ELEMENTARY CONCEPTS IN PHYSICS AND CHEMISTRY



LibreTexts™

Elementary Concepts in Physics and Chemistry

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, 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.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 01/11/2026

TABLE OF CONTENTS

Licensing

1: Chapter 1 - Descriptions in the Physical Sciences

- 1.1: Chemistry as a Science
- 1.2: Expressing Units
- 1.3: Converting Units
- 1.4: Other Units - Temperature and Density
- 1.5: 0.1 Introduction and Review

2: Chapter 2 - Introduction to Chemistry

- 2.1: Prelude to Chemistry
- 2.2: Basic Definitions

3: Chapter 3 - Atoms and Elemental Properties

- 3.1: Prelude to Atoms, Molecules, and Ions
- 3.2: Atomic Theory
- 3.3: Masses of Atoms and Molecules
- 3.4: Organization of the Elements - The Periodic Table
- 3.5: Ions and Ionic Compounds

4: Chapter 4 - Molecules, Bonding, and Forces

- 4.1: Compounds, Lewis Diagrams and Ionic Bonds
- 4.2: Nomenclature of Ionic Compounds
- 4.3: Covalent Bonding
- 4.4: Exceptions to the Octet Rule
- 4.5: Nomenclature of Molecular Compounds
- 4.6: Electronegativity and the Polar Covalent Bond
- 4.7: Intermolecular Forces

5: Chapter 5 - Solution Chemistry

- 5.1: Prelude to Solutions
- 5.2: Definitions
- 5.3: The Mole
- 5.4: Quantitative Units of Concentration
- 5.5: Dilutions and Concentrations
- 5.6: Colligative Properties of Solutions

6: Chapter 6 - Survey of Chemical Reactions

- 6.1: Prelude to Chemical Reactions
- 6.2: The Chemical Equation
- 6.3: The Mole in Chemical Reactions
- 6.4: Stoichiometry
- 6.5: Limiting Reagents

- 6.6: Types of Chemical Reactions - Single and Double Displacement Reactions
- 6.7: Composition, Decomposition, and Combustion Reactions
- 6.8: Neutralization Reactions
- 6.9: Energy
- 6.10: Enthalpy and Chemical Reactions
- 6.11: Chemical Reaction Rate
- 6.12: Collision Theory
- 6.13: Activation Energy
- 6.14: Chemical Equilibrium
- 6.15: Equilibrium Constant
- 6.16: Le Châtelier's Principle

7: Chapter 7 - Survey of Kinematics and Newtonian Motion

- 7.1: Prelude to One-Dimensional Kinematics
- 7.2: Displacement
- 7.3: Vectors, Scalars, and Coordinate Systems
- 7.4: Time, Velocity, and Speed
- 7.5: Acceleration
- 7.6: Kinematics in Two Dimensions - An Introduction
- 7.7: Projectile Motion
- 7.8: Development of Force Concept
- 7.9: Newton’s First Law of Motion- Inertia
- 7.10: Newton’s Second Law of Motion- Concept of a System
- 7.11: Newton’s Third Law of Motion- Symmetry in Forces
- 7.12: Linear Momentum and Force
- 7.13: Impulse

8: Chapter 8 - Energy

- 8.1: Prelude to Work, Energy, and Energy Resources
- 8.2: Work- The Scientific Definition
- 8.3: Kinetic Energy and the Work-Energy Theorem
- 8.4: Gravitational Potential Energy
- 8.5: Conservation of Energy
- 8.6: Nonconservative Forces
- 8.7: Power
- 8.8: Work, Energy, and Power in Humans

9: Chapter 9 - Waves

- 9.1: Free Waves
- 9.2: Sound
- 9.3: Hearing
- 9.4: Speed of Sound, Frequency, and Wavelength
- 9.5: Sound Intensity and Sound Level
- 9.6: Sound Interference and Resonance- Standing Waves in Air Columns
- 9.7: Images by Reflection
- 9.8: Refraction
- 9.9: Wave Optics

10: Chapter 10 - Electricty and Magnetism

- 10.1: Current and Voltage
- 10.2: Parallel and Series Circuits
- 10.3: More About the Magnetic Field
- 10.4: Induced Electric Fields

Index

Glossary

Detailed Licensing

Licensing

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

CHAPTER OVERVIEW

1: Chapter 1 - Descriptions in the Physical Sciences

1.1: Chemistry as a Science

1.2: Expressing Units

1.3: Converting Units

1.4: Other Units - Temperature and Density

1.5: 0.1 Introduction and Review

1: Chapter 1 - Descriptions in the Physical Sciences is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

1.1: Chemistry as a Science

Learning Objective

- Learn what science is and how it works.

Chemistry is a branch of science. Although science itself is difficult to define exactly, the following definition can serve as a starting point. Science is the process of knowing about the natural universe through observation and experiment. Science is not the only process of knowing (e.g., the ancient Greeks simply sat and *thought*), but it has evolved over more than 350 years into the best process that humanity has devised, to date, to learn about the universe around us.

The process of science is usually stated as the *scientific method*, which is rather naively described as follows:

1. state a hypothesis,
2. test the hypothesis, and
3. refine the hypothesis

In actuality, the process is not that simple. (For example, a scientist does not go into their lab every day and exclaim, "I am going to state a hypothesis today and spend the day testing it!") The process is not that simple because science and scientists have a body of knowledge that has already been identified as coming from the highest level of understanding, and most scientists build from that body of knowledge.

An educated guess about how the natural universe works is called a **hypothesis**. A scientist who is familiar with how part of the natural universe works—say, a chemist—is interested in furthering that knowledge. That person makes a reasonable guess—a hypothesis—that is designed to see if the universe works in a new way as well. Here's an example of a hypothesis: "if I mix one part of hydrogen with one part of oxygen, I can make a substance that contains both elements."

> *For a hypothesis to be termed a scientific hypothesis, it has to be something that can be supported or refuted through carefully crafted experimentation or observation.*

Most good hypotheses are grounded in previously understood knowledge and represent a testable extension of that knowledge. The scientist then devises ways to test if that hypothesis is correct or not. That is, the scientist plans experiments. **Experiments are tests of the natural universe to see if a guess (hypothesis) is correct**. An experiment to test our previous hypothesis would be to actually mix hydrogen and oxygen and see what happens. Most experiments include observations of small, well-defined parts of the natural universe designed to see results of the experiments.

A Scientific Hypothesis

A hypothesis is often written in the form of an if/then statement that gives a possibility (if) and explains what may happen because of the possibility (then). For example, if eating elemental sulfur repels ticks, then someone that is eating sulfur every day will not get ticks.



Why do we have to do experiments? Why do we have to test? Because the natural universe is not always so obvious, experiments are necessary. For example, it is fairly obvious that if you drop an object from a height, it will fall. Several hundred years ago (coincidentally, near the inception of modern science), the concept of gravity explained that test. However, is it obvious that the

entire natural universe is composed of only about 115 fundamental chemical building blocks called elements? This wouldn't seem true if you looked at the world around you and saw all the different forms matter can take. In fact, the concept of *the element* is only about 200 years old, and the last naturally occurring element was identified about 80 years ago. It took decades of tests and millions of experiments to establish what the elements actually are. These are just two examples; a myriad of such examples exists in chemistry and science in general.

When enough evidence has been collected to establish a general principle of how the natural universe works, the evidence is summarized in a theory. A theory is a general statement that explains a large number of observations. "All matter is composed of atoms" is a general statement, a theory, that explains many observations in chemistry. A theory is a very powerful statement in science. There are many statements referred to as "the theory of _____" or the "_____ theory" in science (where the blanks represent a word or concept). When written in this way, theories indicate that science has an overwhelming amount of evidence of its correctness. We will see several theories in the course of this text.

A specific statement that is thought to never be violated by the entire natural universe is called a law. A scientific law is the highest understanding of the natural universe that science has and is thought to be inviolate. The fact that all matter attracts all other matter —the law of gravitation—is one such law. Note that the terms *theory* and *law* used in science have slightly different meanings from those in common usage; where theory is often used to mean hypothesis ("I have a theory…"), and a law is an arbitrary limitation that can be broken but with potential consequences (such as speed limits). Here again, science uses these terms differently, and it is important to apply their proper definitions when you use these words in science. (Figure **1.1.1**)

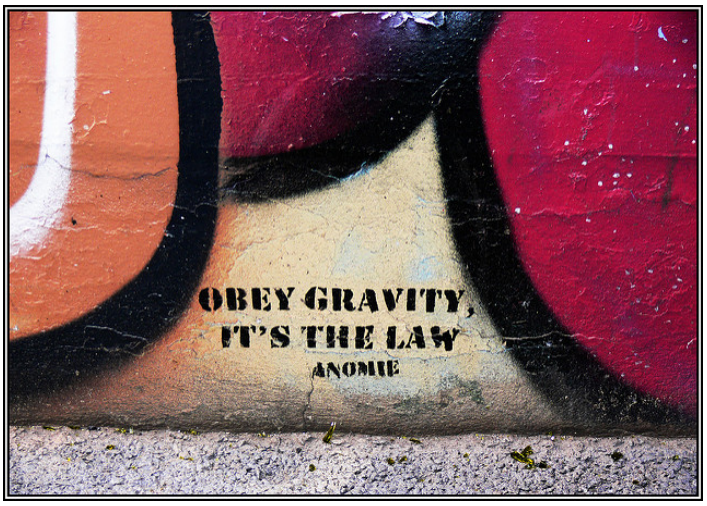


Figure **1.1.1**: Defining a law. Does this graffiti mean "law" the way science defines "law"? (CC BY-SA-NC-ND; Koppenbadger).

There is an additional phrase in our definition of science: "the natural universe." Science is concerned *only* with the natural universe. What is the natural universe? It's anything that occurs around us, well, naturally. Stars, planets, the appearance of life on earth; as well as how animals, plants, and other matter function are all part of the natural universe. Science is concerned with that—and *only* that.

Of course, there are other things that concern us. For example, is the English language part of science? Most of us can easily answer no; English is not science. English is certainly worth knowing (at least for people in predominantly English-speaking countries), but why isn't it science? English, or any human language, is not science because ultimately it is *contrived*; it is made up. Think of it: the word spelled b-l-u-e represents a certain color, and we all agree what color that is. But what if we used the word h-a-r-d to describe that color? (Figure **1.1.2**) That would be fine—as long as everyone agreed. Anyone who has learned a second language must initially wonder why a certain word is used to describe a certain concept; ultimately, the speakers of that language agreed that a particular word would represent a particular concept. It was contrived.

That doesn't mean language isn't worth knowing. It is very important in society. But it's not *science*. Science deals only with what occurs naturally.

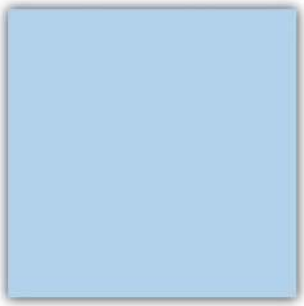


Figure 1.1.2: English Is Not Science. How would you describe this color? Blue or hard? Either way, you're not doing science.

Example 1.1.1: Identifying Science

Which of the following fields would be considered science?

a. geology, the study of the earth
b. ethics, the study of morality
c. political science, the study of governance
d. biology, the study of living organisms

Solution

a. Because the earth is a natural object, the study of it is indeed considered part of science.
b. Ethics is a branch of philosophy that deals with right and wrong. Although these are useful concepts, they are not science.
c. There are many forms of government, but all are created by humans. Despite the fact that the word *science* appears in its name, political science is not true science.
d. Living organisms are part of the natural universe, so the study of them is part of science.

Exercise 1.1.1

Which is part of science, and which is not?

a. dynamics, the study of systems that change over time
b. aesthetics, the concept of beauty

Answer A

science

Answer B

not science

The field of science has gotten so big that it is common to separate it into more specific fields. First, there is mathematics, the language of science. All scientific fields use mathematics to express themselves—some more than others. Physics and astronomy are scientific fields concerned with the fundamental interactions between matter and energy. Chemistry, as defined previously, is the study of the interactions of matter with other matter and with energy. Biology is the study of living organisms, while geology is the study of the earth. Other sciences can be named as well. Understand that these fields are not always completely separate; the boundaries between scientific fields are not always readily apparent. A scientist may be labeled a biochemist if he or she studies the chemistry of biological organisms.

Finally, understand that science can be either qualitative or quantitative. Qualitative implies a description of the quality of an object. For example, physical properties are generally qualitative descriptions: sulfur is yellow, your math book is heavy, or that statue is pretty. A quantitative description represents the specific amount of something; it means knowing how much of something is present, usually by counting or measuring it. Some quantitative descriptions include: 25 students in a class, 650 pages in a book, or a velocity of 66 miles per hour. Quantitative expressions are very important in science; they are also very important in chemistry.

Example 1.1.2: qualitative vs. quantitative Descriptions

Identify each statement as either a qualitative description or a quantitative description.

a. Gold metal is yellow.
b. A ream of paper has 500 sheets in it.
c. The weather outside is snowy.
d. The temperature outside is 24 degrees Fahrenheit.

Solution

a. Because we are describing a physical property of gold, this statement is qualitative.
b. This statement mentions a specific amount, so it is quantitative.
c. The word *snowy* is a description of how the day is; therefore, it is a qualitative statement.
d. In this case, the weather is described with a specific quantity—the temperature. Therefore, it is quantitative.

Exercise 1.1.2

Are these qualitative or quantitative statements?

a. Roses are red, and violets are blue.
b. Four score and seven years ago….

Answer A

qualitative

Answer B

quantitative

Food and Drink Application: Carbonated Beverages

Some of the simple chemical principles discussed in this chapter can be illustrated with carbonated beverages: sodas, beer, and sparkling wines. Each product is produced in a different way, but they all have one thing in common: they are solutions of carbon dioxide dissolved in water.

Carbon dioxide is a compound composed of carbon and oxygen. Under normal conditions, it is a gas. If you cool it down enough, it becomes a solid known as dry ice. Carbon dioxide is an important compound in the cycle of life on earth.

Even though it is a gas, carbon dioxide can dissolve in water, just like sugar or salt can dissolve in water. When that occurs, we have a homogeneous mixture, or a solution, of carbon dioxide in water. However, very little carbon dioxide can dissolve in water. If the atmosphere were pure carbon dioxide, the solution would be only about 0.07% carbon dioxide. In reality, the air is only about 0.03% carbon dioxide, so the amount of carbon dioxide in water is reduced proportionally.

However, when soda and beer are made, manufacturers do two important things: they use pure carbon dioxide gas, and they use it at very high pressures. With higher pressures, more carbon dioxide can dissolve in the water. When the soda or beer container is sealed, the high pressure of carbon dioxide gas remains inside the package. (Of course, there are more ingredients in soda and beer besides carbon dioxide and water.)

When you open a container of soda or beer, you hear a distinctive *hiss* as the excess carbon dioxide gas escapes. But something else happens as well. The carbon dioxide in the solution comes out of solution as a bunch of tiny bubbles. These bubbles impart a pleasing sensation in the mouth, so much so that the soda industry sold over *225 billion* servings of soda in the United States alone in 2009.

Some sparkling wines are made in the same way—by forcing carbon dioxide into regular wine. Some sparkling wines (including champagne) are made by sealing a bottle of wine with some yeast in it. The yeast *ferments*, a process by which the yeast converts sugars into energy and excess carbon dioxide. The carbon dioxide produced by the yeast dissolves in the wine.

Then, when the champagne bottle is opened, the increased pressure of carbon dioxide is released, and the drink bubbles just like an expensive glass of soda.



Figure **1.1.3**: Carbonated Beverages © Thinkstock. Soda, beer, and sparkling wine take advantage of the properties of a solution of carbon dioxide in water.

Soda (left), beer (middle) and sparkling wine (right) images next to one another.

Key Takeaways

- Science is a process of knowing about the natural universe through observation and experiment.
- Scientists go through a rigorous process to determine new knowledge about the universe; this process is generally referred to as the scientific method.
- Science is broken down into various fields, of which chemistry is one.
- Science, including chemistry, is both qualitative and quantitative.

1.1: Chemistry as a Science is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **1.3: Chemistry as a Science** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

1.2: Expressing Units

Learning Objective

- Learn the units that go with various quantities
- Express units using their abbreviations
- Make new units by combining numerical prefixes with units

A number indicates "how much," but the unit indicates "of what." The "of what" is important when communicating a quantity. For example, if you were to ask a friend how close you are to Lake Erie and your friend says "six," then your friend isn't giving you complete information. Six *what*? Six miles? Six inches? Six city blocks? The actual distance to the lake depends on what units you use.

Chemistry, like most sciences, uses the International System of Units, or SI for short. (The letters *SI* stand for the French "le Système International d'unités.") SI specifies certain units for various types of quantities, based on seven fundamental units. We will use most of the fundamental units in chemistry. Initially, we will deal with three fundamental units. The **meter** (m) is the SI unit of length. It is a little longer than a yard (Figure **1.2.1**). The SI unit of mass is the **kilogram** (kg), which is about 2.2 pounds (lb). The SI unit of time is the **second** (s).

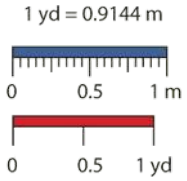


Figure **1.2.1**: The Meter. The SI standard unit of length, the meter, is a little longer than a yard.

