the charge through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta PE = q\Delta V$.

So to find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000\mathrm{C}$ and $\Delta = 12.0\mathrm{V}$. The total energy delivered by the motorcycle battery is

$$\Delta \mathrm{PE}_{cycle} = (5000\mathrm{C})(12.0\mathrm{V}) \tag{7.6.8}$$

$$= (5000\mathrm{C})(12.0\mathrm{J/C}) \tag{7.6.9}$$

$$= 6.00 \times 10^{4}\mathrm{J}. \tag{7.6.10}$$

Similarly, for the car battery, $q = 60,000\mathrm{C}$ and

$$\Delta \mathrm{PE}_{car} = (60,000\mathrm{C})(12.0\mathrm{V}) \tag{7.6.11}$$

$$= 7.20 \times 10^{5}\mathrm{J}. \tag{7.6.12}$$

Discussion

While voltage and energy are related, they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a low car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (A) through whatever circuitry is involved and attract them to their positive terminals (B) as shown in Figure 7.6.2. The change in potential is $\Delta V = V_B - V_A = +12\mathrm{V}$ and the charge q is negative, so that $\Delta \mathrm{PE} = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from A to B.

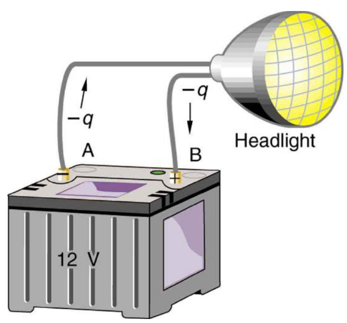


Figure 7.6.2: A battery moves negative charge from its negative terminal through a headlight to its positive terminal. Appropriate combinations of chemicals in the battery separate charges so that the negative terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative. Inside the battery, both positive and negative charges move.

Example 7.6.2: How Many Electrons Move through a Headlight Each Second?

When a 12.0 V car battery runs a single 30.0 W headlight, how many electrons pass through it each second?

Strategy

To find the number of electrons, we must first find the charge that moved in 1.00 s. The charge moved is related to voltage and energy through the equation $\Delta \mathrm{PE} = q\Delta V$. A 30.0 W lamp uses 30.0 joules per second. Since the battery loses energy, we have $\Delta \mathrm{PE} = -30.0J$ and, since the electrons are going from the negative terminal to the positive, we see that $\Delta V = +12.0V$.

Solution

To find the charge q moved, we solve the equation $\Delta\text{PE} = q\Delta V$:

$$q = \frac{\Delta\text{PE}}{\Delta V}. \tag{7.6.13}$$

Entering the values for ΔPE and ΔV, we get

$$q = \frac{-30.0\text{J}}{+12.0\text{V}} = \frac{-30.0\text{J}}{+12.0\text{J/C}} = -2.50\text{C}. \tag{7.6.14}$$

The number of electrons n_e is the total charge divided by the charge per electron. That is,

$$n_e = \frac{-2.50\text{C}}{-1.60 \times 10^{-19}\text{C/e}^-} = 1.56 \times 10^{19}\text{electrons}. \tag{7.6.15}$$

Discussion

This is a very large number. It is no wonder that we do not ordinarily observe individual electrons with so many being present in ordinary systems. In fact, electricity had been in use for many decades before it was determined that the moving charges in many circumstances were negative. Positive charge moving in the opposite direction of negative charge often produces identical effects; this makes it difficult to determine which is moving or whether both are moving.

The Electron Volt

The energy per electron is very small in macroscopic situations like that in the previous example—a tiny fraction of a joule. But on a submicroscopic scale, such energy per particle (electron, proton, or ion) can be of great importance. For example, even a tiny fraction of a joule can be great enough for these particles to destroy organic molecules and harm living tissue. The particle may do its damage by direct collision, or it may create harmful x rays, which can also inflict damage. It is useful to have an energy unit related to submicroscopic effects. Figure $7.6.3$ shows a situation related to the definition of such an energy unit. An electron is accelerated between two charged metal plates as it might be in an old-model television tube or oscilloscope. The electron is given kinetic energy that is later converted to another form—light in the television tube, for example. (Note that downhill for the electron is uphill for a positive charge.) Since energy is related to voltage by $\Delta PE = q\Delta V$ we can think of the joule as a coulomb-volt.

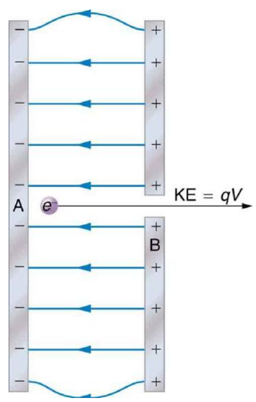


Figure $7.6.3$: A typical electron gun accelerates electrons using a potential difference between two metal plates. The energy of the electron in electron volts is numerically the same as the voltage between the plates. For example, a 5000 V potential difference produces 5000 eV electrons.

On the submicroscopic scale, it is more convenient to define an energy unit called the **electron volt** (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

$$1\text{ev} = (1.60 \times 10^{-19}\text{C})(1\text{V}) = (1.60 \times 10^{-19}\text{C})(1\text{J/C}) \tag{7.6.16}$$

$$= 1.60 \times 10^{-19} J. \tag{7.6.17}$$

ELECTRON VOLT

On the submicroscopic scale, it is more convenient to define an energy unit called the electron volt (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

$$1\text{eV} = (1.60 \times 10^{-19}\text{C})(1\text{V}) = (1.60 \times 10^{-19}\text{C})(1\text{J/C}) \tag{7.6.18}$$

$$= 1.60 \times 10^{-19}\text{C} \tag{7.6.19}$$

An electron accelerated through a potential difference of 1 V is given an energy of 1 eV. It follows that an electron accelerated through 50 V is given 50 eV. A potential difference of 100,000 V (100 kV) will give an electron an energy of 100,000 eV (100 keV), and so on. Similarly, an ion with a double positive charge accelerated through 100 V will be given 200 eV of energy. These simple relationships between accelerating voltage and particle charges make the electron volt a simple and convenient energy unit in such circumstances.

CONNECTIONS: ENERGY UNITS

The electron volt (eV) is the most common energy unit for submicroscopic processes. This will be particularly noticeable in the chapters on modern physics. Energy is so important to so many subjects that there is a tendency to define a special energy unit for each major topic. There are, for example, calories for food energy, kilowatt-hours for electrical energy, and therms for natural gas energy.

The electron volt is commonly employed in submicroscopic processes—chemical valence energies and molecular and nuclear binding energies are among the quantities often expressed in electron volts. For example, about 5 eV of energy is required to break up certain organic molecules. If a proton is accelerated from rest through a potential difference of 30 kV, it is given an energy of 30 keV (30,000 eV) and it can break up as many as 6000 of these molecules ($30,000\text{eV} \div 5\text{eV}$ per molecule $=$ 6000 molecules). Nuclear decay energies are on the order of 1 MeV (1,000,000 eV) per event and can, thus, produce significant biological damage.

Conservation of Energy

The total energy of a system is conserved if there is no net addition (or subtraction) of work or heat transfer. For conservative forces, such as the electrostatic force, conservation of energy states that mechanical energy is a constant.

Mechanical energy is the sum of the kinetic energy and potential energy of a system; that is, $KE + PE = \text{ constant }$. A loss of PE of a charged particle becomes an increase in its KE. Here PE is the electric potential energy. Conservation of energy is stated in equation form as

$$\text{KE} + \text{PE} = \text{constant} \tag{7.6.20}$$

or

$$\text{KE}_i + \text{PE}_i = \text{KE}_f + \text{PE}_f, \tag{7.6.21}$$

where i and f stand for initial and final conditions. As we have found many times before, considering energy can give us insights and facilitate problem solving.

Example 7.6.3: Electrical Potential Energy Converted to Kinetic Energy

Calculate the final speed of a free electron accelerated from rest through a potential difference of 100 V. (Assume that this numerical value is accurate to three significant figures.)

Strategy

We have a system with only conservative forces. Assuming the electron is accelerated in a vacuum, and neglecting the gravitational force (we will check on this assumption later), all of the electrical potential energy is converted into kinetic energy. We can identify the initial and final forms of energy to be $\text{KE}_i = 0, \text{KE}_f = \frac{1}{2}mv^2, \text{PE}_i = qV, \text{ and } \text{PE}_f = 0$.

Solution

Conservation of energy states that

$$\text{KE}_i + \text{PE}_i = \text{KE}_f + \text{PE}_f \tag{7.6.22}$$

Entering the forms identified above, we obtain

$$qV = \frac{mv^2}{2}. \tag{7.6.23}$$

We solve this for v:

$$v = \sqrt{\frac{2qV}{m}}. \tag{7.6.24}$$

Entering values for q, V, and m gives

$$v = \sqrt{\frac{2(-1.60 \times 10^{-19}\mathrm{C})(-100\mathrm{J/C})}{9.11 \times 10^{-31}\mathrm{kg}}} \tag{7.6.25}$$

$$= 5.93 \times 10^6 \mathrm{m/s}. \tag{7.6.26}$$

Discussion

Note that both the charge and the initial voltage are negative, as in Figure. From the discussions in Electric Charge and Electric Field, we know that electrostatic forces on small particles are generally very large compared with the gravitational force. The large final speed confirms that the gravitational force is indeed negligible here. The large speed also indicates how easy it is to accelerate electrons with small voltages because of their very small mass. Voltages much higher than the 100 V in this problem are typically used in electron guns. Those higher voltages produce electron speeds so great that relativistic effects must be taken into account. That is why a low voltage is considered (accurately) in this example.

Summary

- Electric potential is potential energy per unit charge.
- The potential difference between points A and B, $V_\mathrm{B} - V_\mathrm{A}$, defined to be the change in potential energy of a charge q moved from A to B, is equal to the change in potential energy divided by the charge, Potential difference is commonly called voltage, represented by the symbol ΔV.

$$\Delta V = \frac{\Delta \mathrm{PE}}{q} \text{ and } \Delta \mathrm{PE} = q\Delta V.$$

- An electron volt is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

$$1\mathrm{eV} = (1.60 \times 10^{-19}\mathrm{C})(1\mathrm{V}) = (1.60 \times 10^{-19}\mathrm{C})(1\mathrm{J/C})$$
$$= 1.60 \times 10^{-19}\mathrm{J}.$$

- Mechanical energy is the sum of the kinetic energy and potential energy of a system, that is, $\mathrm{KE} + \mathrm{PE}$ This sum is a constant.

Glossary

electric potential

potential energy per unit charge

potential difference (or voltage)

change in potential energy of a charge moved from one point to another, divided by the charge; units of potential difference are joules per coulomb, known as volt

electron volt

the energy given to a fundamental charge accelerated through a potential difference of one volt

mechanical energy

sum of the kinetic energy and potential energy of a system; this sum is a constant

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of

Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

CHAPTER OVERVIEW

8: Electric Current and Resistance

8.1: Prelude to Electric Current, Resistance, and Ohm's Law

The flicker of numbers on a handheld calculator, nerve impulses carrying signals of vision to the brain, an ultrasound device sending a signal to a computer screen, the brain sending a message for a baby to twitch its toes, an electric train pulling its load over a mountain pass, a hydroelectric plant sending energy to metropolitan and rural users—these and many other examples of electricity involve *electric current, the movement of charge*. Humankind has indeed harnessed electricity, the basis of technology, to improve our quality of life. Whereas the previous two chapters concentrated on static electricity and the fundamental force underlying its behavior, the next few chapters will be devoted to electric and magnetic phenomena involving current. In addition to exploring applications of electricity, we shall gain new insights into nature—in particular, the fact that all magnetism results from electric current.

Figure 20.1.1 Electric energy in massive quantities is transmitted from this hydroelectric facility, the Srisailam power station located along the Krishna River in India, by the movement of charge—that is, by electric current. (credit: Chintohere, Wikimedia Commons)

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

8.2: Current

Learning Objectives

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

- Define electric current, ampere, and drift velocity
- Describe the direction of charge flow in conventional current.
- Use drift velocity to calculate current and vice versa.

Electric Current

Electric current is defined to be the rate at which charge flows. A large current, such as that used to start a truck engine, moves a large amount of charge in a small time, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge over a long period of time. In equation form, **electric current** I is defined to be

$$I = \frac{\Delta Q}{\Delta t}, \tag{8.2.1}$$

where ΔQ is the amount of charge passing through a given area in time Δt. As in previous chapters, initial time is often taken to be zero, in which case $\Delta t = t$. (8.2.1). The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). From Equation 8.2.1, we see that an ampere is one coulomb per second:

$$1A = 1C/s \tag{8.2.2}$$

Not only are fuses and circuit breakers rated in amperes (or amps), so are many electrical appliances.

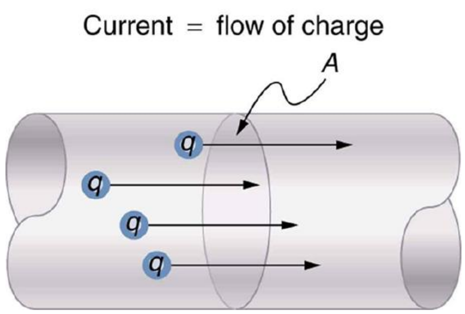


Figure 8.2.1: The rate of flow of charge is current. An ampere is the flow of one coulomb through an area in one second.

Example 8.2.1: Calculating Currents: Current in a Truck Battery and a Handheld Calculator

a. What is the current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine?
b. How long does it take 1.00 C of charge to flow through a handheld calculator if a 0.300-mA current is flowing?

Strategy

We can use the definition of current in the equation $I = \Delta Q/\Delta t$ to find the current in part (a), since charge and time are given. In part (b), we rearrange the definition of current and use the given values of charge and current to find the time required.

Solution (a)

Entering the given values for charge and time into the definition of current gives

$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} \\ &= \frac{720C}{4.00s} \\ &= 180C/s \\ &= 180A. \end{aligned}$$

Discussion (a)

This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these "starter motors" are fairly large because large frictional forces need to be overcome when setting something in motion.

Solution (b)

Solving the relationship $I = \Delta Q/\Delta t$ for time Δt, and entering the known values for charge and current gives

$$\begin{aligned}\Delta t &= \frac{\Delta Q}{I} \\ &= \frac{1.00 C}{0.300\times 10^{-3} C/s} \\ &= 3.33\times 10^{3} s.\end{aligned}$$

Discussion (b)

This time is slightly less than an hour. The small current used by the hand-held calculator takes a much longer time to move a smaller charge than the large current of the truck starter. So why can we operate our calculators only seconds after turning them on? It's because calculators require very little energy. Such small current and energy demands allow handheld calculators to operate from solar cells or to get many hours of use out of small batteries. Remember, calculators do not have moving parts in the same way that a truck engine has with cylinders and pistons, so the technology requires smaller currents.

Figure $8.2.2$ shows a simple circuit and the standard schematic representation of a battery, conducting path, and load (a resistor). Schematics are very useful in visualizing the main features of a circuit. A single schematic can represent a wide variety of situations. The schematic in Figure $8.2.2b$, for example, can represent anything from a truck battery connected to a headlight lighting the street in front of the truck to a small battery connected to a penlight lighting a keyhole in a door. Such schematics are useful because the analysis is the same for a wide variety of situations. We need to understand a few schematics to apply the concepts and analysis to many more situations.

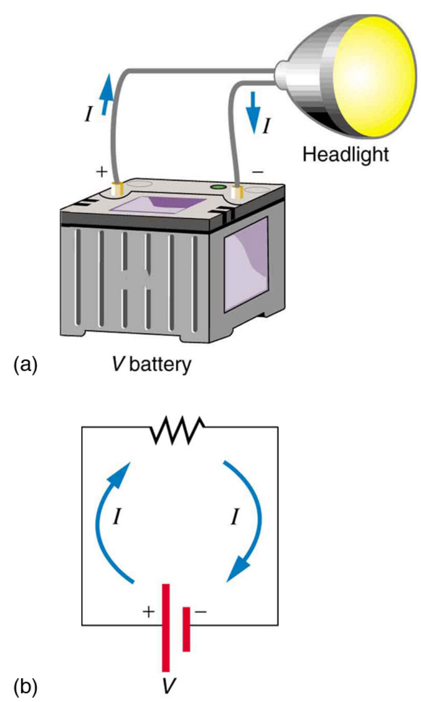


Figure $8.2.2$: (a) A simple electric circuit. A closed path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by the two parallel red lines, conducting wires are shown as straight lines, and the zigzag represents the load. The schematic represents a wide variety of similar circuits.

Note that the direction of current flow in Figure $8.2.2$ is from positive to negative. *The direction of conventional current is the direction that positive charge would flow.* Depending on the situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator used for nuclear research can produce a current of pure positive charges, such as protons. Figure $8.2.3$ illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American politician and scientist Benjamin Franklin in the 1700s. He named the type of charge associated with electrons negative, long before they were known to carry current in so many situations. Franklin, in fact, was totally unaware of the small-scale structure of electricity.

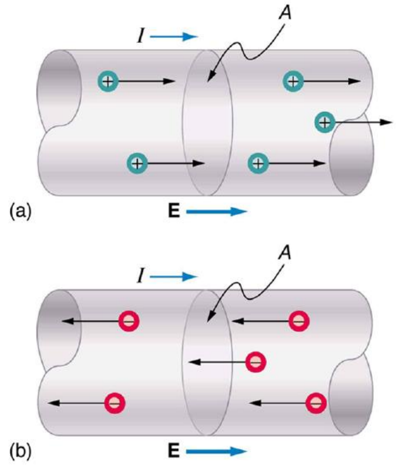


Figure $8.2.3$: Current I is the rate at which charge moves through an area A, such as the cross-section of a wire. Conventional current is defined to move in the direction of the electric field. (a) Positive charges move in the direction of the electric field and the same direction as conventional current. (b) Negative charges move in the direction opposite to the electric field. Conventional current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

It is important to realize that there is an electric field in conductors responsible for producing the current, as illustrated in Figure $8.2.3$ Unlike static electricity, where a conductor in equilibrium cannot have an electric field in it, conductors carrying a current have an electric field and are not in static equilibrium. An electric field is needed to supply energy to move the charges.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION-ELECTRIC CURRENT ILLUSTRATION

Find a straw and little peas that can move freely in the straw. Place the straw flat on a table and fill the straw with peas. When you pop one pea in at one end, a different pea should pop out the other end. This demonstration is an analogy for an electric current. Identify what compares to the electrons and what compares to the supply of energy. What other analogies can you find for an electric current?

Note that the flow of peas is based on the peas physically bumping into each other; electrons flow due to mutually repulsive electrostatic forces.

Example $8.2.2$: Calculating the Number of Electrons that Move through a Calculator

If the 0.300-mA current through the calculator mentioned in the example is carried by electrons, how many electrons per second pass through it?

Strategy

The current calculated in the previous example was defined for the flow of positive charge. For electrons, the magnitude is the same, but the sign is opposite, $I_{electrons} = -0.300 \times 10^{-3} C/s$. Since each electron (e^{-}) has a charge of $-1.60 \times 10^{-19} C$, we can convert the current in coulombs per second to electrons per second.

Solution:

Starting with the definition of current, we have

$$
\begin{aligned}
I_{electrons} &= \frac{\Delta Q_{electrons}}{\Delta t} \\
&= \frac{-0.300\times 10^{-3} C}{s}.
\end{aligned}
$$

We divide this by the charge per electron, so that

$$
\begin{aligned}
\frac{e}{s} &= \frac{-0.300\times 10{-}3C}{s}\times\frac{1e}{-1.60\times 10^{-19}C} \\
&= 1.88\times 10^{15}\frac{e}{s}.
\end{aligned}
$$

Discussion:

There are so many charged particles moving, even in small currents, that individual charges are not noticed, just as individual water molecules are not noticed in water flow. Even more amazing is that they do not always keep moving forward like soldiers in a parade. Rather they are like a crowd of people with movement in different directions but a general trend to move forward. There are lots of collisions with atoms in the metal wire and, of course, with other electrons.

Drift Velocity

Electrical signals are known to move very rapidly. Telephone conversations carried by currents in wires cover large distances without noticeable delays. Lights come on as soon as a switch is flicked. Most electrical signals carried by currents travel at speeds on the order of $10^8 m/s$, a significant fraction of the speed of light. Interestingly, the individual charges that make up the current move *much* more slowly on average, typically drifting at speeds on the order of $10^{-4} m/s$. How do we reconcile these two speeds, and what does it tell us about standard conductors?

The high speed of electrical signals results from the fact that the force between charges acts rapidly at a distance. Thus, when a free charge is forced into a wire, as in $8.2.4$, the incoming charge pushes other charges ahead of it, which in turn push on charges farther down the line. The density of charge in a system cannot easily be increased, and so the signal is passed on rapidly. The resulting electrical shock wave moves through the system at nearly the speed of light. To be precise, this rapidly moving signal or shock wave is a rapidly propagating change in electric field.

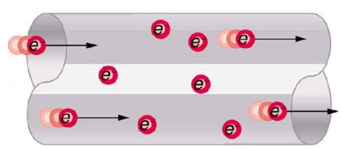


Figure $8.2.4$: When charged particles are forced into this volume of a conductor, an equal number are quickly forced to leave. The repulsion between like charges makes it difficult to increase the number of charges in a volume. Thus, as one charge enters, another leaves almost immediately, carrying the signal rapidly forward.

Good conductors have large numbers of free charges in them. In metals, the free charges are free electrons. $8.2.5$ shows how free electrons move through an ordinary conductor. The distance that an individual electron can move between collisions with atoms or other electrons is quite small. The electron paths thus appear nearly random, like the motion of atoms in a gas. But there is an electric field in the conductor that causes the electrons to drift in the direction shown (opposite to the field, since they are negative). The **drift velocity** v_d i s the average velocity of the free charges. Drift velocity is quite small, since there are so many free charges. If we have an estimate of the density of free electrons in a conductor, we can calculate the drift velocity for a given current. The larger the density, the lower the velocity required for a given current.

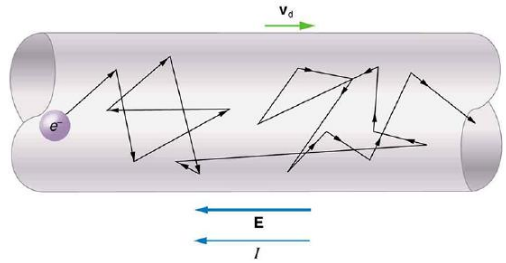


Figure $8.2.5$: Free electrons moving in a conductor make many collisions with other electrons and atoms. The path of one electron is shown. The average velocity of the free charges is called the drift velocity, v_d, and it is in the direction opposite to the electric field for electrons. The collisions normally transfer energy to the conductor, requiring a constant supply of energy to maintain a steady current.

CONDUCTION OF ELECTRICITY AND HEAT

Good electrical conductors are often good heat conductors, too. This is because large numbers of free electrons can carry electrical current and can transport thermal energy.

The free-electron collisions transfer energy to the atoms of the conductor. The electric field does work in moving the electrons through a distance, but that work does not increase the kinetic energy (nor speed, therefore) of the electrons. The work is transferred to the conductor's atoms, possibly increasing temperature. Thus a continuous power input is required to keep a current flowing. An exception, of course, is found in superconductors, for reasons we shall explore in a later chapter. Superconductors can have a steady current without a continual supply of energy—a great energy savings. In contrast, the supply of energy can be useful, such as in a lightbulb filament. The supply of energy is necessary to increase the temperature of the tungsten filament, so that the filament glows.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION--FILAMENT OBSERVATIONS

Find a lightbulb with a filament. Look carefully at the filament and describe its structure. To what points is the filament connected?

We can obtain an expression for the relationship between current and drift velocity by considering the number of free charges in a segment of wire, as illustrated in Figure $8.2.6$. *The number of free charges per unit volume* is given the symbol n and depends on the material. The shaded segment has a volume Ax, so that the number of free charges in it is nAx. The charge ΔQ in this segment is thus $qnAx$, where q is the amount of charge on each carrier. (Recall that for electrons, q is $-1.60 \times 10^{-19} C$.) Current is charge moved per unit time; thus, if all the original charges move out of this segment in time Δt, the current is

$$I = \frac{\Delta Q}{\Delta t} = \frac{qnAx}{\Delta t}. \tag{8.2.3}$$

Note that $x/\Delta t$ is the magnitude of the drift velocity, v_d, since the charges move an average distance x in a time Δt. Rearranging terms gives

$$I = nqAv_d, \tag{8.2.4}$$

where I is the current through a wire of cross-sectional area A made of a material with a free charge density n. The carriers of the current each have charge q and move with a drift velocity of magnitude v_d.

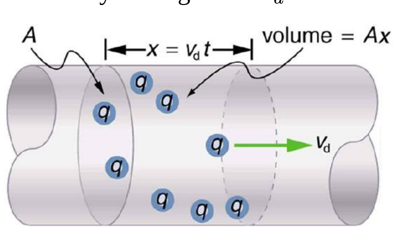


Figure $8.2.6$: All the charges in the shaded volume of this wire move out in a time t, having a drift velocity of magnitude $v_d = x/t$. See text for further discussion.

Note that simple drift velocity is not the entire story. The speed of an electron is much greater than its drift velocity. In addition, not all of the electrons in a conductor can move freely, and those that do might move somewhat faster or slower than the drift velocity.

So what do we mean by free electrons? Atoms in a metallic conductor are packed in the form of a lattice structure. Some electrons are far enough away from the atomic nuclei that they do not experience the attraction of the nuclei as much as the inner electrons do. These are the free electrons. They are not bound to a single atom but can instead move freely among the atoms in a "sea" of electrons. These free electrons respond by accelerating when an electric field is applied. Of course as they move they collide with the atoms in the lattice and other electrons, generating thermal energy, and the conductor gets warmer. In an insulator, the organization of the atoms and the structure do not allow for such free electrons.

Example 8.2.3: Calculating Drift Velocity in a Common Wire

Calculate the drift velocity of electrons in a 12-gauge copper wire (which has a diameter of 2.053 mm) carrying a 20.0-A current, given that there is one free electron per copper atom. (Household wiring often contains 12-gauge copper wire, and the maximum current allowed in such wire is usually 20 A.) The density of copper is $8.80 \times 10^{3} kg/m^{3}$.

Strategy

We can calculate the drift velocity using the equation $I = nqAv_{d}$. The current $I = 20.0 A$ is given, and $q = -1.60 \times 10^{-19} C$ is the charge of an electron. We can calculate the area of a cross-section of the wire using the formula $A = \pi r^{2}$, where r is one-half the given diameter, 2.053 mm. We are given the density of copper, $8.80 \times 10^{3} kg/m^{3}$, and the periodic table shows that the atomic mass of copper is 63.54 g/mol. We can use these two quantities along with Avogadro's number, $6.02 \times 10^{23} atoms/mol$, to determine n, the number of free electrons per cubic meter.

Solution

First, calculate the density of free electrons in copper. There is one free electron per copper atom. Therefore, is the same as the number of copper atoms per m^{3}. We can now find n as follows:

$$\begin{aligned} n &= \frac{1e^{-}}{atom} \times \frac{6.02 \times 10^{23} atoms}{mol} \times \frac{1 mol}{63.54 g} \times \frac{1000 g}{kg} \times \frac{8.80 \times 10^{3} kg}{1 m^{3}} \\ &= 8.342 \times 10^{28} e^{-}/m^{3}. \end{aligned}$$

The cross-sectional area of the wire is

\begin{align*} A &= \pi r^{2}

= \pi \left(\dfrac{2.053 \times 10^{-3} m}{2} \right)^{2}\\[5pt] &= 3.310 \times 10^{-6}m^{2}.\end{align*}

Rearranging $I = nqAv_{d}$ to isolate drift velocity gives

$$\begin{aligned} v_{d} &= \frac{I}{nqA} \\ &= \frac{20.0 A}{(8.342 \times 10^{28}/m^{3})(-1.60 \times 10^{-19} C)(3.310 \times 10^{-6} m^{2})} \\ &= -4.53 \times 10^{-4} m/s. \end{aligned}$$

Discussion

The minus sign indicates that the negative charges are moving in the direction opposite to conventional current. The small value for drift velocity (on the order of $10^{-4} m/s$) confirms that the signal moves on the order of 10^{12} times faster (about $10^{8} m/s$) than the charges that carry it.

Summary

- Electric current I is the rate at which charge flows, given by

$$I = \frac{\Delta Q}{\Delta t,} \tag{8.2.5}$$

  where ΔQ is the amount of charge passing through an area in time Δt.
- The direction of conventional current is taken as the direction in which positive charge moves.

- The SI unit for current is the ampere (A), where $1A = 1C/s$.
- Current is the flow of free charges, such as electrons and ions.
- Drift velociy v_d is the average speed at which these charges move.
- Current I is proportional to drift velocity v_d, as expressed in the relationship $I = nqAv_d$. Here, I is the current through a wire of cross-sectional area A. The wire's material as a free-charge density n, and each carrier has charge q and a drift velocity v_d.
- Electrical signals travel at speeds about 10^{12} times greater than the drift velocity of free electrons.

Glossary

electric current

the rate at which charge flows, $I = \Delta Q/\Delta t$

ampere

(amp) the SI unit for current; 1 A = 1 C/s

drift velocity

the average velocity at which free charges flow in response to an electric field

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

8.3: Ohm's Law - Resistance and Simple Circuits

Learning Objectives

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

- Explain the origin of Ohm's law.
- Calculate voltages, currents, or resistances with Ohm's law.
- Explain what an ohmic material is.
- Describe a simple circuit.

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—which are necessary to maintain a current. All such devices create a potential difference and are loosely referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electric field. The electric field in turn exerts force on charges, causing current.

Ohm's Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:

$$I \propto V. \tag{8.3.1}$$

This important relationship is known as **Ohm's law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction—an experimentally observed phenomenon. Such a linear relationship doesn't always occur.

Resistance and Simple Circuits

If voltage drives current, what impedes it? The electric property that impedes current (crudely similar to friction and air resistance) is called resistance R. Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or

$$I \propto \frac{1}{R}. \tag{8.3.2}$$

Thus, for example, current is cut in half if resistance doubles. Combining the relationships of current to voltage and current to resistance gives

$$I = \frac{V}{R}. \tag{8.3.3}$$

This relationship is also called Ohm's law. Ohm's law in this form really defines resistance for certain materials. Ohm's law (like Hooke's law) is not universally valid. The many substances for which Ohm's law holds are called **ohmic**. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I. An object that has simple resistance is called a *resistor*, even if its resistance is small. The unit for resistance is an **ohm** and is given the symbol Ω (upper case Greek omega). Rearrranging $I = V/R$ gives $R = V/I$, and so the units of resistance are 1 ohm = 1 volt per ampere:

$$1\Omega = 1\frac{V}{A}. \tag{8.3.4}$$

Figure $8.3.1$ shows the schematic for a simple circuit. A **simple circuit** has a single voltage source and a single resistor. The wires connecting the voltage source to the resistor can be assumed to have negligible resistance, or their resistance can be included in R.

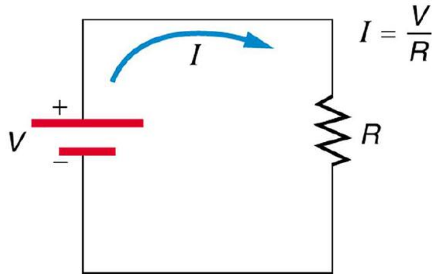


Figure 8.3.1: A simple electric circuit in which a closed path for current to flow is supplied by conductors (usually metal wires) connecting a load to the terminals of a battery, represented by the red parallel lines. The zigzag symbol represents the single resistor and includes any resistance in the connections to the voltage source.

Example 8.3.1: Calculating Resistance: An Automobile Headlight:

What is the resistance of an automobile headlight through which 2.50 A flows when 12.0 V is applied to it?

Strategy

We can rearrange Ohm's law as stated by $I = V/R$ and use it to find the resistance.