To express a quantity, you need to combine a number with a unit. If you have a length that is 2.4 m, then you express that length as simply 2.4 m. A time of 15,000 s can be expressed as 1.5×10^4 s in scientific notation.

Sometimes, a given unit is not an appropriate size to easily express a quantity. For example, the width of a human hair is very small, and it doesn't make much sense to express it in meters. SI also defines a series of *numerical prefixes*, referring to multiples or fractions of a fundamental unit, to make a unit more conveniently sized for a specific quantity. Table **1.2.1** lists the prefixes, their abbreviations, and their multiplicative factors. Some of the prefixes, such as kilo-, mega-, and giga-, represent more than one of the fundamental unit, while other prefixes, such as centi-, milli-, and micro-, represent fractions of the original unit. Note, too, that once again we are using powers of 10. Each prefix is a multiple of or fraction of a power of 10.

Table **1.2.1**: Multiplicative Prefixes for SI Units

Prefix	Abbreviation	Multiplicative Amount	Power of Ten
giga-	G	1,000,000,000 ×	10^9 ×
mega-	M	1,000,000 ×	10^6 ×
kilo-	k	1,000 ×	10^3 ×
deci-	d	1/10 ×	×
centi-	c	1/100 ×	×
milli-	m	1/1,000 ×	×
micro-	μ*	1/1,000,000 ×	×
nano-	n	1/1,000,000,000 ×	×
pico-	p	1/1,000,000,000,000 ×	x
* The letter *μ* is the Greek letter lowercase equivalent to an m and is called "mu" (pronounced "myoo").			

To use the fractions to generate new units, simply combine the prefix with the unit itself; the abbreviation for the new unit is the combination of the abbreviation for the prefix and the abbreviation of the unit. For example, the kilometer (km) is 1,000 × meter, or 1,000 m. Thus, 5 kilometers (5 km) is equal to 5,000 m. Similarly, a millisecond (ms) is 1/1,000 × second, or one-thousandth of a second. Thus, 25 ms is 25 thousandths of a second. You will need to become proficient in combining prefixes and units. (You may recognize that one of our fundamental units, the kilogram, automatically has a prefix-unit combination. The word *kilogram* means 1,000 g.)

In addition to the fundamental units, SI also allows for derived units based on a fundamental unit or units. There are many derived units used in science. For example, the derived unit for area comes from the idea that area is defined as width times height. Because both width and height are lengths, they both have the fundamental unit of meter, so the unit of area is meter × meter, or meter^2 (m^2). This is sometimes spoken as "square meters." A unit with a prefix can also be used to derive a unit for area, so we can also have cm^2, mm^2, or km^2 as acceptable units for area.

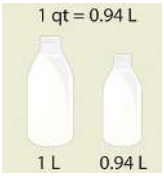


Figure **1.2.2**: The Liter. The SI unit of volume, the liter, is slightly larger than 1 quart.

Volume is defined as length times width times height, so it has units of meter × meter × meter, or meter^3 (m^3)—sometimes spoken as "cubic meters." The cubic meter is a rather large unit, however, so another unit is defined that is somewhat more manageable: the liter (L). A liter is 1/1,000th of a cubic meter and is a little more than 1 quart in volume (Figure **1.2.2**). Prefixes can also be used with the liter unit, so we can speak of milliliters (1/1,000th of a liter; mL) and kiloliters (1,000 L; kL).

Another definition of a liter is one-tenth of a meter cubed. Because one-tenth of a meter is 10 cm, then a liter is equal to 1,000 cm^3 (Figure **1.2.3**). Because 1 L equals 1,000 mL, we conclude that 1 mL equals 1 cm^3; thus, these units are interchangeable.

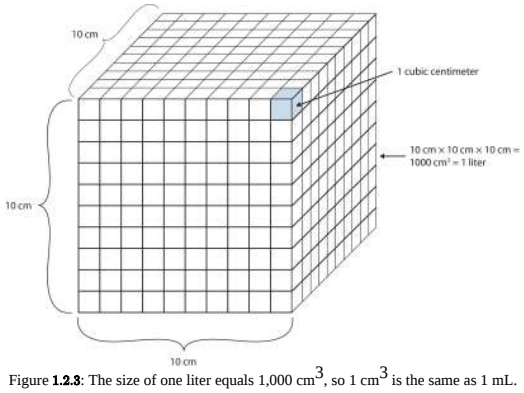


Figure **1.2.3**: The size of one liter equals 1,000 cm^3, so 1 cm^3 is the same as 1 mL.

Units are not only multiplied together—they can also be divided. For example, if you are traveling at one meter for every second of time elapsed, your velocity is 1 meter per second, or 1 m/s. The word *per* implies division, so velocity is determined by dividing a distance quantity by a time quantity. Other units for velocity include kilometers per hour (km/h) or even micrometers per nanosecond (μm/ns). Later, we will see other derived units that can be expressed as fractions.

Example 1.2.1

a. A human hair has a diameter of about 6.0×10^{-5} m. Suggest an appropriate unit for this measurement and write the diameter of a human hair in terms of that unit.
b. What is the velocity of a car if it goes 25 m in 5.0 s?

Solution

a. The scientific notation 10^{-5} is close to 10^{-6}, which defines the micro- prefix. Let us use micrometers as the unit for hair diameter. The number 6.0×10^{-5} can be written as 60×10^{-6}, and a micrometer is 10^{-6} m, so the diameter of a human hair is about 60 μm.
b. If velocity is defined as a distance quantity divided by a time quantity, then velocity is 25 meters/5.0 seconds. Dividing the numbers gives us 25/5.0 = 5.0, and dividing the units gives us meters/second, or m/s. The velocity is 5.0 m/s.

Exercise 1.2.1

a. Express the volume of an Olympic-sized swimming pool, 2,500,000 L, in more appropriate units.
b. A common garden snail moves about 6.1 m in 30 min. What is its velocity in meters per minute (m/min)?

Answer a

- 2.5 ML

Answer b

- 0.203 m/min

Key Takeaways

- Numbers tell "how much," and units tell "of what."
- Chemistry uses a set of fundamental units and derived units from SI units.
- Chemistry uses a set of prefixes that represent multiples or fractions of units.
- Units can be multiplied and divided to generate new units for quantities.

1.2: Expressing Units is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.3: Expressing Units** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.



https://chem.libretexts.org/@go/page/200103

1.3: Converting Units

Learning Objective

- Convert from one unit to another unit of the same type.

In Section 2.2, we showed some examples of how to replace initial units with other units of the same type to get a numerical value that is easier to comprehend. In this section, we will formalize the process.

Consider a simple example: how many feet are there in 4 yards? Most people will almost automatically answer that there are 12 feet in 4 yards. How did you make this determination? Well, if there are 3 feet in 1 yard and there are 4 yards, then there are 4 × 3 = 12 feet in 4 yards.

This is correct, of course, but it is informal. Let us formalize it in a way that can be applied more generally. We know that 1 yard (yd) equals 3 feet (ft):

$$1\,yd = 3\,ft$$

In math, this expression is called an *equality*. The rules of algebra say that you can change (i.e., multiply or divide or add or subtract) the equality (as long as you do not divide by zero) and the new expression will still be an equality. For example, if we divide both sides by 2, we get:

$$\frac{1}{2}\,yd = \frac{3}{2}\,ft$$

We see that one-half of a yard equals 3/2, or one and a half, feet—something we also know to be true—so the above equation is still an equality. Going back to the original equality, suppose we divide both sides of the equation by 1 yard (number **and** unit):

$$\frac{1\,yd}{1\,yd} = \frac{3\,ft}{1\,yd}$$

The expression is still an equality, by the rules of algebra. The left fraction equals 1. It has the same quantity in the numerator and the denominator, so it must equal 1. The quantities in the numerator and denominator cancel, both the number *and* the unit:

$$\frac{1\,yd}{1\,yd} = \frac{3\,ft}{1\,yd}$$

When everything cancels in a fraction, the fraction reduces to 1:

$$1 = \frac{3\,ft}{1\,yd}$$

Conversion Factors

We have an expression that equals 1.

$$\frac{3\,ft}{1\,yd} = 1$$

This is a strange way to write 1, but it makes sense: 3 ft equal 1 yd, so the quantities in the numerator and denominator are the same quantity, just expressed with different units.

The expression

$$\frac{3\,ft}{1\,yd}$$

is called a conversion factor and it is used to formally change the unit of a quantity into another unit. (The process of converting units in such a formal fashion is sometimes called *dimensional analysis* or the *factor label method*.)

To see how this happens, let us start with the original quantity: 4 yd.

Now let us multiply this quantity by 1. When you multiply anything by 1, you do not change the value of the quantity. Rather than multiplying by just 1, let us write 1 as:

$$\frac{3\,ft}{1\,yd}$$

$$4\,yd \times \frac{3\,ft}{1\,yd}$$

The 4 yd term can be thought of as 4yd/1; that is, it can be thought of as a fraction with 1 in the denominator. We are essentially multiplying fractions. If the same thing appears in the numerator and denominator of a fraction, they cancel. In this case, what cancels is the unit *yard*:

$$4\,yd \times \frac{3\,ft}{1\,yd}$$

That is all that we can cancel. Now, multiply and divide all the numbers to get the final answer:

$$\frac{4 \times 3\,ft}{1} = \frac{12\,ft}{1} = 12\,ft$$

Again, we get an answer of 12 ft, just as we did originally. But in this case, we used a more formal procedure that is applicable to a variety of problems.

How many millimeters are in 14.66 m? To answer this, we need to construct a conversion factor between millimeters and meters and apply it correctly to the original quantity. We start with the definition of a millimeter, which is:

$$1\,mm = \frac{1}{1000\,m}$$

The 1/1000 is what the prefix *milli-* means. Most people are more comfortable working without fractions, so we will rewrite this equation by bringing the 1,000 into the numerator of the other side of the equation:

$$1000\,mm = 1\,m$$

Now we construct a conversion factor by dividing one quantity into both sides. But now a question arises: which quantity do we divide by? It turns out that we have two choices, and the two choices will give us different conversion factors, both of which equal 1:

$$\frac{1000\,mm}{1000\,mm} = \frac{1\,m}{1000\,mm}$$

or

$$\frac{1000\,mm}{1\,m} = \frac{1\,m}{1\,m}$$

$$1 = \frac{1\,m}{1000\,mm}$$

or

$$\frac{1000\,mm}{1\,m} = 1$$

Which conversion factor do we use? The answer is based on *what unit you want to get rid of in your initial quantity*. The original unit of our quantity is meters, which we want to convert to millimeters. Because the original unit is assumed to be in the numerator, to get rid of it, we want the meter unit in the *denominator*; then they will cancel. Therefore, we will use the second conversion factor. Canceling units and performing the mathematics, we get:

$$14.66m \times \frac{1000\,mm}{1\,m} = 14660\,mm$$

Note how m cancels, leaving mm, which is the unit of interest.

The ability to construct and apply proper conversion factors is a very powerful mathematical technique in chemistry. You need to master this technique if you are going to be successful in this and future courses.

Example 1.3.1

a. Convert 35.9 kL to liters.
b. Convert 555 nm to meters.

Solution

a. We will use the fact that 1 kL = 1,000 L. Of the two conversion factors that can be defined, the one that will work is 1000L/1kL. Applying this conversion factor, we get:

$$35.9\,kL \times \frac{1000\,L}{1\,kL} = 35,900\,L$$

b. We will use the fact that 1 nm = 1/1,000,000,000 m, which we will rewrite as 1,000,000,000 nm = 1 m, or 10^9 nm = 1 m. Of the two possible conversion factors, the appropriate one has the nm unit in the denominator:

$$\frac{1\,m}{10^9\,nm}$$

Applying this conversion factor, we get:

$$555\,nm \times \frac{1m}{10^9 nm} = 0.000000555\,m = 5.55 \times 10^{-7}\,m$$

In the final step, we expressed the answer in scientific notation.

Exercise 1.3.1

a. Convert 67.08 µL to liters.
b. Convert 56.8 m to kilometers.

Answer a

6.708×10^{-5} L

Answer b

5.68×10^{-2} km

What if we have a derived unit that is the product of more than one unit, such as m^2? Suppose we want to convert square meters to square centimeters? The key is to remember that m^2 means m × m, which means we have *two* meter units in our derived unit. That means we have to include *two* conversion factors, one for each unit. For example, to convert 17.6 m^2 to square centimeters, we perform the conversion as follows:

$$\begin{aligned} 17.6m^2 &= 17.6(m \times m) \times \frac{100cm}{1m} \times \frac{100cm}{1m} \\ &= 176000\,cm \times cm \\ &= 1.76 \times 10^5\,cm^2 \end{aligned} \tag{1.3.1}$$

Example 1.3.2

How many cubic centimeters are in 0.883 m^3?

Solution

With an exponent of 3, we have three length units, so by extension we need to use three conversion factors between meters and centimeters. Thus, we have:

$$0.883 m^3 \times \frac{100\,cm}{1\,m} \times \frac{100\,cm}{1\,m} \times \frac{100\,cm}{1\,m} = 883000\,cm^3 = 8.83 \times 10^5\,cm^3$$

You should demonstrate to yourself that the three meter units do indeed cancel.

? Exercise 1.3.2

How many cubic millimeters are present in 0.0923 m^3?

Answer

9.23×10^7 mm^3

Suppose the unit you want to convert is in the denominator of a derived unit—what then? Then, in the conversion factor, the unit you want to remove must be in the *numerator*. This will cancel with the original unit in the denominator and introduce a new unit in the denominator. The following example illustrates this situation.

✓ Example 1.3.3

Convert 88.4 m/min to meters/second.

Solution

We want to change the unit in the denominator from minutes to seconds. Because there are 60 seconds in 1 minute (60 s = 1 min), we construct a conversion factor so that the unit we want to remove, minutes, is in the numerator: 1min/60s. Apply and perform the math:

$$\frac{88.4 m}{min} \times \frac{1\,min}{60\,s} = 1.47 \frac{m}{s}$$

Notice how the 88.4 automatically goes in the numerator. That's because any number can be thought of as being in the numerator of a fraction divided by 1.

? Exercise 1.3.3

Convert 0.203 m/min to meters/second.

Answer

0.00338 m/s

or

3.38×10^{-3} m/s

Sometimes there will be a need to convert from one unit with one numerical prefix to another unit with a different numerical prefix. How do we handle those conversions? Well, you could memorize the conversion factors that interrelate all numerical prefixes. Or you can go the easier route: first convert the quantity to the base unit—the unit with no numerical prefix—using the definition of the original prefix. Then, convert the quantity in the base unit to the desired unit using the definition of the second prefix. You can do the conversion in two separate steps or as one long algebraic step. For example, to convert 2.77 kg to milligrams:

$$2.77\,kg \times \frac{1000\,g}{1\,kg} = 2770\,g$$

(convert to the base units of grams)

$$2770\,g \times \frac{1000\,mg}{1\,g} = 2770000\,mg = 2.77 \times 10^6\,mg$$

(convert to desired unit)

Alternatively, it can be done in a single multi-step process:

$$2.77 \cancel{kg} \times \frac{1000 \cancel{g}}{1 \cancel{kg}} \times \frac{1000\,mg}{1 \cancel{g}} = 2770000\,mg$$

$$= 2.77 \times 10^6\,mg \tag{1.3.2}$$

You get the same answer either way.

Example 1.3.4

How many nanoseconds are in 368.09 µs?

Solution

You can either do this as a one-step conversion from microseconds to nanoseconds or convert to the base unit first and then to the final desired unit. We will use the second method here, showing the two steps in a single line. Using the definitions of the prefixes *micro-* and *nano-*,

$$368.0\,\mu s \times \frac{1\,s}{1000000\,\mu s} \times \frac{1000000000}{1\,s} = 3.6809 \times 10^5\,ns$$

Exercise 1.3.4

How many milliliters are in 607.8 kL?

Answer

6.078×10^8 mL

When considering the significant figures of a final numerical answer in a conversion, there is one important case where a number does not impact the number of significant figures in a final answer: the so-called **exact number**. An exact number is a number from a defined relationship, not a measured one. For example, the prefix *kilo-* means 1,000-*exactly* 1,000, no more or no less. Thus, in constructing the conversion factor:

$$\frac{1000\,g}{1\,kg}$$

neither the 1,000 nor the 1 enter into our consideration of significant figures. The numbers in the numerator and denominator are defined exactly by what the prefix *kilo-* means. Another way of thinking about it is that these numbers can be thought of as having an infinite number of significant figures, such as:

$$\frac{1000.0000000000\ldots\,g}{1.0000000000\ldots\,kg}$$

The other numbers in the calculation will determine the number of significant figures in the final answer.

Example 1.3.5

A rectangular plot in a garden has the dimensions 36.7 cm by 128.8 cm. What is the area of the garden plot in square meters? Express your answer in the proper number of significant figures.

Solution

Area is defined as the product of the two dimensions, which we then have to convert to square meters, and express our final answer to the correct number of significant figures—which in this case will be three.

$$36.7\,cm \times 128.8\,cm \times \frac{1\,m}{100\,cm} \times \frac{1\,m}{100\,cm} = 0.472696\,m^2 = 0.473\,m^2$$

The 1 and 100 in the conversion factors do not affect the determination of significant figures because they are exact numbers, defined by the centi- prefix.

? Exercise 1.3.5

What is the volume of a block in cubic meters with the dimensions 2.1 cm × 34.0 cm × 118 cm?