Solution:

Rearranging Equation 8.3.3 and substituting known values gives

$$\begin{aligned} R &= \frac{V}{I} \\ &= \frac{12.0V}{2.50A} \\ &= 4.80\Omega. \end{aligned}$$

Discussion:

This is a relatively small resistance, but it is larger than the cold resistance of the headlight. As we shall see, resistance of metals usually *increases* with *increasing* temperature, and so the bulb has a lower resistance when it is first switched on and will draw considerably more current during its brief warm-up period.

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12}\Omega$ or more. A dry person may have a hand-to-foot resistance of $10^{5}\Omega$, whereas the resistance of the human heart is about $10^{3}\Omega$. A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5}\Omega$, and superconductors have no resistance at all (they are non-ohmic). Resistance is related to the shape of an object and the material of which it is composed, as will be seen in Resistance and Resistivity.

Additional insight is gained by solving $I = V/R$ for V, yielding

$$V = IR. \tag{8.3.5}$$

The expression for V can be interpreted as the *voltage drop across a resistor produced by the low of current* I. The phrase IR *drop* is often used for this voltage. For instance, the headlight in the example has an IR drop of 12.0 V. If voltage is measured at various points in a circuit, it will be seen to increase at the voltage source and decrease at the resistor. Voltage is similar to fluid pressure. The voltage source is like a pump, creating a pressure difference, causing current—the flow of charge. The resistor is like a pipe that reduces pressure and limits flow because of its resistance. Conservation of energy has important consequences here. The voltage source supplies energy (causing an electric field and a current), and the resistor converts it to another form (such as thermal energy). In a simple circuit (one with a single simple resistor), the voltage supplied by the source equals the voltage drop across the resistor, since $PE = q\Delta V$, and the same q flows through each. Thus the energy supplied by the voltage source and the energy converted by the resistor are equal (Figure 8.3.2).

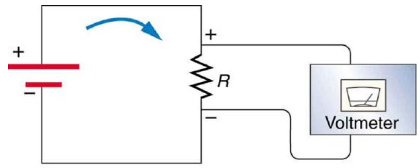


Figure 8.3.2: The voltage drop across a resistor in a simple circuit equals the voltage output of the battery.

MAKING CONNECTIONS: CONSERVATION OF ENERGY

In a simple electrical circuit, the sole resistor converts energy supplied by the source into another form. Conservation of energy is evidenced here by the fact that all of the energy supplied by the source is converted to another form by the resistor alone. We will find that conservation of energy has other important applications in circuits and is a powerful tool in circuit analysis.

Summary

- A simple circuit is one in which there is a single voltage source and a single resistance.
- One statement of Ohm's law gives the relationship between current I, voltage V, and resistance R in an simple circuit to be $I=\frac{V}{R}$.
- Resistance has units of ohms (Ω), related to volts and amperes by $1\Omega=1V/A$.
- There is a voltage or IR drop across a resistor, caused by the current flowing through it, given by $V=IR$.

Glossary

Ohm's law

an empirical relation stating that the current *I* is proportional to the potential difference *V*, ∝ *V*; it is often written as *I* = *V*/*R*, where *R* is the resistance

resistance

the electric property that impedes current; for ohmic materials, it is the ratio of voltage to current, *R* = *V*/*I*

ohm

the unit of resistance, given by 1Ω = 1 V/A

ohmic

a type of a material for which Ohm's law is valid

simple circuit

a circuit with a single voltage source and a single resistor

Contributors and Attributions

8.4: Resistance and Resistivity

Learning Objectives

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

- Explain the concept of resistivity.
- Use resistivity to calculate the resistance of specified configurations of material.
- Use the thermal coefficient of resistivity to calculate the change of resistance with temperature.

Material and Shape Dependence of Resistance

The resistance of an object depends on its shape and the material of which it is composed. The cylindrical resistor in Figure 1 is easy to analyze, and, by so doing, we can gain insight into the resistance of more complicated shapes. As you might expect, the cylinder's electric resistance R is directly proportional to its length L, similar to the resistance of a pipe to fluid flow. The longer the cylinder, the more collisions charges will make with its atoms. The greater the diameter of the cylinder, the more current it can carry (again similar to the flow of fluid through a pipe). In fact, R is inversely proportional to the cylinder's cross-sectional area A.

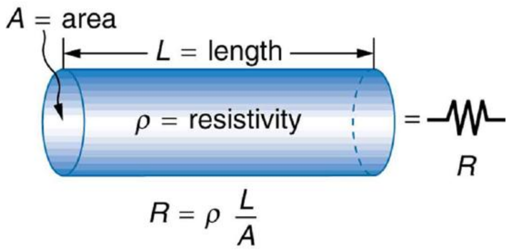


Figure 8.4.1: A uniform cylinder of length L and cross-sectional area A. Its resistance to the flow of current is similar to the resistance posed by a pipe to fluid flow. The longer the cylinder, the greater its resistance. The larger its cross-sectional area A, the smaller its resistance.

For a given shape, the resistance depends on the material of which the object is composed. Different materials offer different resistance to the flow of charge. We define the **resistivity** ρ of a substance so that the **resistance** R of an object is directly proportional to ρ. Resistivity ρ is an *intrinsic* property of a material, independent of its shape or size. The resistance R of a uniform cylinder of length L, of cross-sectional area A, and made of a material with resistivity ρ, is

$$R = \frac{\rho L}{A}. \tag{8.4.1}$$

The table below gives representative values of ρ. The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivities. Conductors have the smallest resistivities, and insulators have the largest; semiconductors have intermediate resistivities. Conductors have varying but large free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as will be explored in later chapters.

Table 8.4.1 gives representative values of ρ. The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivities. Conductors have the smallest resistivities, and insulators have the largest; semiconductors have intermediate resistivities. Conductors have varying but large free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as will be explored in later chapters.

Table 8.4.1: Resistivities ρ of Various materials at 20ºC

Material	**Resistivity** ρ (Ω·m)
Conductors	

Material	Resistivity ρ (Ω·m)
Silver	$1.59x10^{-8}$
Copper	$1.72x10^{-8}$
Gold	$2.44x10^{-8}$
Aluminum	$2.65x10^{-8}$
Tungsten	$5.6x10^{-8}$
Iron	$9.71x10^{-8}$
Platinum	$10.6x10^{-8}$
Steel	$20x10^{-8}$
Lead	$22x10^{-8}$
Manganin (Cu, Mn, Ni alloy)	$44x10^{-8}$
Constantan (Cu, Ni alloy)	$49x10^{-8}$
Mercury	$96x10^{-8}$
Nichrome (Ni, Fe, Cr alloy)	$100x10^{-8}$
Semiconductors	
Carbon (pure)	$3.5x10^{-5}$
Carbon	$(3.5-6.0)x10^{-5}$
Germanium (pure)	$600x10^{-3}$
Germanium	$(1-600)x10^{-3}$
Silicon (pure)	2300
Silion	0.1–2300
Insulators	
Amber	$5x10^{14}$
Glass	$10^{9}-10^{14}$
Lucite	$>10^{13}$
Mica	$10^{11}-10^{15}$
Quartz (fused)	$75x10^{16}$
Rubber (hard)	$10^{13}-10^{16}$
Sulfur	10^{15}
Teflon	$>10^{13}$
Wood	$10^{8}-10^{11}$

Example 8.4.1:Calculating Resistor Diameter: A Headlight Filament

A car headlight filament is made of tungsten and has a cold resistance of **0.350Ω**. If the filament is a cylinder 4.00 cm long (it may be coiled to save space), what is its diameter?

Strategy

We can rearrange the equation $R = \frac{\rho L}{A}$ to find the cross-sectional area A of the filament from the given information. Then its diameter can be found by assuming it has a circular cross-section.

Solution

The cross-sectional area, found by rearranging the expression for the resistance of a cylinder given in $R = \frac{\rho L}{A}$, is

$$A = \frac{\rho L}{R}.$$

Substituting the given values, and taking ρ from Table 8.4.1, yields

$$A = \frac{(5.6\times 10^{-8}\Omega\cdot m)(4.00\times 10^{-2}m)}{0.350\Omega} = 6.40\times 10^{-9}m^2 .$$

The area of a circle is related to its diameter D by

$$A = \frac{\pi D^2}{4}.$$

Solving for the diameter D, and substituting the value found for A, gives

$$D = 2(\frac{A}{p})^{\frac{1}{2}} = 2(\frac{6.40\times 10^{-9}m^2}{3.14})^{\frac{1}{2}} = 9.0\times 10^{-5}m .$$

Discussion

The diameter is just under a tenth of a millimeter. It is quoted to only two digits, because ρ is known to only two digits.

Temperature Variation of Resistance

The resistivity of all materials depends on temperature. Some even become superconductors (zero resistivity) at very low temperatures. (See Figure 2.) Conversely, the resistivity of conductors increases with increasing temperature. Since the atoms vibrate more rapidly and over larger distances at higher temperatures, the electrons moving through a metal make more collisions, effectively making the resistivity higher. Over relatively small temperature changes (about $100°C$ or less), resistivity ρ varies with temperature change ΔT as expressed in the following equation

$$\rho = \rho_0 \left(1 + \alpha\Delta T\right), \tag{8.4.2}$$

where ρ_0 is the original resistivity and α is the **temperature coefficient of resistivity.** (See the values of α in the table below.) For larger temperature changes, α may vary or a nonlinear equation may be needed to find ρ. Note that α is positive for metals, meaning their resistivity increases with temperature. Some alloys have been developed specifically to have a small temperature dependence. Manganin (which is made of copper, manganese and nickel), for example, has α close to zero (to three digits on the scale in the table), and so its resistivity varies only slightly with temperature. This is useful for making a temperature-independent resistance standard, for example.

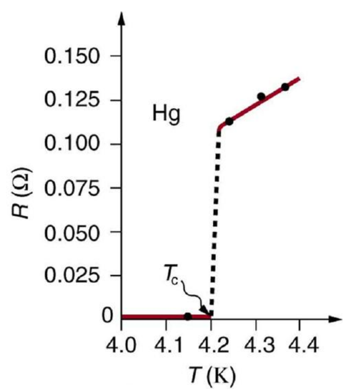


Figure 8.4.2: The resistance of a sample of mercury is zero at very low temperatures—it is a superconductor up to about 4.2 K. Above that critical temperature, its resistance makes a sudden jump and then increases nearly linearly with temperature.

Table 8.4.2: Tempature Coefficients of Resistivity α

Material	Coefficient α(1/°C)
Conductors	

Material	Coefficient α(1/°C)
Silver	$3.8x10^{-3}$
Copper	$3.9x10^{-3}$
Gold	$3.4x10^{-3}$
Aluminum	$3.9x10^{-3}$
Tungsten	$4.5x10^{-3}$
Iron	$5.0x10^{-3}$
Platinum	$3.93x10^{-3}$
Lead	$3.9x10^{-3}$
Manganin (Cu, Mn, Ni alloy)	$0.000x10^{-3}$
Constantan (Cu, Ni, alloy)	$0.002x10^{-3}$
Mercury	$0.89x10^{-3}$
Nichrome (Ni, Fe, Cr alloy)	$0.4x10^{-3}$
Semiconductors	
Carbon (pure)	$-0.5x10^{-3}$
Germanium (pure)	$-50x10^{-3}$
Silicon (pure)	$-70x10^{-3}$

Note also that α is negative for the semiconductors listed in the table, meaning that their resistivity decreases with increasing temperature. They become better conductors at higher temperature, because increased thermal agitation increases the number of free charges available to carry current. This property of decreasing ρ with temperature is also related to the type and amount of impurities present in the semiconductors.

The resistance of an object also depends on temperature, since R_0 is directly proportional to ρ. For a cylinder we know $R = \rho L/A$, and so, if L and A do not change greatly with temperature, R will have the same temperature dependence as ρ. (Examination of the coefficients of linear expansion shows them to be about two orders of magnitude less than typical temperature coefficients of resistivity, and so the effect of temperature on L and A is about two orders of magnitude less than on ρ.) Thus,

$$R = R_0 \left(1 + \alpha \Delta T\right) \tag{8.4.3}$$

is the temperature dependence of the resistance of an object, where R_0 is the original resistance and R is the resistance after a temperature change ΔT. Numerous thermometers are based on the effect of temperature on resistance. (See Figure 3.) One of the most common is the thermistor, a semiconductor crystal with a strong temperature dependence, the resistance of which is measured to obtain its temperature. The device is small, so that it quickly comes into thermal equilibrium with the part of a person it touches.

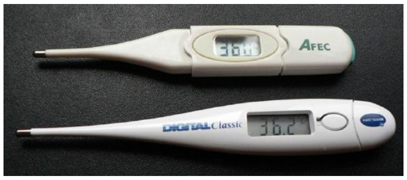


Figure 8.4.3:These familiar thermometers are based on the automated measurement of a thermistor's temperature-dependent resistance. (credit: Biol, Wikimedia Commons)

Example 8.4.2:Calculating Resistance: Hot-Filament Resistance:

Although caution must be used in applying $\rho = \rho_0 \left(1 + \alpha \Delta T\right)$ and $R = R_0 \left(1 + \alpha \Delta T\right)$ for temperature changes greater than $100^{\circ} C$, for tungsten the equations work reasonably well for very large temperature changes. What, then, is the resistance of the tungsten filament in the previous example if its temperature is increased from room temperature ($20^{\circ}C$) to a typical operating temperature of $2850^{\circ}C$?

Strategy

This is a straightforward application of $R = R_0 (1 + \alpha \Delta T)$, since the original resistance of the filament was givent o eb $R_0 = 0.350 \Omega$, and the temperature change is $\Delta T = 2830^{\circ}C$.

Solution

The hot resistance R is obtained by entering known values into the above equation:

$$\begin{aligned} R &= R_0 (1 + \alpha \Delta T) \\ &= (0.350 \Omega) \left[1 + \left(4.5 \times 10^{-3} / ^{\circ}C\right)\right] \\ &= 4.8 \Omega \end{aligned}$$

Discussion

This value is consistent with the headlight resistance example in 20.3.

PHET EXPLORATIONS: RESISTANCE IN A WIRE

Learn about the physics of resistance in a wire. Change its resistivity, length, and area to see how they affect the wire's resistance. The sizes of the symbols in the equation change along with the diagram of a wire.

Figure 8.4.4: Resistance in a Wire

Summary

- The resistance R of a cylinder of length L and cross-sectional area A is $R = \frac{\rho L}{A}$, where ρ is the resistivity of the material.
- Values of ρ in the Table show that materials fall into three groups—*conductors, semiconductors, and insulators*.
- Temperature affects resistivity; for relatively small temperature changes ΔT, resistivity is $\rho = \rho_0 (1 + \alpha \Delta T), where \backslash (\rho_0$ is the original resistivity and α is the temperature coefficient of resistivity.
- The Table gives values for α, the temperature coefficient of resistivity.
- The resistance R of an object also varies with temperature: $R = R_0 (1 + \alpha \Delta T)$, where R_0 is the original resistance, and R is the resistance after the temperature change.

Footnotes

1 Values depend strongly on amounts and types of impurities

2 Values at 20°C.

Glossary

resistivity

an intrinsic property of a material, independent of its shape or size, directly proportional to the resistance, denoted by ρ

temperature coefficient of resistivity

an empirical quantity, denoted by α, which describes the change in resistance or resistivity of a material with temperature

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

8.5: 20.4 Electric Power and Energy

Learning Objectives

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

- Calculate the power dissipated by a resistor and power supplied by a power supply.
- Calculate the cost of electricity under various circumstances.

Power in Electric Circuits

Power is associated by many people with electricity. Knowing that power is the rate of energy use or energy conversion, what is the expression for **electric power**? Power transmission lines might come to mind. We also think of lightbulbs in terms of their power ratings in watts. Let us compare a 25-W bulb with a 60-W bulb (Figure $8.5.1a$.) Since both operate on the same voltage, the 60-W bulb must draw more current to have a greater power rating. Thus the 60-W bulb's resistance must be lower than that of a 25-W bulb. If we increase voltage, we also increase power. For example, when a 25-W bulb that is designed to operate on 120 V is connected to 240 V, it briefly glows very brightly and then burns out. Precisely how are voltage, current, and resistance related to electric power?

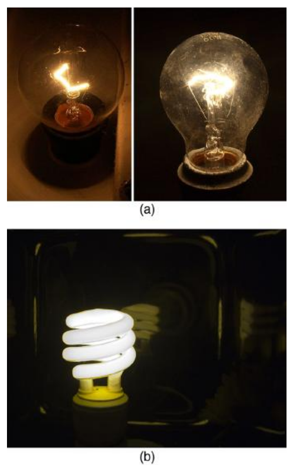


***Figure* $8.5.1$:** *(a) Which of these lightbulbs, the 25-W bulb (upper left) or the 60-W bulb (upper right), has the higher resistance? Which draws more current? Which uses the most energy? Can you tell from the color that the 25-W filament is cooler? Is the brighter bulb a different color and if so why? (credits: Dickbauch, Wikimedia Commons; Greg Westfall, Flickr) (b) This compact fluorescent light (CFL) puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit: dbgg1979, Flickr)*

Electric energy depends on both the voltage involved and the charge moved. This is expressed most simply as $PE = qV$, where q is the charge moved and V is the voltage (or more precisely, the potential difference the charge moves through). Power is the rate at which energy is moved, and so electric power is

$$P = \frac{PE}{t} = \frac{qV}{t}. \tag{8.5.1}$$

Recognizing that current is $I = q/t$ (note that $\Delta t = t$ here), the expression for power becomes

$$P = IV. \tag{8.5.2}$$

Electric power (P) is simply the product of current times voltage. Power has familiar units of watts. Since the SI unit for potential energy (PE) is the joule, power has units of joules per second, or watts. Thus, $1 A \cdot V = 1 W$. For example, cars often have one or more auxiliary power outlets with which you can charge a cell phone or other electronic devices. These outlets may be rated at 20 A, so that the circuit can deliver a maximum power $P = IV = (20 A)(12 V) = 240 W$. In some applications, electric power may be expressed as volt-amperes or even kilovolt-amperes ($1 kA \cdot V = 1 kW$).

To see the relationship of power to resistance, we combine Ohm's law with $P = IV$. Substituting Ohm Law ($I = V/R$) into Equation $8.5.2$ gives

$$P = (V/R)\,V = V^2/R. \tag{8.5.3}$$

Similarly, substituting $V = IR$ gives

$$P = (IR) = I^2 R. \tag{8.5.4}$$

Three expressions for electric power are listed together here for convenience:

$$P = IV \tag{8.5.5}$$

$$P = \frac{V^2}{R} \tag{8.5.6}$$

$$P = I^2 R. \tag{8.5.7}$$

Note that the first equation is always valid, whereas the other two can be used only for resistors. In a simple circuit, with one voltage source and a single resistor, the power supplied by the voltage source and that dissipated by the resistor are identical. (In more complicated circuits, P can be the power dissipated by a single device and not the total power in the circuit.)

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature its resistance is higher, too.

Example 8.5.1: Calculating Power Dissipation and Current: Hot and Cold Power

(a) Consider the examples given in 20.3 and 20.4. Then find the power dissipated by the car headlight in these examples, both when it is hot and when it is cold.

Strategy

For the hot headlight, we know voltage and current, so we can use $P = IV$ to find the power. For the cold headlight, we know the voltage and resistance, so we can use $P = V^2/R$ to find the power.

Solution

Entering the known values of current and voltage for the hot headlight, we obtain

$$P = IV = (2.50A)\,(12.0V) = 30.0W. \tag{8.5.8}$$

The cold resistance was 0.350Ω, and so the power it uses when first switched on is

$$P = \frac{V^2}{R} = \frac{(12.0V)^2}{0.350\Omega} = 411W. \tag{8.5.9}$$

Discussion

The 30 W dissipated by the hot headlight is typical. But the 411 W when cold is surprisingly higher. The initial power quickly decreases as the bulb's temperature increases and its resistance increases.

(b) What current does it draw when cold?

Solution

The current when the bulb is cold can be found several different ways. We rearrange one of the power equations, $P = I^2R$, and enter known values, obtaining

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{411W}{0.350\Omega}} = 34.3A. \tag{8.5.10}$$

Discussion

The cold current is remarkably higher than the steady-state value of 2.50 A, but the current will quickly decline to that value as the bulb's temperature increases. Most fuses and circuit breakers (used to limit the current in a circuit) are designed to tolerate very high currents briefly as a device comes on. In some cases, such as with electric motors, the current remains high for several seconds, necessitating special "slow blow" fuses.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P = E/t$, we see that

$$E = Pt \tag{8.5.11}$$

is the energy used by a device using power P for a time interval t. For example, the more lightbulbs burning, the greater P used; the longer they are on, the greater t is. The energy unit on electric bills is the killowatt-hour ($kW \cdot h$), consistent with the relationship $E = Pt$. It is easy to estimate the cost of operating electric appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted to joules. You can prove to yourself that $1 kW \cdot h = 3.6 \times 10^6 J$.

The electric energy (E) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This will not only reduce the cost, but it will also result in a reduced impact on the environment. Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, while the number for commercial establishments is closer to 40%. Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFL). (See Figure 1b.) Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.) The heat transfer from these CFLs is less, and they last up to 10 times longer. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last 5 times longer than CFLs. However, their cost is still high.

Making Connections: Energy, Power, and Time

The relationship $E = Pt$ is one that you will find useful in many different contexts. The energy your body uses in exercise is related to the power level and duration of your activity, for example. The amount of heating by a power source is related to the power level and time it is applied. Even the radiation dose of an X-ray image is related to the power and time of exposure.

Example 8.5.2: Calculating the Cost Effectiveness of Compact Fluorescent Lights (CFL)

(a) If the cost of electricity in your area is 12 cents per kWh, what is the total cost (capital plus operation) of using a 60-W incandescent bulb for 1000 hours (the lifetime of that bulb) if the bulb cost 25 cents?

Strategy

To find the operating cost, we first find the energy used in kilowatt-hours and then multiply by the cost per kilowatt-hour.

Solution

The energy used in kilowatt-hours is found by entering the power and time into the expression for energy:

$$E = Pt = (60 W)(1000 h) = 60,000 W \cdot h. \tag{8.5.12}$$

In kilowatt-hours, this is \]E = 60.0 kW \cdot h.\] Now the electricity cost is

$$cost = (60.0 kW \cdot h)(\$0.12/kW \cdot h) = \$7.20. \tag{8.5.13}$$

The total cost will be $7.20 for 1000 hours (about one-half year at 5 hours per day).

(b) If we replace this bulb with a compact fluorescent light that provides the same light output, but at one-quarter the wattage, and which costs $1.50 but lasts 10 times longer (10,000 hours), what will that total cost be?

Solution

Since the CFL uses only 15 W and not 60 W, the electricity cost will be $7.20/4 = $1.80. The CFL will last 10 times longer than the incandescent, so that the investment cost will be 1/10 of the bulb cost for that time period of use, or 0.1($1.50) = $0.15. Therefore, the total cost will be $1.95 for 1000 hours.

Discussion

Therefore, it is much cheaper to use the CFLs, even though the initial investment is higher. The increased cost of labor that a business must include for replacing the incandescent bulbs more often has not been figured in here.

Making Connections: Take-Home Experiment—Electrical Energy Use Inventory

1) Make a list of the power ratings on a range of appliances in your home or room. Explain why something like a toaster has a higher rating than a digital clock. Estimate the energy consumed by these appliances in an average day (by estimating their time of use). Some appliances might only state the operating current. If the household voltage is 120 V, then use $P = IV$.

2) Check out the total wattage used in the rest rooms of your school's floor or building. (You might need to assume the long fluorescent lights in use are rated at 32 W.) Suppose that the building was closed all weekend and that these lights were left on from 6 p.m. Friday until 8 a.m. Monday. What would this oversight cost? How about for an entire year of weekends?

Summary

- Electric power P is the rate (in watts) that energy is supplied by a source or dissipated by a device.
- Three expressions for electrical power are

$$P = IV, \tag{8.5.14}$$

$$P = \frac{V^2}{R}, \tag{8.5.15}$$

and

$$P = I^2 R. \tag{8.5.16}$$

- The energy used by a device with a power P over a time t is $E = Pt$.

Glossary

electric power

the rate at which electrical energy is supplied by a source or dissipated by a device; it is the product of current times voltage

Contributors and Attributions

CHAPTER OVERVIEW

9: Magnetism

Magnetism is a class of physical phenomena that are mediated by magnetic fields. Electric currents and the magnetic moments of elementary particles give rise to a magnetic field, which acts on other currents and magnetic moments. Every material is influenced to some extent by a magnetic field.

Thumbnail: Magnetic field of an ideal cylindrical magnet with its axis of symmetry inside the image plane. The magnetic field is represented by magnetic field lines, which show the direction of the field at different points. (CC-SA-BY-3.0; Geek3).

Contributors and Attributions

9.1: Prelude to Magnetism

One evening, an Alaskan sticks a note to his refrigerator with a small magnet. Through the kitchen window, the Aurora Borealis glows in the night sky. This grand spectacle is shaped by the same force that holds the note to the refrigerator.

Figure 9.1.1: The magnificent spectacle of the Aurora Borealis, or northern lights, glows in the northern sky above Bear Lake near Eielson Air Force Base, Alaska. Shaped by the Earth's magnetic field, this light is produced by radiation spewed from solar storms. (credit: Senior Airman Joshua Strang, via Flickr)

People have been aware of magnets and magnetism for thousands of years. The earliest records date to well before the time of Christ, particularly in a region of Asia Minor called Magnesia (the name of this region is the source of words like *magnetic*). Magnetic rocks found in Magnesia, which is now part of western Turkey, stimulated interest during ancient times. A practical application for magnets was found later, when they were employed as navigational compasses. The use of magnets in compasses resulted not only in improved long-distance sailing, but also in the names of "north" and "south" being given to the two types of magnetic poles.

Today magnetism plays many important roles in our lives. Physicists' understanding of magnetism has enabled the development of technologies that affect our everyday lives. The iPod in your purse or backpack, for example, wouldn't have been possible without the applications of magnetism and electricity on a small scale.

The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology. The 2007 Nobel Prize in Physics went to Albert Fert from France and Peter Grunberg from Germany for this discovery of *giant magnetoresistance* and its applications to computer memory.

All electric motors, with uses as diverse as powering refrigerators, starting cars, and moving elevators, contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Hundreds of millions of dollars are spent annually on magnetic containment of fusion as a future energy source. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is used to explain atomic energy levels, cosmic rays, and charged particles trapped in the Van Allen belts. Once again, we will find all these disparate phenomena are linked by a small number of underlying physical principles.

Figure 9.1.2: Engineering of technology like iPods would not be possible without a deep understanding magnetism. (credit: Jesse! S?, Flickr)

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of

Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

-

9.2: Magnets

Learning Objectives

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

- Describe the difference between the north and south poles of a magnet.
- Describe how magnetic poles interact with each other.

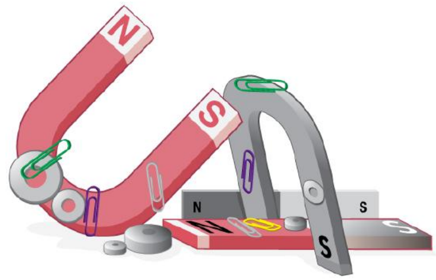


Figure 9.2.1: Magnets come in various shapes, sizes, and strengths. All have both a north pole and a south pole. There is never an isolated pole (a monopole).

All magnets attract iron, such as that in a refrigerator door. However, magnets may attract or repel other magnets. Experimentation shows that all magnets have two poles. If freely suspended, one pole will point toward the north. The two poles are thus named the north magnetic pole and the south magnetic pole (or more properly, north-seeking and south-seeking poles, for the attractions in those directions).

UNIVERSAL CHARACTERISTICS OF MAGNETS AND MAGNET POLES

It is a universal characteristic of all magnets that *like poles repel and unlike poles attract*. (Note the similarity with electrostatics: unlike charges attract and like charges repel.)

Further experimentation shows that it is *impossible to separate north and south poles* in the manner that + and − charges can be separated.

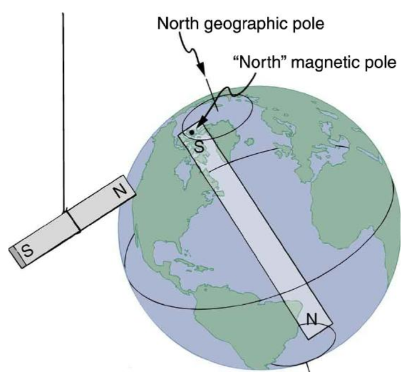


Figure 9.2.2: One end of a bar magnet is suspended from a thread that points toward north. The magnet's two poles are labeled N and S for north-seeking and south-seeking poles, respectively.

MISCONCEPTION ALERT: EARTH'S GEOGRAPHIC NORTH POLE HIDES AN S

The Earth acts like a very large bar magnet with its south-seeking pole near the geographic North Pole. That is why the north pole of your compass is attracted toward the geographic north pole of the Earth—because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole! Confusion arises because the geographic term "North Pole" has come to be used (incorrectly) for the magnetic pole that is near the North Pole. Thus, "North magnetic pole" is actually a misnomer—it should be called the South magnetic pole.

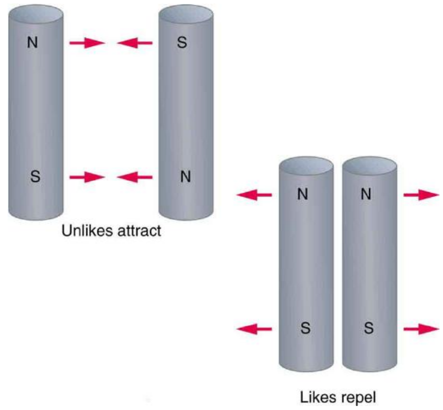


Figure 9.2.3: Unlike poles attract, whereas like poles repel.

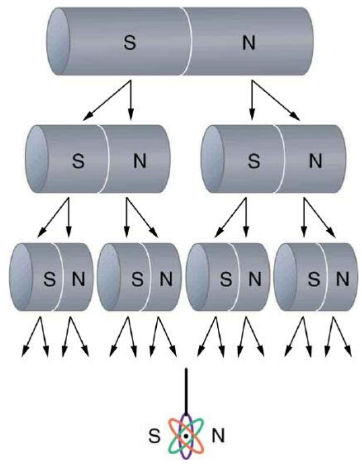


Figure 9.2.4: North and south poles always occur in pairs. Attempts to separate them result in more pairs of poles. If we continue to split the magnet, we will eventually get down to an iron atom with a north pole and a south pole—these, too, cannot be separated.

The fact that magnetic poles always occur in pairs of north and south is true from the very large scale—for example, sunspots always occur in pairs that are north and south magnetic poles—all the way down to the very small scale. Magnetic atoms have both a north pole and a south pole, as do many types of subatomic particles, such as electrons, protons, and neutrons.

MAKING CONNECTIONS: TAKE-HOME EXPERIMENT -- REFRIGERATOR MAGNETS

We know that like magnetic poles repel and unlike poles attract. See if you can show this for two refrigerator magnets. Will the magnets stick if you turn them over? Why do they stick to the door anyway? What can you say about the magnetic properties of the door next to the magnet? Do refrigerator magnets stick to metal or plastic spoons? Do they stick to all types of metal?

Summary

- Magnetism is a subject that includes the properties of magnets, the effect of the magnetic force on moving charges and currents, and the creation of magnetic fields by currents.
- There are two types of magnetic poles, called the north magnetic pole and south magnetic pole.
- North magnetic poles are those that are attracted toward the Earth's geographic north pole.
- Like poles repel and unlike poles attract.
- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.