Answer

0.0084 m^3

Chemistry is Everywhere: The Gimli Glider

On July 23, 1983, an Air Canada Boeing 767 jet had to glide to an emergency landing at Gimli Industrial Park Airport in Gimli, Manitoba, because it unexpectedly ran out of fuel during flight. There was no loss of life in the course of the emergency landing, only some minor injuries associated in part with the evacuation of the craft after landing. For the remainder of its operational life (the plane was retired in 2008), the aircraft was nicknamed "the Gimli Glider."



The Gimli Glider is the Boeing 767 that ran out of fuel and glided to safety at Gimli Airport. The aircraft ran out of fuel because of confusion over the units used to express the amount of fuel. Source: Photo courtesy of Will F., (CC BY-SA 2.5; Aero Icarus).

The 767 took off from Montreal on its way to Ottawa, ultimately heading for Edmonton, Canada. About halfway through the flight, all the engines on the plane began to shut down because of a lack of fuel. When the final engine cut off, all electricity (which was generated by the engines) was lost; the plane became, essentially, a powerless glider. Captain Robert Pearson was an experienced glider pilot, although he had never flown a glider the size of a 767. First Officer Maurice Quintal quickly determined that the aircraft would not be able make it to Winnipeg, the next large airport. He suggested his old Royal Air Force base at Gimli Station, one of whose runways was still being used as a community airport. Between the efforts of the pilots and the flight crew, they managed to get the airplane safely on the ground (although with buckled landing gear) and all passengers off safely.

What happened? At the time, Canada was transitioning from the older English system to the metric system. The Boeing 767s were the first aircraft whose gauges were calibrated in the metric system of units (liters and kilograms) rather than the English system of units (gallons and pounds). Thus, when the fuel gauge read 22,300, the gauge meant kilograms, but the ground crew mistakenly fueled the plane with 22,300 *pounds* of fuel. This ended up being just less than half of the fuel needed to make the trip, causing the engines to quit about halfway to Ottawa. Quick thinking and extraordinary skill saved the lives of 61 passengers and 8 crew members—an incident that would not have occurred if people were watching their units.

Key Takeaways

- Units can be converted to other units using the proper conversion factors.
- Conversion factors are constructed from equalities that relate two different units.
- Conversions can be a single step or multi-step.

- Unit conversion is a powerful mathematical technique in chemistry that must be mastered.
- Exact numbers do not affect the determination of significant figures.

1.3: Converting Units is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.5: Converting Units** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

1.4: Other Units - Temperature and Density

Learning Objective

- Learn about the various temperature scales that are commonly used in chemistry.
- Define density and use it as a conversion factor.

There are other units in chemistry that are important, and we will cover others over the course of the entire book. One of the fundamental quantities in science is temperature. Temperature is a measure of the average amount of energy of motion, or *kinetic energy*, a system contains. Temperatures are expressed using scales that use units called degrees, and there are several temperature scales in use. In the United States, the commonly used temperature scale is the *Fahrenheit scale* (symbolized by °F and spoken as "degrees Fahrenheit"). On this scale, the freezing point of liquid water (the temperature at which liquid water turns to solid ice) is 32°F, and the boiling point of water (the temperature at which liquid water turns to steam) is 212°F.

Science also uses other scales to express temperature. The Celsius scale (symbolized by °C and spoken as "degrees Celsius") is a temperature scale where 0°C is the freezing point of water and 100°C is the boiling point of water; the scale is divided into 100 divisions between these two landmarks and extended higher and lower. By comparing the Fahrenheit and Celsius scales, a conversion between the two scales can be determined:

$$^{\circ}\mathrm{C} = (^{\circ}\mathrm{F}\text{-}32) \times \tfrac{5}{9} \tag{1.4.1}$$

$$^{\circ}\mathrm{F} = (^{\circ}\mathrm{C} \times \tfrac{9}{5}) + 32 \tag{1.4.2}$$

Example 1.4.1: Conversions

a. What is 98.6 °F in degrees Celsius?
b. What is 25.0 °C in degrees Fahrenheit?

Solution

a. Using Equation 1.4.1, we have

$$\begin{aligned} ^{\circ}C &= (98.6 - 32) \times \frac{5}{9} \\ &= 66.6 \times \frac{5}{9} \\ &= 37.0^{\circ}C \end{aligned}$$

b. Using Equation 1.4.2, we have

$$\begin{aligned} ^{\circ}F &= \left(25.0 \times \frac{9}{5}\right) + 32 \\ &= 45.0 + 32 \\ &= 77.0^{\circ}F \end{aligned}$$

Exercise 1.4.1

a. Convert 0 °F to degrees Celsius.
b. Convert 212 °C to degrees Fahrenheit.

Answer a

−17.8 °C

Answer b

414 °F

The fundamental unit of temperature (another fundamental unit of science, bringing us to four) in SI is the kelvin (K). The Kelvin temperature scale (note that the name of the scale capitalizes the word *Kelvin*, but the unit itself is lowercase) uses degrees that are the same size as the Celsius degree, but the numerical scale is shifted up by 273.15 units. That is, the conversion between the Kelvin and Celsius scales is as follows:

$$K = {}^\circ C + 273.15$$

For most purposes, it is acceptable to use 273 instead of 273.15. Note that the Kelvin scale does not use the word *degrees*; a temperature of 295 K is spoken of as "two hundred ninety-five kelvins" and not "two hundred ninety-five degrees Kelvin."

The reason that the Kelvin scale is defined this way is because there exists a minimum possible temperature called absolute zero (zero kelvins). The Kelvin temperature scale is set so that 0 K is absolute zero, and temperature is counted upward from there. Normal room temperature is about 295 K, as seen in the following example.

Example 1.4.2: Room Temperature

If normal room temperature is 72.0°F, what is room temperature in degrees Celsius and kelvin?

Solution

First, we use the formula to determine the temperature in degrees Celsius:

$$\begin{aligned} {}^\circ C &= (72.0 - 32) \times \frac{5}{9} \\ &= 40.0 \times \frac{5}{9} \\ &= 22.2^\circ C \end{aligned}$$

Then we use the appropriate formula above to determine the temperature in the Kelvin scale:

$$\begin{aligned} K &= 22.2^\circ C + 273.15 \\ &= 295.4 K \end{aligned}$$

So, room temperature is about 295 K.

Exercise 1.4.2

What is 98.6°F on the Kelvin scale?

Answer

310.2 K

Figure **1.4.1** compares the three temperature scales. Note that science uses the Celsius and Kelvin scales almost exclusively; virtually no practicing chemist expresses laboratory-measured temperatures with the Fahrenheit scale. In fact, the United States is one of the few countries in the world that still uses the Fahrenheit scale on a daily basis. The other two countries are Liberia and Myanmar (formerly Burma). People driving near the borders of Canada or Mexico may pick up local radio stations on the other side of the border that express the daily weather in degrees Celsius, so do not get confused by their weather reports.

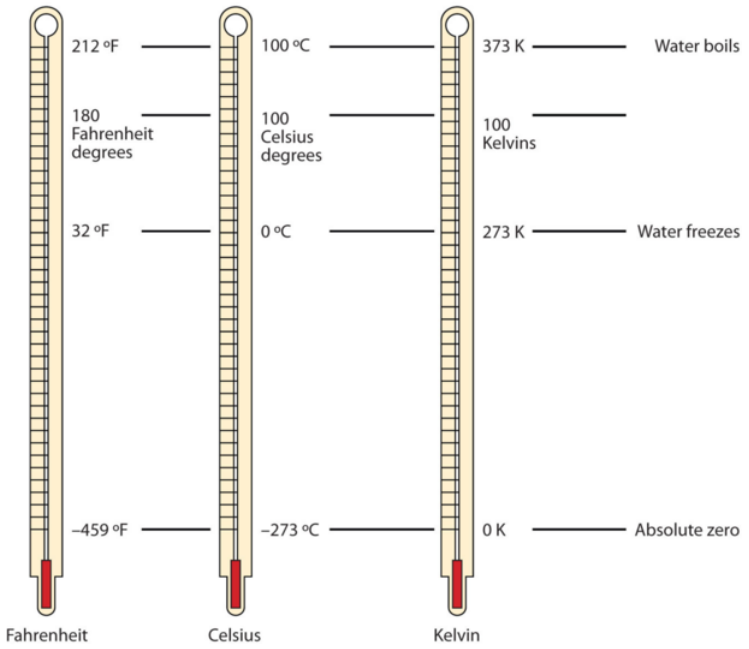


Figure 1.4.1: Fahrenheit, Celsius, and Kelvin Temperatures. A comparison of the three temperature scales.

Density

Density is a physical property that is defined as a substance's mass divided by its volume:

$$density = \frac{mass}{volume} \Rightarrow d = \frac{m}{v}$$

Density is usually a measured property of a substance, so its numerical value affects the significant figures in a calculation. Notice that density is defined in terms of two dissimilar units, mass and volume. That means that density overall has derived units, just like velocity. Common units for density include g/mL, g/cm^3, g/L, kg/L, and even kg/m^3. Densities for some common substances are listed in Table 1.4.1.

Table 1.4.1: Densities of Some Common Substances

Substance	Density (g/mL or g/cm^3)
water	1.0
gold	19.3
mercury	13.6
air	0.0012
cork	0.22–0.26
aluminum	2.7
iron	7.87

Because of how it is defined, density can act as a conversion factor for switching between units of mass and volume. For example, suppose you have a sample of aluminum that has a volume of 7.88 cm^3. How can you determine what mass of aluminum you have without measuring it? You can use the volume to calculate it. If you multiply the given volume by the known density (from Table 1.4.1), the volume units will cancel and leave you with mass units, telling you the mass of the sample:

$$7.88 \cancel{cm^3} \times \frac{2.7\,g}{\cancel{cm^3}} = 21\,g \text{ of aluminium}$$

where we have limited our answer to two significant figures.

Example 1.4.3: Mercury

What is the mass of 44.6 mL of mercury?

Solution

Use the density from Table **1.4.1** "Densities of Some Common Substances" as a conversion factor to go from volume to mass:

$$44.6 \cancel{mL} \times \frac{13.6\,g}{\cancel{mL}} = 607\,g$$

The mass of the mercury is 607 g.

Exercise 1.4.3

What is the mass of 25.0 cm^3 of iron?

Answer

197 g

Density can also be used as a conversion factor to convert mass to volume—but care must be taken. We have already demonstrated that the number that goes with density normally goes in the numerator when density is written as a fraction. Take the density of gold, for example:

$$d = 19.3\,g/mL = \frac{19.3\,g}{mL}$$

Although this was not previously pointed out, it can be assumed that there is a 1 in the denominator:

$$d = 19.3\,g/mL = \frac{19.3\,g}{mL}$$

That is, the density value tells us that we have 19.3 grams for every 1 milliliter of volume, and the 1 is an exact number. When we want to use density to convert from mass to volume, the numerator and denominator of density need to be switched—that is, we must take the *reciprocal* of the density. In so doing, we move not only the units, but also the numbers:

$$\frac{1}{d} = \frac{1\,mL}{19.3\,g}$$

This reciprocal density is still a useful conversion factor, but now the mass unit will cancel and the volume unit will be introduced. Thus, if we want to know the volume of 45.9 g of gold, we would set up the conversion as follows:

$$45.9 \cancel{g} \times \frac{1\,mL}{19.3\,\cancel{g}} = 2.38\,mL$$

Note how the mass units cancel, leaving the volume unit, which is what we are looking for.

Example 1.4.4: Calculating Volume from Density

A cork stopper from a bottle of wine has a mass of 3.78 g. If the density of cork is 0.22 g/cm^3, what is the volume of the cork?

Solution

To use density as a conversion factor, we need to take the reciprocal so that the mass unit of density is in the denominator. Taking the reciprocal, we find:

$$\frac{1}{d} = \frac{1\,cm^3}{0.22\,g}$$

We can use this expression as the conversion factor. So

$$3.78 \cancel{g} \times \frac{1\,cm^3}{0.22\,\cancel{g}} = 17.2\,cm^3$$

? Exercise 1.4.4

What is the volume of 3.78 g of gold?

Answer

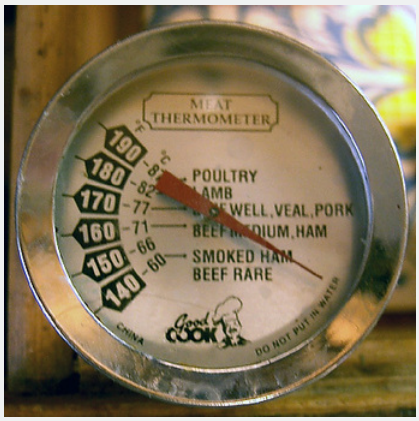
0.196 cm^3

Care must be used with density as a conversion factor. Make sure the mass units are the same, or the volume units are the same, before using density to convert to a different unit. Often, the unit of the given quantity must be first converted to the appropriate unit before applying density as a conversion factor.

Food and Drink Application: Cooking Temperatures

Because degrees Fahrenheit is the common temperature scale in the United States, kitchen appliances, such as ovens, are calibrated in that scale. A cool oven may be only 150°F, while a cake may be baked at 350°F and a chicken roasted at 400°F. The broil setting on many ovens is 500°F, which is typically the highest temperature setting on a household oven.

People who live at high altitudes, typically 2,000 ft above sea level or higher, are sometimes urged to use slightly different cooking instructions on some products, such as cakes and bread, because water boils at a lower temperature the higher in altitude you go, meaning that foods cook slower. For example, in Cleveland water typically boils at 212°F (100°C), but in Denver, the Mile-High City, water boils at about 200°F (93.3°C), which can significantly lengthen cooking times. Good cooks need to be aware of this.



A meat thermometer with a dial. Notice the markings for Fahrenheit (outer scale) and Celsius (inner scale) temperatures. Recipes for cooking food in an oven can use very different numbers, depending on the country you're in. (CC BY2.0 Bev Sykes)

At the other end is pressure cooking. A pressure cooker is a closed vessel that allows steam to build up additional pressure, which increases the temperature at which water boils. A good pressure cooker can get to temperatures as high as 252°F (122°C); at these temperatures, food cooks much faster than it normally would. Great care must be used with pressure cookers because of the high pressure and high temperature. (When a pressure cooker is used to sterilize medical instruments, it is called an *autoclave*.)

Other countries use the Celsius scale for everyday purposes. Therefore, oven dials in their kitchens are marked in degrees Celsius. It can be confusing for US cooks to use ovens abroad—a 425°F oven in the United States is equivalent to a 220°C oven in other countries. These days, many oven thermometers are marked with both temperature scales.

Key Takeaways

- Chemistry uses the Celsius and Kelvin scales to express temperatures.
- A temperature on the Kelvin scale is the Celsius temperature plus 273.15.
- The minimum possible temperature is absolute zero and is assigned 0 K on the Kelvin scale.
- Density relates the mass and volume of a substance.
- Density can be used to calculate volume from a given mass or mass from a given volume.

1.4: Other Units - Temperature and Density is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.6: Other Units - Temperature and Density** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

1.5: 0.1 Introduction and Review

If you drop your shoe and a coin side by side, they hit the ground at the same time. Why doesn't the shoe get there first, since gravity is pulling harder on it? How does the lens of your eye work, and why do your eye's muscles need to squash its lens into different shapes in order to focus on objects nearby or far away? These are the kinds of questions that physics tries to answer about the behavior of light and matter, the two things that the universe is made of.

The Scientific Method

Until very recently in history, no progress was made in answering questions like these. Worse than that, the *wrong* answers written by thinkers like the ancient Greek physicist Aristotle were accepted without question for thousands of years. Why is it that scientific knowledge has progressed more since the Renaissance than it had in all the preceding millennia since the beginning of recorded history? Undoubtedly the industrial revolution is part of the answer. Building its centerpiece, the steam engine, required improved techniques for precise construction and measurement. (Early on, it was considered a major advance when English machine shops learned to build pistons and cylinders that fit together with a gap narrower than the thickness of a penny.) But even before the industrial revolution, the pace of discovery had picked up, mainly because of the introduction of the modern scientific method. Although it evolved over time, most scientists today would agree on something like the following list of the basic principles of the scientific method:

(1) *Science is a cycle of theory and experiment.* Scientific theories are created to explain the results of experiments that were created under certain conditions. A successful theory will also make new predictions about new experiments under new conditions. Eventually, though, it always seems to happen that a new experiment comes along, showing that under certain conditions the theory is not a good approximation or is not valid at all. The ball is then back in the theorists' court. If an experiment disagrees with the current theory, the theory has to be changed, not the experiment.

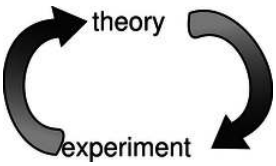


Figure 1.5.1: Science is a cycle of theory and experiment


Figure 1.5.2: A satirical drawing of an alchemist's laboratory. H. Cock, after a drawing by Peter Brueghel the Elder (16th century).

(2) *Theories should both predict and explain.* The requirement of predictive power means that a theory is only meaningful if it predicts something that can be checked against experimental measurements that the theorist did not already have at hand. That is, a theory should be testable. Explanatory value means that many phenomena should be accounted for with few basic principles. If you answer every “why” question with “because that's the way it is,” then your theory has no explanatory value. Collecting lots of data without being able to find any basic underlying principles is not science.

(3) *Experiments should be reproducible.* An experiment should be treated with suspicion if it only works for one person, or only in one part of the world. Anyone with the necessary skills and equipment should be able to get the same results from the same experiment. This implies that science transcends national and ethnic boundaries; you can be sure that nobody is doing actual science who claims that their work is “Aryan, not Jewish,” “Marxist, not bourgeois,” or “Christian, not atheistic.” An experiment cannot be reproduced if it is secret, so science is necessarily a public enterprise.

As an example of the cycle of theory and experiment, a vital step toward modern chemistry was the experimental observation that the chemical elements could not be transformed into each other, e.g., lead could not be turned into gold. This led to the theory that chemical reactions consisted of rearrangements of the elements in different combinations, without any change in the identities of the elements themselves. The theory worked for hundreds of years, and was confirmed experimentally over a wide range of pressures and temperatures and with many combinations of elements. Only in the twentieth century did we learn that one element could be trans-formed into one another under the conditions of extremely high pressure and temperature existing in a nuclear bomb or inside a star. That observation didn't completely invalidate the original theory of the immutability of the elements, but it showed that it was only an approximation, valid at ordinary temperatures and pressures.

self-check:

A psychic conducts seances in which the spirits of the dead speak to the participants. He says he has special psychic powers not possessed by other people, which allow him to “channel” the communications with the spirits. What part of the scientific method is being violated here?

The scientific method as described here is an idealization, and should not be understood as a set procedure for doing science. Scientists have as many weaknesses and character flaws as any other group, and it is very common for scientists to try to discredit other people's experiments when the results run contrary to their own favored point of view. Successful science also has more to do with luck, intuition, and creativity than most people realize, and the restrictions of the scientific method do not stifle individuality and self-expression any more than the fugue and sonata forms stifled Bach and Haydn. There is a recent tendency among social scientists to go even further and to deny that the scientific method even exists, claiming that science is no more than an arbitrary social system that determines what ideas to accept based on an in-group's criteria. I think that's going too far. If science is an arbitrary social ritual, it would seem difficult to explain its effectiveness in building such useful items as airplanes, CD players, and sewers. If alchemy and astrology were no less scientific in their methods than chemistry and astronomy, what was it that kept them from producing anything useful?

Discussion Questions

withintro{Consider whether or not the scientific method is being applied in the following examples. If the scientific method is not being applied, are the people whose actions are being described performing a useful human activity, albeit an unscientific one? }

◊ Acupuncture is a traditional medical technique of Asian origin in which small needles are inserted in the patient's body to relieve pain. Many doctors trained in the west consider acupuncture unworthy of experimental study because if it had therapeutic effects, such effects could not be explained by their theories of the nervous system. Who is being more scientific, the western or eastern practitioners?

◊ Goethe, a German poet, is less well known for his theory of color. He published a book on the subject, in which he argued that scientific apparatus for measuring and quantifying color, such as prisms, lenses and colored filters, could not give us full insight into the ultimate meaning of color, for instance the cold feeling evoked by blue and green or the heroic sentiments inspired by red. Was his work scientific?

◊ A child asks why things fall down, and an adult answers “because of gravity.” The ancient Greek philosopher Aristotle explained that rocks fell because it was their nature to seek out their natural place, in contact with the earth. Are these explanations scientific?

◊ Buddhism is partly a psychological explanation of human suffering, and psychology is of course a science. The Buddha could be said to have engaged in a cycle of theory and experiment, since he worked by trial and error, and even late in his life he asked his followers to challenge his ideas. Buddhism could also be considered reproducible, since the Buddha told his followers they could find enlightenment for themselves if they followed a certain course of study and discipline. Is Buddhism a scientific pursuit?

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

This page titled 1.5: 0.1 Introduction and Review is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

- **1.1: Introduction and Review** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

CHAPTER OVERVIEW

2: Chapter 2 - Introduction to Chemistry

2.1: Prelude to Chemistry
2.2: Basic Definitions

2: Chapter 2 - Introduction to Chemistry is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

2.1: Prelude to Chemistry

If you are reading these words, you are likely starting a chemistry course. Get ready for a fantastic journey through a world of wonder, delight, and knowledge. One of the themes of this book is "chemistry is everywhere," and indeed it is; you would not be alive if it were not for chemistry, because your body is a big chemical machine.



Figure **2.1.1** © Thinkstock

6 images showing a volcanic eruption, large dinner spread, explosion, red wine, a pool of large fish, and molecules are laid out together.

If you do not believe it, do not worry. Every chapter in this book contains examples that will show you how chemistry is, in fact, everywhere. So enjoy the ride, and enjoy chemistry.

2.1: Prelude to Chemistry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **1.1: Prelude to Chemistry** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

2.2: Basic Definitions

Learning Objective

- Learn the basic terms used to describe matter

The definition of chemistry—the study of the interactions of matter with other matter and with energy—uses some terms that should also be defined. We start the study of chemistry by defining basic terms.

Matter

Matter is anything that has mass and takes up space. A book is matter, a computer is matter, food is matter, and dirt in the ground is matter. Sometimes matter may be difficult to identify. For example, air is matter, but because it is so thin compared to other matter (e.g., a book, a computer, food, and dirt), we sometimes forget that air has mass and takes up space. Things that are not matter include thoughts, ideas, emotions, and hopes.

Example 2.2.1

Which of the following is matter and not matter?

a. a hot dog
b. love
c. a tree

Solution

a. A hot dog has mass and takes up space, so it is matter.
b. Love is an emotion, and emotions are not matter.
c. A tree has mass and takes up space, so it is matter.

Exercise 2.2.1

Which of the following is matter and not matter?

a. the moon
b. an idea for a new invention

Answer a

The moon is matter.

Answer b

The invention itself may be matter, but the idea for it is not.

To understand matter and how it changes, we need to be able to describe matter. There are two basic ways to describe matter: physical properties and chemical properties.

Physical properties

Physical properties are characteristics that describe matter as it exists. Some physical characteristics of matter are shape, color, size, and temperature. An important physical property is the **phase** (or **state**) of matter. The three fundamental phases of matter are solid, liquid, and gas (Figure 2.2.1).



Figure 2.2.1: The Phases of Matter. Chemistry recognizes three fundamental phases of matter: solid (left), liquid (middle), and gas (right). (CC BY-SA 3.0; Spirit469). The solid state depicted is an ice sculpture of an angel, the liquid state is a drop of water, and the gas is clouds made of water vapor.

Chemical Properties

Chemical properties are characteristics of matter that describe how matter changes form in the presence of other matter. Does a sample of matter burn? Burning is a chemical property. Does it behave violently when put in water? This reaction is a chemical property as well (Figure 2.2.2). In the following chapters, we will see how descriptions of physical and chemical properties are important aspects of chemistry.



Figure 2.2.2: Chemical Properties. The fact that this match burns is a chemical property of the match. (Sebastian Ritter (Rise0011)).

Physical Change

A physical change occurs when a sample of matter changes one or more of its physical properties. For example, a solid may melt (Figure 2.2.3), or alcohol in a thermometer may change volume as the temperature changes. A physical change does not affect the chemical composition of matter.



Figure 2.2.2: Physical Changes: The solid ice melts into liquid water—a physical change. A time-lapse animation of ice cubes melting in a glass over 50 minutes. (Public Domain; Moussa).

Chemical Change

Chemical change is the process of demonstrating a chemical property, such as the burning match in Figure 2.2.2 "Chemical Properties". As the matter in the match burns, its chemical composition changes, and new forms of matter with new physical properties are created. Note that chemical changes are frequently accompanied by physical changes, as the new matter will likely have different physical properties from the original matter.

Example 2.2.2

Describe each process as a physical change or a chemical change.

a. Water in the air turns into snow.
b. A person's hair is cut.
c. Bread dough becomes fresh bread in an oven.

Solution

a. Because the water is going from a gas phase to a solid phase, this is a physical change.
b. Your long hair is being shortened. This is a physical change.
c. Because of the oven's temperature, chemical changes are occurring in the bread dough to make fresh bread. These are chemical changes. (In fact, a lot of cooking involves chemical changes.)

Exercise 2.2.2

Identify each process as a physical change or a chemical change.

a. A fire is raging in a fireplace.
b. Water is warmed to make a cup of coffee.

Answer a

chemical change

Answer b

physical change

Substance

A sample of matter that has the same physical and chemical properties throughout is called a substance. Sometimes the phrase *pure substance* is used, but the word *pure* isn't needed. The definition of the term *substance* is an example of how chemistry has a specific definition for a word that is used in everyday language with a different, vaguer definition. Here, we will use the term *substance* with its strict chemical definition.

> *Chemistry recognizes two different types of substances: elements and compounds.*

Element

An element is the simplest type of chemical substance; it cannot be broken down into simpler chemical substances by ordinary chemical means. There are 118 elements known to science, of which 80 are stable. (The other elements are radioactive, a condition we will consider in Chapter 15.) Each element has its own unique set of physical and chemical properties. Examples of elements include iron, carbon, and gold.

Compound

A compound is a combination of more than one element. The physical and chemical properties of a compound are different from the physical and chemical properties of its constituent elements; that is, it behaves as a completely different substance. There are over 50 million compounds known, and more are being discovered daily. Examples of compounds include water, penicillin, and sodium chloride (the chemical name for common table salt).

Mixtures

Physical combinations of more than one substance are called mixtures. Elements and compounds are not the only ways in which matter can be present. We frequently encounter objects that are physical combinations of more than one element or compound—mixtures. There are two types of mixtures.

Heterogeneous Mixture

A heterogeneous mixture is a mixture composed of two or more substances. It is easy to tell, sometimes by the naked eye, that more than one substance is present.

Homogeneous Mixture/ Solution

A homogeneous mixture is a combination of two or more substances that is so intimately mixed, that the mixture behaves as a single substance. Another word for a homogeneous mixture is a solution. Thus, a combination of salt and steel wool is a heterogeneous mixture because it is easy to see which particles of the matter are salt crystals and which are steel wool. On the other hand, if you take salt crystals and dissolve them in water, it is very difficult to tell that you have more than one substance present just by looking—even if you use a powerful microscope. The salt dissolved in water is a homogeneous mixture, or a solution (Figure **2.2.3**).



Figure **2.2.3**: Types of Mixtures © Thinkstock. On the left, the combination of two substances is a heterogeneous mixture because the particles of the two components look different. On the right, the salt crystals have dissolved in the water so finely that you cannot tell that salt is present. The homogeneous mixture appears like a single substance.

✔ Example 2.2.3

Identify the following combinations as heterogeneous mixtures or homogenous mixtures.

a. soda water (carbon dioxide is dissolved in water)
b. a mixture of iron metal filings and sulfur powder (both iron and sulfur are elements)

Solution

a. Because carbon dioxide is dissolved in water, we can infer from the behavior of salt crystals dissolved in water that carbon dioxide dissolved in water is (also) a homogeneous mixture.
b. Assuming that the iron and sulfur are simply mixed together, it should be easy to see what is iron and what is sulfur, so this is a heterogeneous mixture.

? Exercise 2.2.3

a. the human body
b. an amalgam, a combination of some other metals dissolved in a small amount of mercury

Answer a

heterogeneous mixture

Answer b

homogeneous mixture

There are other descriptors that we can use to describe matter, especially elements. We can usually divide elements into metals and nonmetals, and each set shares certain (but not always all) properties.

Metal

A metal is an element that conducts electricity and heat well and is shiny, silvery, solid, ductile, and malleable. At room temperature, metals are solid (although mercury is a well-known exception). A metal is ductile because it can be drawn into thin wires (a property called *ductility*); and malleable because it can be pounded into thin sheets (a property called *malleability*).

Nonmetal

A non-metal is an element that is brittle when solid, and does not conduct electricity or heat very well. Non-metals cannot be made into thin sheets or wires (Figure **2.2.4**). Nonmetals also exist in a variety of phases and colors at room temperature.

Semi-metals

Some elements have properties of both metals and nonmetals and are called semi-metals (or metalloids). We will see later how these descriptions can be assigned rather easily to various elements.



Figure **2.2.4**: Semimetals © Thinkstock. On the left is some elemental mercury, the only metal that exists as a liquid at room temperature. It has all the other expected properties of a metal. On the right, elemental sulfur is a yellow nonmetal that usually is found as a powder.

Describing Matter Flowchart

"Describing Matter" is a flowchart of the relationships among the different ways of describing matter.

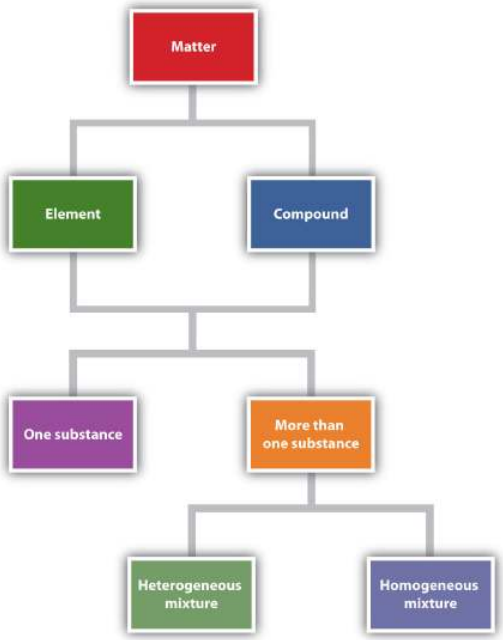


Figure **2.2.5**: Describing Matter. This flowchart shows how matter can be described. Matter forks into element & compound which fork into 1 substance & more than 1 substance. More than 1 substance is homogenous or heterogenous.

Example **2.2.1**: Chemistry is Everywhere: In the Morning

Most people have a morning ritual, a process that they go through every morning to get ready for the day. Chemistry appears in many of these activities.

- If you take a shower or bath in the morning, you probably use soap, shampoo, or both. These items contain chemicals that interact with the oil and dirt on your body and hair to remove them and wash them away. Many of these products also contain chemicals that make you smell good; they are called *fragrances*.
- When you brush your teeth in the morning, you usually use toothpaste, a form of soap, to clean your teeth. Toothpastes typically contain tiny, hard particles called *abrasives* that physically scrub your teeth. Many toothpastes also contain fluoride, a substance that chemically interacts with the surface of the teeth to help prevent cavities.
- Perhaps you take vitamins, supplements, or medicines every morning. Vitamins and other supplements contain chemicals your body needs in small amounts to function properly. Medicines are chemicals that help combat diseases and promote health.
- Perhaps you make some fried eggs for breakfast. Frying eggs involves heating them enough so that a chemical reaction occurs to cook the eggs.
- After you eat, the food in your stomach is chemically reacted so that the body (mostly the intestines) can absorb food, water, and other nutrients.
- If you drive or take the bus to school or work, you are using a vehicle that probably burns gasoline, a material that burns fairly easily and provides energy to power the vehicle. Recall that burning is a chemical change.

These are just a few examples of how chemistry impacts your everyday life. And we haven't even made it to lunch yet!



Figure 2.2.6: Chemistry in Real Life © Thinkstock. Examples of chemistry can be found everywhere—in personal hygiene products, food, and motor vehicles. Personal hygiene products (left), food (middle), and motor vehicles (right) images next to one another.

Key Takeaways

- Chemistry is the study of matter and its interactions with other matter and energy.
- Matter is anything that has mass and takes up space.
- Matter can be described in terms of physical properties and chemical properties.
- Physical properties and chemical properties of matter can change.
- Matter is composed of elements and compounds.
- Combinations of different substances are called mixtures.
- Elements can be described as metals, nonmetals, and semi-metals.

2.2: Basic Definitions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **1.2: Basic Definitions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

CHAPTER OVERVIEW

3: Chapter 3 - Atoms and Elemental Properties

3.1: Prelude to Atoms, Molecules, and Ions

3.2: Atomic Theory

3.3: Masses of Atoms and Molecules

3.4: Organization of the Elements - The Periodic Table

3.5: Ions and Ionic Compounds

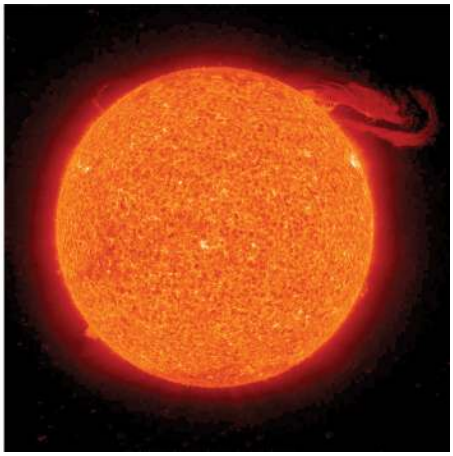
3: Chapter 3 - Atoms and Elemental Properties is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

3.1: Prelude to Atoms, Molecules, and Ions

Although not an SI unit, the angstrom (Å) is a useful unit of length. It is one ten-billionth of a meter, or 10^{-10} m. Why is it a useful unit? The ultimate particles that compose all matter are about 10^{-10} m in size, or about 1 Å. This makes the angstrom a natural—though not approved—unit for describing these particles.

The angstrom unit is named after Anders Jonas Ångström, a nineteenth-century Swedish physicist. Ångström's research dealt with light being emitted by glowing objects, including the sun. Ångström studied the brightness of the different colors of light that the sun emitted and was able to deduce that the sun is composed of the same kinds of matter that are present on the earth. By extension, we now know that all matter throughout the universe is similar to the matter that exists on our own planet.