Glossary

north magnetic pole

the end or the side of a magnet that is attracted toward Earth's geographic north pole

south magnetic pole

the end or the side of a magnet that is attracted toward Earth's geographic south pole

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

9.3: Ferromagnets and Electromagnets

Learning Objectives

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

- Define ferromagnet.
- Describe the role of magnetic domains in magnetization.
- Explain the significance of the Curie temperature.
- Describe the relationship between electricity and magnetism.

Ferromagnets

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. Such materials are called ferromagnetic, after the Latin word for iron, *ferrum*. A group of materials made from the alloys of the rare earth elements are also used as strong and permanent magnets; a popular one is neodymium. Other materials exhibit weak magnetic effects, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets (the way iron is attracted to magnets), they can also be magnetized themselves—that is, they can be induced to be magnetic or made into permanent magnets.

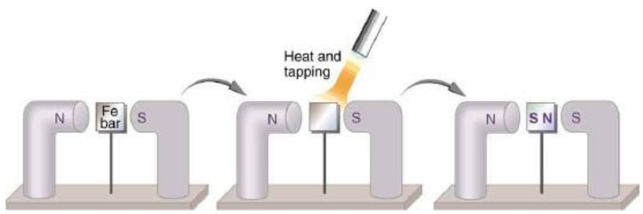


Figure 9.3.1: An unmagnetized piece of iron is placed between two magnets, heated, and then cooled, or simply tapped when cold. The iron becomes a permanent magnet with the poles aligned as shown: its south pole is adjacent to the north pole of the original magnet, and its north pole is adjacent to the south pole of the original magnet. Note that there are attractive forces between the magnets.

When a magnet is brought near a previously unmagnetized ferromagnetic material, it causes local magnetization of the material with unlike poles closest, as in Figure 1. (This results in the attraction of the previously unmagnetized material to the magnet.) What happens on a microscopic scale is illustrated in Figure 2. The regions within the material called domains act like small bar magnets. Within domains, the poles of individual atoms are aligned. Each atom acts like a tiny bar magnet. Domains are small and randomly oriented in an unmagnetized ferromagnetic object. In response to an external magnetic field, the domains may grow to millimeter size, aligning themselves as shown in Figure 2b. This induced magnetization can be made permanent if the material is heated and then cooled, or simply tapped in the presence of other magnets.

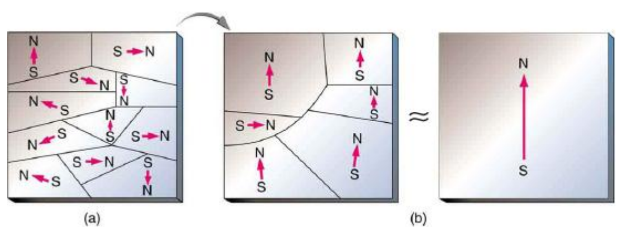


Figure 9.3.2: (a) An unmagnetized piece of iron (or other ferromagnetic material) has randomly oriented domains. (b) When magnetized by an external field, the domains show greater alignment, and some grow at the expense of others. Individual atoms are aligned within domains; each atom acts like a tiny bar magnet.

Conversely, a permanent magnet can be demagnetized by hard blows or by heating it in the absence of another magnet. Increased thermal motion at higher temperature can disrupt and randomize the orientation and the size of the domains. There is a well-defined temperature for ferromagnetic materials, which is called the Curie temperature, above which they cannot be magnetized. The Curie temperature for iron is 1043 K $(770^{\circ}C)$, which is well above room temperature. There are several elements and alloys that have Curie temperatures much lower than room temperature and are ferromagnetic only below those temperatures.

Electromagnets

Early in the 19th century, it was discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of charges had any connection with magnets. Electromagnetism is the use of electric current to make magnets. These temporarily induced magnets are called electromagnets. Electromagnets are employed for everything from a wrecking yard crane that lifts scrapped cars to controlling the beam of a 90-km-circumference particle accelerator to the magnets in medical imaging machines (See Figure 3).

Figure 9.3.3: Instrument for magnetic resonance imaging (MRI). The device uses a superconducting cylindrical coil for the main magnetic field. The patient goes into this “tunnel” on the gurney. (credit: Bill McChesney, Flickr)

Figure 4 shows that the response of iron filings to a current-carrying coil and to a permanent bar magnet. The patterns are similar. In fact, electromagnets and ferromagnets have the same basic characteristics—for example, they have north and south poles that cannot be separated and for which like poles repel and unlike poles attract.

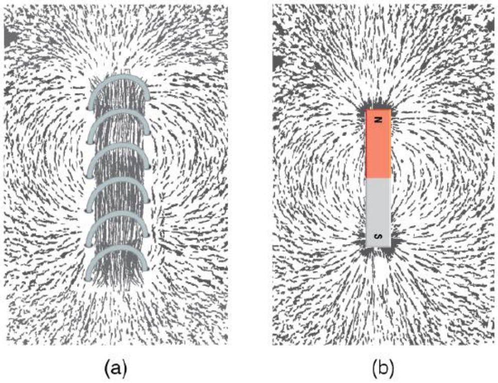


Figure 9.3.4: Iron filings near (a) a current-carrying coil and (b) a magnet act like tiny compass needles, showing the shape of their fields. Their response to a current-carrying coil and a permanent magnet is seen to be very similar, especially near the ends of the coil and the magnet.

Combining a ferromagnet with an electromagnet can produce particularly strong magnetic effects. (See Figure 5.) Whenever strong magnetic effects are needed, such as lifting scrap metal, or in particle accelerators, electromagnets are enhanced by ferromagnetic materials. Limits to how strong the magnets can be made are imposed by coil resistance (it will overheat and melt at sufficiently high current), and so superconducting magnets may be employed. These are still limited, because superconducting properties are destroyed by too great a magnetic field.

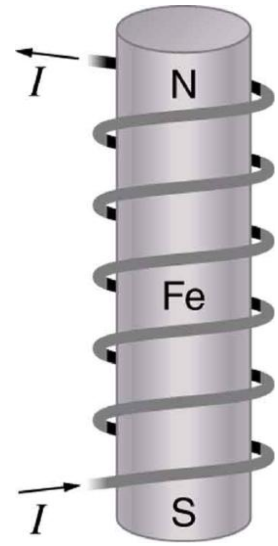


Figure 9.3.5: An electromagnet with a ferromagnetic core can produce very strong magnetic effects. Alignment of domains in the core produces a magnet, the poles of which are aligned with the electromagnet.

Figure 6 shows a few uses of combinations of electromagnets and ferromagnets. Ferromagnetic materials can act as memory devices, because the orientation of the magnetic fields of small domains can be reversed or erased. Magnetic information storage on videotapes and computer hard drives are among the most common applications. This property is vital in our digital world.

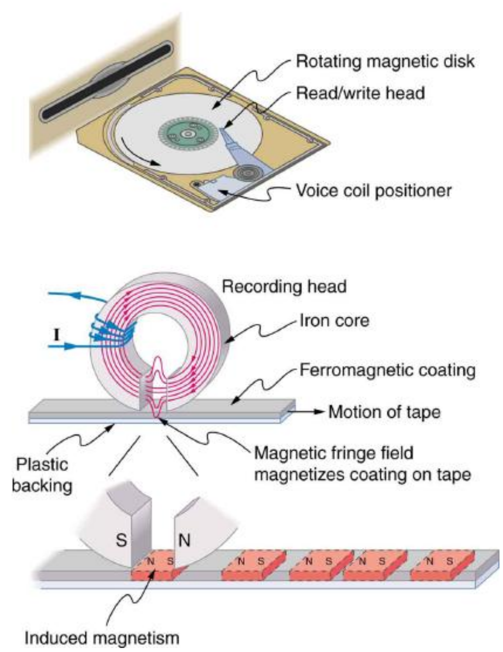


Figure 9.3.6: An electromagnet induces regions of permanent magnetism on a floppy disk coated with a ferromagnetic material. The information stored here is digital (a region is either magnetic or not); in other applications, it can be analog (with a varying strength), such as on audiotapes.

Current: The Source of All Magnetism

An electromagnet creates magnetism with an electric current. In later sections we explore this more quantitatively, finding the strength and direction of magnetic fields created by various currents. But what about ferromagnets? Figure 7 shows models of how electric currents create magnetism at the submicroscopic level. (Note that we cannot directly observe the paths of individual electrons about atoms, and so a model or visual image, consistent with all direct observations, is made. We can directly observe the electron’s orbital angular momentum, its spin momentum, and subsequent magnetic moments, all of which are explained with electric-current-creating subatomic magnetism.) Currents, including those associated with other submicroscopic particles like protons, allow us to explain ferromagnetism and all other magnetic effects. Ferromagnetism, for example, results from an internal cooperative alignment of electron spins, possible in some materials but not in others.

Crucial to the statement that electric current is the source of all magnetism is the fact that it is impossible to separate north and south magnetic poles. (This is far different from the case of positive and negative charges, which are easily separated.) A current loop always produces a magnetic dipole—that is, a magnetic field that acts like a north pole and south pole pair. Since isolated north and south magnetic poles, called magnetic monopoles, are not observed, currents are used to explain all magnetic effects. If magnetic monopoles did exist, then we would have to modify this underlying connection that all magnetism is due to electrical current. There is no known reason that magnetic monopoles should not exist—they are simply never observed—and so searches at the subnuclear level continue. If they do *not* exist, we would like to find out why not. If they *do* exist, we would like to see evidence of them.

ELECTRIC CURRENTS AND MAGNETISM:

Electric current is the source of all magnetism.

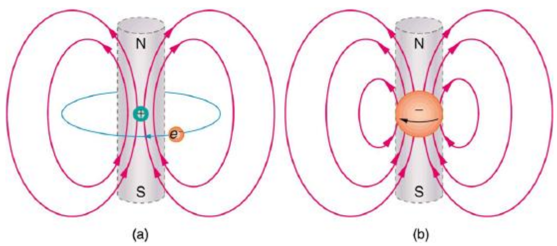


Figure 9.3.7: (a) In the planetary model of the atom, an electron orbits a nucleus, forming a closed-current loop and producing a magnetic field with a north pole and a south pole. (b) Electrons have spin and can be crudely pictured as rotating charge, forming a current that produces a magnetic field with a north pole and a south pole. Neither the planetary model nor the image of a spinning electron is completely consistent with modern physics. However, they do provide a useful way of understanding phenomena.

PHET EXPLORATIONS: MAGNETS AND ELECTROMAGNETS

Explore the interactions between a compass and bar magnet. Discover how you can use a battery and wire to make a magnet! Can you make it a stronger magnet? Can you make the magnetic field reverse?

Summary

- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.
- All magnetism is created by electric current.
- Ferromagnetic materials, such as iron, are those that exhibit strong magnetic effects.
- The atoms in ferromagnetic materials act like small magnets (due to currents within the atoms) and can be aligned, usually in millimeter-sized regions called domains.
- Domains can grow and align on a larger scale, producing permanent magnets. Such a material is magnetized, or induced to be magnetic.
- Above a material's Curie temperature, thermal agitation destroys the alignment of atoms, and ferromagnetism disappears.
- Electromagnets employ electric currents to make magnetic fields, often aided by induced fields in ferromagnetic materials.

Glossary

ferromagnetic

materials, such as iron, cobalt, nickel, and gadolinium, that exhibit strong magnetic effects

magnetized

to be turned into a magnet; to be induced to be magnetic

domains

regions within a material that behave like small bar magnets

Curie temperature

the temperature above which a ferromagnetic material cannot be magnetized

electromagnetism

the use of electrical currents to induce magnetism

electromagnet

an object that is temporarily magnetic when an electrical current is passed through it

magnetic monopoles

an isolated magnetic pole; a south pole without a north pole, or vice versa (no magnetic monopole has ever been observed)

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

9.4: Magnetic Fields and Magnetic Field Lines

Learning Objectives

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

- Define magnetic field and describe the magnetic field lines of various magnetic fields.

Einstein is said to have been fascinated by a compass as a child, perhaps musing on how the needle felt a force without direct physical contact. His ability to think deeply and clearly about action at a distance, particularly for gravitational, electric, and magnetic forces, later enabled him to create his revolutionary theory of relativity. Since magnetic forces act at a distance, we define a magnetic field to represent magnetic forces. The pictorial representation of magnetic field lines is very useful in visualizing the strength and direction of the magnetic field. As shown in Figure $9.4.1$, the direction of magnetic field lines is defined to be the direction in which the north end of a compass needle points. The magnetic field is traditionally called the *B*-field.

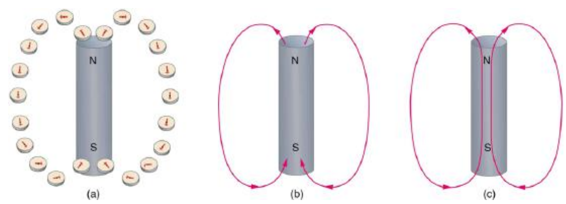


Figure $9.4.1$: Magnetic field lines are defined to have the direction that a small compass points when placed at a location. (a) If small compasses are used to map the magnetic field around a bar magnet, they will point in the directions shown: away from the north pole of the magnet, toward the south pole of the magnet. (Recall that the Earth's north magnetic pole is really a south pole in terms of definitions of poles on a bar magnet.) (b) Connecting the arrows gives continuous magnetic field lines. The strength of the field is proportional to the closeness (or density) of the lines. (c) If the interior of the magnet could be probed, the field lines would be found to form continuous closed loops.

Small compasses used to test a magnetic field will not disturb it. (This is analogous to the way we tested electric fields with a small test charge. In both cases, the fields represent only the object creating them and not the probe testing them.) Figure $9.4.2$ shows how the magnetic field appears for a current loop and a long straight wire, as could be explored with small compasses. A small compass placed in these fields will align itself parallel to the field line at its location, with its north pole pointing in the direction of *B*. Note the symbols used for field into and out of the paper.

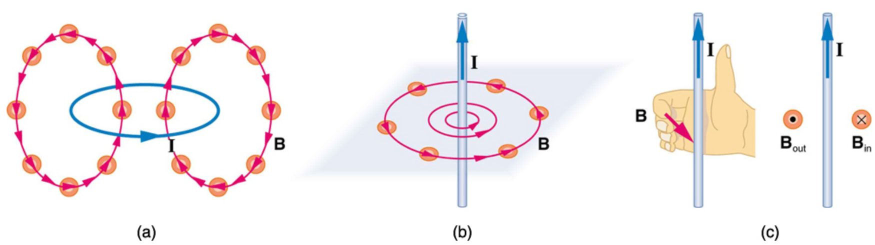


Figure $9.4.2$: Small compasses could be used to map the fields shown here. (a) The magnetic field of a circular current loop is similar to that of a bar magnet. (b) A long and straight wire creates a field with magnetic field lines forming circular loops. (c) When the wire is in the plane of the paper, the field is perpendicular to the paper. Note that the symbols used for the field pointing inward (like the tail of an arrow) and the field pointing outward (like the tip of an arrow).

MAKING CONNECTIONS: CONCEPT OF A FIELD

A field is a way of mapping forces surrounding any object that can act on another object at a distance without apparent physical connection. The field represents the object generating it. Gravitational fields map gravitational forces, electric fields map electrical forces, and magnetic fields map magnetic forces.

Extensive exploration of magnetic fields has revealed a number of hard-and-fast rules. We use magnetic field lines to represent the field (the lines are a pictorial tool, not a physical entity in and of themselves). The properties of magnetic field lines can be summarized by these rules:

1. The direction of the magnetic field is tangent to the field line at any point in space. A small compass will point in the direction of the field line.

2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without beginning or end. They go from the north pole to the south pole.

The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which begin and end on the positive and negative charges. If magnetic monopoles existed, then magnetic field lines would begin and end on them.

Summary

- Magnetic fields can be pictorially represented by magnetic field lines, the properties of which are as follows:
  - The field is tangent to the magnetic field line.
  - Field strength is proportional to the line density.
  - Field lines cannot cross.
  - Field lines are continuous loops.

Glossary

magnetic field
the representation of magnetic forces

***B*-field**
another term for magnetic field

magnetic field lines
the pictorial representation of the strength and the direction of a magnetic field

direction of magnetic field lines
the direction that the north end of a compass needle points

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

9.5: Magnetic Field Strength- Force on a Moving Charge in a Magnetic Field

Learning Objectives

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

- Describe the effects of magnetic fields on moving charges.
- Use the right hand rule 1 to determine the velocity of a charge, the direction of the magnetic field, and the direction of the magnetic force on a moving charge.
- Calculate the magnetic force on a moving charge.

What is the mechanism by which one magnet exerts a force on another? The answer is related to the fact that all magnetism is caused by current, the flow of charge. *Magnetic fields exert forces on moving charges*, and so they exert forces on other magnets, all of which have moving charges.

Right Hand Rule 1

The magnetic force on a moving charge is one of the most fundamental known. Magnetic force is as important as the electrostatic or Coulomb force. Yet the magnetic force is more complex, in both the number of factors that affects it and in its direction, than the relatively simple Coulomb force. The magnitude of the magnetic force F on a charge q moving at a speed v in a direction that is at right angles to a magnetic field of strength B is given by

$$F = qvB \tag{9.5.1}$$

This force is often called the **Lorentz force**. In fact, this is how we define the magnetic field strength B--in in terms of the force on a charged particle moving in a magnetic field. The SI unit for magnetic field strength B is called the **tesla** (T) after the eccentric but brilliant inventor Nikola Tesla (1856–1943). The tesla relates to other SI units as follows:

$$1\,T = \frac{1\,N}{C \cdot m/s} = \frac{1\,N}{A \cdot m} \tag{9.5.2}$$

(note that C/s = A).

Another smaller unit, called the gauss (G), where $1G = 10^{-4}T$, is sometimes used. The strongest permanent magnets have fields near 2 T; superconducting electromagnets may attain 10 T or more. The Earth’s magnetic field on its surface is only about $5 \times 10^{-5}T$, or 0.5 G.

The *direction* of the magnetic force $\mathbf{F}$ is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$, as determined by the right hand rule 1 (or RHR-1), which is illustrated in Figure $9.5.1$. RHR-1 states that, to determine the direction of the magnetic force on a positive moving charge, you point the thumb of the right hand in the direction of v, the fingers in the direction of $\mathbf{B}$, and a perpendicular to the palm points in the direction of $\mathbf{F}$. One way to remember this is that there is one velocity, and so the thumb represents it. There are many field lines, and so the fingers represent them. The force is in the direction you would push with your palm. The force on a negative charge is in exactly the opposite direction to that on a positive charge.

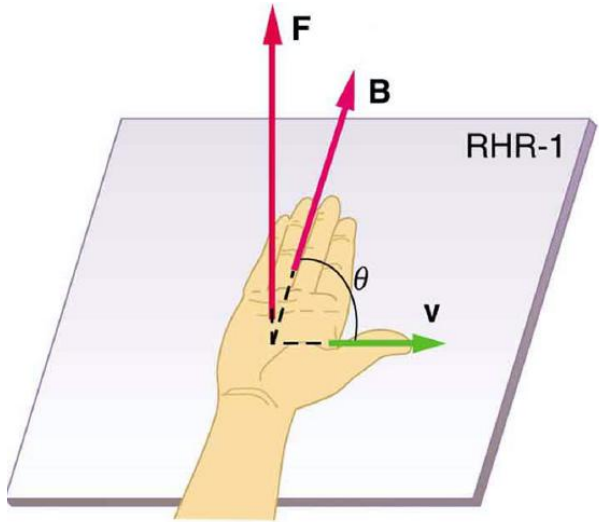


Figure 9.5.1: Magnetic fields exert forces on moving charges. This force is one of the most basic known. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$ and follows right hand rule–1 (RHR-1) as shown. The magnitude of the force is proportional to q, v, B, and the sine of the angle between $\mathbf{v}$ and $\mathbf{B}$.

MAKING CONNECTIONS: CHARGES AND MAGNETS

There is no magnetic force on static charges. However, there is a magnetic force on moving charges. When charges are stationary, their electric fields do not affect magnets. But, when charges move, they produce magnetic fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic fields emerges—each affects the other.

Summary

- Magnetic fields exert a force on a moving charge q, the magnitude of which is

$$F = qvB,$$

.

- The SI unit for magnetic field strength B is the tesla (T), which is related to other units by

$$1T = \frac{1N}{C\cdot m/s} = \frac{1N}{A\cdot m}.$$

- The *direction* of the force on a moving charge is given by right hand rule 1 (RHR-1): Point the thumb of the right hand in the direction of v, the fingers in the direction of B, and a perpendicular to the palm points in the direction of F.
- The force is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$. Since the force is zero if $\mathbf{v}$ is parallel to $\mathbf{B}$, charged particles often follow magnetic field lines rather than cross them.

Glossary

right hand rule 1 (RHR-1)

the rule to determine the direction of the magnetic force on a positive moving charge: when the thumb of the right hand points in the direction of the charge's velocity v and the fingers point in the direction of the magnetic field B then the force on the charge is perpendicular and away from the palm; the force on a negative charge is perpendicular and into the palm

Lorentz force

the force on a charge moving in a magnetic field

tesla

T, the SI unit of the magnetic field strength; $1T = \frac{1N}{A\cdot m}$

magnetic force

the force on a charge produced by its motion through a magnetic field; the Lorentz force

gauss

G, the unit of the magnetic field strength; $1 G = 10^{-4} T$ 10–4T

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

9.6: Magnetic Force on a Current-Carrying Conductor

Learning Objectives

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

- Describe the effects of a magnetic force on a current-carrying conductor.
- Calculate the magnetic force on a current-carrying conductor.

Because charges ordinarily cannot escape a conductor, the magnetic force on charges moving in a conductor is transmitted to the conductor itself.

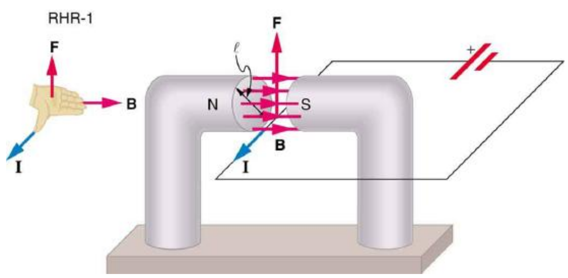


Figure 9.6.1: The magnetic field exerts a force on a current-carrying wire in a direction given by the right hand rule 1 (the same direction as that on the individual moving charges). This force can easily be large enough to move the wire, since typical currents consist of very large numbers of moving charges.

It can therefore be shown that the force on a current-carrying conductor is given by the expression

$$F = IlB \tag{9.6.1}$$

where I is the current, l is the length of the conductor, and B is the strength of the magnetic field. The direction of this force is given by RHR-1, with the thumb in the direction of the current I. Then, with the fingers in the direction of B, a perpendicular to the palm points in the direction of F, as in Figure 2.

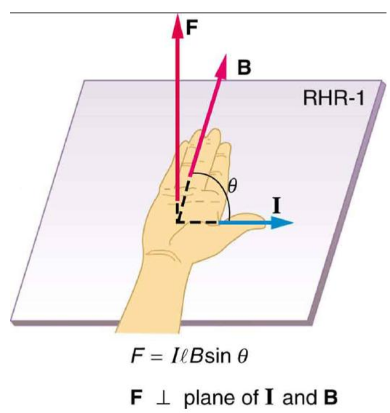


Figure 9.6.2: The force on a current-carrying wire in a magnetic field is $F = \pi B \sin\theta$. Its direction is given by RHR-1.

Magnetic force on current-carrying conductors is used to convert electric energy to work. (Motors are a prime example—they employ loops of wire and are considered in the next section.) Magnetohydrodynamics (MHD) is the technical name given to a clever application where magnetic force pumps fluids without moving mechanical parts (Figure 9.6.3).

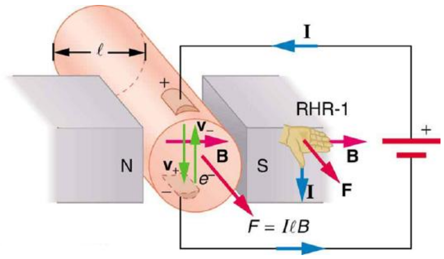


Figure 9.6.3: Magnetohydrodynamics. The magnetic force on the current passed through this fluid can be used as a nonmechanical pump.

A strong magnetic field is applied across a tube and a current is passed through the fluid at right angles to the field, resulting in a force on the fluid parallel to the tube axis as shown. The absence of moving parts makes this attractive for moving a hot, chemically active substance, such as the liquid sodium employed in some nuclear reactors. Experimental artificial hearts are testing with this technique for pumping blood, perhaps circumventing the adverse effects of mechanical pumps. (Cell membranes, however, are affected by the large fields needed in MHD, delaying its practical application in humans.) MHD propulsion for nuclear submarines has been proposed, because it could be considerably quieter than conventional propeller drives. The deterrent value of nuclear submarines is based on their ability to hide and survive a first or second nuclear strike. As we slowly disassemble our nuclear weapons arsenals, the submarine branch will be the last to be decommissioned because of this ability (Figure 9.6.4). Existing MHD drives are heavy and inefficient—much development work is needed.

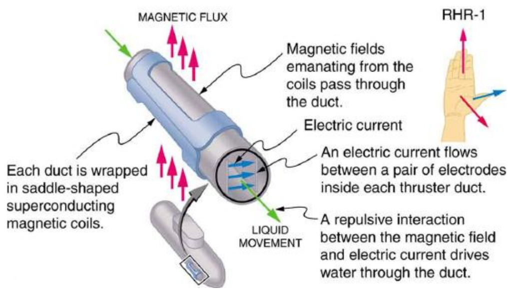


Figure 9.6.4:An MHD propulsion system in a nuclear submarine could produce significantly less turbulence than propellers and allow it to run more silently. The development of a silent drive submarine was dramatized in the book and the film The Hunt for Red October.

Summary

- The magnetic force on current-carrying conductors is given by

$$F = IlB \tag{9.6.2}$$

where I is the current, and l the length of a straight conductor in a uniform magnetic field B. The force follows RHR-1 with the thumb in the direction of I.

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

CHAPTER OVERVIEW

10: Electromagnetic Induction, AC Circuits, and Electrical Technologies

Historically, it was very shortly after Oersted discovered currents cause magnetic fields that other scientists asked the following question: Can magnetic fields cause currents? The answer was soon found by experiment to be yes. In 1831, some 12 years after Oersted's discovery, the English scientist Michael Faraday (1791–1862) and the American scientist Joseph Henry (1797–1878) independently demonstrated that magnetic fields can produce currents. The basic process of generating emfs (electromotive force) and, hence, currents with magnetic fields is known as **induction**; this process is also called magnetic induction to distinguish it from charging by induction, which utilizes the Coulomb force.

Thumbnail: Small cheap inductor. (CC-SA-BY 3.0; FDominec).

10.1: Prelude to Electromagnetic Induction, AC Circuits and Electrical Technologies

Nature's displays of symmetry are beautiful and alluring. A butterfly's wings exhibit an appealing symmetry in a complex system. (See Figure 2.) The laws of physics display symmetries at the most basic level—these symmetries are a source of wonder and imply deeper meaning. Since we place a high value on symmetry, we look for it when we explore nature. The remarkable thing is that we find it.

Figure 10.1.1: These wind turbines in the Thames Estuary in the UK are an example of induction at work. Wind pushes the blades of the turbine, spinning a shaft attached to magnets. The magnets spin around a conductive coil, inducing an electric current in the coil, and eventually feeding the electrical grid. (credit: modification of work by Petr Kratochvil)

Figure 10.1.2: Physics, like this butterfly, has inherent symmetries. (credit: Thomas Bresson)

The hint of symmetry between electricity and magnetism found in the preceding chapter will be elaborated upon in this chapter. Specifically, we know that a current creates a magnetic field. If nature is symmetric here, then perhaps a magnetic field can create a current. The Hall effect is a voltage caused by a magnetic force. That voltage could drive a current. Historically, it was very shortly after Oersted discovered currents cause magnetic fields that other scientists asked the following question: Can magnetic fields cause currents? The answer was soon found by experiment to be yes. In 1831, some 12 years after Oersted's discovery, the English scientist Michael Faraday (1791–1862) and the American scientist Joseph Henry (1797–1878) independently demonstrated that magnetic fields can produce currents. The basic process of generating emfs (electromotive force) and, hence, currents with magnetic fields is known as **induction**; this process is also called magnetic induction to distinguish it from charging by induction, which utilizes the Coulomb force.

Today, currents induced by magnetic fields are essential to our technological society. The ubiquitous generator—found in automobiles, on bicycles, in nuclear power plants, and so on—uses magnetism to generate current. Other devices that use magnetism to induce currents include pickup coils in electric guitars, transformers of every size, certain microphones, airport security gates, and damping mechanisms on sensitive chemical balances. Not so familiar perhaps, but important nevertheless, is that the behavior of AC circuits depends strongly on the effect of magnetic fields on currents.

Glossary

induction

(magnetic induction) the creation of emfs and hence currents by magnetic fields

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

This page titled 10.1: Prelude to Electromagnetic Induction, AC Circuits and Electrical Technologies is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **23.0: Prelude to Electromagnetic Induction, AC Circuits and Electrical Technologies** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.
- **Current page** by OpenStax is licensed CC BY 4.0.

10.2: Induced Emf and Magnetic Flux

Learning Objectives

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

- Calculate the flux of a uniform magnetic field through a loop of arbitrary orientation.
- Describe methods to produce an electromotive force (emf) with a magnetic field or magnet and a loop of wire.

The apparatus used by Faraday to demonstrate that magnetic fields can create currents is illustrated in Figure $10.2.1$. When the switch is closed, a magnetic field is produced in the coil on the top part of the iron ring and transmitted to the coil on the bottom part of the ring. The galvanometer is used to detect any current induced in the coil on the bottom. It was found that each time the switch is closed, the galvanometer detects a current in one direction in the coil on the bottom. (You can also observe this in a physics lab.) Each time the switch is opened, the galvanometer detects a current in the opposite direction. Interestingly, if the switch remains closed or open for any length of time, there is no current through the galvanometer. *Closing and opening the switch* induces the current. It is the *change* in magnetic field that creates the current. More basic than the current that flows is the *emf* that causes it. The current is a result of an *emf induced by a changing magnetic field*, whether or not there is a path for current to flow.

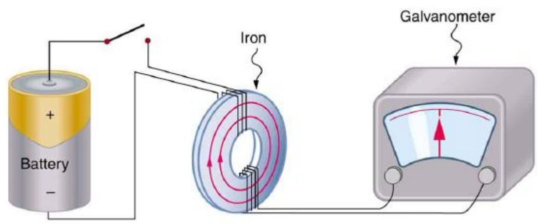


Figure 10.2.1: Faraday's apparatus for demonstrating that a magnetic field can produce a current. A change in the field produced by the top coil induces an emf and, hence, a current in the bottom coil. When the switch is opened and closed, the galvanometer registers currents in opposite directions. No current flows through the galvanometer when the switch remains closed or open.

An experiment easily performed and often done in physics labs is illustrated in Figure $10.2.2$. An emf is induced in the coil when a bar magnet is pushed in and out of it. Emfs of opposite signs are produced by motion in opposite directions, and the emfs are also reversed by reversing poles. The same results are produced if the coil is moved rather than the magnet—it is the relative motion that is important. The faster the motion, the greater the emf, and there is no emf when the magnet is stationary relative to the coil.