Anders Jonas Ångstrom, a Swedish physicist, studied the light coming from the sun. His contributions to science were sufficient to have a tiny unit of length named after him, the angstrom, which is one ten-billionth of a meter.

Source: Photo of the sun courtesy of NASA's Solar Dynamics Observatory.

3.1: Prelude to Atoms, Molecules, and Ions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.1: Prelude to Atoms, Molecules, and Ions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

3.2: Atomic Theory

Learning Objectives

- State the modern atomic theory.
- Learn how atoms are constructed.

The smallest piece of an element that maintains the identity of that element is called an atom. Individual atoms are extremely small. It would take about fifty million atoms in a row to make a line that is 1 cm long. The period at the end of a printed sentence has several million atoms in it. Atoms are so small that it is difficult to believe that all matter is made from atoms—but it is.

The concept that atoms play a fundamental role in chemistry is formalized by the modern atomic theory, first stated by John Dalton, an English scientist, in 1808. It consists of three parts:

1. All matter is composed of atoms.
2. Atoms of the same element are the same; atoms of different elements are different.
3. Atoms combine in whole-number ratios to form compounds.

These concepts form the basis of chemistry. Although the word *atom* comes from a Greek word that means "indivisible," we understand now that atoms themselves are composed of smaller parts called *subatomic particles*. The first part to be discovered was the electron, a tiny subatomic particle with a negative charge. It is often represented as e^-, with the right superscript showing the negative charge. Later, two larger particles were discovered. The proton is a more massive (but still tiny) subatomic particle with a positive charge, represented as p^+. The neutron is a subatomic particle with about the same mass as a proton, but no charge. It is represented as either n or n^0. We now know that all atoms of all elements are composed of electrons, protons, and (with one exception) neutrons. Table **3.2.1** summarizes the properties of these three subatomic particles.

Table **3.2.1**: Properties of the Three Subatomic Particles

Name	Symbol	Mass (approx.; kg)	Charge
Proton	p^+	1.6×10^{-27}	1+
Neutron	n, n^0	1.6×10^{-27}	none
Electron	e^-	9.1×10^{-31}	1−

How are these particles arranged in atoms? They are not arranged at random. Experiments by Ernest Rutherford in England in the 1910s pointed to a nuclear model with atoms that has the protons and neutrons in a central nucleus with the electrons in orbit about the nucleus. The relatively massive protons and neutrons are collected in the center of an atom, in a region called the nucleus of the atom (plural *nuclei*). The electrons are outside the nucleus and spend their time orbiting in space about the nucleus. (Figure **3.2.1**).

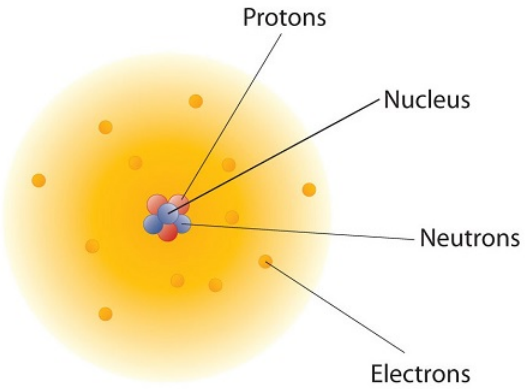


Figure **3.2.1**: The Structure of the Atom. Atoms have protons and neutrons in the center, making the nucleus, while the electrons orbit the nucleus.

The modern atomic theory states that atoms of one element are the same, while atoms of different elements are different. What makes atoms of different elements different? The fundamental characteristic that all atoms of the same element share is the *number of protons*. All atoms of hydrogen have one and only one proton in the nucleus; all atoms of iron have 26 protons in the nucleus. This number of protons is so important to the identity of an atom that it is called the atomic number. The number of protons in an atom is the atomic number of the element. Thus, hydrogen has an atomic number of 1, while iron has an atomic number of 26. Each element has its own characteristic atomic number.

Atoms of the same element can have different numbers of neutrons, however. Atoms of the same element (i.e., atoms with the same number of protons) with different numbers of neutrons are called isotopes. Most naturally occurring elements exist as isotopes. For example, most hydrogen atoms have a single proton in their nucleus. However, a small number (about one in a million) of hydrogen atoms have a proton and a neutron in their nuclei. This particular isotope of hydrogen is called deuterium. A very rare form of hydrogen has one proton and two neutrons in the nucleus; this isotope of hydrogen is called tritium. The sum of the number of protons and neutrons in the nucleus is called the mass number of the isotope.

Neutral atoms have the same number of electrons as they have protons, so their overall charge is zero. However, as we shall see later, this will not always be the case.

✔ Example 3.2.1

a. The most common carbon atoms have six protons and six neutrons in their nuclei. What are the atomic number and the mass number of these carbon atoms?
b. An isotope of uranium has an atomic number of 92 and a mass number of 235. What are the number of protons and neutrons in the nucleus of this atom?

Solution

a. If a carbon atom has six protons in its nucleus, its atomic number is 6. If it also has six neutrons in the nucleus, then the mass number is 6 + 6, or 12.
b. If the atomic number of uranium is 92, then that is the number of protons in the nucleus. Because the mass number is 235, then the number of neutrons in the nucleus is 235 − 92, or 143.

? Exercise 3.2.1

The number of protons in the nucleus of a tin atom is 50, while the number of neutrons in the nucleus is 68. What are the atomic number and the mass number of this isotope?

Answer

Atomic number = 50, mass number = 118

When referring to an atom, we simply use the element's name: the term *sodium* refers to the element as well as an atom of sodium. But it can be unwieldy to use the name of elements all the time. Instead, chemistry defines a symbol for each element. The atomic symbol is a one- or two-letter representation of the name of an element. By convention, the first letter of an element's symbol is always capitalized, while the second letter (if present) is lowercase. Thus, the symbol for hydrogen is H, the symbol for sodium is Na, and the symbol for nickel is Ni. Most symbols come from the English name of the element, although some symbols come from an element's Latin name. (The symbol for sodium, Na, comes from its Latin name, *natrium*.) Table **3.2.2** lists some common elements and their symbols. You should memorize the symbols in Table **3.2.2**, as this is how we will be representing elements throughout chemistry.

Table **3.2.2**: Names and Symbols of Common Elements

Element Name	Symbol	Element Name	Symbol
Aluminum	Al	Mercury	Hg
Argon	Ar	Molybdenum	Mo
Arsenic	As	Neon	Ne
Barium	Ba	Nickel	Ni
Beryllium	Be	Nitrogen	N
Bismuth	Bi	Oxygen	O
Boron	B	Palladium	Pd
Bromine	Br	Phosphorus	P
Calcium	Ca	Platinum	Pt
Carbon	C	Potassium	K
Chlorine	Cl	Radium	Ra
Chromium	Cr	Radon	Rn
Cobalt	Co	Rubidium	Rb
Copper	Cu	Scandium	Sc
Fluorine	F	Selenium	Se
Gallium	Ga	Silicon	Si
Germanium	Ge	Silver	Ag
Gold	Au	Sodium	Na
Helium	He	Strontium	Sr
Hydrogen	H	Sulfur	S
Iodine	I	Tantalum	Ta
Iridium	Ir	Tin	Sn
Iron	Fe	Titanium	Ti
Krypton	Kr	Tungsten	W
Lead	Pb	Uranium	U
Lithium	Li	Xenon	Xe
Magnesium	Mg	Zinc	Zn
Manganese	Mn	Zirconium	Zr

The elements are grouped together in a special chart called the periodic table of all the elements. A simple periodic table is shown in Figure **3.2.2**, while one may view a more extensive periodic table from another source. The elements on the periodic table are listed in order of ascending atomic number. The periodic table has a special shape that will become important to us when we consider the organization of electrons in atoms (Chapter 8). One immediate use of the periodic table helps us identify metals and nonmetals. Nonmetals are in the upper right-hand corner of the periodic table, on one side of the heavy line splitting the right-side part of the chart. All other elements are metals.

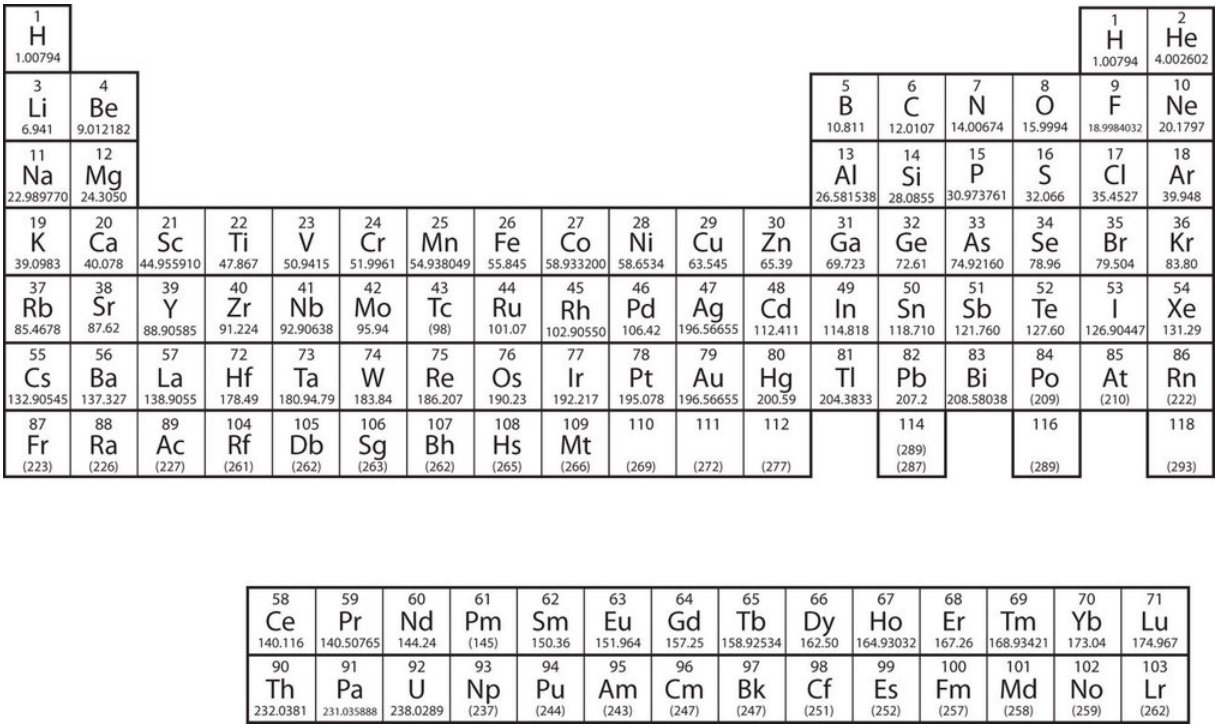


Figure 3.2.1: A Simple Periodic Table

There is an easy way to represent isotopes using the atomic symbols. We use the construction:

$$^{A}_{Z}X$$

where X is the symbol of the element, A is the mass number, and Z is the atomic number. Thus, for the isotope of carbon that has 6 protons and 6 neutrons, the symbol is:

$$^{12}_{6}C$$

where C is the symbol for the element, 6 represents the atomic number, and 12 represents the mass number.

Example 3.2.2

a. What is the symbol for an isotope of uranium that has an atomic number of 92 and a mass number of 235?

b. How many protons and neutrons are in $^{56}_{26}Fe$

Solution

a. The symbol for this isotope is $^{235}_{92}U$

b. This iron atom has 26 protons and 56 − 26 = 30 neutrons.

Exercise 3.2.2

How many protons are in $^{23}_{11}Na$

Answer

11 protons

It is also common to state the mass number after the name of an element to indicate a particular isotope. *Carbon-12* represents an isotope of carbon with 6 protons and 6 neutrons, while *uranium-238* is an isotope of uranium that has 146 neutrons.

Key Takeaways

- Chemistry is based on the modern atomic theory, which states that all matter is composed of atoms.
- Atoms themselves are composed of protons, neutrons, and electrons.
- Each element has its own atomic number, which is equal to the number of protons in its nucleus.
- Isotopes of an element contain different numbers of neutrons.
- Elements are represented by an atomic symbol.
- The periodic table is a chart that organizes all the elements.

3.2: Atomic Theory is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.2: Atomic Theory** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

3.3: Masses of Atoms and Molecules

Learning Objective

- Express the masses of atoms and molecules.

Because matter is defined as anything that has mass and takes up space, it should not be surprising to learn that atoms and molecules have mass.

Individual atoms and molecules, however, are very small, and the masses of individual atoms and molecules are also very small. For macroscopic objects, we use units such as grams and kilograms to state their masses, but these units are much too big to comfortably describe the masses of individual atoms and molecules. Another scale is needed.

Atomic Mass Unit

The atomic mass unit (u; some texts use amu, but this older style is no longer accepted) is defined as one-twelfth of the mass of a carbon-12 atom, an isotope of carbon that has six protons and six neutrons in its nucleus. By this scale, the mass of a proton is 1.00728 u, the mass of a neutron is 1.00866 u, and the mass of an electron is 0.000549 u. There will not be much error if you estimate the mass of an atom by simply counting the total number of protons and neutrons in the nucleus (i.e., identify its mass number) and ignore the electrons. Thus, the mass of carbon-12 is about 12 u, the mass of oxygen-16 is about 16 u, and the mass of uranium-238 is about 238 u. More exact masses are found in scientific references—for example, the exact mass of uranium-238 is 238.050788 u, so you can see that we are not far off by using the whole-number value as the mass of the atom.

What is the mass of an element? This is somewhat more complicated because most elements exist as a mixture of isotopes, each of which has its own mass. Thus, although it is easy to speak of the mass of an atom, when talking about the mass of an element, we must take the isotopic mixture into account.

Atomic Mass

The atomic mass of an element is a weighted average of the masses of the isotopes that compose an element. What do we mean by a weighted average? Well, consider an element that consists of two isotopes, 50% with mass 10 u and 50% with mass 11 u. A weighted average is found by multiplying each mass by its fractional occurrence (in decimal form) and then adding all the products. The sum is the weighted average and serves as the formal atomic mass of the element. In this example, we have the following:

0.50 × 10 u	= 5.0 u
0.50 × 11 u	= 5.5 u
Sum	= 10.5 u = the atomic mass of our element

Note that no atom in our hypothetical element has a mass of 10.5 u; rather, that is the average mass of the atoms, weighted by their percent occurrence.

This example is similar to a real element. Boron exists as about 20% boron-10 (five protons and five neutrons in the nuclei) and about 80% boron-11 (five protons and six neutrons in the nuclei). The atomic mass of boron is calculated similarly to what we did for our hypothetical example, but the percentages are different:

0.20 × 10 u	= 2.0 u
0.80 × 11 u	= 8.8 u
Sum	= 10.8 u = the atomic mass of boron

Thus, we use 10.8 u for the atomic mass of boron.

Virtually all elements exist as mixtures of isotopes, so atomic masses may vary significantly from whole numbers. Table **3.3.1** lists the atomic masses of some elements. The atomic masses in Table **3.3.1** are listed to three decimal places where possible, but in most cases, only one or two decimal places are needed. Note that many of the atomic masses, especially the larger ones, are not very close to whole numbers. This is, in part, the effect of an increasing number of isotopes as the atoms increase in size. (The record number is 10 isotopes for tin.)

Table **3.3.1**: Selected Atomic Masses of Some Elements

Element Name	Atomic Mass (u)	Element Name	Atomic Mass (u)
Aluminum	26.981	Molybdenum	95.94
Argon	39.948	Neon	20.180
Arsenic	74.922	Nickel	58.693
Barium	137.327	Nitrogen	14.007
Beryllium	9.012	Oxygen	15.999
Bismuth	208.980	Palladium	106.42
Boron	10.811	Phosphorus	30.974
Bromine	79.904	Platinum	195.084
Calcium	40.078	Potassium	39.098
Carbon	12.011	Radium	n/a
Chlorine	35.453	Radon	n/a
Cobalt	58.933	Rubidium	85.468
Copper	63.546	Scandium	44.956
Fluorine	18.998	Selenium	78.96
Gallium	69.723	Silicon	28.086
Germanium	72.64	Silver	107.868
Gold	196.967	Sodium	22.990
Helium	4.003	Strontium	87.62
Hydrogen	1.008	Sulfur	32.065
Iodine	126.904	Tantalum	180.948
Iridium	192.217	Tin	118.710
Iron	55.845	Titanium	47.867
Krypton	83.798	Tungsten	183.84
Lead	207.2	Uranium	238.029
Lithium	6.941	Xenon	131.293
Magnesium	24.305	Zinc	65.409
Manganese	54.938	Zirconium	91.224
Mercury	200.59	Molybdenum	95.94

Note: Atomic mass is given to three decimal places, if known.

Molecular Mass

Now that we understand that atoms have mass, it is easy to extend the concept to the mass of molecules. The molecular mass is the sum of the masses of the atoms in a molecule. This may seem like a trivial extension of the concept, but it is important to count the number of each type of atom in the molecular formula. Also, although each atom in a molecule is a particular isotope, we use the weighted average, or atomic mass, for each atom in the molecule.

For example, if we were to determine the molecular mass of dinitrogen trioxide, N_2O_3, we would need to add the atomic mass of nitrogen two times with the atomic mass of oxygen three times:

2 N masses = 2 × 14.007 u	= 28.014 u
3 O masses = 3 × 15.999 u	= 47.997 u
Total	= 76.011 u = the molecular mass of N_2O_3

We would not be far off if we limited our numbers to one or even two decimal places.

Example 3.3.1

What is the molecular mass of each substance?

a. NBr_3
b. C_2H_6

Solution

Add one atomic mass of nitrogen and three atomic masses of bromine:

Solutions to Example 3.4.1

1 N mass	= 14.007 u
3 Br masses = 3 × 79.904 u	= 239.712 u
Total	= 253.719 u = the molecular mass of NBr_3

Add two atomic masses of carbon and six atomic masses of hydrogen:

Solutions to Example 3.4.1

2 C masses = 2 × 12.011 u	= 24.022 u
6 H masses = 6 × 1.008 u	= 6.048 u
Total	= 30.070 u = the molecular mass of C_2H_6

The compound C_2H_6 also has a common name—ethane.

Exercise 3.3.1

What is the molecular mass of each substance?

a. SO_2
b. Extra close brace or missing open brace

Answer a

64.063 u

Answer b

87.968 u

https://chem.libretexts.org/@go/page/200113

Chemistry is Everywhere: Sulfur Hexafluoride

On March 20, 1995, the Japanese terrorist group Aum Shinrikyo (Sanskrit for "Supreme Truth") released some sarin gas in the Tokyo subway system; twelve people were killed, and thousands were injured (Figure **3.3.2*a***). Sarin (molecular formula $C_4H_{10}FPO_2$) is a nerve toxin that was first synthesized in 1938. It is regarded as one of the most deadly toxins known, estimated to be about 500 times more potent than cyanide. Scientists and engineers who study the spread of chemical weapons such as sarin (yes, there are such scientists) would like to have a similar less dangerous chemical to study, indeed one that is nontoxic, so they are not at risk themselves.

Sulfur hexafluoride is used as a model compound for sarin. SF_6 (a molecular model of which is shown Figure **3.3.2*b***) has a similar molecular mass (about 146 u) as sarin (about 140 u), so it has similar physical properties in the vapor phase. Sulfur hexafluoride is also very easy to accurately detect, even at low levels, and it is not a normal part of the atmosphere, so there is little potential for contamination from natural sources. Consequently, SF_6 is also used as an aerial tracer for ventilation systems in buildings. It is nontoxic and very chemically inert, so workers do not have to take special precautions other than watching for asphyxiation.

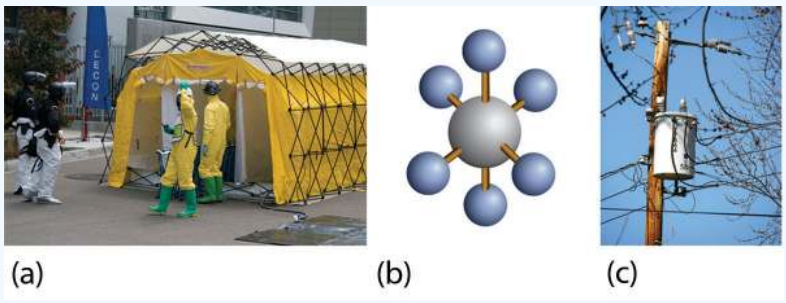


Figure **3.3.2**: Sarin and Sulfur Hexafluoride © Thinkstock. (a) Properly protected workers clear out the Tokyo subway after the nerve toxin sarin was released. (b) A molecular model of SF_6. (c) A high-voltage electrical switchgear assembly that would be filled with SF_6 as a spark suppressant.

Sulfur hexafluoride also has another interesting use: a spark suppressant in high-voltage electrical equipment. High-pressure SF_6 gas is used in place of older oils that may have contaminants that are not environmentally friendly (Figure **3.3.2*c***).

Key Takeaways

- The atomic mass unit (u) is a unit that describes the masses of individual atoms and molecules.
- The atomic mass is the weighted average of the masses of all isotopes of an element.
- The molecular mass is the sum of the masses of the atoms in a molecule.

3.3: Masses of Atoms and Molecules is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.4: Masses of Atoms and Molecules** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

3.4: Organization of the Elements - The Periodic Table

The number of protons in the nucleus of an element is called the atomic number of that element. Chemists typically place elements in order of increasing atomic numbers in a special arrangement that is called the periodic table.

The periodic table is not simply a grid of elements arranged numerically. In the periodic table, the elements are arranged in horizontal rows called periods (numbered in blue) and vertically into columns called groups. These groups are numbered by two, somewhat conflicting, schemes. In the simplest presentation, favored by the International Union of Pure and Applied Chemistry (**IUPAC**), the groups are simply numbered 1-18. The convention in much of the world, however, is to number the first two groups 1A and 2A, the last six groups 3A-8A; the middle ten groups are then numbered 1B-8B (but not in that order!). While the IUPAC numbering appears much simpler, in this text we will use the current USA nomenclature (1A-8A). The reason for this choice will become more apparent in Chapter 3 when we discuss "valence" and electron configuration in more detail. The actual layout of the periodic table is based on the grouping of the elements according to chemical properties. For example, elements in each Group of the periodic table (each vertical column) will share many of the same chemical properties. As we discuss the properties of elements and the ways they combine with other elements, the reasons for this particular arrangement of the periodic table will become more obvious.

As you can see, each element in the periodic table is represented by a box containing the chemical symbol, the atomic number (the number of protons in the nucleus) and the atomic mass of the element. Remember that the atomic mass is the weighted average of the masses of all of the natural isotopes of the particular element.

Periodic tables are often colored, or shaded, to distinguish groups of elements that have similar properties or chemical reactivity. The broadest classification is into metals, metalloids (or semi-metals) and nonmetals. The elements in Groups 1A – 8A are called the representative elements and the elements in Groups 3 - 15 are called the transition metals. the metallic elements are shown in purple. Metals are solids (except for mercury), can conduct electricity and are usually malleable (can be rolled into sheets) and are ductile (can be drawn into wires). Metals are usually separated into the main group metals (the elements colored purple in Groups 1A - 5A) and the transition metals (in Groups 3 – 15). Nonmetals (yellow in the Figure) do not conduct electricity (with the exception of carbon in the form of graphite) and have a variety of physical states (some are solids, some liquids and some gasses). Two important subclasses of nonmetals are the halogens (Group 7A) and the inert gasses (or noble gasses; Group 8A). At the border between metals and nonmetals lie the elements boron, silicon, germanium, astatine, antimony and tellurium. These elements share physical properties of metals and nonmetals and are called metalloids, or semi-metals. The common semiconductors silicon and germanium are in this group and it is their unique electrical properties that make transistors and other solid-state devices possible. Later in this book we will see that the position of elements in the periodic table also correlates with their chemical reactivity.

3.4: Organization of the Elements - The Periodic Table is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **1.2: Organization of the Elements - The Periodic Table** by Paul R. Young is licensed CC BY-SA 4.0. Original source: https://en.wikibooks.org/wiki/Introductory_Chemistry_Online.

3.5: Ions and Ionic Compounds

Learning Objectives

- Know how ions form.
- Learn the characteristic charges that ions have.
- Construct a proper formula for an ionic compound.
- Generate a proper name for an ionic compound.

So far, we have discussed elements and compounds that are electrically neutral. They have the same number of electrons as protons, so the negative charges of the electrons are balanced by the positive charges of the protons. However, this is not always the case. Electrons can move from one atom to another; when they do, species with overall electric charges are formed. Such species are called ions. Species with overall positive charges are termed **cations**, while species with overall negative charges are called **anions**. Remember that ions are formed only when *electrons* move from one atom to another; a proton never moves from one atom to another. Compounds formed from positive and negative ions are **ionic compounds**.

Individual atoms can gain or lose electrons. When they do, they become *monatomic* ions. When atoms gain or lose electrons, they usually gain or lose a characteristic number of electrons and so take on a characteristic overall charge. Table **3.5.1** lists some common ions in terms of how many electrons they lose (making cations) or gain (making anions). There are several things to notice about the ions in Table **3.5.1**. First, each element that forms a cation is a metal, except for one (hydrogen), while each element that forms an anion is a nonmetal. This is actually one of the chemical properties of metals and nonmetals: metals tend to form cations, while nonmetals tend to form anions. Second, most atoms form ions of a single characteristic charge. When sodium atoms form ions, they always form a 1+ charge, never a 2+ or 3+ or even 1− charge. Thus, if you commit the information in Table **3.5.1** to memory, you will always know what charges most atoms form.

Table **3.5.1**: Monatomic Ions of Various Charges

Ions formed by losing a single electron	H^+
Na^+	
K^+	
Rb^+	
Ag^+	
Au^+	
Ions formed by losing two electrons	Mg^{2+}
	Ca^{2+}
	Sr^{2+}
	Fe^{2+}
	Co^{2+}
	Ni^{2+}
	Cu^{2+}
	Zn^{2+}
	Sn^{2+}
	Hg^{2+}
	Pb^{2+}
Ions formed by losing three electrons	Sc^{3+}
	Fe^{3+}
	Co^{3+}
	Ni^{3+}
	Au^{3+}
	Al^{3+}
	Cr^{3+}
Ions formed by losing four electrons	Ti^{4+}
	Sn^{4+}
	Pb^{4+}
Ions formed by gaining a single electron	F^-
	Cl^-
	Br^-
	I^-
Ions formed by gaining two electrons	O^{2-}
	S^{2-}
	Se^{2-}
Ions formed by gaining three electrons	N^{3-}
	P^{3-}

Third, there are some exceptions to the previous point. A few elements, all of which are metals, can form more than one possible charge. For example, iron atoms can form 2+ cations or 3+ cations. Cobalt is another element that can form more than one possible charged ion (2+ and 3+), while lead can form 2+ or 4+ cations. Unfortunately, there is little understanding which two charges a metal atom may take, so it is best to just memorize the possible charges a particular element can have.

Note the convention for indicating an ion. The magnitude of the charge is listed as a right superscript next to the symbol of the element. If the charge is a single positive or negative one, the number 1 is not written; if the magnitude of the charge is greater than 1, then the number is written *before* the + or − sign. An element symbol without a charge written next to it is assumed to be the uncharged atom.

Naming an ion is straightforward. For a cation, simply use the name of the element and add the word *ion* (or if you want to be more specific, add *cation*) after the element's name. So Na^+ is the sodium ion; Ca^{2+} is the calcium ion. If the element has more than one possible charge, the value of the charge comes after the element name and before the word *ion*. Thus, Fe^{2+} is the iron two ion, while Fe^{3+} is the iron three ion. In print, we use roman numerals in parentheses to represent the charge on the ion; so these two iron ions would be represented as the iron(II) cation and the iron(III) cation, respectively.

For a monatomic anion, use the stem of the element name and append the suffix *-ide* to it, and then add *ion*. This is similar to how we named molecular compounds. Thus, Cl^- is the chloride ion, and N^{3-} is the nitride ion.

Example 3.5.1

Name each species.

a. O^{2-}
b. Co
c. Co^{2+}

Solution

a. This species has a 2− charge on it, so it is an anion. Anions are named using the stem of the element name with the suffix *-ide* added. This is the oxide anion.
b. Because this species has no charge, it is an atom in its elemental form. This is cobalt.
c. In this case, there is a 2+ charge on the atom, so it is a cation. We note from Table **3.5.1** that cobalt cations can have two possible charges, so the name of the ion must specify which charge the ion has. This is the cobalt(II) cation.

Exercise 3.5.1

Name each species.

a. P^{3-}
b. Sr^{2+}

Answers

a. the phosphide anion
b. the strontium cation

Ionic Formulas

Chemical formulas for ionic compounds are called ionic formulas. A proper ionic formula has a cation and an anion in it; an ionic compound is never formed between two cations only or two anions only. The key to writing proper ionic formulas is simple: the total positive charge must balance the total negative charge. Because the charges on the ions are characteristic, sometimes we have to have more than one of a cation or an anion to balance the overall positive and negative charges. It is conventional to use the lowest ratio of ions that are needed to balance the charges.

For example, consider the ionic compound between Na^+ and Cl^-. Each ion has a single charge, one positive and one negative, so we need only one ion of each to balance the overall charge. When writing the ionic formula, we follow two additional conventions: (1) write the formula for the cation first and the formula for the anion second, but (2) do not write the charges on the ions. Thus, for

the compound between Na^+ and Cl^-, we have the ionic formula NaCl (Figure **3.5.1**). The formula Na_2Cl_2 also has balanced charges, but the convention is to use the lowest ratio of ions, which would be one of each. (Remember from our conventions for writing formulas that we do not write a 1 subscript if there is only one atom of a particular element present.) For the ionic compound between magnesium cations (Mg^{2+}) and oxide anions (O^{2-}), again we need only one of each ion to balance the charges. By convention, the formula is MgO.



Figure **3.5.1**: The ionic compound **NaCl** is table salt and is very common. (Public Domain; Pexels.)

For the ionic compound between Mg^{2+} ions and Cl^- ions, we now consider the fact that the charges have different magnitudes: 2+ on the magnesium ion and 1− on the chloride ion. To balance the charges with the lowest number of ions possible, we need to have two chloride ions to balance the charge on the one magnesium ion. Rather than write the formula MgClCl, we combine the two chloride ions and write it with a 2 subscript: $MgCl_2$.

What is the formula $MgCl_2$ telling us? There are two chloride ions in the formula. Although chlorine as an element is a diatomic molecule, Cl_2, elemental chlorine is not part of this ionic compound. The chlorine is in the form of a negatively charged *ion*, not the neutral *element*. The 2 subscript is in the ionic formula because we need two Cl^- ions to balance the charge on one Mg^{2+} ion.

✔ Example 3.5.2

Write the proper ionic formula for each of the two given ions.

a. Ca^{2+} and Cl^-
b. Al^{3+} and F^-
c. Al^{3+} and O^{2-}

Solution

a. We need two Cl^- ions to balance the charge on one Ca^{2+} ion, so the proper ionic formula is $CaCl_2$.
b. We need three F^- ions to balance the charge on the Al^{3+} ion, so the proper ionic formula is AlF_3.
c. With Al^{3+} and O^{2-}, note that neither charge is a perfect multiple of the other. This means we have to go to a least common multiple, which in this case will be six. To get a total of 6+, we need two Al^{3+} ions; to get 6−, we need three O^{2-} ions. Hence the proper ionic formula is Al_2O_3.

? Exercise 3.5.2

Write the proper ionic formulas for each of the two given ions.

a. Fe^{2+} and S^{2-}
b. Fe^{3+} and S^{2-}

Answers

a. FeS
b. Fe_2S_3

Naming ionic compounds is simple: combine the name of the cation and the name of the anion, in both cases omitting the word *ion*. *Do not use numerical prefixes if there is more than one ion necessary to balance the charges.* NaCl is sodium chloride, a combination of the name of the cation (sodium) and the anion (chloride). MgO is magnesium oxide. $MgCl_2$ is magnesium chloride—*not* magnesium dichloride.

When naming ionic compounds whose cations can have more than one possible charge, we must also include the charge, in parentheses and in roman numerals, as part of the name. Hence FeS is iron(II) sulfide, while Fe_2S_3 is iron(III) sulfide. Again, no numerical prefixes appear in the name. The number of ions in the formula is dictated by the need to balance the positive and negative charges.

Example 3.5.3

Name each ionic compound.

a. $CaCl_2$
b. AlF_3
c. Co_2O_3

Solution

a. Using the names of the ions, this ionic compound is named calcium chloride. It is *not* calcium(II) chloride, because calcium forms only one cation when it forms an ion, and it has a characteristic charge of 2+.
b. The name of this ionic compound is aluminum fluoride.
c. We know that cobalt can have more than one possible charge; we just need to determine what it is. Oxide always has a 2− charge, so with three oxide ions, we have a total negative charge of 6−. This means that the two cobalt ions have to contribute 6+, which for two cobalt ions means that each one is 3+. Therefore, the proper name for this ionic compound is cobalt(III) oxide.

Exercise 3.5.3

Name each ionic compound.

a. Sc_2O_3
b. AgCl

Answers

a. scandium oxide
b. silver chloride

How do you know whether a formula—and by extension, a name—is for a molecular compound or for an ionic compound? Molecular compounds form between nonmetals and nonmetals, while ionic compounds form between metals and nonmetals. The periodic table can be used to determine which elements are metals and nonmetals.

There also exists a group of ions that contain more than one atom. These are called **polyatomic ions**. Table 3.5.2 lists the formulas, charges, and names of some common polyatomic ions. Only one of them, the ammonium ion, is a cation; the rest are anions. Most of them also contain oxygen atoms, so sometimes they are referred to as ***oxyanions***. Some of them, such as nitrate and nitrite, and sulfate and sulfite, have very similar formulas and names, so care must be taken to get the formulas and names correct. Note that the -ite polyatomic ion has one less oxygen atom in its formula than the -ate ion but with the same ionic charge.

Table 3.5.2: Common Polyatomic Ions

Name	Formula and Charge		Name	Formula and Charge
ammonium	NH_4^+		hydroxide	OH^-
acetate	$C_2H_3O_2^-$, or CH_3COO^-		nitrate	NO_3^-
bicarbonate (hydrogen carbonate)	HCO_3^-		nitrite	NO_2^-
bisulfate (hydrogen sulfate)	HSO_4^-		peroxide	O_2^{2-}
carbonate	CO_3^{2-}		perchlorate	ClO_4^-
chlorate	ClO_3^-		phosphate	PO_4^{3-}
chromate	CrO_4^{2-}		sulfate	SO_4^{2-}
cyanide	CN^-		sulfite	SO_3^{2-}
dichromate	$Cr_2O_7^{2-}$		triiodide	I_3^-

The naming of ionic compounds that contain polyatomic ions follows the same rules as the naming for other ionic compounds: simply combine the name of the cation and the name of the anion. Do not use numerical prefixes in the name if there is more than one polyatomic ion; the only exception to this is if the name of the ion itself contains a numerical prefix, such as dichromate or triiodide.