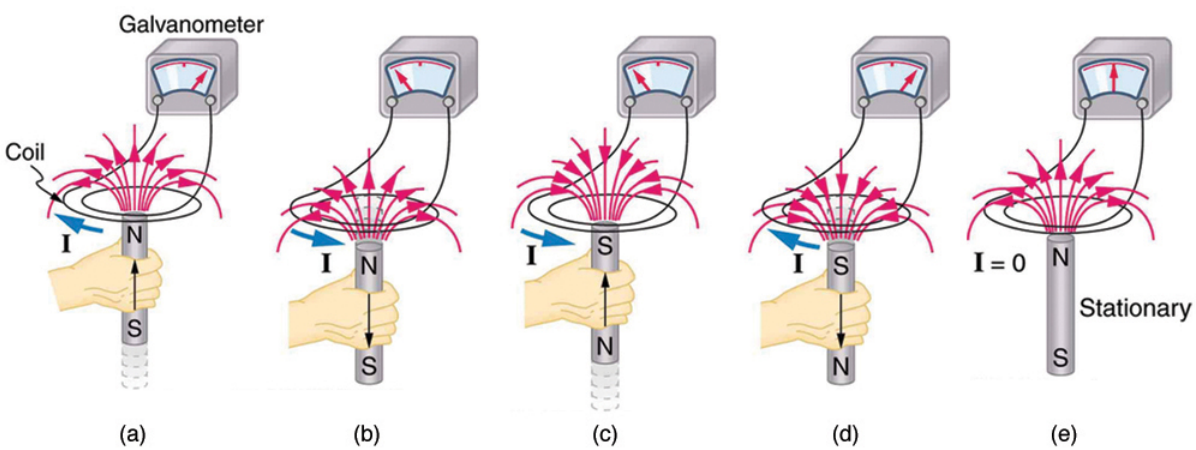


Figure 10.2.2: Movement of a magnet relative to a coil produces emfs as shown. The same emfs are produced if the coil is moved relative to the magnet. The greater the speed, the greater the magnitude of the emf, and the emf is zero when there is no motion.

The method of inducing an emf used in most electric generators is shown in Figure $10.2.3$. A coil is rotated in a magnetic field, producing an alternating current emf, which depends on rotation rate and other factors that will be explored in later sections. Note that the generator is remarkably similar in construction to a motor (another symmetry).

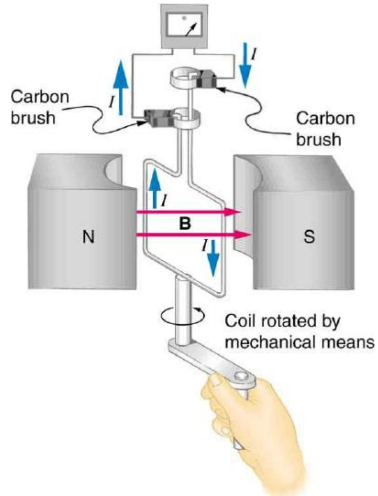


Figure 10.2.3: Rotation of a coil in a magnetic field produces an emf. This is the basic construction of a generator, where work done to turn the coil is converted to electric energy. Note the generator is very similar in construction to a motor.

So we see that changing the magnitude or direction of a magnetic field produces an emf. Experiments revealed that there is a crucial quantity called the **magnetic flux**, Φ, given by

$$\Phi = B\perp A, \tag{10.2.1}$$

where B is the magnetic field strength over an area A, at an angle θ with the perpendicular to the area as shown in Figure $10.2.4$

Any change in magnetic flux Φ induces an emf. This process is defined to be **electromagnetic induction**. Units of magnetic flux Φ are $T \cdot m^2$. As seen in Figure 4, B_{\perp}\) is the component of B perpendicular to the area A. Thus magnetic flux is $\Phi = B_{\perp} A$, the product of the area and the component of the magnetic field perpendicular to it.

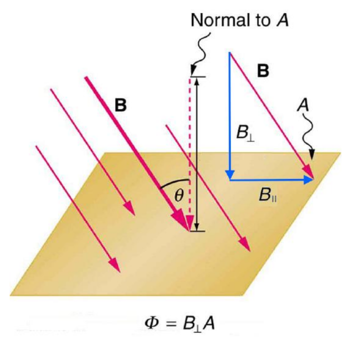


Figure 10.2.4: Magnetic flux Φ is related to the magnetic field and the area over which it exists. The flux $\Phi = B\perp A$ is related to induction; any change in Φ induces an emf.

All induction, including the examples given so far, arises from some change in magnetic flux Φ. For example, Faraday changed B and hence Φ when opening and closing the switch in his apparatus (shown in Figure $10.2.1$). This is also true for the bar magnet and coil shown in Figure $10.2.2$ When rotating the coil of a generator, the angle θ and, hence, Φ is changed. Just how great an emf and what direction it takes depend on the change in Φ and how rapidly the change is made, as examined in the next section.

Summary

- The crucial quantity in induction is magnetic flux Φ, defined to be $\Phi = B\perp A$, where B is the magnetic field strength over an area A at an angle θ with the perpendicular to the area.
- Units of the magnetic flux Φ are $T \cdot m^2$.
- Any change in magnetic flux Φ induces an emf—the process is defined to be electromagnetic induction.

Glossary

magnetic flux

the amount of magnetic field going through a particular area, calculated with $\Phi = BAcos\theta$ where B is the magnetic field strength over an area A at an angle θ with the perpendicular to the area

electromagnetic induction

the process of inducing an emf (voltage) with a change in magnetic flux

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

10.3: Faraday’s Law of Induction- Lenz’s Law

Learning Objectives

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

- Calculate emf, current, and magnetic fields using Faraday’s Law.
- Explain the physical results of Lenz’s Law

Faraday’s and Lenz’s Law

Faraday’s experiments showed that the emf induced by a change in magnetic flux depends on only a few factors. First, emf is directly proportional to the change in flux $\Delta\Phi$. Second, emf is greatest when the change in time Δt is smallest—that is, emf is inversely proportional to Δt. Finally, if a coil has N turns, an emf will be produced that is N times greater than for a single coil, so that emf is directly proportional to N. The equation for the emf induced by a change in magnetic flux is

$$emf = -N\frac{\Delta\Phi}{\Delta t}. \tag{10.3.1}$$

This relationship is known as **Faraday's law of induction**. The units for emf are volts, as is usual.

The minus sign in Faraday’s law of induction is very important. The minus means that *the emf creates a current I and magnetic field B that oppose the change in flux* $\Delta\Phi$ *-- this is known as Lenz's law*. The direction (given by the minus sign) of the emfis so important that it is called **Lenz's law** after the Russian Heinrich Lenz (1804–1865), who, like Faraday and Henry,independently investigated aspects of induction. Faraday was aware of the direction, but Lenz stated it so clearly that he is credited for its discovery. (See Figure 1.)

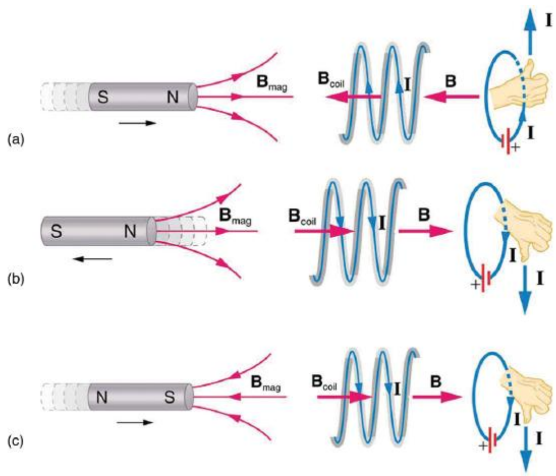


Figure 10.3.1: (a) When this bar magnet is thrust into the coil, the strength of the magnetic field increases in the coil. The current induced in the coil creates another field, in the opposite direction of the bar magnet’s to oppose the increase. This is one aspect of Lenz's law--induction opposes any charge in flux. (b) and (c) are two other situations. Verify for yourself that the direction of the induced B_{coil} shown indeed opposes the change in flux and that the current direction shown is consistent with RHR-2.

PROBLEM-SOLVING STRATEGY FOR LENZ'S LAW:

To use Lenz’s law to determine the directions of the induced magnetic fields, currents, and emfs:

1. Make a sketch of the situation for use in visualizing and recording directions.
2. Determine the direction of the magnetic field B.
3. Determine whether the flux is increasing or decreasing.
4. Now determine the direction of the induced magnetic field B. It opposes the change in flux by adding or subtracting from the original field.
5. Use RHR-2 to determine the direction of the induced current I that is responsible for the induced magnetic field B.
6. The direction (or polarity) of the induced emf will now drive a current in this direction and can be represented as current emerging from the positive terminal of the emf and returning to its negative terminal.

For practice, apply these steps to the situations shown in Figure 1 and to others that are part of the following text material.

Applications of Electromagnetic Induction

There are many applications of Faraday's Law of induction, as we will explore in this chapter and others. At this juncture, let us mention several that have to do with data storage and magnetic fields. A very important application has to do with audio and video *recording tapes*. A plastic tape, coated with iron oxide, moves past a recording head. This recording head is basically a round iron ring about which is wrapped a coil of wire—an electromagnet (Figure 2). A signal in the form of a varying input current from a microphone or camera goes to the recording head. These signals (which are a function of the signal amplitude and frequency) produce varying magnetic fields at the recording head. As the tape moves past the recording head, the magnetic field orientations of the iron oxide molecules on the tape are changed thus recording the signal. In the playback mode, the magnetized tape is run past another head, similar in structure to the recording head. The different magnetic field orientations of the iron oxide molecules on the tape induces an emf in the coil of wire in the playback head. This signal then is sent to a loudspeaker or video player.

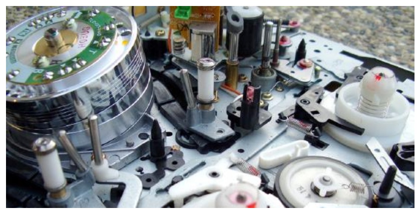


Figure 10.3.2: Recording and playback heads used with audio and video magnetic tapes. (credit: Steve Jurvetson)

Similar principles apply to computer hard drives, except at a much faster rate. Here recordings are on a coated, spinning disk. Read heads historically were made to work on the principle of induction. However, the input information is carried in digital rather than analog form – a series of 0's or 1's are written upon the spinning hard drive. Today, most hard drive readout devices do not work on the principle of induction, but use a technique known as *giant magnetoresistance*. (The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology.) Another application of induction is found on the magnetic stripe on the back of your personal credit card as used at the grocery store or the ATM machine. This works on the same principle as the audio or video tape mentioned in the last paragraph in which a head reads personal information from your card.

Another application of electromagnetic induction is when electrical signals need to be transmitted across a barrier. Consider the *cochlear implant* shown below. Sound is picked up by a microphone on the outside of the skull and is used to set up a varying magnetic field. A current is induced in a receiver secured in the bone beneath the skin and transmitted to electrodes in the inner ear. Electromagnetic induction can be used in other instances where electric signals need to be conveyed across various media.

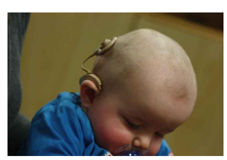


Figure 10.3.3: Electromagnetic induction used in transmitting electric currents across mediums. The device on the baby's head induces an electrical current in a receiver secured in the bone beneath the skin. (credit: Bjorn Knetsch)

Another contemporary area of research in which electromagnetic induction is being successfully implemented (and with substantial potential) is transcranial magnetic simulation. A host of disorders, including depression and hallucinations can be traced to irregular localized electrical activity in the brain. In *transcranial magnetic stimulation*, a rapidly varying and very localized magnetic field is placed close to certain sites identified in the brain. Weak electric currents are induced in the identified sites and can result in recovery of electrical functioning in the brain tissue.

Sleep apnea ("the cessation of breath") affects both adults and infants (especially premature babies and it may be a cause of sudden infant deaths [SID]). In such individuals, breath can stop repeatedly during their sleep. A cessation of more than 20 seconds can be very dangerous. Stroke, heart failure, and tiredness are just some of the possible consequences for a person having sleep apnea. The concern in infants is the stopping of breath for these longer times. One type of monitor to alert parents when a child is not breathing uses electromagnetic induction. A wire wrapped around the infant's chest has an alternating current running through it. The expansion and contraction of the infant's chest as the infant breathes changes the area through the coil. A pickup coil located nearby has an alternating current induced in it due to the changing magnetic field of the initial wire. If the child stops breathing, there will be a change in the induced current, and so a parent can be alerted.

MAKING CONNECTIONS: CONSERVATION OF ENERGY:

Lenz's law is a manifestation of the conservation of energy. The induced emf produces a current that opposes the change in flux, because a change in flux means a change in energy. Energy can enter or leave, but not instantaneously. Lenz's law is a consequence. As the change begins, the law says induction opposes and, thus, slows the change. In fact, if the induced emf were in the same direction as the change in flux, there would be a positive feedback that would give us free energy from no apparent source—conservation of energy would be violated.

Example 10.3.1: Calculating Emf: How Great is the Induced Emf?

Calculate the magnitude of the induced emf when the magnet in Figure 1a is thrust into the coil, given the following information: the single loop coil has a radius of 6.00 cm and the average value of $B\cos\theta$ (this is given, since the bar magnet's field is complex) increases from 0.0500 T to 0.250 T in 0.100 s.

Strategy:

To find the *magnitute* of emf, we use Faraday's law of induction as stated by $emf = -N\frac{\Delta\Phi}{\Delta t}$, but without the minus sign that indicates direction:

$$emf = N\frac{\Delta\Phi}{\Delta t}. \tag{10.3.2}$$

Solution:

We are given that $N = 1$ and $\Delta t = 0.100s$ but we must determine the change in flux $\Delta\Phi$ before we can find emf. Since the area of the loop is fixed, we see that

$$\Delta\Phi\left(BA\cos\theta\right) = A\Delta\left(B\cos\theta\right). \tag{10.3.3}$$

Now $\Delta\left(B\cos\theta\right) = 0.200T$, since it was given that $B\cos\theta$ changes from 0.0500 to 0.250 T. The area of the loop is $A = \pi r^2 = (3.14\ldots)\left(0.060m\right)^2 = 1.13\times10^{-2}m^2$. Thus,

$$\Delta\Phi = \left(1.13\times10^{-2}m^2\right)\left(0.200T\right). \tag{10.3.4}$$

Entering the determined values into the expression for emf gives

$$Emf = N\frac{\Delta\Phi}{\Delta t} = \frac{\left(1.13\times10^{-2}m^2\right)\left(0.200T\right)}{0.100s} = 22.6mV. \tag{10.3.5}$$

Discussion:

While this is an easily measured voltage, it is certainly not large enough for most practical applications. More loops in the coil, a stronger magnet, and faster movement make induction the practical source of voltages that it is.

Summary

- Faraday's law of induction states that the emfinduced by a change in magnetic flux is

$$emf = N\frac{\Delta\Phi}{\Delta t} \tag{10.3.6}$$

  when flux changes by $\Delta\Phi$ in a time $Deltat$.
- If emf is induced in a coil, N is its number of turns.
- The minus sign means that the emf creates a current I and magnetic field B that *oppose the change in flux* $\Delta\Phi$ -- this opposition is known as Lenz's law.

Glossary

Faraday's law of induction

the means of calculating the emf in a coil due to changing magnetic flux, given by $emf = -N\frac{\Delta\Phi}{\Delta t}$

Lenz's law

the minus sign in Faraday's law, signifying that the emf induced in a coil opposes the change in magnetic flux

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

10.4: Motional Emf

Learning Objectives

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

- Calculate emf, force, magnetic field, and work due to the motion of an object in a magnetic field.

As we have seen, any change in magnetic flux induces an emf opposing that change—a process known as induction. Motion is one of the major causes of induction. For example, a magnet moved toward a coil induces an emf, and a coil moved toward a magnet produces a similar emf. In this section, we concentrate on motion in a magnetic field that is stationary relative to the Earth, producing what is loosely called *motional emf.*

One situation where motional emf occurs is known as the Hall effect and has already been examined. Charges moving in a magnetic field experience the magnetic force $F = qvB\sin\theta$, which moves opposite charges in opposite directions and produces an $emf = Blv$. We saw that the Hall effect has applications, including measurements of B and v. We will now see that the Hall effect is one aspect of the broader phenomenon of induction, and we will find that motional emf can be used as a power source.

Consider the situation shown in Figure $10.4.1$. A rod is moved at a speed v along a pair of conducting rails separated by a distance l in a uniform magnetic field B. The rails are stationary relative to B and are connected to a stationary resistor R. The resistor could be anything from a light bulb to a voltmeter. Consider the area enclosed by the moving rod, rails, and resistor. B is perpendicular to this area, and the area is increasing as the rod moves. Thus the magnetic flux enclosed by the rails, rod, and resistor is increasing. When flux changes, an emf is induced according to Faraday's law of induction.

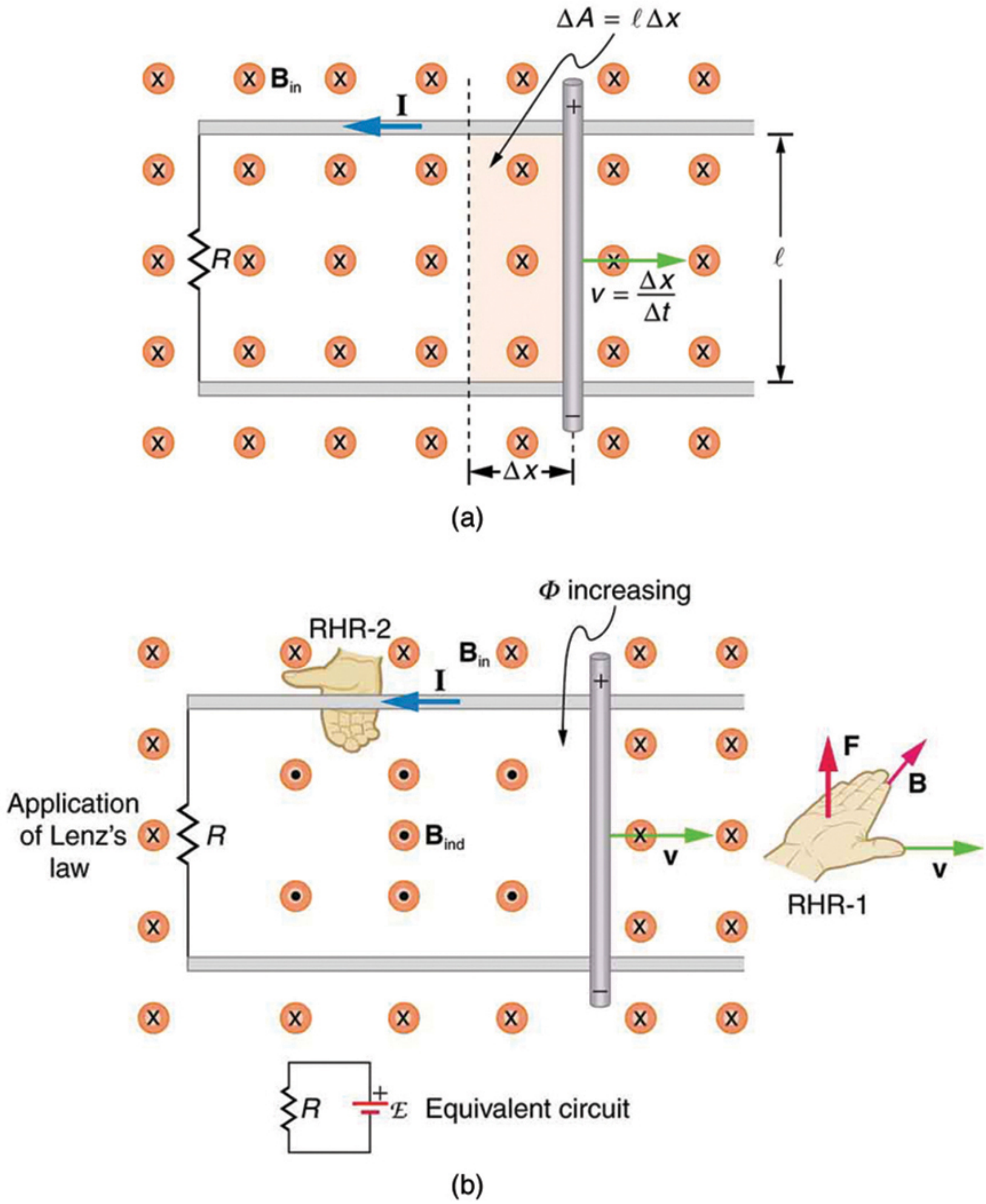


Figure 10.4.1: (a) A motional $emf = Blv$ is induced between the rails when this rod moves to the right in the uniform magnetic field. The magnetic field B is into the page, perpendicular to the moving rod and rails and, hence, to the area enclosed by them. (b) Lenz's law gives the directions of the induced field and current, and the polarity of the induced emf. Since the flux is increasing, the induced field is in the opposite direction, or out of the page. RHR-2 gives the current direction shown, and the polarity of the rod will drive such a current. RHR-1 also indicates the same polarity for the rod. (Note that the script E symbol used in the equivalent circuit at the bottom of part (b) represents emf.)

To find the magnitude of emf induced along the moving rod, we use Faraday's law of induction without the sign:

$$emf = N\frac{\Delta\Phi}{\Delta t}. \tag{10.4.1}$$

Here and below, "emf" implies the magnitude of the emf. In this equation, $N = 1$ and the flux $\Phi = BA\cos\theta$. We have $\theta = 0^\circ$ and $\cos\theta = 1$, since B is perpendicular to A. Now $\Delta\Phi = \Delta(BA) = B\Delta A$, since B is uniform. Note that the area swept out by the rod is $\Delta A = l\Delta x$. Entering these quantities into the expression for emf yields

$$emf = \frac{B\Delta A}{\Delta t} = B\frac{l\Delta x}{\Delta t}. \tag{10.4.2}$$

Finally, note that $\Delta x/\Delta t = v$, the velocity of the rod. Entering this into the last expression shows that

$$emf = Blv(B, l, and \quad v \quad perpendicular) \tag{10.4.3}$$

is the motional emf. This is the same expression given for the Hall effect previously..

MAKING CONNECTIONS: UNIFICATION OF FORCES:

There are many connections between the electric force and the magnetic force. The fact that a moving electric field produces a magnetic field and, conversely, a moving magnetic field produces an electric field is part of why electric and magnetic forces are now considered to be different manifestations of the same force. This classic unification of electric and magnetic forces into what is called the electromagnetic force is the inspiration for contemporary efforts to unify other basic forces.

To find the direction of the induced field, the direction of the current, and the polarity of the induced emf, we apply Lenz's law as explained in "Faraday's Law of Induction: Lenz's Law" (Figure $10.4.1b$).

Flux is increasing, since the area enclosed is increasing. Thus the induced field must oppose the existing one and be out of the page. And so the RHR-2 requires that *I* be counterclockwise, which in turn means the top of the rod is positive as shown.

Motional emf also occurs if the magnetic field moves and the rod (or other object) is stationary relative to the Earth (or some observer). We have seen an example of this in the situation where a moving magnet induces an emf in a stationary coil. It is the relative motion that is important. What is emerging in these observations is a connection between magnetic and electric fields. A moving magnetic field produces an electric field through its induced emf. We already have seen that a moving electric field produces a magnetic field—moving charge implies moving electric field and moving charge produces a magnetic field.

Motional emfs in the Earth's weak magnetic field are not ordinarily very large, or we would notice voltage along metal rods, such as a screwdriver, during ordinary motions. For example, a simple calculation of the motional emf of a 1 m rod moving at 3.0 m/s perpendicular to the Earth's field gives $emf = Blv = \left(5.0 \times 10^{-5} T\right)(1.0m)(3.0m/s) = 150\mu V$. This small value is consistent with experience. There is a spectacular exception, however. In 1992 and 1996, attempts were made with the space shuttle to create large motional emfs. The Tethered Satellite was to be let out on a 20 km length of wire as shown in Figure 2, to create a 5 kV emf by moving at orbital speed through the Earth's field. This emf could be used to convert some of the shuttle's kinetic and potential energy into electrical energy if a complete circuit could be made. To complete the circuit, the stationary ionosphere was to supply a return path for the current to flow. (The ionosphere is the rarefied and partially ionized atmosphere at orbital altitudes. It conducts because of the ionization. The ionosphere serves the same function as the stationary rails and connecting resistor in Figure 1, without which there would not be a complete circuit.) Drag on the current in the cable due to the magnetic force $F = IlB \sin\theta$ does the work that reduces the shuttle's kinetic and potential energy and allows it to be converted to electrical energy. The tests were both unsuccessful. In the first, the cable hung up and could only be extended a couple of hundred meters; in the second, the cable broke when almost fully extended. The example below indicates feasibility in principle.

Example 10.4.1: Calculating the Large Motional Emf of an Object in Orbit

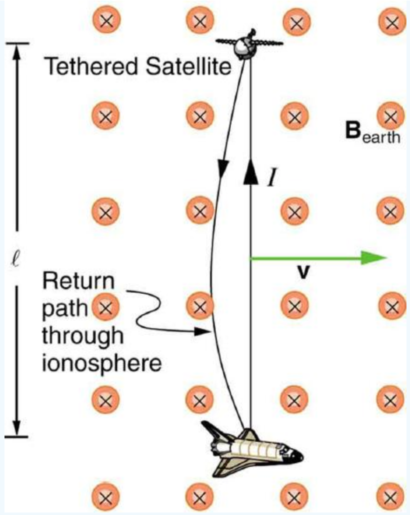


Figure 10.4.2: Motional emf as electrical power conversion for the space shuttle is the motivation for the Tethered Satellite experiment. A 5 kV emf was predicted to be induced in the 20 km long tether while moving at orbital speed in the Earth's magnetic field. The circuit is completed by a return path through the stationary ionosphere.

Calculate the motional emf induced along a 20.0 km long conductor moving at an orbital speed of 7.80 km/s perpendicular to the Earth's $5.00 \times 10^{-5} T$ magnetic field.

Strategy:

This is a straightforward application of the expression for motional emf--$emf = Blv$.

Solution:

Entering the given values into $emf = Blv$ gives

$$emf = Blv \tag{10.4.4}$$

$$= \left(5.00 \times 10^{-5} T\right) \left(2.0 \times 10^{4} m\right) \left(7.80 \times 10^{3} m/s\right) \tag{10.4.5}$$

$$= 7.80 \times 10^{3} V. \tag{10.4.6}$$

Discussion:

The value obtained is greater than the 5 kV measured voltage for the shuttle experiment, since the actual orbital motion of the tether is not perpendicular to the Earth's field. The 7.80 kV value is the maximum emf obtained when $\theta = 90^\circ$ and $\sin\theta = 1$.

Summary

- An emf induced by motion relative to a magnetic field B is called a *motional emf* and is given by

$$emf = Blv \left(B, l, and \quad v \quad perpendicular\right), \tag{10.4.7}$$

where l is the length of the object moving at speed v relative to the field.

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

10.5: Electric Generators

Learning Objectives

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

- Calculate the emf induced in a generator.
- Calculate the peak emf which can be induced in a particular generator system.

Electric generators induce an emf by rotating a coil in a magnetic field, as briefly discussed in "Induced Emf and Magnetic Flux." We will now explore generators in more detail. Consider the following example.

The emf calculated in the example is the average over one-fourth of a revolution. What is the emf at any given instant? It varies with the angle between the magnetic field and a perpendicular to the coil. We can get an expression for emf as a function of time by considering the motional emf on a rotating rectangular coil of width w and height l in a uniform magnetic field, as illustrated in Figure $10.5.2$

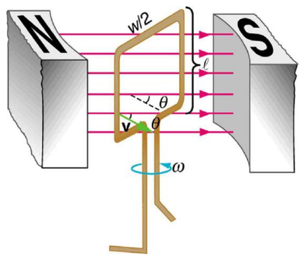


Figure 10.5.2: A generator with a single rectangular coil rotated at constant angular velocity in a uniform magnetic field produces an emf that varies sinusoidally in time. Note the generator is similar to a motor, except the shaft is rotated to produce a current rather than the other way around.

Charges in the wires of the loop experience the magnetic force, because they are moving in a magnetic field, and a current is generated, which in turn results in an emf across the loop. As the loop turns, the angle between the loop and the magnetic field oscillates continuously between 90 and 0 degrees, and which means the induced emf in the loop also oscillates. As it oscillates, the emf never exceeds a particular maximum value. This is called the **peak emf**. Figure $10.5.3$ shows a graph of emf as a function of time; this corresponds to what is commonly referred to as AC voltage.

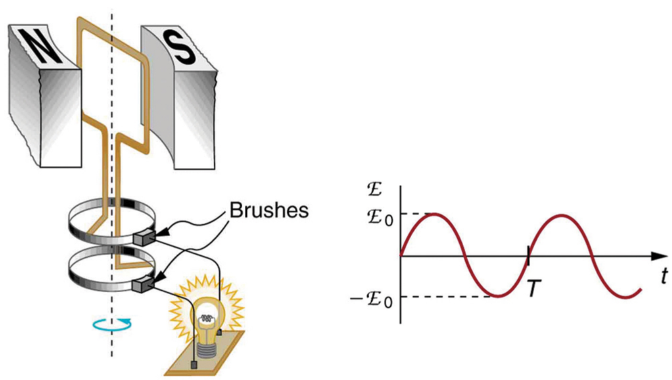


Figure 10.5.3:The emf of a generator is sent to a light bulb with the system of rings and brushes shown. The graph gives the emf of the generator as a function of time. emf_0 is the peak emf. Note that the script E stands for emf.

The greater the number of coils, the larger their area, and the stronger the field, the greater the output voltage. Additionally, the faster the generator is spun, the greater the peak emf. This is noticeable on bicycle generators—at least the cheaper varieties. One of the authors as a juvenile found it amusing to ride his bicycle fast enough to burn out his lights, until he had to ride home lightless one dark night.

Figure shows a scheme by which a generator can be made to produce pulsed DC. More elaborate arrangements of multiple coils and split rings can produce smoother DC, although electronic rather than mechanical means are usually used to make ripple-free DC.

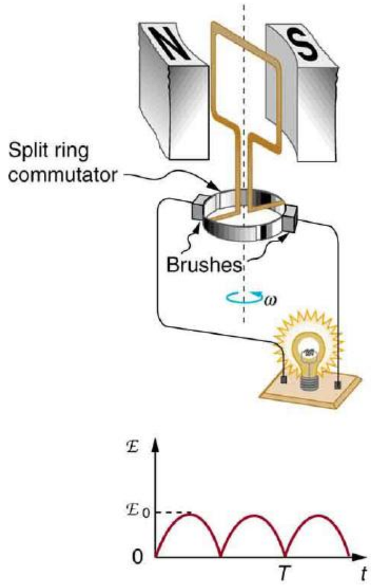


Figure* $10.5.4$*: Split rings, called commutators, produce a pulsed DC emf output in this configuration.

In real life, electric generators look a lot different than the figures in this section, but the principles are the same. The source of mechanical energy that turns the coil can be falling water (hydropower), steam produced by the burning of fossil fuels, or the kinetic energy of wind. $10.5.5$ shows a cutaway view of a steam turbine; steam moves over the blades connected to the shaft, which rotates the coil within the generator.

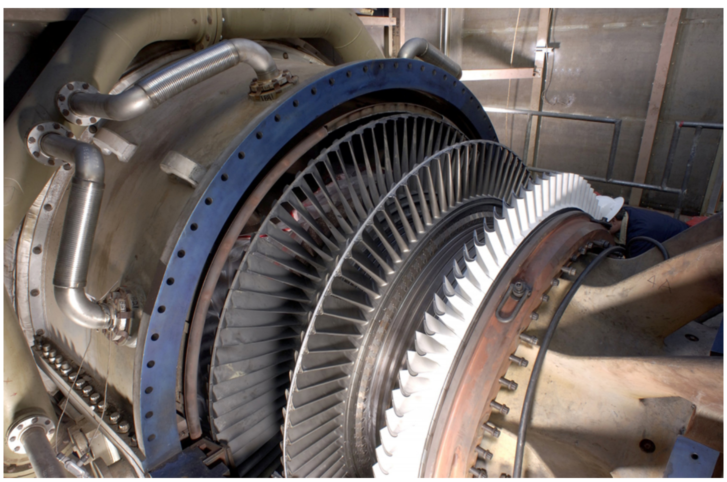


Figure 10.5.5: Steam turbine/generator. The steam produced by burning coal impacts the turbine blades, turning the shaft which is connected to the generator. (credit: Nabonaco, Wikimedia Commons)

Generators illustrated in this section look very much like the motors illustrated previously. This is not coincidental. In fact, a motor becomes a generator when its shaft rotates. Certain early automobiles used their starter motor as a generator. In Back Emf, we shall further explore the action of a motor as a generator.