Writing the formulas of ionic compounds has one important difference. If more than one polyatomic ion is needed to balance the overall charge in the formula, enclose the formula of the polyatomic ion in parentheses and write the proper numerical subscript to the right and *outside* of the parentheses. Thus, the formula between calcium ions, Ca^{2+}, and nitrate ions, NO_3^-, is properly written $Ca(NO_3)_2$, not $CaNO_{32}$ or CaN_2O_6. Use parentheses where required. The name of this ionic compound is simply calcium nitrate.

✔ Example 3.5.4

Write the proper formula and give the proper name for each ionic compound formed between the two listed ions.

a. NH_4^+ and S^{2-}
b. Al^{3+} and PO_4^{3-}
c. Fe^{2+} and PO_4^{3-}

Solution

a. Because the ammonium ion has a 1+ charge and the sulfide ion has a 2− charge, we need two ammonium ions to balance the charge on a single sulfide ion. Enclosing the formula for the ammonium ion in parentheses, we have $(NH_4)_2S$. The compound's name is ammonium sulfide.
b. Because the ions have the same magnitude of charge, we need only one of each to balance the charges. The formula is $AlPO_4$, and the name of the compound is aluminum phosphate.
c. Neither charge is an exact multiple of the other, so we have to go to the least common multiple of 6. To get 6+, we need three iron(II) ions, and to get 6−, we need two phosphate ions. The proper formula is $Fe_3(PO_4)_2$, and the compound's name is iron(II) phosphate.

? Exercise 3.5.4

Write the proper formula and give the proper name for each ionic compound formed between the two listed ions.

a. NH_4^+ and PO_4^{3-}
b. Co^{3+} and NO_2^-

Answers

a. $(NH_4)_3PO_4$, ammonium phosphate

b. $Co(NO_2)_3$, cobalt(III) nitrite

Food and Drink Application: Sodium in Your Food

The element sodium, at least in its ionic form as Na^+, is a necessary nutrient for humans to live. In fact, the human body is approximately 0.15% sodium, with the average person having one-twentieth to one-tenth of a kilogram in their body at any given time, mostly in fluids outside cells and in other bodily fluids.

Sodium is also present in our diet. The common table salt we use on our foods is an ionic sodium compound. Many processed foods also contain significant amounts of sodium added to them as a variety of ionic compounds. Why are sodium compounds used so much? Usually sodium compounds are inexpensive, but, more importantly, most ionic sodium compounds dissolve easily. This allows processed food manufacturers to add sodium-containing substances to food mixtures and know that the compound will dissolve and distribute evenly throughout the food. Simple ionic compounds such as sodium nitrite ($NaNO_2$) are added to cured meats, such as bacon and deli-style meats, while a compound called sodium benzoate is added to many packaged foods as a preservative. Table **3.5.3** is a partial list of some sodium additives used in food. Some of them you may recognize after reading this chapter. Others you may not recognize, but they are all ionic sodium compounds with some negatively charged ion also present.

Table **3.5.3**: Some Sodium Compounds Added to Food

Sodium Compound	Use in Food
Sodium acetate	preservative, acidity regulator
Sodium adipate	food acid
Sodium alginate	thickener, vegetable gum, stabilizer, gelling agent, emulsifier
Sodium aluminum phosphate	acidity regulator, emulsifier
Sodium aluminosilicate	anti-caking agent
Sodium ascorbate	antioxidant
Sodium benzoate	preservative
Sodium bicarbonate	mineral salt
Sodium bisulfite	preservative, antioxidant
Sodium carbonate	mineral salt
Sodium carboxymethylcellulose	emulsifier
Sodium citrates	food acid
Sodium dehydroacetate	preservative
Sodium erythorbate	antioxidant
Sodium erythorbin	antioxidant
Sodium ethyl para-hydroxybenzoate	preservative
Sodium ferrocyanide	anti-caking agent
Sodium formate	preservative
Sodium fumarate	food acid
Sodium gluconate	stabilizer
Sodium hydrogen acetate	preservative, acidity regulator
Sodium hydroxide	mineral salt
Sodium lactate	food acid
Sodium malate	food acid
Sodium metabisulfite	preservative, antioxidant, bleaching agent
Sodium methyl para-hydroxybenzoate	preservative
Sodium nitrate	preservative, color fixative
Sodium nitrite	preservative, color fixative
Sodium orthophenyl phenol	preservative
Sodium propionate	preservative
Sodium propyl para-hydroxybenzoate	preservative
Sodium sorbate	preservative
Sodium stearoyl lactylate	emulsifier

Sodium Compound	Use in Food
Sodium succinates	acidity regulator, flavor enhancer
Sodium salts of fatty acids	emulsifier, stabilizer, anti-caking agent
Sodium sulfite	mineral salt, preservative, antioxidant
Sodium sulfite	preservative, antioxidant
Sodium tartrate	food acid
Sodium tetraborate	preservative

The use of so many sodium compounds in prepared and processed foods has alarmed some physicians and nutritionists. They argue that the average person consumes too much sodium from his or her diet. The average person needs only about 500 mg of sodium every day; most people consume more than this—up to 10 times as much. Some studies have implicated increased sodium intake with high blood pressure; newer studies suggest that the link is questionable. However, there has been a push to reduce the amount of sodium most people ingest every day: avoid processed and manufactured foods, read labels on packaged foods (which include an indication of the sodium content), avoid oversalting foods, and use other herbs and spices besides salt in cooking.



Figure **3.5.2**: Food labels include the amount of sodium per serving. This particular label shows that there are 75 mg of sodium in one serving of this particular food item.

Key Takeaways

- Ions form when atoms lose or gain electrons.
- Ionic compounds have positive ions and negative ions.
- Ionic formulas balance the total positive and negative charges.
- Ionic compounds have a simple system of naming.
- Groups of atoms can have an overall charge and make ionic compounds.

3.5: Ions and Ionic Compounds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.5: Ions and Ionic Compounds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

CHAPTER OVERVIEW

4: Chapter 4 - Molecules, Bonding, and Forces

4.1: Compounds, Lewis Diagrams and Ionic Bonds
4.2: Nomenclature of Ionic Compounds
4.3: Covalent Bonding
4.4: Exceptions to the Octet Rule
4.5: Nomenclature of Molecular Compounds
4.6: Electronegativity and the Polar Covalent Bond
4.7: Intermolecular Forces

4: Chapter 4 - Molecules, Bonding, and Forces is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

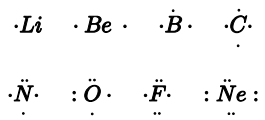
4.1: Compounds, Lewis Diagrams and Ionic Bonds

If we take two or more atoms and bond them together chemically so that they now behave as a single substance, we have made a chemical compound. We will see that the process of bonding actually involves either the *sharing*, or the net *transfer*, of electrons from one atom to another. The two types of bonding are **covalent**, for the sharing of electrons between atoms, and **ionic**, for the net transfer of electrons between atoms. Covalent or ionic bonding will determine the *type* of compound that will be formed.

In Chapter 1, we used atomic theory to describe the structure of the fluorine atom. We said that neutral fluorine has nine protons in its nucleus (an atomic number of 9), nine electrons surrounding the nucleus (to make it neutral), and the most common isotope has ten neutrons in its nucleus, for a mass number of 19. Further, we said that the nine electrons exist in two energy levels; the first energy level contains two electrons and is written $1s^2$. The second energy level contains seven electrons, distributed as $2s^2\ 2p^5$. The *outermost* electron level in any atom is referred to as the **valence shell**. For the representative elements (remember, this includes all of the elements except for the transition metals), the number of electrons in the valence shell corresponds to the Group number of the element in the periodic table. Group 1A elements will have one valence electron, Group 6A elements will have six valence electrons, and so on. Fluorine is a Group 7A element and has seven valence electrons. We can show the electron configuration for fluorine using a **Lewis diagram** (or *electron-dot structure*), named after the American chemist G. N. Lewis, who proposed the concepts of electron shells and valence electrons. In a Lewis diagram, the electrons in the valence shell are shown as small "dots" surrounding the atomic symbol for the element.



When more than four electrons are present in the valence shell, they are shown as pairs when writing the Lewis diagram (but never more than pairs). Lewis diagrams for the atoms in the second period are shown below:



As you look at the dot-structures please understand that it makes *no difference* where you place the electrons, or the electron pairs, around the symbol, as long as pairs are shown whenever there are four or more valence electrons.

If you examine the Lewis diagram for neon (Ne) above, you will see that the valence shell is *filled*; that is, there are eight electrons in the valence shell. Elements in Group 8A of the periodic table are called *noble gasses*; they are very stable and do not routinely combine with other elements to form compounds (although today, many compounds containing noble gasses are known). Modern bonding theory tells us that this stability arises because the valence shell in the noble gasses is completely filled. When the valence shell is *not* full, theory suggests that atoms will transfer or share electrons with other atoms in order to achieve a filled valence shell… that is, the electron configuration of the noble gasses. Chemical bonding can then be viewed as a quest by atoms to acquire (or lose) enough electrons so that their valence shells are filled, that is, to achieve a "noble gas configuration". This is often referred to as the "**octet rule**"; the desire for elements to obtain eight electrons in the valence shell (except of course for helium where the noble gas configuration is *two* valence electrons).

Atoms can achieve a noble gas configuration by two methods; the transfer of electrons from their valence shells to another atom, or by sharing electrons with another atom. If you examine the Lewis diagram for lithium (Li), you will see that it has only one valence electron. If lithium was to *transfer* this electron to another atom, it would be left with two electrons in the 1*s*-orbital (denoted as $1s^2$). This is the same electron configuration as helium (He), and so by losing this electron, lithium has achieved a *noble gas configuration*. Because electrons carry a negative charge, the loss of this electron leaves lithium with a single positive charge. This is the **lithium cation** and it is shown as Li^+.

Returning to fluorine (F), in order to achieve the $2s^2\ 2p^6$ configuration of neon (Ne), fluorine needs to *gain* one valence electron. Because fluorine has *gained* one electron, it now has one negative charge. This is the fluoride anion and it is shown as F^-. The transfer of electrons in order to achieve a noble gas configuration is the process known as **ionic bonding**, and this will be covered in more detail later in this chapter.

Exercise 4.1.1

a. Sodium and chlorine are both third-period elements. Draw Lewis diagrams for each of these elements.
b. What number of electrons would chlorine have to gain in order to achieve a “noble gas configuration”? What would be the charge on chlorine?
c. What number of electrons would Na have to lose to obtain the noble gas configuration of Ne with *eight* valence electrons? What charge would Na have?

4.1: Compounds, Lewis Diagrams and Ionic Bonds is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **3.1: Compounds, Lewis Diagrams and Ionic Bonds** by Paul R. Young is licensed CC BY-SA 4.0. Original source: https://en.wikibooks.org/wiki/Introductory_Chemistry_Online.

4.2: Nomenclature of Ionic Compounds

The simplest ionic compounds consist of a single type of cation associated with a single type of anion. Nomenclature for these compounds is trivial; the cation is named first, followed by the anion. If the anion is a single element, the suffix *ide* is added to the root name of the element.

When you are constructing names for ionic compounds, you do not use "multipliers" to indicate how many cations or anions are present in the compound. For example NaI is named *sodium iodide;* Na_2S is named *sodium sulfide;* $CaCl_2$ is named *calcium chloride.* The chemist reading the name is assumed to have sufficient knowledge to pair the elements properly based on their common valence states. There are exceptions to this simple nomenclature, however. Many transition metals exist as more than one type of cation. Thus, iron exists as Fe^{2+} and Fe^{3+} cations (they are referred to as "oxidation states", and will be covered in detail in Chapter 5). When you are naming an ionic compound containing iron, it is necessary to indicate which oxidation state the metal has. For metals, the oxidation state is the same as the charge. Thus Fe^{2+} in a compound with chloride would have a formula $FeCl_2$ and would be named iron (II) chloride, with the oxidation state (the charge on the iron) appearing as a Roman numeral in parenthesis after the cation. The cation Fe^{3+} paired with oxygen would have the formula Fe_2O_3 and would have the name iron (III) oxide.

The procedure for naming ionic compounds contain polyatomic ions is identical to that described above for simple ions. Thus, $CaCO_3$ is named *calcium carbonate;* Na_2SO_4 is named *sodium sulfate;* $(NH_4)_2HPO_4$ (a compound with two polyatomic ions) is named *ammonium hydrogen phosphate;* and Pb^{2+} paired with SO_4^{2-}, $PbSO_4$ is named *lead (II) sulfate.*

Example 4.2.1:

Write a correct chemical formula for each of the following ionic compounds:

a. Calcium bromide
b. Aluminum oxide
c. Copper (II) chloride
d. Iron (III) oxide

Solution

a. Calcium is 2+, bromide is 1-; $CaBr_2$.
b. Aluminum is 3+, oxide is 2-; Al_2O_3.
c. From the oxidation state that is given, copper is 2+, chloride is 1-; $CuCl_2$.
d. From the oxidation state, iron is 3+, oxide is 2-; Fe_2O_3.

Example 4.2.1:

Write a proper chemical name for each of the following ionic compounds:

a. Li_2S
b. CaO
c. $NiCl_2$
d. FeO

Solution

a. We don't use multipliers, so this is simply lithium sulfide.
b. This is simply calcium oxide.
c. We don't have to specify an oxidation state for nickel, so this is nickel chloride.
d. We must specify that iron is 2+ in this compound; iron (II) oxide.

? Exercise 4.2.1

Write a correct chemical formula for each of the following ionic compounds:

a. Sodium phosphide
b. Iron (II) nitrite
c. Calcium hydrogen phosphate
d. Chromium (III) oxide

? Exercise 4.2.1

Write a proper chemical name for each of the following ionic compounds:

a. NaBr
b. $CuCl_2$
c. $Fe(NO_3)_3$
d. $(NH_4)_3PO_4$

4.2: Nomenclature of Ionic Compounds is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **3.10: Nomenclature of Ionic Compounds** by Paul R. Young is licensed CC BY-SA 4.0. Original source: https://en.wikibooks.org/wiki/Introductory_Chemistry_Online.

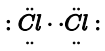
4.3: Covalent Bonding

A second method by which atoms can achieve a filled valence shell is by *sharing* valence electrons with another atom. Thus fluorine, with one unpaired valence electron, can share that electron with an unshared electron on another fluorine to form the compound, F_2 in which the two shared electrons form a chemical bond holding the two fluorine atoms together. When you do this, each fluorine now has the equivalent of eight electrons in its valence shell; three unshared pairs and one pair that is *shared* between the two atoms. Note that when you are counting electrons, the electrons that are shared in the covalent bond are counted for each atom, individually. A chemical bond formed by *sharing* electrons between atoms is called a **covalent bond**. When two or more atoms are bonded together utilizing covalent bonds, the compound is referred to as a **molecule**.

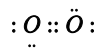
There is a simple method, given below, that we can use to construct Lewis diagrams for diatomic and for polyatomic molecules:

- Begin by adding up all of the *valence electrons* in the molecule. For F_2, each fluorine has seven, giving a total of 14 valence electrons.
- Next, draw your *central atom*. For a diatomic molecule like F_2, both atoms are the same, but if several different atoms are present, the *central atom* will be to the *left* (or lower) in the periodic table.
- Next, draw the other atoms around the *central atom*, placing two electrons *between* the atoms to form a covalent bond.
- Distribute the remaining valence electrons, as pairs, around each of the outer atoms, so that they all are surrounded by *eight electrons.*
- Place any remaining electrons on the *central atom.*
- If the central atom is *not* surrounded by an octet of electrons, construct *multiple bonds* with the outer atoms until *all* atoms have a complete octet.
- If there are an *odd number* of valence electrons in the molecule, leave the remaining single electron on the *central atom.*

Let's apply these rules for the Lewis diagram for chlorine gas, Cl_2. There are 14 valence electrons in the molecule. Both atoms are the same, so we draw them next to each other and place two electrons between them to form the covalent bond. Of the twelve remain electrons, we now place six around one chlorine (to give an octet) and then place the other six around the other chlorine (our *central atom*). Checking, we see that each atom is surrounded by an octet of valence electrons, and so our structure is complete.



All of the Group 7A elements (the halogens), have valence shells with seven electrons and all of the common halogens exist in nature as diatomic molecules; fluorine, F_2; chlorine, Cl_2; bromine, Br_2 and iodine I_2 (astatine, the halogen in the sixth period, is a short-lived radioactive element and its chemical properties are poorly understood). Nitrogen and oxygen, Group 5A and 6A elements, respectively, also exists in nature as diatomic molecules (N_2 and O_2). Let's consider **oxygen**; oxygen has six valence electrons (a Group 6A element). Following the logic that we used for chlorine, we draw the two atoms and place one pair of electrons between them, leaving 10 valence electrons. We place three pairs on one oxygen atom, and the remaining two pairs on the second (our *central atom*). Because we only have six valence electrons surrounding the second oxygen atom, we must move one pair from the other oxygen and form a second covalent bond (a **double bond**) between the two atoms. Doing this, each atom now has an octet of valence electrons.