Summary

- An electric generator rotates a coil in a magnetic field, inducing an emf that varies in time
- The faster the rotation, the bigger the peak value of the induced emf

Glossary

electric generator

a device for converting mechanical work into electric energy; it induces an emf by rotating a coil in a magnetic field

peak emf

the maximum value attained by the induced emf as the coil is rotated in the magnetic field

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

CHAPTER OVERVIEW

11: Geometric Optics

Geometrical optics describes light propagation in terms of rays, which is useful in approximating the paths along which light propagates in certain classes of circumstances. Geometrical optics does not account for certain optical effects such as diffraction and interference.

Thumbnail: Parallel light rays entering a diverging lens from the right seem to come from the focal point on the right.

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

11.1: Prelude to Geometric Optics

Geometric Optics

Light from this page or screen is formed into an image by the lens of your eye, much as the lens of the camera that made this photograph. Mirrors, like lenses, can also form images that in turn are captured by your eye.

Figure 11.1.1: Image seen as a result of reflection of light on a plane smooth surface. (credit: NASA Goddard Photo and Video, via Flickr)

Our lives are filled with light. Through vision, the most valued of our senses, light can evoke spiritual emotions, such as when we view a magnificent sunset or glimpse a rainbow breaking through the clouds. Light can also simply amuse us in a theater, or warn us to stop at an intersection. It has innumerable uses beyond vision. Light can carry telephone signals through glass fibers or cook a meal in a solar oven. Life itself could not exist without light's energy. From photosynthesis in plants to the sun warming a cold-blooded animal, its supply of energy is vital.

Figure 11.1.2: Double Rainbow over the bay of Pocitos in Montevideo, Uruguay. (credit: Madrax, Wikimedia Commons)

We already know that visible light is the type of electromagnetic waves to which our eyes respond. That knowledge still leaves many questions regarding the nature of light and vision. What is color, and how do our eyes detect it? Why do diamonds sparkle? How does light travel? How do lenses and mirrors form images? These are but a few of the questions that are answered by the study of optics. Optics is the branch of physics that deals with the behavior of visible light and other electromagnetic waves. In particular, optics is concerned with the generation and propagation of light and its interaction with matter. What we have already learned about the generation of light in our study of heat transfer by radiation will be expanded upon in later topics, especially those on atomic physics. Now, we will concentrate on the propagation of light and its interaction with matter.

It is convenient to divide optics into two major parts based on the size of objects that light encounters. When light interacts with an object that is several times as large as the light's wavelength, its observable behavior is like that of a ray; it does not prominently display its wave characteristics. We call this part of optics "geometric optics." This chapter will concentrate on such situations. When light interacts with smaller objects, it has very prominent wave characteristics, such as constructive and destructive interference. "Wave Optics" will concentrate on such situations.

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney).

11.2: The Ray Aspect of Light

Learning Objectives

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

- List the ways by which light travels from a source to another location.

There are three ways in which light can travel from a source to another location (Figure $11.2.1$). It can come directly from the source through empty space, such as from the Sun to Earth. Or light can travel through various media, such as air and glass, to the person. Light can also arrive after being reflected, such as by a mirror. In all of these cases, light is modeled as traveling in straight lines called rays. Light may change direction when it encounters objects (such as a mirror) or in passing from one material to another (such as in passing from air to glass), but it then continues in a straight line or as a ray. The word **ray** comes from mathematics and here means a straight line that originates at some point. It is acceptable to visualize light rays as laser rays (or even science fiction depictions of ray guns).

Definition: RAY

The word “ray” comes from mathematics and here means a straight line that originates at some point.

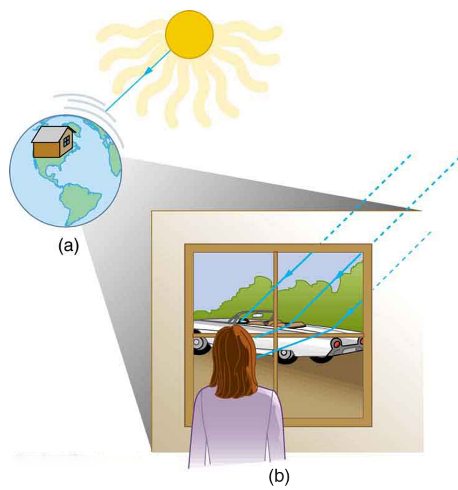


Figure $11.2.1$: Three methods for light to travel from a source to another location. (a) Light reaches the upper atmosphere of Earth traveling through empty space directly from the source. (b) Light can reach a person in one of two ways. It can travel through media like air and glass. It can also reflect from an object like a mirror. In the situations shown here, light interacts with objects large enough that it travels in straight lines, like a ray.

Experiments, as well as our own experiences, show that when light interacts with objects several times as large as its wavelength, it travels in straight lines and acts like a ray. Its wave characteristics are not pronounced in such situations. Since the wavelength of light is less than a micron (a thousandth of a millimeter), it acts like a ray in the many common situations in which it encounters objects larger than a micron. For example, when light encounters anything we can observe with unaided eyes, such as a mirror, it acts like a ray, with only subtle wave characteristics. We will concentrate on the ray characteristics in this chapter.

Since light moves in straight lines, changing directions when it interacts with materials, it is described by geometry and simple trigonometry. This part of optics, where the ray aspect of light dominates, is therefore called geometric optics. There are two laws that govern how light changes direction when it interacts with matter. These are the law of reflection, for situations in which light bounces off matter, and the law of refraction, for situations in which light passes through matter.

Definition: GEOMETRIC OPTICS

The part of optics dealing with the ray aspect of light is called geometric optics.

Summary

- A straight line that originates at some point is called a ray.
- The part of optics dealing with the ray aspect of light is called geometric optics.
- Light can travel in three ways from a source to another location: (1) directly from the source through empty space; (2) through various media; (3) after being reflected from a mirror.

Glossary

ray

straight line that originates at some point

geometric optics

part of optics dealing with the ray aspect of light

Contributors and Attributions

11.3: The Law of Reflection

Learning Objectives

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

- Explain reflection of light from polished and rough surfaces.

Whenever we look into a mirror, or squint at sunlight glinting from a lake, we are seeing a reflection. When you look at this page, too, you are seeing light reflected from it. Large telescopes use reflection to form an image of stars and other astronomical objects.

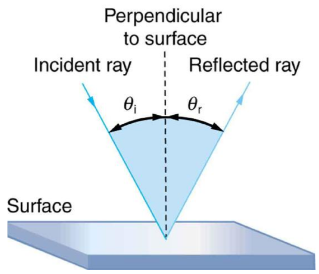


Figure 11.3.1: The law of reflection states that the angle of reflection equals the angle of incidence -- $\theta_r = \theta_i$. The angles are measured relative to the perpendicular to the surface at the point where the ray strikes the surface.

The law of reflection is illustrated in Figure $11.3.1$, which also shows how the angles are measured relative to the perpendicular to the surface at the point where the light ray strikes. We expect to see reflections from smooth surfaces, but Figure $11.3.2$ illustrates how a rough surface reflects light. Since the light strikes different parts of the surface at different angles, it is reflected in many different directions, or diffused.

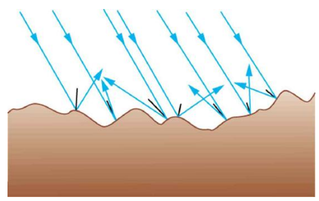


Figure 11.3.2: Light is diffused when it reflects from a rough surface. Here many parallel rays are incident, but they are reflected at many different angles since the surface is rough.

Diffused light is what allows us to see a sheet of paper from any angle, as illustrated in Figure $11.3.3a$. Many objects, such as people, clothing, leaves, and walls, have rough surfaces and can be seen from all sides. A mirror, on the other hand, has a smooth surface (compared with the wavelength of light) and reflects light at specific angles, as illustrated in Figure $11.3.3b$. When the moon reflects from a lake, as shown in Figure $11.3.4$, a combination of these effects takes place.

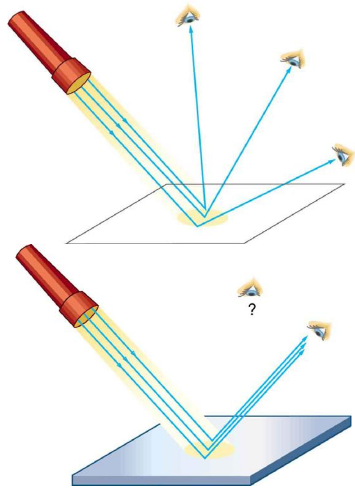


Figure 11.3.3: (left) When a sheet of paper is illuminated with many parallel incident rays, it can be seen at many different angles, because its surface is rough and diffuses the light. (right) A mirror illuminated by many parallel rays reflects them in only one direction, since its surface is very smooth. Only the observer at a particular angle will see the reflected light.

Figure 11.3.4: Moonlight is spread out when it is reflected by the lake, since the surface is shiny but uneven. (credit: Diego Torres Silvestre, Flickr)

The law of reflection is very simple: The angle of reflection equals the angle of incidence.

Definition: THE LAW OF REFLECTION

The angle of reflection equals the angle of incidence.

When we see ourselves in a mirror, it appears that our image is actually behind the mirror. This is illustrated in Figure 11.3.5. We see the light coming from a direction determined by the law of reflection. The angles are such that our image is exactly the same distance behind the mirror as we stand away from the mirror. If the mirror is on the wall of a room, the images in it are all behind the mirror, which can make the room seem bigger. Although these mirror images make objects appear to be where they cannot be (like behind a solid wall), the images are not figments of our imagination. Mirror images can be photographed and videotaped by instruments and look just as they do with our eyes (optical instruments themselves). The precise manner in which images are formed by mirrors and lenses will be treated in later sections of this chapter

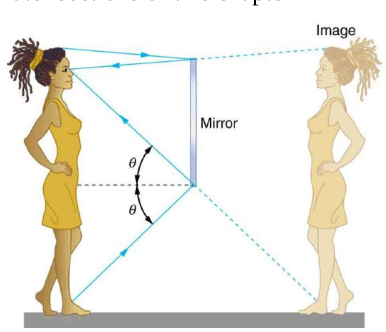


Figure 11.3.5: Our image in a mirror is behind the mirror. The two rays shown are those that strike the mirror at just the correct angles to be reflected into the eyes of the person. The image appears to be in the direction the rays are coming from when they enter the eyes.

TAKE-HOME EXPERIMENT: LAW OF REFLECTION

Take a piece of paper and shine a flashlight at an angle at the paper, as shown in Figure $11.3.5a$. Now shine the flashlight at a mirror at an angle. Do your observations confirm the predictions in Figure $11.3.3$? Shine the flashlight on various surfaces and determine whether the reflected light is diffuse or not. You can choose a shiny metallic lid of a pot or your skin. Using the mirror and flashlight, can you confirm the law of reflection? You will need to draw lines on a piece of paper showing the incident and reflected rays. (This part works even better if you use a laser pencil.)

Summary

- The angle of reflection equals the angle of incidence.
- A mirror has a smooth surface and reflects light at specific angles.
- Light is diffused when it reflects from a rough surface.
- Mirror images can be photographed and videotaped by instruments.

Glossary

mirror

smooth surface that reflects light at specific angles, forming an image of the person or object in front of it

law of reflection

angle of reflection equals the angle of incidence

Contributors and Attributions

11.4: The Law of Refraction

Learning Objectives

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

- Determine the index of refraction, given the speed of light in a medium.

It is easy to notice some odd things when looking into a fish tank. For example, you may see the same fish appearing to be in two different places (Figure $11.4.1$). This is because light coming from the fish to us changes direction when it leaves the tank, and in this case, it can travel two different paths to get to our eyes. The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called **refraction**. Refraction is responsible for a tremendous range of optical phenomena, from the action of lenses to voice transmission through optical fibers.

Definition: REFRACTION

The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called refraction.

SPEED OF LIGHT

The speed of light c not only affects refraction, it is one of the central concepts of Einstein's theory of relativity. As the accuracy of the measurements of the speed of light were improved, c was found not to depend on the velocity of the source or the observer. However, the speed of light does vary in a precise manner with the material it traverses. These facts have far-reaching implications, as we will see in "Special Relativity." It makes connections between space and time and alters our expectations that all observers measure the same time for the same event, for example. The speed of light is so important that its value in a vacuum is one of the most fundamental constants in nature as well as being one of the four fundamental SI units.

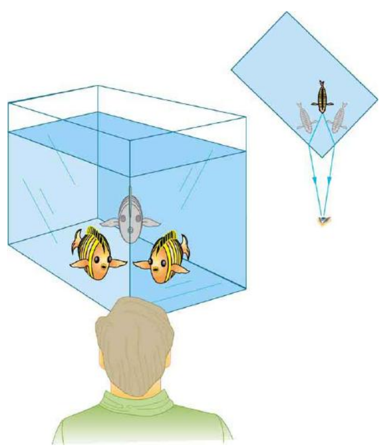


Figure 11.4.1: Looking at the fish tank as shown, we can see the same fish in two different locations, because light changes directions when it passes from water to air. In this case, the light can reach the observer by two different paths, and so the fish seems to be in two different places. This bending of light is called refraction and is responsible for many optical phenomena.

Why does light change direction when passing from one material (medium) to another? It is because light changes speed when going from one material to another. So before we study the law of refraction, it is useful to discuss the speed of light and how it varies in different media.

The Speed of Light

Early attempts to measure the speed of light, such as those made by Galileo, determined that light moved extremely fast, perhaps instantaneously. The first real evidence that light traveled at a finite speed came from the Danish astronomer Ole Roemer in the late 17th century. Roemer had noted that the average orbital period of one of Jupiter's moons, as measured from Earth, varied depending on whether Earth was moving toward or away from Jupiter. He correctly concluded that the apparent change in period was due to the change in distance between Earth and Jupiter and the time it took light to travel this distance. From his 1676 data, a value of the speed of light was calculated to be $2.26 \times 10^{8} m/s$ (only 25% different from today's accepted value). In more recent times, physicists have measured the speed of light in numerous ways and with increasing accuracy. One particularly direct method, used in 1887 by the American physicist Albert Michelson (1852–1931), is illustrated in Figure $11.4.2$. Light reflected from a rotating set of mirrors was reflected from a stationary mirror 35 km away and returned to the rotating mirrors. The time for the light to travel can be determined by how fast the mirrors must rotate for the light to be returned to the observer's eye.

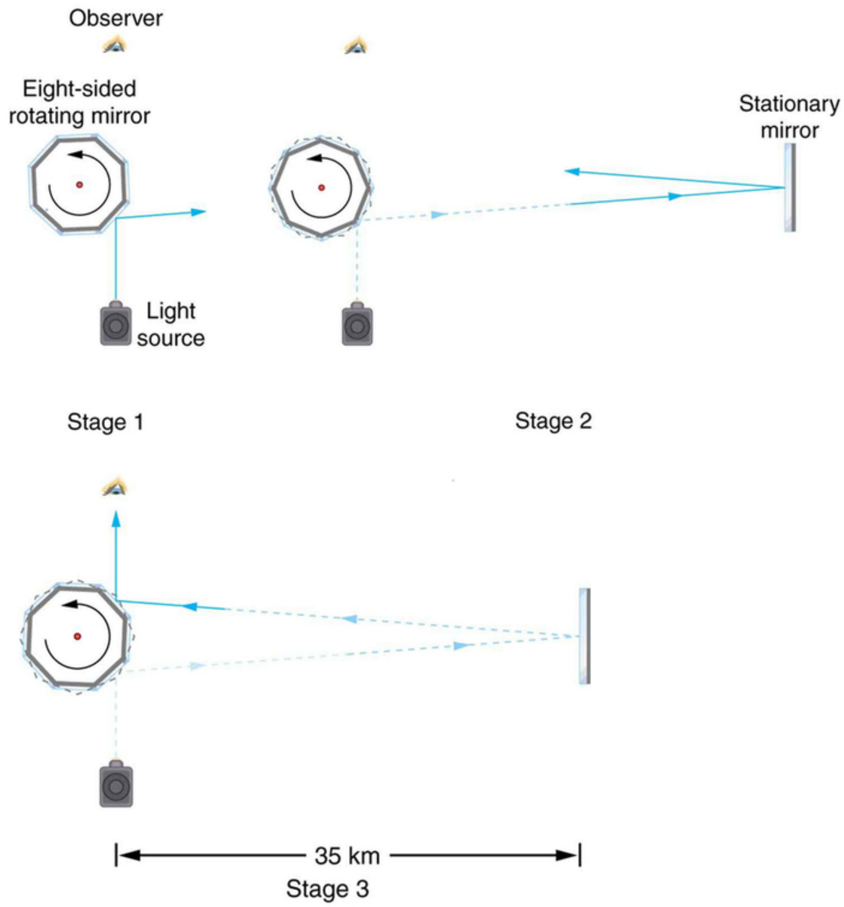


Figure 11.4.2: A schematic of early apparatus used by Michelson and others to determine the speed of light. As the mirrors rotate, the reflected ray is only briefly directed at the stationary mirror. The returning ray will be reflected into the observer's eye only if the next mirror has rotated into the correct position just as the ray returns. By measuring the correct rotation rate, the time for the round trip can be measured and the speed of light calculated. Michelson's calculated value of the speed of light was only 0.04% different from the value used today.

The speed of light is now known to great precision. In fact, the speed of light in a vacuum c is so important that it is accepted as one of the basic physical quantities and has the fixed value.

VALUE OF THE SPEED OF LIGHT

$$c \equiv 2.99792458 \times 10^{8} \tag{11.4.1}$$

$$\sim 3.00 \times 10^{8} m/s \tag{11.4.2}$$

The approximate value of $3.00 \times 10^{8} m/s$ is used whenever three-digit accuracy is sufficient. The speed of light through matter is less than it is in a vacuum, because light interacts with atoms in a material. The speed of light depends strongly on the type of material, since its interaction with different atoms, crystal lattices, and other substructures varies.

Definition: INDEX OF REFRACTION

We define the *index of refraction* n of a material to be

$$n = \frac{c}{v}, \tag{11.4.3}$$

where v is the observed speed of light in the material. Since the speed of light is always less than c in matter and equals c only in a vacuum, the index of refraction is always greater than or equal to one. That is, $n > 1$.

Table $11.4.1$ gives the indices of refraction for some representative substances. The values are listed for a particular wavelength of light, because they vary slightly with wavelength. (This can have important effects, such as colors produced by a prism.) Note that for gases, n is close to 1.0. This seems reasonable, since atoms in gases are widely separated and light travels at c in the vacuum between atoms. It is common to take $n = 1$ for gases unless great precision is needed. Although the speed of light v in a medium varies considerably from its value c in a vacuum, it is still a large speed.

Table 11.4.1: Index of Refraction in Various Media

Medium	**n**
Gases at $0°C, 1atm$	
Air	1.000293
Carbon dioxide	1.00045
Hydrogen	1.000139
Oxygen	1.000271
Liquids at 20ºC	
Benzene	1.501
Carbon disulfide	1.628
Carbon tetrachloride	1.461
Ethanol	1.361
Glycerine	1.473
Water, fresh	1.333
Solids at 20ºC	
Diamond	2.419
Fluorite	1.434
Glass, crown	1.52
Glass, flint	1.66
Ice at 20ºC	1.309
Polystyrene	1.49
Plexiglas	1.51
Quartz, crystalline	1.544
Quartz, fused	1.458
Sodium chloride	1.544
Zircon	1.923

Example 11.4.1: Speed of Light in Matter

Calculate the speed of light in zircon, a material used in jewelry to imitate diamond.

Strategy:

The speed of light in a material, v, can be calculated from the index of refraction n of the material using the equation $n = c/v$.

Solution

The equation for index of refraction (Equation 11.4.3) can be rearranged to determine v

$$v = \frac{c}{n}.$$

The index of refraction for zircon is given as 1.923 in Table $11.4.1$, and c is given in the equation for speed of light. Entering these values in the last expression gives

$$\begin{aligned} v &= \frac{3.00 \times 10^8 m/s}{1.923} \\ &= 1.56 \times 10^8 m/s. \end{aligned}$$

Discussion:

This speed is slightly larger than half the speed of light in a vacuum and is still high compared with speeds we normally experience. The only substance listed in Table $11.4.1$ that has a greater index of refraction than zircon is diamond. We shall see later that the large index of refraction for zircon makes it sparkle more than glass, but less than diamond.

Law of Refraction

Figure $11.4.3$ shows how a ray of light changes direction when it passes from one medium to another. As before, the angles are measured relative to a perpendicular to the surface at the point where the light ray crosses it. (Some of the incident light will be reflected from the surface, but for now we will concentrate on the light that is transmitted.) The change in direction of the light ray depends on how the speed of light changes. The change in the speed of light is related to the indices of refraction of the media involved. In the situations shown in Figure $11.4.3$, medium 2 has a greater index of refraction than medium 1. This means that the speed of light is less in medium 2 than in medium 1. Note that as shown in Figure $11.4.3a$, the direction of the ray moves closer to the perpendicular when it slows down. Conversely, as shown in Figure $11.4.3b$, the direction of the ray moves away from the perpendicular when it speeds up. The path is exactly reversible. In both cases, you can imagine what happens by thinking about pushing a lawn mower from a footpath onto grass, and vice versa. Going from the footpath to grass, the front wheels are slowed and pulled to the side as shown. This is the same change in direction as for light when it goes from a fast medium to a slow one. When going from the grass to the footpath, the front wheels can move faster and the mower changes direction as shown. This, too, is the same change in direction as for light going from slow to fast.

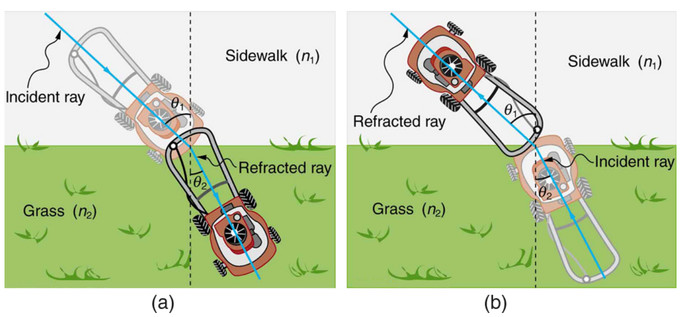


Figure $11.4.3$: The change in direction of a light ray depends on how the speed of light changes when it crosses from one medium to another. The speed of light is greater in medium 1 than in medium 2 in the situations shown here. (a) A ray of light moves closer to the perpendicular when it slows down. This is analogous to what happens when a lawn mower goes from a footpath to grass. (b) A ray of light moves away from the perpendicular when it speeds up. This is analogous to what happens when a lawn mower goes from grass to footpath. The paths are exactly reversible.

The amount that a light ray changes its direction depends both on the incident angle and the amount that the speed changes. For a ray at a given incident angle, a large change in speed causes a large change in direction, and thus a large change in angle.

TAKE-HOME EXPERIMENT: A BROKEN PENCIL

A classic observation of refraction occurs when a pencil is placed in a glass half filled with water. Do this and observe the shape of the pencil when you look at the pencil sideways, that is, through air, glass, water. Explain your observations. Draw ray diagrams for the situation.

Summary

- The changing of a light ray's direction when it passes through variations in matter is called refraction.
- The speed of light in vacuuum $c = 2.99792458 \times 10^{8} \sim 3.00 \times 10^{8} m/s$
- Index of refraction $n = \frac{c}{v}$, where v is the speed of light in the material, c is the speed of light in vacuum, and n is the index of refraction.

Glossary

refraction
changing of a light ray's direction when it passes through variations in matter

index of refraction
for a material, the ratio of the speed of light in vacuum to that in the material

Contributors and Attributions

11.5: Dispersion - Rainbows and Prisms

Learning Objectives

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

- Explain the phenomenon of dispersion and discuss its advantages and disadvantages.

Everyone enjoys the spectacle of a rainbow glimmering against a dark stormy sky. How does sunlight falling on clear drops of rain get broken into the rainbow of colors we see? The same process causes white light to be broken into colors by a clear glass prism or a diamond. (See Figure 1.)

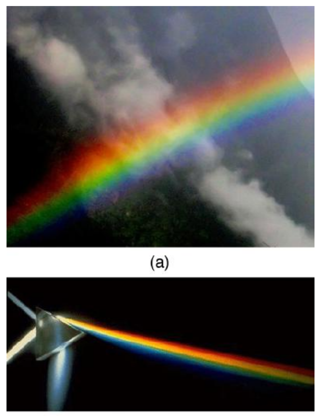


Figure 11.5.1: The colors of the rainbow (a) and those produced by a prism (b) are identical. (credit: Alfredo55, Wikimedia Commons; NASA)

We see about six colors in a rainbow -- red, orange, yellow, green, blue, and violet; sometimes indigo is listed, too. Those colors are associated with different wavelengths of light, as shown in Figure 1. When our eye receives pure-wavelength light, we tend to see only one of the six colors, depending on wavelength. The thousands of other hues we can sense in other situations are our eye's response to various mixtures of wavelengths. White light, in particular, is a fairly uniform mixture of all visible wavelengths. Sunlight, considered to be white, actually appears to be a bit yellow because of its mixture of wavelengths, but it does contain all visible wavelengths. The sequence of colors in rainbows is the same sequence as the colors plotted versus wavelength in Figure 2. What this implies is that white light is spread out according to wavelength in a rainbow. **Dispersion** is defined as the spreading of white light into its full spectrum of wavelengths. More technically, dispersion occurs whenever there is a process that changes the direction of light in a manner that depends on wavelength. Dispersion, as a general phenomenon, can occur for any type of wave and always involves wavelength-dependent processes.

DISPERSION

Dispersion is defined to be the spreading of white light into its full spectrum of wavelengths.

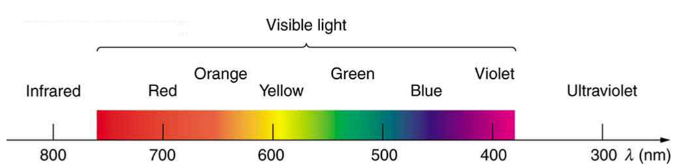


Figure 11.5.2: Even though rainbows are associated with seven colors, the rainbow is a continuous distribution of colors according to wavelengths.

Refraction is responsible for dispersion in rainbows and many other situations. The angle of refraction depends on the index of refraction, as we saw in "The Law of Refraction." We know that the index of refraction n depends on the medium. But for a given medium, n also depends on wavelength (Table $11.5.1$). Note that, for a given medium, n increases as wavelength decreases and is greatest for violet light. Thus violet light is bent more than red light, as shown for a prism in Figure 3 and the light is dispersed into the same sequence of wavelengths as seen in 1 and 2.

MAKING CONNECTIONS: DISPERSION

Any type of wave can exhibit dispersion. Sound waves, all types of electromagnetic waves, and water waves can be dispersed according to wavelength. Dispersion occurs whenever the speed of propagation depends on wavelength, thus separating and spreading out various wavelengths. Dispersion may require special circumstances and can result in spectacular displays such as in the production of a rainbow. This is also true for sound, since all frequencies ordinarily travel at the same speed. If you listen to sound through a long tube, such as a vacuum cleaner hose, you can easily hear it is dispersed by interaction with the tube. Dispersion, in fact, can reveal a great deal about what the wave has encountered that disperses its wavelengths. The dispersion of electromagnetic radiation from outer space, for example, has revealed much about what exists between the stars -- the so-called empty space.

Table 11.5.1: Index of Refraction n in Selected Media at Various Wavelengths

Medium	Red (660 nm)	Orange (610 nm)	Yellow (580 nm)	Green (550 nm)	Blue (470 nm)	Violet (410 nm)
Water	1.331	1.332	1.333	1.335	1.338	1.342
Diamond	2.410	2.415	2.417	2.426	2.444	2.458
Glass, crown	1.512	1.514	1.518	1.519	1.524	1.530
Glass, flint	1.662	1.665	1.667	1.674	1.684	1.698
Polustyrene	1.488	1.490	1.492	1.493	1.499	1.506
Quartz, fused	1.455	1.456	1.458	1.459	1.462	1.468

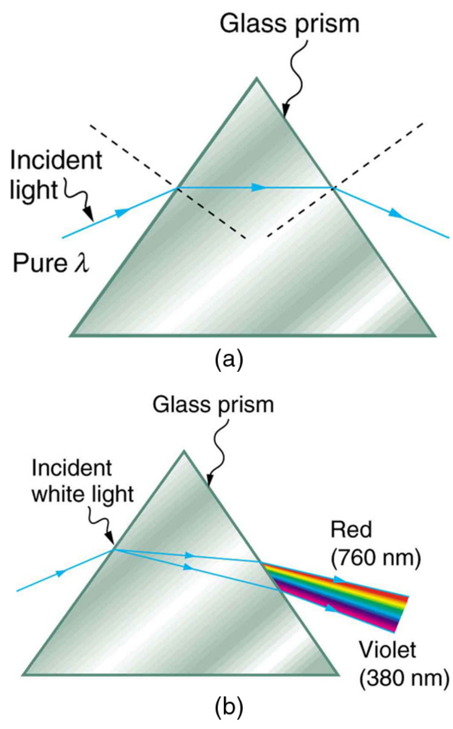


Figure 11.5.3: (a) A pure wavelength of light falls onto a prism and is refracted at both surfaces. (b) White light is dispersed by the prism (shown exaggerated). Since the index of refraction varies with wavelength, the angles of refraction vary with wavelength. A sequence of red to violet is produced, because the index of refraction increases steadily with decreasing wavelength.

Rainbows are produced by a combination of refraction and reflection. You may have noticed that you see a rainbow only when you look away from the sun. Light enters a drop of water and is reflected from the back of the drop, as shown in Figure 4. The light is refracted both as it enters and as it leaves the drop. Since the index of refraction of water varies with wavelength, the light is dispersed, and a rainbow is observed, as shown in Figure 5a. (There is no dispersion caused by reflection at the back surface, since the law of reflection does not depend on wavelength.) The actual rainbow of colors seen by an observer depends on the myriad of rays being refracted and reflected toward the observer's eyes from numerous drops of water. The effect is most spectacular when the background is dark, as in stormy weather, but can also be observed in waterfalls and lawn sprinklers. The arc of a rainbow comes from the need to be looking at a specific angle relative to the direction of the sun, as illustrated in Figure 5b. (If there are two reflections of light within the water drop, another "secondary" rainbow is produced. This rare event produces an arc that lies above the primary rainbow arc -- see Figure 5c.)

RAINBOWS

Rainbows are produced by a combination of refraction and reflection.

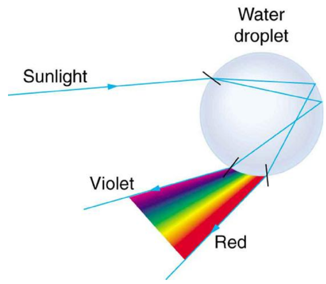


Figure 11.5.4: Part of the light falling on this water drop enters and is reflected from the back of the drop. This light is refracted and dispersed both as it enters and as it leaves the drop.

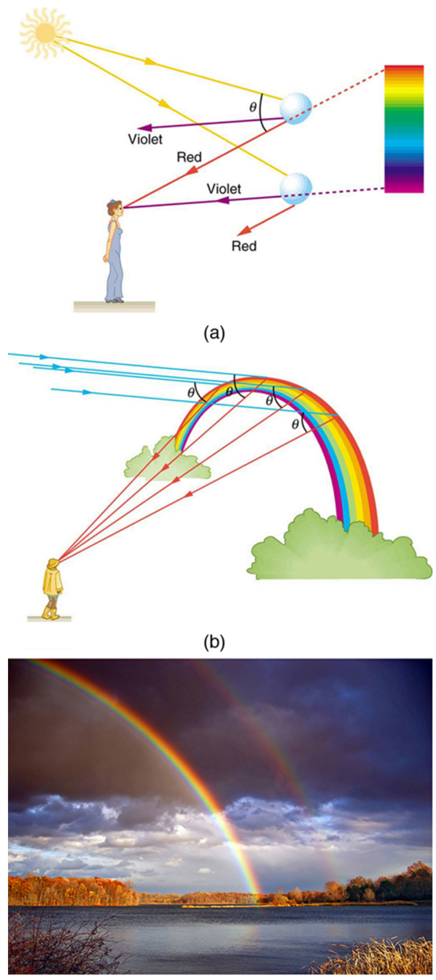


Figure 11.5.5: (a) Different colors emerge in different directions, and so you must look at different locations to see the various colors of a rainbow. (b) The arc of a rainbow results from the fact that a line between the observer and any point on the arc must make the correct angle with the parallel rays of sunlight to receive the refracted rays. (c) Double rainbow. (credit: Nicholas, Wikimedia Commons)

Dispersion may produce beautiful rainbows, but it can cause problems in optical systems. White light used to transmit messages in a fiber is dispersed, spreading out in time and eventually overlapping with other messages. Since a laser produces a nearly pure wavelength, its light experiences little dispersion, an advantage over white light for transmission of information. In contrast, dispersion of electromagnetic waves coming to us from outer space can be used to determine the amount of matter they pass through. As with many phenomena, dispersion can be useful or a nuisance, depending on the situation and our human goals.

PHET EXPLORATIONS: GEOMETRIC OPTICS

How does a lens form an image? See how light rays are refracted by a lens. Watch how the image changes when you adjust the focal length of the lens, move the object, move the lens, or move the screen.

Summary

- The spreading of white light into its full spectrum of wavelengths is called dispersion.
- Rainbows are produced by a combination of refraction and reflection and involve the dispersion of sunlight into a continuous distribution of colors.
- Dispersion produces beautiful rainbows but also causes problems in certain optical systems.

Glossary

dispersion
spreading of white light into its full spectrum of wavelengths

rainbow
dispersion of sunlight into a continuous distribution of colors according to wavelength, produced by the refraction and reflection of sunlight by water droplets in the sky

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

11.6: Image Formation by Lenses

Learning Objectives

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

- List the rules for ray tracking for thin lenses.
- Illustrate the formation of images using the technique of ray tracking.
- Determine power of a lens given the focal length.

Lenses are found in a huge array of optical instruments, ranging from a simple magnifying glass to the eye to a camera's zoom lens. In this section, we will use the law of refraction to explore the properties of lenses and how they form images.

The word *lens* derives from the Latin word for a lentil bean, the shape of which is similar to the convex lens in Figure 1. The convex lens shown has been shaped so that all light rays that enter it parallel to its axis cross one another at a single point on the opposite side of the lens. (The axis is defined to be a line normal to the lens at its center, as shown in Figure 1.) Such a lens is called a **converging (or convex) lens** for the converging effect it has on light rays. An expanded view of the path of one ray through the lens is shown, to illustrate how the ray changes direction both as it enters and as it leaves the lens. Since the index of refraction of the lens is greater than that of air, the ray moves towards the perpendicular as it enters and away from the perpendicular as it leaves. (This is in accordance with the law of refraction.) Due to the lens's shape, light is thus bent toward the axis at both surfaces. The point at which the rays cross is defined to be the **focal point** F of the lens. The distance from the center of the lens to its focal point is defined to be the **focal length f** of the lens. Figure 2 shows how a converging lens, such as that in a magnifying glass, can converge the nearly parallel light rays from the sun to a small spot.

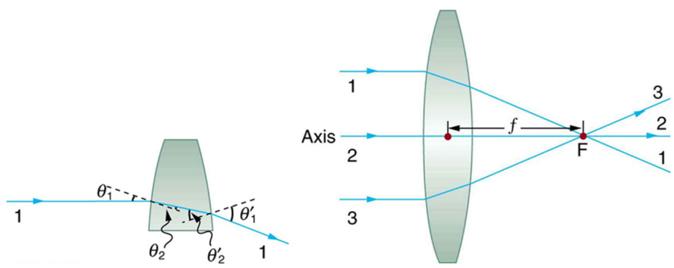


Figure 11.6.1: Rays of light entering a converging lens parallel to its axis converge at its focal point F. (Ray 2 lies on the axis of the lens.) The distance from the center of the lens to the focal point is the lens's focal length f. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Definition: CONVERGING OR CONVEX LENS

The lens in which light rays that enter it parallel to its axis cross one another at a single point on the opposite side with a converging effect is called converging lens.

Definition: FOCAL POINT F

The point at which the light rays cross is called the focal point F of the lens.

Definition: FOCAL LENGTH f

The distance from the center of the lens to its focal point is called focal length f.

Figure 11.6.2: Sunlight focused by a converging magnifying glass can burn paper. Light rays from the sun are nearly parallel and cross at the focal point of the lens. The more powerful the lens, the closer to the lens the rays will cross.

The greater effect a lens has on light rays, the more powerful it is said to be. For example, a powerful converging lens will focus parallel light rays closer to itself and will have a smaller focal length than a weak lens. The light will also focus into a smaller and more intense spot for a more powerful lens. The **power P** of a lens is defined to be the inverse of its focal length:

Definition: POWER P

The power P of a lens is defined to be the inverse of its focal length. In equation form, this is

$$P = \frac{1}{f}. \tag{11.6.1}$$

where f is the focal length of the lens, which must be given in meters (and not cm or mm). The power of a lens P has the unit diopters (D), provided that the focal length is given in meters. That is $1D = 1/m$ or $1m^{-1}$. (Note that this power (optical power, actually) is not the same as power in watts defined in "Work, Energy, and Energy Resources." It is a concept related to the effect of optical devices on light.) Optometrists prescribe common spectacles and contact lenses in units of diopters.

Example 11.6.1: What is the Power of a Common Magnifying Glass?

Suppose you take a magnifying glass out on a sunny day and you find that it concentrates sunlight to a small spot 8.00 cm away from the lens. What are the focal length and power of the lens?

Strategy:

The situation here is the same as those shown in Figure 1 and Figure 2. The Sun is so far away that the Sun's rays are nearly parallel when they reach Earth. The magnifying glass is a convex (or converging) lens, focusing the nearly parallel rays of sunlight. Thus the focal length of the lens is the distance from the lens to the spot, and its power is the inverse of this distance (in m).

Solution

The focal length of the lens is the distance from the center of the lens to the spot, given to be 8.00 cm. Thus,

$$f = 8.00cm. \tag{11.6.2}$$

To find the power of the lens, we must first convert the focal length to meters; then, we substitute this value into the equation for power. This gives

$$P = \frac{1}{f} = \frac{1}{0.0800m} = 12.5D. \tag{11.6.3}$$

Discussion:

This is a relatively powerful lens. The power of a lens in diopters should not be confused with the familiar concept of power in watts. It is an unfortunate fact that the word "power" is used for two completely different concepts. If you examine a prescription for eyeglasses, you will note lens powers given in diopters. If you examine the label on a motor, you will note energy consumption rate given as a power in watts.

Figure $11.6.3$ shows a concave lens and the effect it has on rays of light that enter it parallel to its axis (the path taken by ray 2 in the figure is the axis of the lens). The concave lens is a **diverging lens**, because it causes the light rays to bend away (diverge) from its axis. In this case, the lens has been shaped so that all light rays entering it parallel to its axis appear to originate from the same point, F, defined to be the focal point of a diverging lens. The distance from the center of the lens to the focal point is again called the focal length f of the lens. Note that the focal length and power of a diverging lens are defined to be negative. For example, if the distance to F in Figure 3 is 5.00 cm, then the focal length is $f = -5.00 cm$ and the power of the lens is $P = -20D$. An expanded view of the path of one ray through the lens is shown in the figure to illustrate how the shape of the lens, together with the law of refraction, causes the ray to follow its particular path and be diverged.

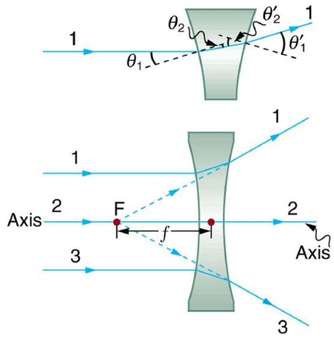


Figure $11.6.3$: Rays of light entering a diverging lens parallel to its axis are diverged, and all appear to originate at its focal point F. The dashed lines are not rays -- they indicate the directions from which the rays appear to come. The focal length f of a diverging lens is negative. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Definition: DIVERGING LENS

A lens that causes the light rays to bend away from its axis is called a diverging lens.

As noted in the initial discussion of the law of refraction in "The Law of Refraction," the paths of light rays are exactly reversible. This means that the direction of the arrows could be reversed for all of the rays in Figures $11.6.1$ and Figure $11.6.3$. For example, if a point light source is placed at the focal point of a convex lens, as shown in Figure 4, parallel light rays emerge from the other side.

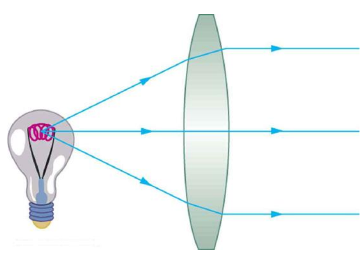


Figure 1. This technique is used in lighthouses and sometimes in traffic lights to produce a directional beam of light from a source that emits light in all directions.

Ray Tracing and Thin Lenses

Ray tracing is the technique of determining or following (tracing) the paths that light rays take. For rays passing through matter, the law of refraction is used to trace the paths. Here we use ray tracing to help us understand the action of lenses in situations ranging from forming images on film to magnifying small print to correcting nearsightedness. While ray tracing for complicated lenses, such as those found in sophisticated cameras, may require computer techniques, there is a set of simple rules for tracing rays through thin lenses. A **thin lens** is defined to be one whose thickness allows rays to refract, as illustrated in Figure $11.6.1$, but does not allow properties such as dispersion and aberrations. An ideal thin lens has two refracting surfaces but the lens is thin enough to assume that light rays bend only once. A thin symmetrical lens has two focal points, one on either side and both at the same distance from the lens (Figure $11.6.5$). Another important characteristic of a thin lens is that light rays through its center are deflected by a negligible amount, as seen in Figure $11.6.6$.

Definition: THIN LENS

A thin lens is defined to be one whose thickness allows rays to refract but does not allow properties such as dispersion and aberrations.

TAKE-HOME EXPERIMENT: A VISIT TO THE OPTICIAN

Look through your eyeglasses (or those of a friend) backward and forward and comment on whether they act like thin lenses.

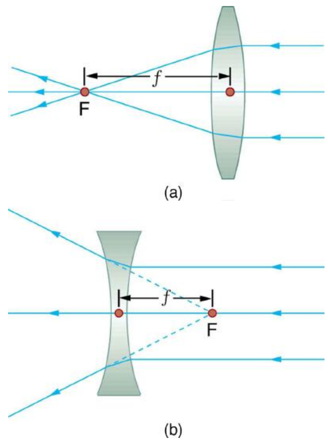


Figure 11.6.5: Thin lenses have the same focal length on either side. (a) Parallel light rays entering a converging lens from the right cross at its focal point on the left. (b) Parallel light rays entering a diverging lens from the right seem to come from the focal point on the right.

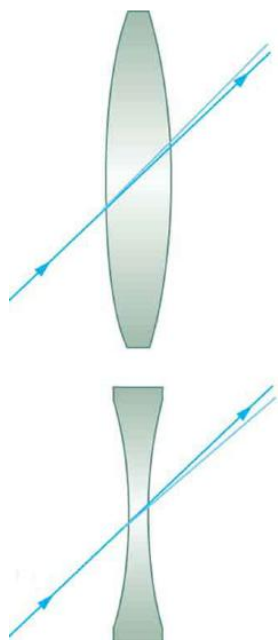


Figure 11.6.6: The light ray through the center of a thin lens is deflected by a negligible amount and is assumed to emerge parallel to its original path (shown as a shaded line).

Using paper, pencil, and a straight edge, ray tracing can accurately describe the operation of a lens. The Ruls for Ray Tracing for thin lenses are based on the illustrations already discussed:

RULES FOR RAY TRACING (for thin lenses)

1. A ray entering a converging lens parallel to its axis passes through the focal point F of the lens on the other side. (See rays 1 and 3 in Figure 11.6.1).
2. A ray entering a diverging lens parallel to its axis seems to come from the focal point F. (See rays 1 and 3 in Figure 11.6.3).

3. A ray passing through the center of either a converging or a diverging lens does not change direction. (See Figure $11.6.6$, and see ray 2 in Figure $11.6.1$ and $11.6.3$).
4. A ray entering a converging lens through its focal point exits parallel to its axis. (The reverse of rays 1 and 3 in Figure $11.6.1$).
5. A ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis. (The reverse of rays 1 and 3 in Figure $11.6.3$).

Image Formation by Thin Lenses

In some circumstances, a lens forms an obvious image, such as when a movie projector casts an image onto a screen. In other cases, the image is less obvious. Where, for example, is the image formed by eyeglasses? We will use ray tracing for thin lenses to illustrate how they form images, and we will develop equations to describe the image formation quantitatively.

Consider an object some distance away from a converging lens, as shown in Figure $11.6.7$. To find the location and size of the image formed, we trace the paths of selected light rays originating from one point on the object, in this case the top of the person's head. The figure shows three rays from the top of the object that can be traced using the ray tracing rules given above. (Rays leave this point going in many directions, but we concentrate on only a few with paths that are easy to trace.) The first ray is one that enters the lens parallel to its axis and passes through the focal point on the other side (rule 1). The second ray passes through the center of the lens without changing direction (rule 3). The third ray passes through the nearer focal point on its way into the lens and leaves the lens parallel to its axis (rule 4). The three rays cross at the same point on the other side of the lens. The image of the top of the person's head is located at this point. All rays that come from the same point on the top of the person's head are refracted in such a way as to cross at the point shown. Rays from another point on the object, such as her belt buckle, will also cross at another common point, forming a complete image, as shown. Although three rays are traced in Figure $11.6.7$, only two are necessary to locate the image. It is best to trace rays for which there are simple ray tracing rules. Before applying ray tracing to other situations, let us consider the example shown in Figure 7 in more detail.

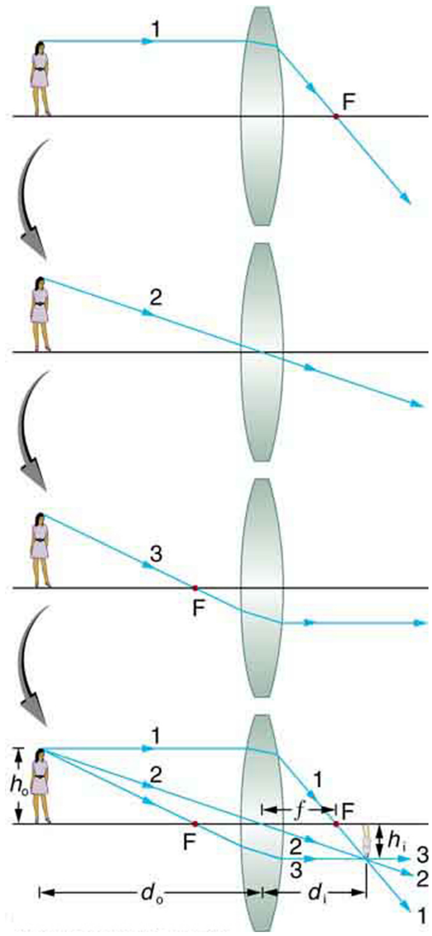
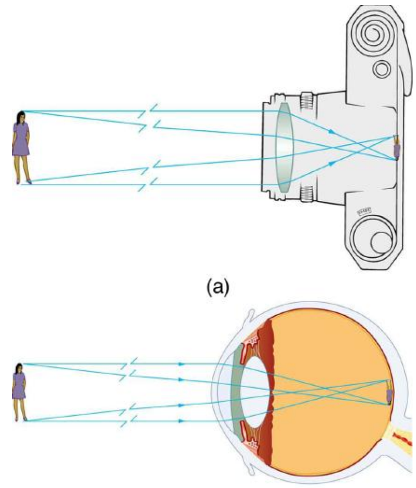


Figure 11.6.7: Ray tracing is used to locate the image formed by a lens. Rays originating from the same point on the object are traced -- the three chosen rays each follow one of the rules for ray tracing, so that their paths are easy to determine. The image is located at the point where the rays cross. In this case, a real image -- one that can be projected on a screen -- is formed.

The image formed in Figure 11.6.7 is a **real image**, meaning that it can be projected. That is, light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye, for example. Figure 11.6.8 shows how such an image would be projected onto film by a camera lens. This figure also shows how a real image is projected onto the retina by the lens of an eye. Note that the image is there whether it is projected onto a screen or not.

Definition: REAL IMAGE

The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.



(a)

(b)

Figure 11.6.8: Real images can be projected. (a) A real image of the person is projected onto film. (b) The converging nature of the multiple surfaces that make up the eye result in the projection of a real image on the retina.

Several important distances appear in Figure 7. We define d_0 to be the object of distance, the distance of an object from the center of a lens. **Image distance d_i** is defined to be the distance of the image from the center of a lens. The height of the object and height of the image are given the symbols h_0 and h_i, respectively. Images that appear upright relative to the object have heights that are positive and those that are inverted have negative heights. Using the rules of ray tracing and making a scale drawing with paper and pencil, like that in Figure 7, we can accurately describe the location and size of an image. But the real benefit of ray tracing is in visualizing how images are formed in a variety of situations. To obtain numerical information, we use a pair of equations that can be derived from a geometric analysis of ray tracing for thin lenses. The **thin lens equations** are

$$\frac{1}{d_0}+\frac{1}{d_i}=\frac{1}{f} \tag{11.6.4}$$

and

$$\frac{h_i}{h_0}=-\frac{d_i}{d_0}=m. \tag{11.6.5}$$

We define the ratio of image height to object height (h_i/h_0) to be the **magnification** m. (The minus sign in the equation above will be discussed shortly.) The thin lens equations are broadly applicable to all situations involving thin lenses (and "thin" mirrors, as we will see later). We will explore many features of image formation in the following worked examples.

Definition: IMAGE DISTANCE

The distance of the image from the center of the lens is called image distance.

Example 11.6.2: Finding the Image of a Light Bulb Filament by Ray Tracing and by the Thin Lens Equations

A clear glass light bulb is placed 0.750 m from a convex lens having a 0.500 m focal length, as shown in the figure. Use ray tracing to get an approximate location for the image. Then use the thin lens equations to calculate

a. the location of the image and
b. its magnification.

Verify that ray tracing and the thin lens equations produce consistent results.

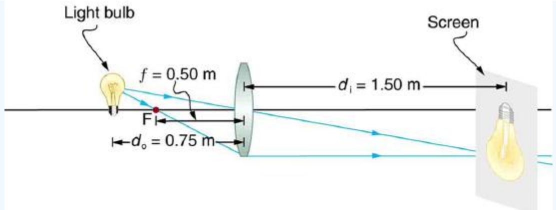


Figure 11.6.9: A light bulb placed 0.750 m from a lens having a 0.500 m focal length produces a real image on a poster board as discussed in the example above. Ray tracing predicts the image location and size.

Strategy and Concept

Since the object is placed farther away from a converging lens than the focal length of the lens, this situation is analogous to those illustrated in Figure 7 and Figure 8. Ray tracing to scale should produce similar results for d_i. Numerical solutions for d_i and m can be obtained using the thin lens equations, noting that $d_o = 0.750m$ and $f = 0.500m$.

Solutions (Ray Tracing)

The ray tracing to scale in Figure 9 shows two rays from a point on the bulb's filament crossing about 1.50 m on the far side of the lens. Thus the image distance d_i is about 1.50 m. Similarly, the image height based on ray tracing is greater than the object height by about a factor of 2, and the image is inverted. Thus m is about –2. The minus sign indicates that the image is inverted.

The thin lens equations can be used to find d_i from the given information:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}. \tag{11.6.6}$$

Rearranging to isolate d_i gives

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}. \tag{11.6.7}$$

Entering known quantities gives a value for $1/d_i$:

$$\frac{1}{d_i} = \frac{1}{0.500m} - \frac{1}{0.750m} = \frac{0.667}{m}. \tag{11.6.8}$$

This must be inverted to find d_i:

$$d_i = \frac{m}{0.667} = 1.50m. \tag{11.6.9}$$

Note that another way to find d_i is to rearrange equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}. \tag{11.6.10}$$

This yields the equation for the image distance as:

$$d_i = \frac{fd_o}{d_o - f}. \tag{11.6.11}$$

Note that there is no inverting here.

The thin lens equations can be used to find the magnification m, since both d_i and d_o are known. Entering their values gives

$$m = -\frac{d_i}{d_o} = -\frac{1.50m}{0.750m} = -2.00. \tag{11.6.12}$$

Discussion

Note that the minus sign causes the magnification to be negative when the image is inverted. Ray tracing and the use of the thin lens equations produce consistent results. The thin lens equations give the most precise results, being limited only by the accuracy of the given information. Ray tracing is limited by the accuracy with which you can draw, but it is highly useful both conceptually and visually.

Real images, such as the one considered in the previous example, are formed by converging lenses whenever an object is farther from the lens than its focal length. This is true for movie projectors, cameras, and the eye. We shall refer to these as *case 1* images. A case 1 image is formed when $d_o > f$ and f is positive, as in Figure $11.6.10a$. (A summary of the three cases or types of image formation appears at the end of this section.)

A different type of image is formed when an object, such as a person's face, is held close to a convex lens. The image is upright and larger than the object, as seen in Figure $11.6.10b$, and so the lens is called a magnifier. If you slowly pull the magnifier away from the face, you will see that the magnification steadily increases until the image begins to blur. Pulling the magnifier even farther away produces an inverted image as seen in Figure $11.6.10a$. The distance at which the image blurs, and beyond which it inverts, is the focal length of the lens. To use a convex lens as a magnifier, the object must be closer to the converging lens than its focal length. This is called a *case 2* image. A case 2 image is formed when $d_o < f$ and f is positive.

(a)

(b)

Figure 11.6.10: (a) When a converging lens is held farther away from the face than the lens's focal length, an inverted image is formed. This is a case 1 image. Note that the image is in focus but the face is not, because the image is much closer to the camera taking this photograph than the face. (credit: DaMongMan, Flickr) (b) A magnified image of a face is produced by placing it closer to the converging lens than its focal length. This is a case 2 image. (credit: Casey Fleser, Flickr)

Figure $11.6.11$ uses ray tracing to show how an image is formed when an object is held closer to a converging lens than its focal length. Rays coming from a common point on the object continue to diverge after passing through the lens, but all appear to originate from a point at the location of the image. The image is on the same side of the lens as the object and is farther away from the lens than the object. This image, like all case 2 images, cannot be projected and, hence, is called a **virtual image**. Light rays only appear to originate at a virtual image; they do not actually pass through that location in space. A screen placed at the location of a virtual image will receive only diffuse light from the object, not focused rays from the lens. Additionally, a screen placed on the opposite side of the lens will receive rays that are still diverging, and so no image will be projected on it. We can see the magnified image with our eyes, because the lens of the eye converges the rays into a real image projected on our retina. Finally, we note that a virtual image is upright and larger than the object, meaning that the magnification is positive and greater than 1.

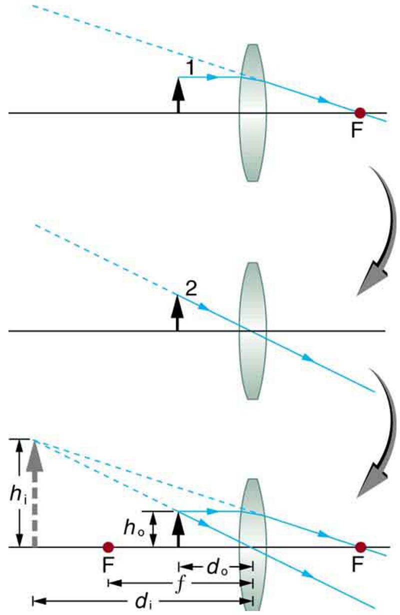


Figure 11.6.11: Ray tracing predicts the image location and size for an object held closer to a converging lens than its focal length. Ray 1 enters parallel to the axis and exits through the focal point on the opposite side, while ray 2 passes through the center of the lens without changing path. The two rays continue to diverge on the other side of the lens, but both appear to come from a common point, locating the upright, magnified, virtual image. This is a case 2 image.

Definition: VIRTUAL IMAGE

An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

Example 11.6.3: Image Produced by a Magnifying Glass

Suppose the book in Figure $11.6.11a$ is held 7.50 cm from a convex lens of focal length 10.0 cm, such as a typical magnifying glass might have. What magnification is produced?

Strategy and Concept

We are given that $d_o = 7.50 cm$ and $f = 10.0 cm$, so we have a situation where the object is placed closer to the lens than its focal length. We therefore expect to get a case 2 virtual image with a positive magnification that is greater than 1. Ray tracing produces an image like that shown in Figure 11 but we will use the thin lens equations to get numerical solutions in this example.

Solution

To find the magnification m, we try to use magnification equation, $m = -d_i/d_o$. We do not have a value for d_i, so that we must first find the location of the image using lens equation. (The procedure is the same as followed in the preceding example, where d_o and f were known.) Rearranging the magnification equation to isolate d_i gives

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}. \tag{11.6.13}$$

Entering known values, we obtain a value for $1/d_i$:

$$\frac{1}{d_i} = \frac{1}{10.0 cm} - \frac{1}{7.50 cm} = \frac{-0.0333}{cm}. \tag{11.6.14}$$

This must be inverted to find d_i:

$$d_i = -\frac{cm}{0.0333} = -30.0 cm. \tag{11.6.15}$$

Now the thin lens equation can be used to find the magnification m, since both d_i and d_o are known. Entering their values gives

$$m=-\frac{d_i}{d_o}=-\frac{-30.0cm}{7.50cm}=4.00. \tag{11.6.16}$$

Discussion

A number of results in this example are true of all case 2 images, as well as being consistent with Figure 11. Magnification is indeed positive (as predicted), meaning the image is upright. The magnification is also greater than 1, meaning that the image is larger than the object—in this case, by a factor of 4. Note that the image distance is negative. This means the image is on the same side of the lens as the object. Thus the image cannot be projected and is virtual. (Negative values of d_i occur for virtual images.) The image is farther from the lens than the object, since the image distance is greater in magnitude than the object distance. The location of the image is not obvious when you look through a magnifier. In fact, since the image is bigger than the object, you may think the image is closer than the object. But the image is farther away, a fact that is useful in correcting farsightedness, as we shall see in a later section.

A third type of image is formed by a diverging or concave lens. Try looking through eyeglasses meant to correct nearsightedness. (Figure 11.6.12). You will see an image that is upright but smaller than the object. This means that the magnification is positive but less than 1. The ray diagram in Figure 13 shows that the image is on the same side of the lens as the object and, hence, cannot be projected—it is a virtual image. Note that the image is closer to the lens than the object. This is a *case 3* image, formed for any object by a negative focal length or diverging lens.

Figure 11.6.12: A car viewed through a concave or diverging lens looks upright. This is a case 3 image. (credit: Daniel Oines, Flickr)

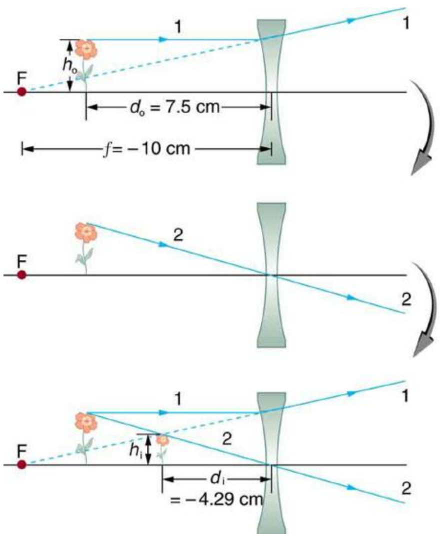


Figure 11.6.13: Ray tracing predicts the image location and size for a concave or diverging lens. Ray 1 enters parallel to the axis and is bent so that it appears to originate from the focal point. Ray 2 passes through the center of the lens without changing path. The two rays appear to come from a common point, locating the upright image. This is a case 3 image, which is closer to the lens than the object and smaller in height.

Example 11.6.4: Image Produced by a Concave Lens

Suppose an object such as a book page is held 7.50 cm from a concave lens of focal length –10.0 cm. Such a lens could be used in eyeglasses to correct pronounced nearsightedness. What magnification is produced?

Strategy and Concept

This example is identical to the preceding one, except that the focal length is negative for a concave or diverging lens. The method of solution is thus the same, but the results are different in important ways.

Solution

To find the magnification m, we must first find the image distance d_i using thin lens equation

$$\frac{1}{d_i}=\frac{1}{f}-\frac{1}{d_o}, \tag{11.6.17}$$

or its alternative rearrangement

$$d_i=\frac{fd_o}{d_o-f}. \tag{11.6.18}$$

We are given that $f=-10.00cm$ and $d_o=7.50cm$. Entering these yields a value for $1/d_i$:

$$\frac{1}{d_i}=\frac{1}{-10.0cm}-\frac{1}{7.50cm}=\frac{-0.2333}{cm}. \tag{11.6.19}$$

This must be inverted to find d_i:

$$d_i=-\frac{cm}{0.2333}=-4.29cm. \tag{11.6.20}$$

Or

$$d_i=\frac{(7.5)(-10)}{(7.5-(-10))}=-75/17.5=-4.29cm. \tag{11.6.21}$$

Now the magnification equation can be used to find the magnification m, since both d_i and d_o are known. Entering their values gives

$$m=-\frac{d_i}{d_o}=-\frac{-4.29cm}{7.50cm}=0.571. \tag{11.6.22}$$

Discussion:

A number of results in this example are true of all case 3 images, as well as being consistent with Figure 13. Magnification is positive (as predicted), meaning the image is upright. The magnification is also less than 1, meaning the image is smaller than the object -- in this case, a little over half its size. The image distance is negative, meaning the image is on the same side of the lens as the object. (The image is virtual.) The image is closer to the lens than the object, since the image distance is smaller in magnitude than the object distance. The location of the image is not obvious when you look through a concave lens. In fact, since the image is smaller than the object, you may think it is farther away. But the image is closer than the object, a fact that is useful in correcting nearsightedness, as we shall see in a later section.

The table summarizes the three types of images formed by single thin lenses. These are referred to as case 1, 2, and 3 images. Convex (converging) lenses can form either real or virtual images (cases 1 and 2, respectively), whereas concave (diverging) lenses can form only virtual images (always case 3). Real images are always inverted, but they can be either larger or smaller than the object. For example, a slide projector forms an image larger than the slide, whereas a camera makes an image smaller than the object being photographed. Virtual images are always upright and cannot be projected. Virtual images are larger than the object only in case 2, where a convex lens is used. The virtual image produced by a concave lens is always smaller than the object -- a case 3 image. We can see and photograph virtual images only by using an additional lens to form a real image.

Table 11.6.1: Three Types of Images Formed By Thin Lenses

Type	Formed When	Image type	d_i	m
Case 1	positive **f** $d_0>f$	real	positive	negative
Case 2	positive **f** $d_0<f$	virtual	negative	positive **m > 1**
Case 3	negative **f**	virtual	negative	positive

In "Image Formation by Mirrors," we shall see that mirrors can form exactly the same types of images as lenses.

TAKE-HOME EXPERIMENT: CONCENTRATING SUNLIGHT

Find several lenses and determine whether they are converging or diverging. In general those that are thicker near the edges are diverging and those that are thicker near the center are converging. On a bright sunny day take the converging lenses outside and try focusing the sunlight onto a piece of paper. Determine the focal lengths of the lenses. Be careful because the paper may start to burn, depending on the type of lens you have selected.

Problem-Solving Strategies for Lenses

- Step 1. Examine the situation to determine that image formation by a lens is involved.
- Step 2. Determine whether ray tracing, the thin lens equations, or both are to be employed. A sketch is very useful even if ray tracing is not specifically required by the problem. Write symbols and values on the sketch.
- Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).
- Step 4. Make alist of what is given or can be inferred from the problem as stated (identify the knowns). It is helpful to determine whether the situation involves a case 1, 2, or 3 image. While these are just names for types of images, they have certain characteristics (given in the table) that can be of great use in solving problems.
- Step 5. If ray tracing is required, use the ray tracing rules listed near the beginning of this section.
- Step 6. Most quantitative problems require the use of the thin lens equations. These are solved in the usual manner by substituting knowns and solving for unknowns. Several worked examples serve as guides.
- Step 7. Check to see if the answer is reasonable: Does it make sense? If you have identified the type of image (case 1, 2, or 3), you should assess whether your answer is consistent with the type of image, magnification, and so on.

MISCONCEPTION ALERT:

We do not realize that light rays are coming from every part of the object, passing through every part of the lens, and all can be used to form the final image. We generally feel the entire lens, or mirror, is needed to form an image. Actually, half a lens will form the same, though a fainter, image.

Summary

- Light rays entering a converging lens parallel to its axis cross one another at a single point on the opposite side.
- For a converging lens, the focal point is the point at which converging light rays cross; for a diverging lens, the focal point is the point from which diverging light rays appear to originate.
- The distance from the center of the lens to its focal point is called the focal length f.
- Power P of a lens is defined to be the inverse of its focal length, $P=\frac{1}{f}$.
- A lens that causes the light rays to bend away from its axis is called a diverging lens.
- Ray tracing is the technique of graphically determining the paths that light rays take.
- The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.
- Thin lens equations are $\frac{1}{d_o}+\frac{1}{d_1}=\frac{1}{f}$ and $\frac{h_1}{h_o}=m$ (magnification).
- The distance of the image from the center of the lens is called image distance.
- An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

Glossary

converging lens

a convex lens in which light rays that enter it parallel to its axis converge at a single point on the opposite side

diverging lens

a concave lens in which light rays that enter it parallel to its axis bend away (diverge) from its axis

focal point

for a converging lens or mirror, the point at which converging light rays cross; for a diverging lens or mirror, the point from which diverging light rays appear to originate

focal length
: distance from the center of a lens or curved mirror to its focal point

magnification
: ratio of image height to object height

power
: inverse of focal length

real image
: image that can be projected

virtual image
: image that cannot be projected

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

11.7: Image Formation by Mirrors

Learning Objectives

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

- Illustrate image formation in a flat mirror.
- Explain with ray diagrams the formation of an image using spherical mirrors.
- Determine focal length and magnification given radius of curvature, distance of object and image.

We only have to look as far as the nearest bathroom to find an example of an image formed by a mirror. Images in flat mirrors are the same size as the object and are located behind the mirror. Like lenses, mirrors can form a variety of images. For example, dental mirrors may produce a magnified image, just as makeup mirrors do. Security mirrors in shops, on the other hand, form images that are smaller than the object. We will use the law of reflection to understand how mirrors form images, and we will find that mirror images are analogous to those formed by lenses.

Figure $11.7.1$ helps illustrate how a flat mirror forms an image. Two rays are shown emerging from the same point, striking the mirror, and being reflected into the observer's eye. The rays can diverge slightly, and both still get into the eye. If the rays are extrapolated backward, they seem to originate from a common point behind the mirror, locating the image. (The paths of the reflected rays into the eye are the same as if they had come directly from that point behind the mirror.) Using the law of reflection -- the angle of reflection equals the angle of incidence -- we can see that the image and object are the same distance from the mirror. This is a virtual image, since it cannot be projected -- the rays only appear to originate from a common point behind the mirror. Obviously, if you walk behind the mirror, you cannot see the image, since the rays do not go there. But in front of the mirror, the rays behave exactly as if they had come from behind the mirror, so that is where the image is situated.

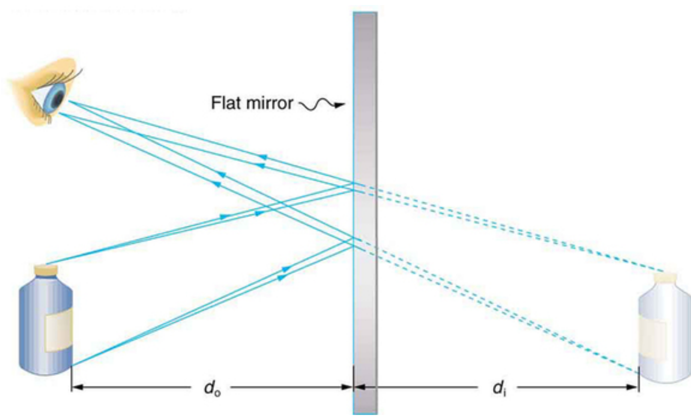


Figure 11.7.1: Two sets of rays from common points on an object are reflected by a flat mirror into the eye of an observer. The reflected rays seem to originate from behind the mirror, locating the virtual image.

Now let us consider the focal length of a mirror -- for example, the concave spherical mirrors in Figure $11.7.2$. Rays of light that strike the surface follow the law of reflection. For a mirror that is large compared with its radius of curvature, as in Figure $11.7.2a$, we see that the reflected rays do not cross at the same point, and the mirror does not have a well-defined focal point. If the mirror had the shape of a parabola, the rays would all cross at a single point, and the mirror would have a well-defined focal point. But parabolic mirrors are much more expensive to make than spherical mirrors. The solution is to use a mirror that is small compared with its radius of curvature, as shown in Figure $11.7.2b$. (This is the mirror equivalent of the thin lens approximation.) To a very good approximation, this mirror has a well-defined focal point at F that is the focal distance f from the center of the mirror. The focal length f of a concave mirror is positive, since it is a converging mirror.

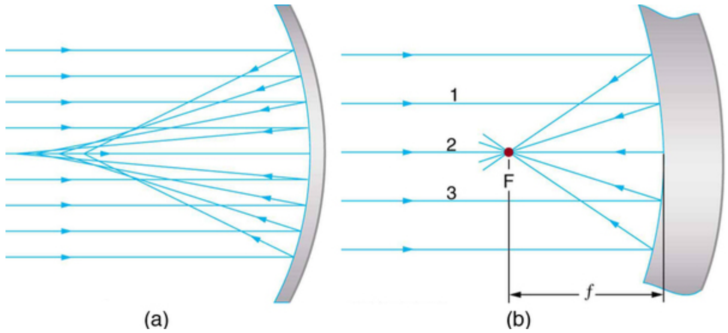


Figure 11.7.2: (a) Parallel rays reflected from a large spherical mirror do not all cross at a common point. (b) If a spherical mirror is small compared with its radius of curvature, parallel rays are focused to a common point. The distance of the focal point from the center of the mirror is its focal length f. Since this mirror is converging, it has a positive focal length.

Just as for lenses, the shorter the focal length, the more powerful the mirror; thus, $P = 1/f$ for a mirror, too. A more strongly curved mirror has a shorter focal length and a greater power. Using the law of reflection and some simple trigonometry, it can be shown that the focal length is half the radius of curvature, or

$$f = \frac{R}{2}, \tag{11.7.1}$$

where R is the radius of curvature of a spherical mirror. The smaller the radius of curvature, the smaller the focal length and, thus, the more powerful the mirror.

The convex mirror shown in Figure 11.7.3 also has a focal point. Parallel rays of light reflected from the mirror seem to originate from the point F at the focal distance f behind the mirror. The focal length and power of a convex mirror are negative, since it is a diverging mirror.

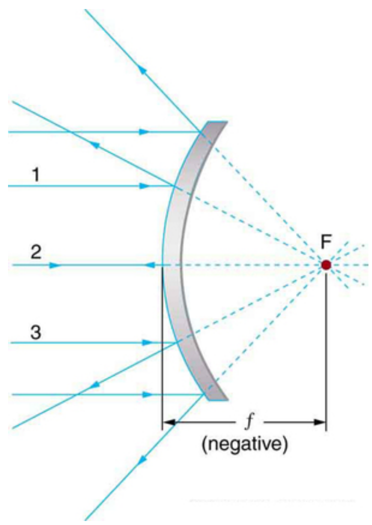


Figure 11.7.3: Parallel rays of light reflected from a convex spherical mirror (small in size compared with its radius of curvature) seem to originate from a well-defined focal point at the focal distance f behind the mirror. Convex mirrors diverge light rays and, thus, have a negative focal length.

Ray tracing is as useful for mirrors as for lenses. The rules for ray tracing for mirrors are based on the illustrations just discussed:

Ray Tracing Rules

1. A ray approaching a concave converging mirror parallel to its axis is reflected through the focal point F of the mirror on the same side. (See rays 1 and 3 in Figure 11.7.2)
2. A ray approaching a convex diverging mirror parallel to its axis is reflected so that it seems to come from the focal point F behind the mirror. (See rays 1 and 3 in Figure 11.7.3).
3. Any ray striking the center of a mirror is followed by applying the law of reflection; it makes the same angle with the axis when leaving as when approaching. (See ray 2 in Figures 11.7.2 and 11.7.3).

4. A ray approaching a concave converging mirror through its focal point is reflected parallel to its axis. (The reverse of rays 1 and 3 in Figure $11.7.3$).
5. A ray approaching a convex diverging mirror by heading toward its focal point on the opposite side is reflected parallel to the axis. (The reverse of rays 1 and 3 in Figure $11.7.3$).

We will use ray tracing to illustrate how images are formed by mirrors, and we can use ray tracing quantitatively to obtain numerical information. But since we assume each mirror is small compared with its radius of curvature, we can use the thin lens equations for mirrors just as we did for lenses.

Consider the situation shown in Figure $11.7.4$, concave spherical mirror reflection, in which an object is placed farther from a concave (converging) mirror than its focal length. That is, f is positive and $d_o > f$, so that we may expect an image similar to the case 1 real image formed by a converging lens. Ray tracing in Figure 4 shows that the rays from a common point on the object all cross at a point on the same side of the mirror as the object. Thus a real image can be projected onto a screen placed at this location. The image distance is positive, and the image is inverted, so its magnification is negative. This is a *case 1 image for mirrors.* It differs from the case 1 image for lenses only in that the image is on the same side of the mirror as the object. It is otherwise identical.

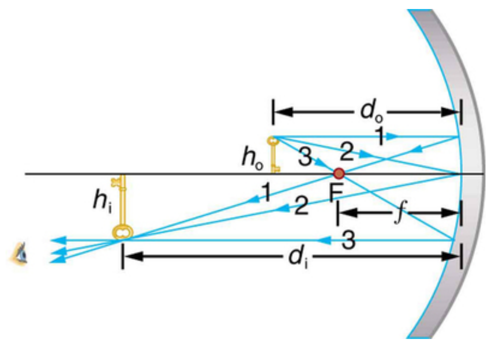


Figure 11.7.4: A case 1 image for a mirror. An object is farther from the converging mirror than its focal length. Rays from a common point on the object are traced using the rules in the text. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 goes through the focal point on the way toward the mirror. All three rays cross at the same point after being reflected, locating the inverted real image. Although three rays are shown, only two of the three are needed to locate the image and determine its height.

Example 11.7.1: A Concave Reflector

Electric room heaters use a concave mirror to reflect infrared (IR) radiation from hot coils. Note that IR follows the same law of reflection as visible light. Given that the mirror has a radius of curvature of 50.0 cm and produces an image of the coils 3.00 m away from the mirror, where are the coils?

Strategy and Concept

We are given that the concave mirror projects a real image of the coils at an image distance $d_i = 3.00 m$. The coils are the object, and we are asked to find their location -- that is, to find the object distance d_o. We are also given the radius of curvature of the mirror, so that its focal length is $f = R/2 = 25.0 cm$ (positive since the mirror is concave or converging). Assuming the mirror is small compared with its radius of curvature, we can use the thin lens equations, to solve this problem.

Solution

Since d_i and f are known, thin lens equation can be used to find d_o:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}.$$

Rearranging to isolate d_o gives

$$\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_i}.$$

Entering known quantities gives a value for $1/d_o$:

$$\frac{1}{d_o} = \frac{1}{0.250m} - \frac{1}{3.00m} = \frac{3.667}{m}.$$

This must be inverted to find d_o:

$$d_o = \frac{1m}{3.667} = 27.3cm.$$

Discussion

Note that the object (the filament) is farther from the mirror than the mirror's focal length. This is a case 1 image ($d_o > f$ and f positive), consistent with the fact that a real image is formed. You will get the most concentrated thermal energy directly in front of the mirror and 3.00 m away from it. Generally, this is not desirable, since it could cause burns. Usually, you want the rays to emerge parallel, and this is accomplished by having the filament at the focal point of the mirror.

Note that the filament here is not much farther from the mirror than its focal length and that the image produced is considerably farther away. This is exactly analogous to a slide projector. Placing a slide only slightly farther away from the projector lens than its focal length produces an image significantly farther away. As the object gets closer to the focal distance, the image gets farther away. In fact, as the object distance approaches the focal length, the image distance approaches infinity and the rays are sent out parallel to one another.

Example 11.7.2: Solar Electric Generating System

One of the solar technologies used today for generating electricity is a device (called a parabolic trough or concentrating collector) that concentrates the sunlight onto a blackened pipe that contains a fluid. This heated fluid is pumped to a heat exchanger, where its heat energy is transferred to another system that is used to generate steam -- and so generate electricity through a conventional steam cycle. Figure $11.7.5$ shows such a working system in southern California. Concave mirrors are used to concentrate the sunlight onto the pipe. The mirror has the approximate shape of a section of a cylinder. For the problem, assume that the mirror is exactly one-quarter of a full cylinder.

a. If we wish to place the fluid-carrying pipe 40.0 cm from the concave mirror at the mirror's focal point, what will be the radius of curvature of the mirror?
b. Per meter of pipe, what will be the amount of sunlight concentrated onto the pipe, assuming the insolation (incident solar radiation) is $0.900kW/m^2$?
c. If the fluid-carrying pipe has a 2.00-cm diameter, what will be the temperature increase of the fluid per meter of pipe over a period of one minute? Assume all the solar radiation incident on the reflector is absorbed by the pipe, and that the fluid is mineral oil.

Strategy

To solve an *Integrated Concept Problem* we must first identify the physical principles involved. Part (a) is related to the current topic. Part (b) involves a little math, primarily geometry. Part (c) requires an understanding of heat and density.

Solution

(a) To a good approximation for a concave or semi-spherical surface, the point where the parallel rays from the sun converge will be at the focal point, so $R = 2f = 80.0cm$.

(b) The insolation is $900W/m^2$. We must find the cross-sectional area A of the concave mirror, since the power delivered is $900W/m^2 \times A$. The mirror in this case is a quarter-section of a cylinder, so the area for a length L of the mirror is $A = \frac{1}{4}(2\pi R)L$. The area for a length of 1.00 m is then

$$\begin{aligned} A &= \frac{\pi}{2}R(1.00m) \\ &= \frac{(3.14)}{2}(0.800m)(1.00m) \\ &= 1.26m^2. \end{aligned}$$

The insolation on the 1.00-m length of pipe is then

$$\left(9.00 \times 10^2 \frac{W}{m^2}\right)\left(1.26 m^2\right) = 1130 W.$$

(c) The increase in temperature is given by $Q = mc\Delta T$. the mass m of the mineral oil in the one-meter section of pipe is

$$\begin{aligned} m &= \rho V \\ &= \rho\pi\left(\frac{d}{2}\right)^2 (1.00 m) \\ &= \left(8.00 \times 10^2 kg/m^3\right)(3.14)(0.0100 m)^2 (1.00 m) \\ &= 0.251 kg. \end{aligned}$$

Therefore, the increase in temperature in one minute is

$$\begin{aligned} \Delta T &= Q/mc \\ &= \frac{(1130 W)(60.0 s)}{(0.251 kg)(1670 J \cdot kg/^\circ C)} \\ &= 162^\circ. \end{aligned}$$

Discussion (c)

An array of such pipes in the California desert can provide a thermal output of 250 MW on a sunny day, with fluids reaching temperatures as high as 400° We are considering only one meter of pipe here, and ignoring heat losses along the pipe.

Figure 11.7.5: Parabolic trough collectors are used to generate electricity in southern California. (credit: kjkolb, Wikimedia Commons)

What happens if an object is closer to a concave mirror than its focal length? This is analogous to a case 2 image for lenses ($d_o < f$ and f positive), which is a magnifier. In fact, this is how makeup mirrors act as magnifiers. Figure $11.7.6a$ uses ray tracing to locate the image of an object placed close to a concave mirror. Rays from a common point on the object are reflected in such a manner that they appear to be coming from behind the mirror, meaning that the image is virtual and cannot be projected. As with a magnifying glass, the image is upright and larger than the object. This is a *case 2 image for mirrors* and is exactly analogous to that for lenses.

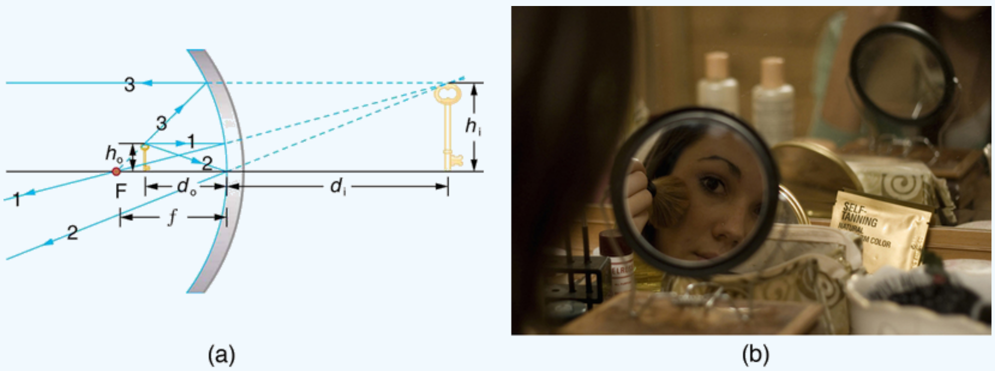
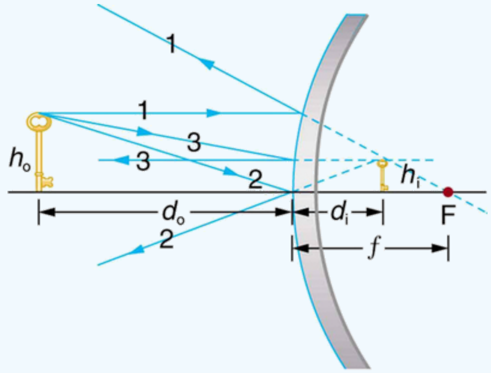


Figure 11.7.6: (a) Case 2 images for mirrors are formed when a converging mirror has an object closer to it than its focal length. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 approaches the mirror as if it came from the focal point. (b) A magnifying mirror showing the reflection. (credit: Mike Melrose, Flickr)

All three rays appear to originate from the same point after being reflected, locating the upright virtual image behind the mirror and showing it to be larger than the object. (b) Makeup mirrors are perhaps the most common use of a concave mirror to produce a larger, upright image.

A convex mirror is a diverging mirror (f is negative) and forms only one type of image. It is a *case 3 image* -- one that is upright and smaller than the object, just as for diverging lenses. Figure $11.7.7a$ uses ray tracing to illustrate the location and size of the case 3 image for mirrors. Since the image is behind the mirror, it cannot be projected and is thus a virtual image. It is also seen to be smaller than the object.



(b)

Figure 11.7.7: uses ray tracing to illustrate the location and size of the case 3 image for mirrors. Since the image is behind the mirror, it cannot be projected and is thus a virtual image. It is also seen to be smaller than the object.

Example 11.7.2: Image in a Convex Mirror

A keratometer is a device used to measure the curvature of the cornea, particularly for fitting contact lenses. Light is reflected from the cornea, which acts like a convex mirror, and the keratometer measures the magnification of the image. The smaller the magnification, the smaller the radius of curvature of the cornea. If the light source is 12.0 cm from the cornea and the image's magnification is 0.0320, what is the cornea's radius of curvature?

Strategy

If we can find the focal length of the convex mirror formed by the cornea, we can find its radius of curvature (the radius of curvature is twice the focal length of a spherical mirror). We are given that the object distance is $d_o = 12.0cm$ and that $m = 0.0320$. We first solve for the image distance d_i, and then for f.

Solution

$m = -d_i/d_o$. Solving this expression for d_i gives

$$d_i = -md_o.$$

Entering known values yields

$$d_i = -(0.0320)(12.0cm) = -0.384cm.$$

$$\frac{1}{f}=\frac{1}{d_o}+\frac{1}{d_i} \tag{11.7.2}$$

Substituting known values,

$$\frac{1}{f}=\frac{1}{12.0cm}+\frac{1}{-0.384cm}=\frac{-2.52}{cm}.$$

This must be inverted to find f.

$$f=\frac{cm}{-2.52}=-0.400cm.$$

The radius of curvature is twice the focal length, so that

$$R=2|f|=0.800cm.$$

Discussion:

Although the focal length f of a convex mirror is defined to be negative, we take the absolute value to give us a positive value for R.

The radius of curvature found here is reasonable for a cornea. The distance from cornea to retina in an adult eye is about 2.0 cm. In practice, many corneas are not spherical, complicating the job of fitting contact lenses. Note that the image distance here is negative, consistent with the fact that the image is behind the mirror, where it cannot be projected. In this section's Problems and Exercises, you will show that for a fixed object distance, the smaller the radius of curvature, the smaller the magnification.

The three types of images formed by mirrors (cases 1, 2, and 3) are exactly analogous to those formed by lenses, as summarized in the table at the end of "Image Formation by Lenses." It is easiest to concentrate on only three types of images -- then remember that concave mirrors act like convex lenses, whereas convex mirrors act like concave lenses.

TAKE-HOME EXPERIMENT: CONCAVE MIRRORS CLOSE TO HOME

Find a flashlight and identify the curved mirror used in it. Find another flashlight and shine the first flashlight onto the second one, which is turned off. Estimate the focal length of the mirror. You might try shining a flashlight on the curved mirror behind the headlight of a car, keeping the headlight switched off, and determine its focal length.

Problem-Solving Strategy for Mirrors

- Step 1. Examine the situation to determine that image formation by a mirror is involved.
- Step 2. Refer to the "Problem-Solving Strategies for Lenses." The same strategies are valid for mirrors as for lenses with one qualification-- use the ray tracing rules for mirrors listed earlier in this section.

Summary

- The characteristics of an image formed by a flat mirror are: (a) The image and object are the same distance from the mirror, (b) The image is a virtual image, and (c) The image is situated behind the mirror.
- Image length is half the radius of curvature.

$$f=\frac{R}{2}$$

- A convex mirror is a diverging mirror and forms only one type of image, namely a virtual image.

Glossary

converging mirror

a concave mirror in which light rays that strike it parallel to its axis converge at one or more points along the axis

diverging mirror

a convex mirror in which light rays that strike it parallel to its axis bend away (diverge) from its axis

law of reflection
: angle of reflection equals the angle of incidence

Contributors and Attributions

CHAPTER OVERVIEW

12: Thermodynamics

Thermodynamics is the branch of science concerned with heat and temperature and their relation to energy and work. It states that the behavior of these quantities is governed by the four laws of thermodynamics, irrespective of the composition or specific properties of the material or system in question. Thermodynamics applies to a wide variety of topics in science and engineering, especially physical chemistry, chemical engineering, and mechanical engineering.

Thumbnail: The Steam engine and gas and oil engines. By John Perry. 1899. This file is in the public domain because its copyright has expired in the United States.

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

12.1: Prelude 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.

Figure 12.1.1. A steam engine uses heat transfer to do work. Tourists regularly ride this narrow-gauge steam engine train near the San Juan Skyway in Durango, Colorado, part of the National Scenic Byways Program. (credit: Dennis Adams)

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.

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

12.2: 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.

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.

Figure 12.2.1: 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)

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

$$\Delta U = Q - W. \tag{12.2.1}$$

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 (Figure 12.2.2). We will now examine Q, W and ΔU further.

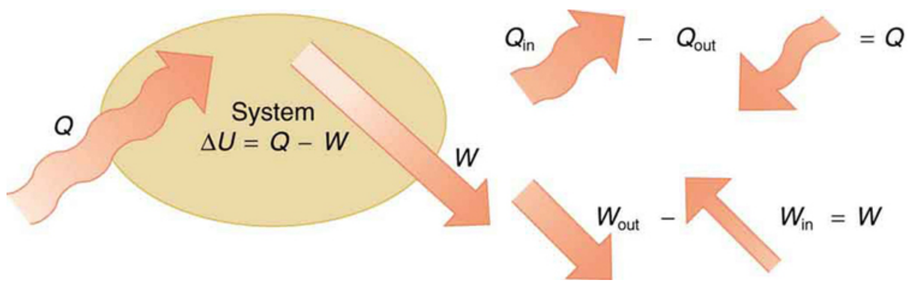


Figure 12.2.2: The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. 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 (Equation 12.2.1).

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 (Equation 12.2.1):

$$\Delta U = Q - W$$

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.

MACROSCOPIC vs. 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

$$\Delta U = U_2 - U_1 \tag{12.2.2}$$

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? Both Q and W *depend on path*, but ΔU does not (Equation 12.2.1). This path independence means that internal energy U is easier to consider than either heat transfer or work done.

Example 12.2.1: 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 (Figure 12.2.3)?

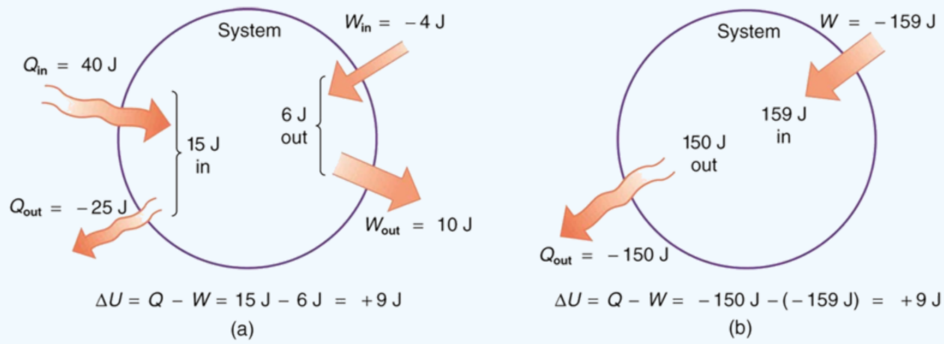


Figure 12.2.3: Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is $\delta U = Q - W = 9.00\,J$. (b) Heat transfer removes 150.00 J from the system while work 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.

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 (Equation 12.2.1).

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

$$\begin{aligned} Q &= 40.00\,J - 25.00\,J \\ &= 15.00\,J \end{aligned}$$

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

$$\begin{aligned} W &= 10.00\,J - 4.00\,J \\ &= 6.00\,J. \end{aligned}$$

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 $Q = -150.00\,J$ and $W = -159.00\,J$, so that

$$\begin{aligned}\Delta U &= Q - W = -150.00 - (-159.00) \\ &= 9.00\,J.\end{aligned}$$

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.

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 thermal equilibrium.

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 $12.2.4$ 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.

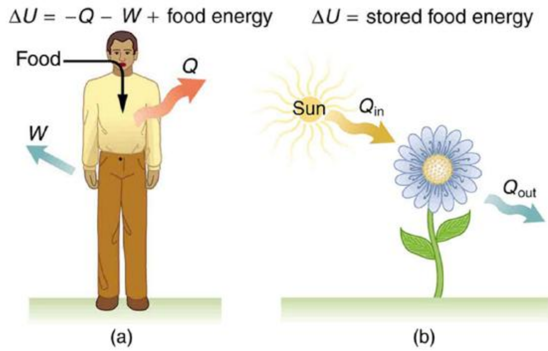


Figure 12.2.4: (a) The first law of thermodynamics applied to metabolism. Heat transferred out of the body Q and work done by the body W remove internal energy, while food intake replaces it. (Food intake may be considered as work done on the body.) (b) Plants convert part of the radiant heat transfer in sunlight to stored chemical energy, a process called photosynthesis.

Summary

The table presents a summary of terms relevant to the first law of thermodynamics.

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.

- 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.

Glossary

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

internal energy

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

human metabolism

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

Contributors and Attributions

12.3: The First Law of Thermodynamics and Some Simple Processes

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.

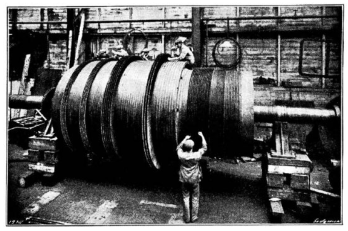


Figure 12.3.1: Beginning with the Industrial Revolution, humans have harnessed power through the use of the first law of thermodynamics, before we even understood it completely. This photo, of a steam engine at the Turbinia Works, dates from 1911, a mere 61 years after the first explicit statement of 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 $12.3.2$shows schematically how the first law of thermodynamics applies to the typical heat engine.

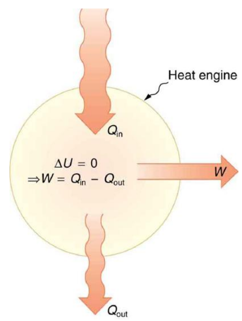


Figure 12.3.2: Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics. It is impossible to devise a system where $Q_{out} = 0$, that is, in which no heat transfer occurs to the environment.

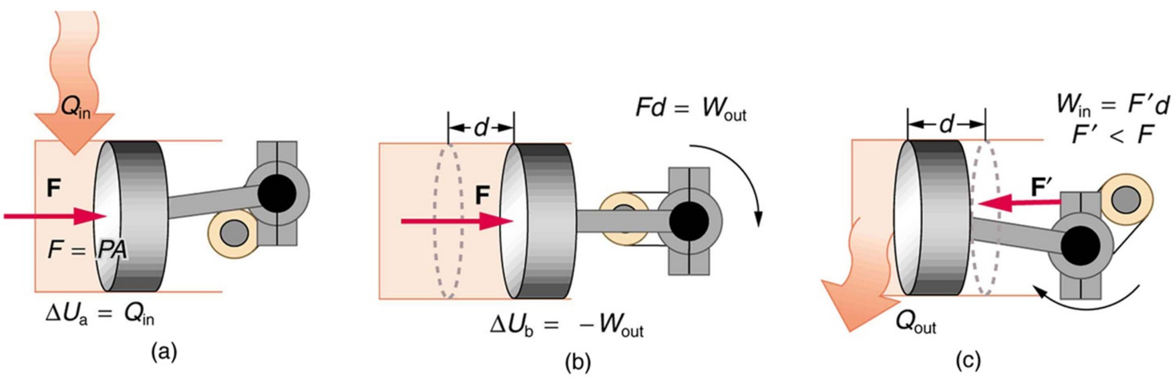


***Figure $12.3.3$:** (a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating that the gas's 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.

Reversible Processes

A process by which a gas does work on a piston at constant pressure is called an **isobaric process**. Such processes are examples of a **thermodynamic process.** A thermodynamic process describes a change that happens to a gas, which results in change in its pressure P, volume V, and/or temperature TTT" id="MathJax-Element-797-Frame" role="presentation" style="position:relative;" tabindex="0">. An **isobaric** process is a thermodynamic process that takes place under constant pressure (so the v

There are three more named thermodynamic processes. These processes are given special names because, like the isobaric process, they occur under some restrictions, which gives them their special pro

An isochoric process is a thermodynamic process in which no change in volume takes place. Because the work done by a gas is proportional to the change in volume, in an isochoric process, no work is

An isothermal process is a thermodynamic process in which no change in temperature takes place. A gas expanding isothermally, for example, does work on the surrounding, but its internal energy (as re

The adiabatic process is, in some sense, the opposite of an isothermal process. In an adiabatic process, no heat transfer takes place (that is $Q = 0$). This may happen because the gas is well-insulated fron

Both isothermal and adiabatic processes 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

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 spontaneo

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

PHET EXPLORATIONS: STATES OF MATTER

Watch different types of molecules form a solid, liquid, or gas in teh States of Matter simuator. Add or remove heat and watch the phase change. Change the temperature or volume of a container and s

Summary

- 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
- 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 stat

Glossary

heat engine
a machine that uses heat transfer to do work

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

isochoric process
a constant-volume process

isothermal process
a constant-temperature process

adiabatic process
a process in which no heat transfer takes place

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

Contributors and Attributions

- Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (Uni

12.4: 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 12.4.1: 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 $12.4.2$)

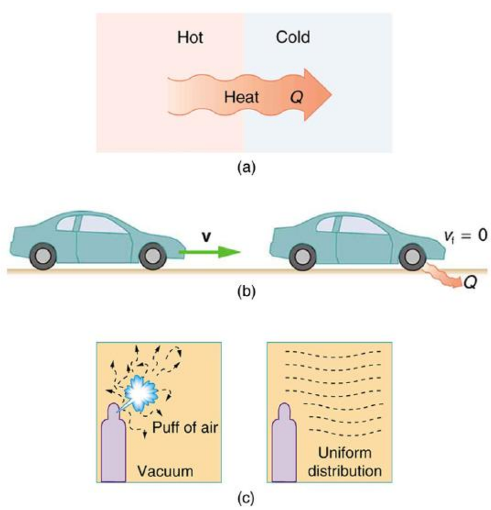


Figure 12.4.2: Examples of one-way processes in nature. (a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum 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 $12.4.3$(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_f, 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.

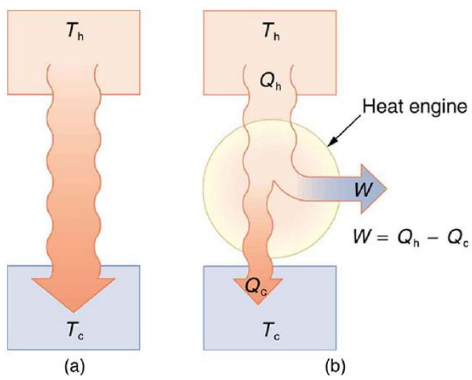


Figure 12.4.3: (a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs Q_h is the 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 $(Q_c = 0)$. 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

$$\Delta U = Q - W, \tag{12.4.1}$$

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

$$0 = Q - W, \tag{12.4.2}$$

so that

$$W = Q. \tag{12.4.3}$$

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

$$W = Q_h - Q_c \,(cyclical\, process), \tag{12.4.4}$$

just as shown schematically in Figure $12.4.3$(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,

$$Eff = \frac{W}{Q_h}. \tag{12.4.5}$$

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

$$Eff = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h} \,(cyclical\, process), \tag{12.4.6}$$

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 12.4.1: 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 $2.50 \times 10^{14}\, J$ of heat transfer from coal and $1.48 \times 10^{1}4\, J$ 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 $2.5 \times 10^{6}\, J$ 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:

$$W = Q_h - Q_c.$$

Substituting the given values:

$$\begin{aligned} W &= 2.50 \times 10^{1}4\, J - 1.48 \times 10^{14}\, J \\ &= 1.02 \times 10^{14}\, J. \end{aligned}$$

Strategy for (b)

The efficiency can be calculated with $Eff = \dfrac{W}{Q_h}$ 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 1.01×10^{14}, and Q_h is given, so the efficiency is

$$Eff = \frac{1.01 \times 10^{14}\ J}{2.50 \times 10^{14}\ J} = 0.408,\ or\ 40.8\% \tag{12.4.7}$$

Strategy for (c)

The daily consumption of coal is calculated using the information that each day there is

$$\frac{2.50 \times 10^{14}\ J}{2.50 \times 10^{6}\ J/kg} = 1.0 \times 10^{8}\ kg.$$

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

$$1.0 \times 10^{8}\ kg\, coal \times \frac{44\ kg\, CO_2}{12\ kg\, coal} = 3.7 \times 10^{8}\ kg\, CO_2.$$

This is 370,000 metric tons of $CO2$ 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 $\mathrm{CO_2}$ per unit energy output (compared to natural gas or oil), making coal the least efficient fossil fuel.

With the information given in Example $12.4.1$, 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 $12.4.4$ 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.

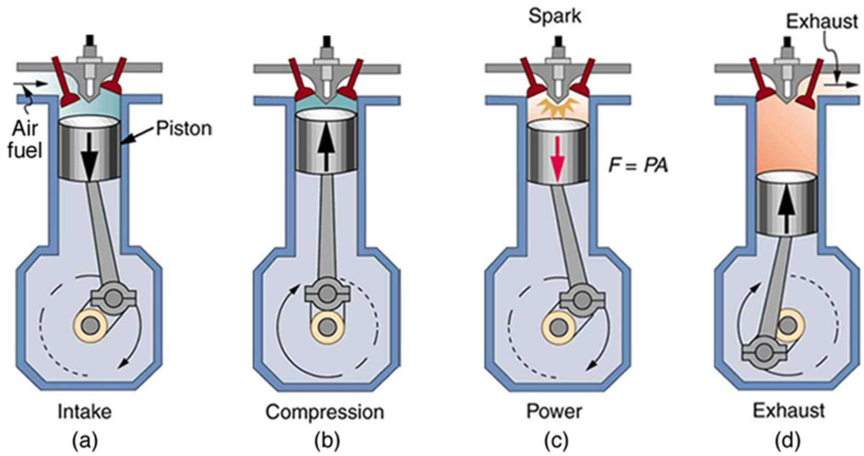


Figure $12.4.4$: In the four-stroke internal combustion gasoline engine, heat transfer into work takes place in the cyclical process shown here. The piston is connected to a rotating crankshaft, which both takes work out of and does work on the gas in the cylinder. (a) Air is mixed with fuel during the 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.

The Otto cycle shown in Figure $12.4.5a$ 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.

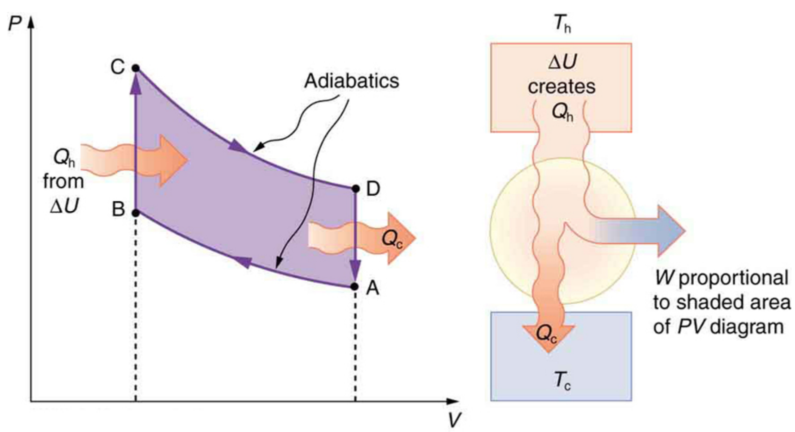


Figure $12.4.5$: PV diagram for a simplified Otto cycle, analogous to that employed in an internal combustion engine. Point A corresponds to the start of the compression stroke of an internal combustion engine. Paths AB and CD are adiabatic and correspond to the compression and power strokes of an internal 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.

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 $12.4.5$. 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 (Figure $12.4.6$. 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.

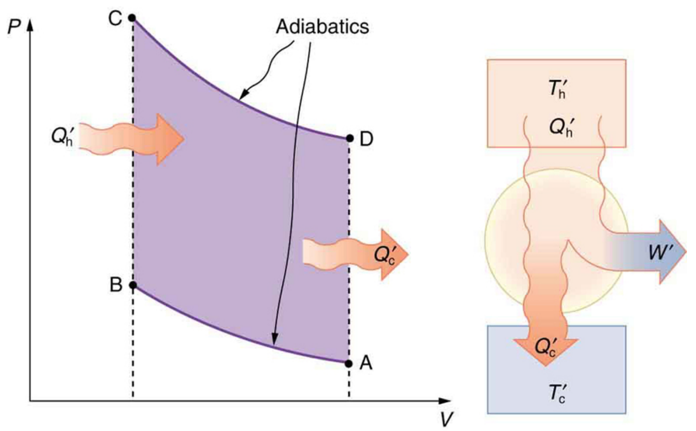


Figure $12.4.5$: This Otto cycle produces a greater work output than the one in Figure 5, because the starting temperature of path CD is higher because the starting temperature of path CD is higher and the starting temperature of path AB is lower. The area inside the loop is greater, corresponding to greater net work output.

Summary

- 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 $W = Q_h - Q_c$, 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 $Eff = \frac{W}{Q_h}$, 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.

Glossary

irreversible process

any process that depends on path direction

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

cyclical process

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

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

Contributors and Attributions

12.5: 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.

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**.

Figure $12.5.1$: 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^oF. 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)

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 $12.5.2$ 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:

$$Eff = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h}. \tag{12.5.1}$$

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

$$Eff_c = 1 - \frac{T_c}{T_h}, \tag{12.5.2}$$

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 $T_c = 0$ - 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.

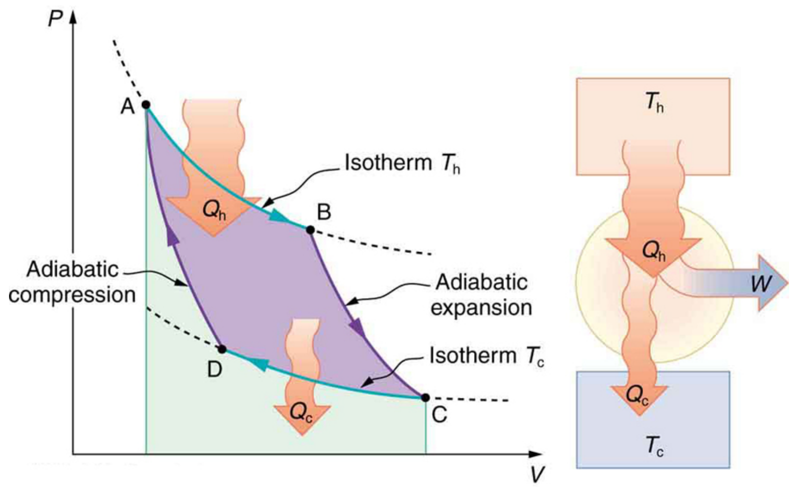


Figure 12.5.2: PV diagram for a Carnot cycle, employing only reversible isothermal and adiabatic processes. Heat transfer Q_h occurs into the working substance during the isothermal path AB, which takes place at constant temperature T_h. Heat transfer Q_c occurs out of the working substance during the 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 12.5.1: Maximum Theoretical Efficiency for a Nuclear Reactor

A nuclear power reactor has pressurized water at 300^oC. (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 12.5.3). Steam, produced in the steam generator, is used to drive the turbine-generators. Eventually the steam is condensed to water at 27^oC and then heated again to start the cycle over. Calculate the maximum theoretical efficiency for a heat engine operating between these two temperatures.

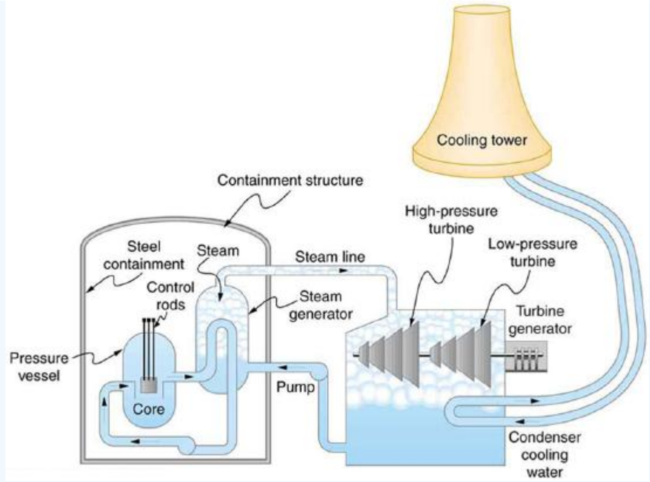


Figure 12.5.3. Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than 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, $Eff_c = 1 - \frac{T_c}{T_h}$ 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^oC and 27^oC, respectively. In kelvins, then, $T_h = 573\,K$ and $T_c = 300\,K$, so that the maximum efficiency is

$$Eff_c = 1 - \frac{T_c}{T_h}. \tag{12.5.3}$$

Thus,

$$Eff_c = 1 - \frac{300\,K}{573\,K} \tag{12.5.4}$$

$$= 0.476,\ or\ 47.6\%. \tag{12.5.5}$$

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 12.5.4 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.

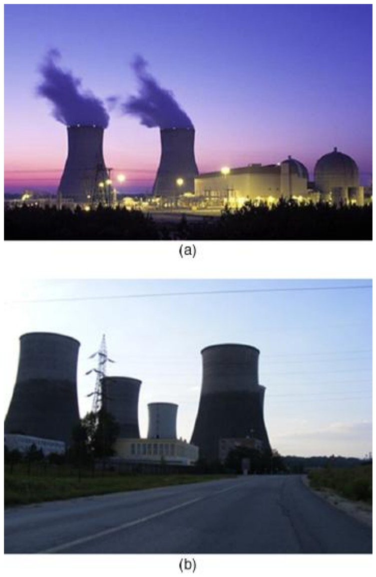


Figure 12.5.4: (a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station. Both have cooling towers in which water evaporates into the environment, representing Q_c. The nuclear reactor, which supplies Q_h, is housed inside the dome-shaped containment buildings. (credit: 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 12.5.5(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 12.5.5(b).

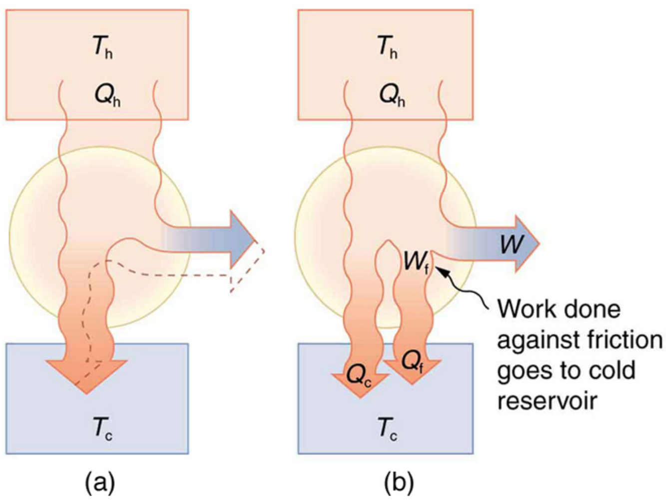


Figure 12.5.5: Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some of its work output into heat transfer to the environment.

Summary

- 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.

Glossary

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

Carnot engine
a heat engine that uses a Carnot cycle

Carnot efficiency
the maximum theoretical efficiency for a heat engine

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

![LibreTexts]

12.6: 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 12.6.1: 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.)

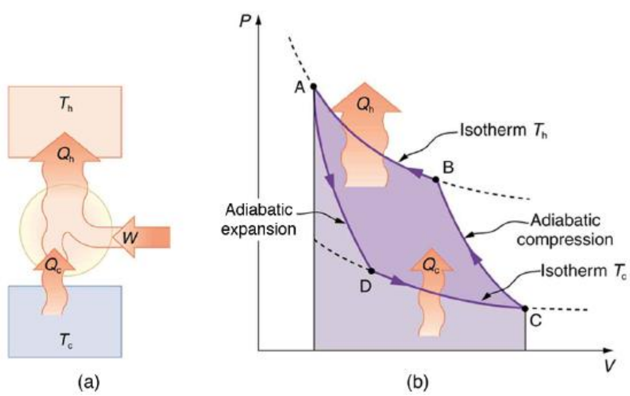


Figure 12.6.3 but reversed, following path ADCBA. The area inside the loop is negative, meaning there is a net work input. There is heat transfer Q_c into the system from a cold reservoir along path DC, and heat transfer Q_h out of the system into a hot reservoir along path BA.

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 12.6.3. 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.

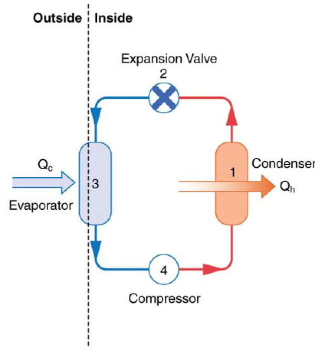


Figure 12.6.3: A simple heat pump has four basic components: (1) condenser, (2) expansion valve, (3) evaporator, and (4) compressor. In the heating mode, heat transfer Q_c occurs to the working fluid in the evaporator (3) from the colder outdoor air, turning it into a gas. The electrically driven compressor (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

$$COP_{hp} = \frac{Q_h}{W}. \tag{12.6.1}$$

Since the efficiency of a heat engine is $Eff = W/Q_h$, wee see that $COP_{hp} = 1/Eff$, 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 $Eff_c = 1 - (T_c?T_h)$ thus, the smaller the temperature difference, the smaller the efficiency and the greater the COP_{hp} (because $COP_{hp} = 1/Eff$). 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.

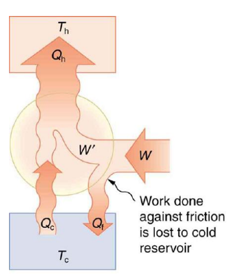


Figure 12.6.4: When a real heat engine is run backward, some of the intended work input W goes into heat transfer before it gets into the heat engine, thereby reducing its coefficient of performance COP_{hp}. In this figure, W' represents the portion of W that goes into the heat pump, while the remainder 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 12.6.1: 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 $COPhp$. (See Figure $12.6.5$) What is the best coefficient of performance possible for such a heat pump, if it has a hot reservoir temperature of 45.0^oC and a cold reservoir temperature of -15.0^oC?

Strategy

A Carnot engine reversed will give the best possible performance as a heat pump. As noted above, $COP_{hp} = 1/Eff$, 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**:**

$$Eff_c = 1 - \frac{T_c}{T_h}. \tag{12.6.2}$$

The temperatures in kelvins are $T_h = 318\,K$ and $T_c = 258\,K$, so that

$$Eff_c = 1 - \frac{258}{318}K = 0.1887. \tag{12.6.3}$$

Thus, from the discussion above,

$$COP_{hp} = \frac{1}{0.1887} = 5.30, \tag{12.6.4}$$

or

$$COP_{hp} = \frac{Q_h}{W} = 5.30, \tag{12.6.5}$$

so that

$$Q_h = 5.30\,W. \tag{12.6.6}$$

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.

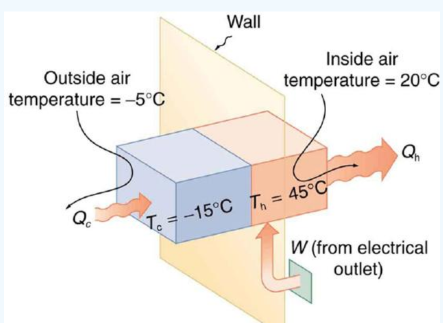


Figure 12.6.5: Heat transfer from the outside to the inside, along with work done to run the pump, takes place in the heat pump of the example above. Note that the cold temperature produced by the heat pump is lower than the outside temperature, so that heat transfer into the working fluid occurs. The 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 12.6.6 shows a heat pump, called a “*reverse cycle*” or “*split-system cooler*” in some countries.

Figure 12.6.6: 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

$$COP_{ref} = \frac{Q_c}{W}. \tag{12.6.7}$$

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

$$COP_{ref} = COP_{hp} - 1 \tag{12.6.8}$$

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 COP 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

$$EER = \frac{Q_c/t_1}{W/t_2}, \tag{12.6.9}$$

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.

Summary

- 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 $COP_{hp} = \frac{Q_h}{W}$.
- A refrigerator is a heat pump; it takes warm ambient air and expands it to chill it.

Glossary

heat pump

a machine that generates heat transfer from cold to hot

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

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

12.7: 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.
- 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 12.7.1: 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

$$\frac{Q_c}{T_c} = \frac{Q_h}{T_h} \tag{12.7.1}$$

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,

$$\Delta S = \left(\frac{Q}{T}\right)_{rev}, \tag{12.7.2}$$

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 $12.7.2$)

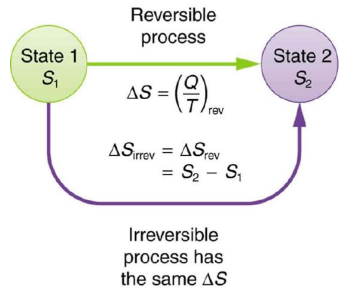


Figure $12.7.2$: When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS whether a hypothetical reversible path is followed or a real irreversible path is taken.

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

$$\Delta S_{tot} = \Delta S_h + \Delta S_c. \tag{12.7.3}$$

Thus, since we know that $Q_h/T_h = Q_c/T_c$ for a Carnot engine,

$$\Delta S_{tot} = -\frac{Q_h}{T_h} + \frac{Q_c}{T_c} = 0. \tag{12.7.4}$$

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 12.7.1: 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 $T_h = 600\,K\,(327^oC)$ to a cold reservoir at $T_c = 250\,K\,(-23^oC)$, assuming there is no temperature change in either reservoir. (See Figure $12.7.3$)

Strategy

How can we calculate the change in entropy for an irreversible process when $\Delta S_{tot} = \Delta S_h + \Delta S_c$ 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 $\Delta S_{tot} = \Delta S_h + \Delta S_c$. First, for the heat transfer from the hot reservoir,

$$\Delta S_h = \frac{-Q_h}{T_h} = \frac{-4000\,J}{600\,K} = -6.67\,J/K. \tag{12.7.5}$$

And for the cold reservoir,

$$\Delta S_c = \frac{Q_c}{T_c} = \frac{4000\ J}{250\ K} = 16.0\ J/K. \tag{12.7.6}$$

Thus the total is

$$\Delta S_{tot} = \Delta S_h + \Delta S_c \tag{12.7.7}$$

$$= (-6.67 + 16.0)\ J/K \tag{12.7.8}$$

$$= 9.33\ J/K. \tag{12.7.9}$$

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.

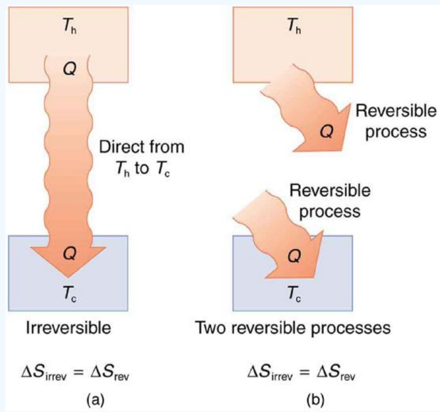


Figure 12.7.3: (a) Heat transfer from a hot object to a cold one is an irreversible process that produces an overall increase in entropy. (b) The same final state and, thus, the same change in entropy is achieved for the objects if reversible heat transfer processes occur between the two objects whose 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 12.7.2: 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 12.7.4)

Strategy

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

Solution (a)

The Carnot efficiency is given by

$$Eff_c = 1 - \frac{T_c}{T_h}. \tag{12.7.10}$$

Substituting the given temperatures yields

$$Eff_c = 1 - \frac{100\ K}{600\ K} = 0.833. \tag{12.7.11}$$

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

$$Eff = \frac{W}{Q_h}. \tag{12.7.12}$$

Solving for W and substituting known terms gives

$$W = Eff_c Q_h \tag{12.7.13}$$

$$= (0.833)(4000\ J) = 3333\ J. \tag{12.7.14}$$

Solution (b)

Similarly,

$$Eff_c = 1 - \frac{T_c}{T_c'} = 1 - \frac{100\ K}{250\ K} = 0.600, \tag{12.7.15}$$

so that

$$W = Eff_c Q_h \tag{12.7.16}$$

$$= (0.600)(4000\ J) = 2400\ J \tag{12.7.17}$$

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.

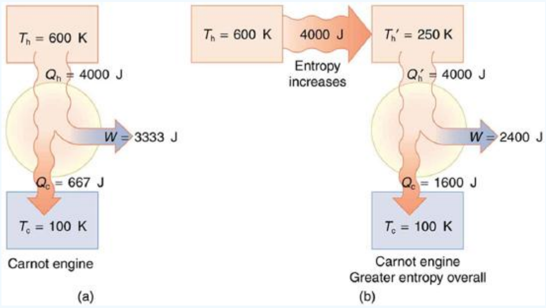


Figure 12.7.4: (a) A Carnot engine working at between 600 K and 100 K has 4000 J of heat transfer and performs 3333 J of work. (b) The 4000 J of heat transfer occurs first irreversibly to a 250 K reservoir and then goes into a Carnot engine. The increase in entropy caused by the heat transfer to a colder 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

$$W_{unavail} = \Delta S \cdot T_0, \tag{12.7.18}$$

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

$$W_{unavail} = (9.33\ J/K)(100\ K) = 933\ J \tag{12.7.19}$$

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 $12.7.5$.) There is a large increase in entropy in the process, as seen in the following example.

Example 12.7.3: Entropy Associated with Disorder

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

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:

$$\Delta S = \frac{Q}{T}. \tag{12.7.20}$$

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

$$Q = mL_f, \tag{12.7.21}$$

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

$$Q = 1.00\, kg)(334\, kJ/kg) = 3.34 \times 10^5\, J. \tag{12.7.22}$$

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

$$\Delta S = \frac{Q}{T} = \frac{3.34 \times 10^5\, J}{T}. \tag{12.7.23}$$

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

$$\Delta S = \frac{3.34 \times 10^5\, J}{273\, K} \tag{12.7.24}$$

$$= 1.22 \times 10^3\, J/K. \tag{12.7.25}$$

Discussion

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

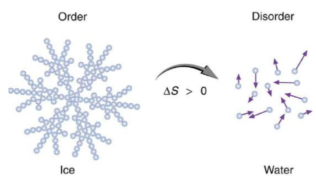


Figure 12.7.5: When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without fixed locations or orientations. Its entropy increases because heat transfer occurs 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

$$\Delta S_{tot} = \Delta S_{syst} + \Delta S_{envir} > 0. \tag{12.7.26}$$

Thus ΔS_{yst} 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, $\Delta S_{tot} = \Delta S_{syst} + \Delta S_{envir} > 0$, 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 $12.7.6$)

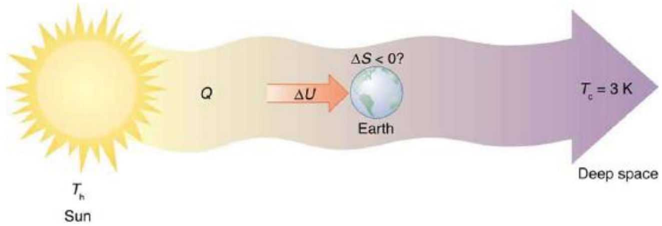


Figure $12.7.6$: Earth's entropy may decrease in the process of intercepting a small part of the heat transfer from the Sun into deep space. Entropy for the entire process increases greatly while Earth becomes more structured with living systems and stored energy in various forms.

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.

PhET Interactive Simulation

$12.7.7$: *Reversible Reaction*

Summary

- 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.

Glossary

entropy

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

change in entropy

the ratio of heat transfer to temperature Q/T

second law of thermodynamics stated in terms of entropy

the total entropy of a system either increases or remains constant; it never decrease

Contributors and Attributions

Paul Peter Urone (Professor Emeritus at California State University, Sacramento) and Roger Hinrichs (State University of New York, College at Oswego) with Contributing Authors: Kim Dirks (University of Auckland) and Manjula Sharma (University of Sydney). This work is licensed by OpenStax University Physics under a Creative Commons Attribution License (by 4.0).

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Survey of Physics

Webpages: 96

All licenses found:

- CC BY 4.0: 76% (73 pages)
- Undeclared: 15.6% (15 pages)
- CC BY-SA 4.0: 8.3% (8 pages)

By Page

- Survey of Physics - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Mass and Inertia - *CC BY-SA 4.0*
    - 1.1: Mass - *CC BY-SA 4.0*
    - 1.2: Equivalence of Gravitational and Inertial Mass - *CC BY-SA 4.0*
    - 1.3: 1.3 Galilean Relativity - *CC BY-SA 4.0*
    - 1.4: A Preview of Some Modern Physics - *CC BY-SA 4.0*
    - 1.5: Footnotes - *CC BY-SA 4.0*
    - 1.6: Problems - *CC BY-SA 4.0*
  - 2: Forces and Motion - *Undeclared*
    - 2.1: Prelude to Dynamics- Newton's Laws of Motion - *CC BY 4.0*
    - 2.2: Development of Force Concept - *CC BY 4.0*
    - 2.3: Newton's First Law of Motion - Inertia - *CC BY 4.0*
    - 2.4: Newton's Second Law of Motion- Concept of a System - *CC BY 4.0*
    - 2.5: Newton's Third Law of Motion- Symmetry in Forces - *CC BY 4.0*
  - 3: Work and Energy - *Undeclared*
    - 3.1: Prelude to Work, Energy, and Energy Resources - *CC BY 4.0*
    - 3.2: Work- The Scientific Definition - *CC BY 4.0*
    - 3.3: Kinetic Energy and the Work-Energy Theorem - *CC BY 4.0*
    - 3.4: Gravitational Potential Energy - *CC BY 4.0*
    - 3.5: Conservation of Energy - *CC BY 4.0*
    - 3.6: Work, Energy, and Power in Humans - *CC BY 4.0*
    - 3.7: World Energy Use - *CC BY 4.0*
  - 4: Momentum - *Undeclared*
    - 4.1: Prelude - *CC BY 4.0*
    - 4.2: Linear Momentum and Force - *CC BY 4.0*
    - 4.3: Impulse - *CC BY 4.0*
    - 4.4: Conservation of Momentum - *CC BY 4.0*
    - 4.5: Elastic Collisions in One Dimension - *CC BY 4.0*
    - 4.6: Inelastic Collisions in One Dimension - *CC BY 4.0*
  - 5: Torque and Angular Momentum - *CC BY 4.0*
    - 5.1: Prelude to Statics and Torque - *CC BY 4.0*
    - 5.2: The First Condition for Equilibrium - *CC BY 4.0*
    - 5.3: The Second Condition for Equilibrium - *CC BY 4.0*
    - 5.4: Simple Machines - *CC BY 4.0*
    - 5.5: Forces and Torques in Muscles and Joints - *CC BY 4.0*
    - 5.6: Prelude to Rotational Motion and Angular Momentum - *CC BY 4.0*
    - 5.7: Angular Momentum and Its Conservation - *CC BY 4.0*
  - 6: Fluid Statics - *Undeclared*
    - 6.1: Prelude to Fluid Statics - *CC BY 4.0*
    - 6.2: What Is a Fluid? - *CC BY 4.0*
    - 6.3: Density - *CC BY 4.0*
    - 6.4: Pressure - *CC BY 4.0*
    - 6.5: Variation of Pressure with Depth in a Fluid - *CC BY 4.0*
    - 6.6: Archimedes' Principle - *CC BY 4.0*
    - 6.7: Cohesion and Adhesion in Liquids - Surface Tension and Capillary Action - *CC BY 4.0*
  - 7: Electricity - *CC BY-SA 4.0*
    - 7.1: Prelude to Electric Charge and Electric Field - *CC BY 4.0*
    - 7.2: Static Electricity and Charge - Conservation of Charge - *CC BY 4.0*
    - 7.3: Conductors and Insulators - *CC BY 4.0*
    - 7.4: Coulomb's Law - *CC BY 4.0*
    - 7.5: Introduction to Electric Potential and Electric Energy - *CC BY 4.0*