Nitrogen has five valence electrons. Sharing one on each atom gives the first intermediate where each nitrogen is surrounded by six electrons (not enough!). Sharing another pair, each nitrogen is surrounded by seven electrons, and finally, sharing the third, we get a structure where each nitrogen is surrounded by eight electrons; a noble gas configuration (or the "octet rule"). Nitrogen is a very stable molecule and relatively unreactive, being held together by a strong **triple covalent bond**.



As we have constructed Lewis diagrams, thus far, we have strived to achieve an octet of electrons around every element. In nature, however, there are many exceptions to the "octet rule". Elements in the first row of the periodic table (hydrogen and helium) can only accommodate *two* valence electrons. Elements below the second row in the periodic table can accommodate, 10, 12 or even

14 valence electrons (we will see an example of this in the next section). Finally, in many cases molecules exist with single unpaired electrons. A classic example of this is oxygen gas (O_2). We have previously drawn the Lewis diagram for oxygen with an oxygen-oxygen double bond. Physical measurements on oxygen, however, suggest that this picture of bonding is not quite accurate. The magnetic properties of oxygen, O_2, are most consistent with a structure having *two unpaired* electrons in the configuration shown below:

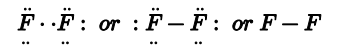


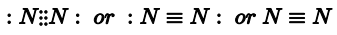
In this Lewis diagram, each oxygen atom is surrounded by *seven* electrons (not eight). This electronic configuration may explain why oxygen is such a *reactive* molecule (reacting with iron, for example, to form rust); the unpaired electrons on the oxygen molecule are readily available to interact with electrons on other elements to form new chemical compounds.

Another notable exception to the "octet rule" is the molecule NO (nitrogen monoxide). Combining one nitrogen (Group 5A) with one oxygen (Group 6A) gives a molecule with *eleven* valence electrons. There is *no way* to arrange eleven electrons without leaving one electron unpaired. Nitric oxide is an extremely reactive molecule (by virtue of its unshared electron) and has been found to play a central role is biochemistry as a reactive, short-lived molecule involved in cellular communication.



As useful as Lewis diagrams can be, chemists tire of drawing little dots and, for a shorthand representation of a covalent bond, a short line (a **line-bond**) is often drawn between the two elements. Whenever you see atoms connected by a line-bond, you are expected to understand that this represents two shared electrons in a covalent bond. Further, the *unshared pairs* of electrons on the bonded atoms are sometimes shown, and sometimes they are omitted. If unshared pairs ore omitted, the chemist reading the structure is assumed to understand that they are present.





4.3: Covalent Bonding is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **3.2: Covalent Bonding** by Paul R. Young is licensed CC BY-SA 4.0. Original source: https://en.wikibooks.org/wiki/Introductory_Chemistry_Online.

4.4: Exceptions to the Octet Rule

Returning briefly to classical Lewis diagrams. Consider the diagram shown below for the molecule SF_4. In constructing this diagram, six valence electrons are placed around the sulfur and seven valence electrons are placed around each fluorine. As we attempt to pair these to form covalent bonds, we note that there are “too many” electrons on the sulfur! We clearly cannot form a covalent bond using three electrons, so we *split* one pair, move the single electrons into bonding position and form bonds with the remaining two fluorines. The “extra pair” of electrons just *sits there* on the sulfur and does not participate directly in the bonding.

This is an example of **valence expansion**. In general, elements *below* the second period in the periodic table (S, Se, Te, etc.) will commonly have 10 – 12 electrons in their valence shells. As in SF_4, these electrons are not directly involved in the formation of covalent bonds, but they affect the overall reactivity of the particular molecule.

In general, all molecular compounds containing elements that appear below the second row in the periodic table are capable of valence expansion and you need to be very careful when you are drawing Lewis diagrams for these compounds. As we saw in Section 3.1 for the molecule nitrogen oxide (NO), stable molecules also exist in which atoms are *not* surrounded by an octet of electrons. Another example of this is the molecule BF_3, which is shown in the following example.

Example 4.4.1: Boron Trifluoride

Construct a Lewis diagram for the molecule BF_3.

Solution

Boron has three valence electrons and fluorine has seven. The central atom in our structure will be boron (it is to the *left* of fluorine in the periodic table). Next, we draw the boron (our central atom) with its’ three electrons and place the three fluorines around the boron with the electrons arranged to form the three covalent bonds. Each of the fluorines have a full octet. The boron, however, is only surrounded by six electrons. Because of this, the boron in BF_3 is a powerful *electron acceptor* and forms strong complexes with electrons from other compounds. In Chapter 8 we will see that this property is called Lewis acidity and BF_3 is a very powerful Lewis acid.

4.4: Exceptions to the Octet Rule is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **3.8: Exceptions to the Octet Rule** by Paul R. Young is licensed CC BY-SA 4.0. Original source: https://en.wikibooks.org/wiki/Introductory_Chemistry_Online.

4.5: Nomenclature of Molecular Compounds

The nomenclature of simple binary molecular compounds (covalently bonded compounds consisting of only two elements) is slightly more complicated than the nomenclature of ionic compounds because *multipliers* must be used to indicate the ratio of the elements in the molecule; the multiplier *mono* is only used for the second element in a compound.

Further, when you are naming a molecular compound, you must also decide which element should be listed first. In general, elements appearing to the *left* or *lower* in the periodic table are listed first in the name. Once you have decided on the order, the second element is named using the element root and *ide*, just like in ionic compounds. Thus, for CCl_4, carbon is to the *left* of chlorine (Group 4A *vs.* Group 5A), so it is listed first. There are four chlorines, so the multiplier *tetra* is used, and the name is **carbon tetrachloride**. Compounds containing hydrogen are generally an exception, and the hydrogen is listed as the first element in the name. Thus, H_2S would be named using the multiplier *di* to indicate that there are two hydrogens and *mono* to indicate that there is only one sulfur, or, dihydrogen monosulfide.

For the molecule SO_2; they are both Group 6A elements, but sulfur is *lower* in the periodic table (Row 3 *vs.* Row 2) so it is *first* in the name. There are two oxygens, so the multiplier is *di* and the name is **sulfur dioxide**.

For the molecule NO; nitrogen is to the *left* of oxygen (Group 6A *vs.* Group 5A) so it is *first* in the name. There is one oxygen, so the multiplier is *mono* and, following the rules, the name would be “nitrogen mon**oo**xide”. In this case, however, the second “o” in the name is *dropped* (to allow for easier pronunciation) and the name is shortened to **nitrogen monoxide**. Distinguish this from another oxide of nitrogen, N_2O_4. Again nitrogen is first and needs the multiplier *di*. There are four oxygens, so the multiplier is *tetra*, but once again the multiplier is shortened (again, the “a” is dropped) and the name is **dinitrogen tetroxide**.

Example 4.5.1:

Write a correct chemical formula for each of the following molecular compounds:

a. Chlorine monofluoride
b. Dihydrogen monosulfile
c. Carbon tetrabromide
d. Bromine

Solution

a. ClF
b. H_2S
c. CBr_4
d. Br_2

Example 4.5.1:

Write a proper chemical name for each of the following molecular compounds:

a. IF
b. PCl_3
c. I_2
d. N_2F_2

Solution

a. Iodine monofluoride
b. Phosphorus trichloride
c. Iodine
d. Dinitrogen difluoride

4.5: Nomenclature of Molecular Compounds is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **3.11: Nomenclature of Molecular Compounds** by Paul R. Young is licensed CC BY-SA 4.0. Original source: https://en.wikibooks.org/wiki/Introductory_Chemistry_Online.

4.6: Electronegativity and the Polar Covalent Bond

If we were to construct a Lewis diagram for molecular hydrogen (H_2), we would pair the single valence electrons on each atom to make a single covalent bond. Each hydrogen would now have two electrons in its valence shell, identical to helium. The mathematical equations chemists use to describe covalent bonding can be solved to predict the regions of space surrounding the molecule in which these electrons are likely to be found. A particularly useful application of these calculations generates a **molecular surface** that is color coded to show *electron density* surrounding the molecule. This type of molecular surface is called an **electrostatic potential map.** When this type of calculation is done for molecules consisting of two (or more) *different* atoms, the results can be strikingly different. Consider the molecule HF. Hydrogen, with one valance electron, can share that electron with fluorine (with seven valence electrons) to form a single covalent bond.

In this electrostatic potential map, blue is used to indicate low electron density (a relative *positive* charge) and red indicates high electron density (a relative *negative* charge); the colors light blue, green, yellow and orange indicate the increasing charge gradient. The molecule HF is clearly very **polar**, meaning that a significant difference in electron density exists across the length of the molecule. The electrostatic potential map for HF contrasts significantly with that for H_2, where the charge was quite symmetrical (a uniform green color). Hydrogen fluoride (HF) can be described as a very *polar* molecule, while hydrogen (H_2) is *nonpolar*.

The origin of the polarization of the HF covalent bond has to do with *electronegativity*, an inherent property of all atoms. Within the periodic table, there is a trend for atoms to attract electrons towards themselves when they are bonded to another atom (as in HF). Atoms that tend to strongly attract electrons have a high electronegativity, relative to atoms that have a relatively low tendency to attract electrons towards themselves. The modern electronegativity scale was devised by Linus Pauling in 1932 and, in the Pauling scale, atoms in the periodic table vary in electronegativity from a low of 0.8 for cesium to a maximum of 4.0 for fluorine.

In the molecule HF, the electronegativity of the hydrogen is 2.2 and fluorine is 4.0. This difference leads to the profound polarization of the HF covalent bond which is apparent in the electrostatic potential map.

Polarizations of covalent bonds also occur in more complex molecules. In water, the oxygen has an electronegativity of 3.5; hydrogen is 2.2. Because of this, each of the H-O bonds is polarized with greater electron density towards the oxygen. Within the molecule, H_2O, the effect of this polarization becomes apparent in the electrostatic potential map, as shown in Figure 3.18. The end of the molecule with the oxygen has a high electron density and the hydrogen ends are electron deficient. We will see in later chapters that the polarization of water, caused by the difference in electronegativities, gives water the special properties that allows it to dissolve ionic compounds, and basically support life as we know it. Within organic chemistry (the study of carbon-containing molecular compounds), you will appreciate that the relative reactivity of organic molecules with each other is largely dependent on the polarization of covalent bonds in these molecules.

4.6: Electronegativity and the Polar Covalent Bond is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **3.7: Electronegativity and the Polar Covalent Bond** by Paul R. Young is licensed CC BY-SA 4.0. Original source: https://en.wikibooks.org/wiki/Introductory_Chemistry_Online.

4.7: Intermolecular Forces

Learning Objective

- Relate phase to intermolecular forces.

Why does a substance have the phase that it does? The preferred phase of a substance at a given set of conditions is a balance between the energy of the particles and intermolecular forces (or intermolecular interactions) between the particles. If the forces between particles are strong enough, the substance is a liquid or, if stronger, a solid. If the forces between particles are weak and sufficient energy is present, the particles separate from each other, so the gas phase is the preferred phase. The energy of the particles is mostly determined by temperature, so temperature is the main variable that determines what phase is stable at any given point.

What forces define intermolecular interactions? There are several. A force present in all substances with electrons is the **dispersion force** (sometimes called the *London dispersion force*, after the physicist Fritz London, who first described this force in the early 1900s). This interaction is caused by the instantaneous position of an electron in a molecule, which temporarily makes that point of the molecule negatively charged and the rest of the molecule positively charged. In an instant, the electron is now somewhere else, but the fleeting imbalance of electric charge in the molecule allows molecules to interact with each other. As you might expect, the greater the number of electrons in a species, the stronger the dispersion force; this partially explains why smaller molecules are gases and larger molecules are liquids and solids at the same temperature. (Mass is a factor as well.)

Molecules with a permanent dipole moment experience **dipole-dipole interactions,** which are generally stronger than dispersion forces if all other things are equal. The oppositely charged ends of a polar molecule, which have partial charges on them, attract each other (Figure **4.7.1**). Thus, a polar molecule such as CH_2Cl_2 has a significantly higher boiling point (313 K, or 40°C) than a nonpolar molecule like CF_4 (145 K, or −128°C), even though it has a lower molar mass (85 g/mol vs. 88 g/mol).

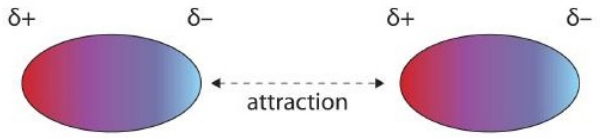


Figure **4.7.1**: Dipole-Dipole Interactions. Oppositely charged ends of polar molecules attract each other.
2 ovals with partial positive left & right sides are attracted to each other. The right of oval 1 is attracted to the left of oval 2.

An unusually strong form of dipole-dipole interaction is called **hydrogen bonding**. Hydrogen bonding is found in molecules with an H atom bonded to an N atom, an O atom, or an F atom. Such covalent bonds are very polar, and the dipole-dipole interaction between these bonds in two or more molecules is strong enough to create a new category of intermolecular force. Hydrogen bonding is the reason water has unusual properties. For such a small molecule (its molar mass is only 18 g/mol), H_2O has relatively high melting and boiling points. Its boiling point is 373 K (100°C), while the boiling point of a similar molecule, H_2S, is 233 K (−60°C). This is because H_2O molecules experience hydrogen bonding, while H_2S molecules do not. This strong attraction between H_2O molecules requires additional energy to separate the molecules in the condensed phase, so its boiling point is higher than would be expected. Hydrogen bonding is also responsible for water's ability as a solvent, its high heat capacity, and its ability to expand when freezing; the molecules line up in such a way that there is extra space between the molecules, increasing its volume in the solid state (Figure **4.7.2**).

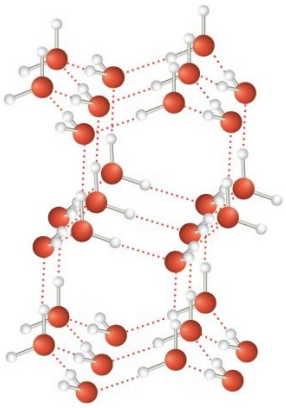


Figure **4.7.2**: Hydrogen Bonding. When water solidifies, hydrogen bonding between the molecules forces the molecules to line up in a way that creates empty space between the molecules, increasing the overall volume of the solid. This is why ice is less dense than liquid water.

✔ Example 4.7.1

Identify the most significant intermolecular force in each substance.

a. C_3H_8
b. CH_3OH
c. H_2S

Solution

1. Although C–H bonds are polar, they are only minimally polar. The most significant intermolecular force for this substance would be dispersion forces.
2. This molecule has an H atom bonded to an O atom, so it will experience hydrogen bonding.
3. Although this molecule does not experience hydrogen bonding, the Lewis electron dot diagram and VSEPR indicate that it is bent, so it has a permanent dipole. The most significant force in this substance is dipole-dipole interaction.

? Exercise 4.7.1

Identify the most significant intermolecular force in each substance.

a. HF
b. HCl

Answers

a. hydrogen bonding
b. dipole-dipole interactions

The preferred phase a substance adopts can change with temperature. At low temperatures, most substances are solids (only helium is predicted to be a liquid at absolute zero). As the temperature increases, those substances with very weak intermolecular forces become gases directly (in a process called *sublimation*, which will be discussed in Section 10.3). Substances with weak interactions can become liquids as the temperature increases. As the temperature increases even more, the individual particles will have so much energy that the intermolecular forces are overcome, so the particles separate from each other, and the substance becomes a gas (assuming that their chemical bonds are not so weak that the compound decomposes from the high temperature). Although it is difficult to predict the temperature ranges for which solid, liquid, or gas is the preferred phase for any random substance, all substances progress from solid to liquid to gas (in that order) as temperature increases.

Summary

- All substances experience dispersion forces between their particles.
- Substances that are polar experience dipole-dipole interactions.

- Substances with covalent bonds between an H atom and N, O, or F atoms experience hydrogen bonding.
- The preferred phase of a substance depends on the strength of the intermolecular force and the energy of the particles.

4.7: Intermolecular Forces is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.2: Intermolecular Forces** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

CHAPTER OVERVIEW

5: Chapter 5 - Solution Chemistry

5.1: Prelude to Solutions
5.2: Definitions
5.3: The Mole
5.4: Quantitative Units of Concentration
5.5: Dilutions and Concentrations
5.6: Colligative Properties of Solutions

5: Chapter 5 - Solution Chemistry is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

5.1: Prelude to Solutions

More than 70% of the earth's surface is covered by a very important solution—seawater. It is likely that without seawater, no life would exist on Earth.

At its simplest, seawater is mostly H_2O. However, about 3.5% of seawater is dissolved solids—mostly NaCl—but other ions as well. Table **5.1.1** lists the percentage by mass of the various ions in seawater. Because it is highly likely that life on Earth originated in the oceans, it should not be surprising that many bodily fluids resemble seawater, especially blood. Table **5.1.1** also lists the percentage by mass of ions in a typical sample of blood.

Table **5.1.1**: Percentage by Mass of Ions in Seawater and Blood

Ion	Percentage in Seawater	Percentage in Blood
Na^+	2.36	0.322
Cl^-	1.94	0.366
Mg^{2+}	0.13	0.002
SO_4^{2-}	0.09	-
K^+	0.04	0.016
Ca^{2+}	0.04	0.0096
HCO_3^-	0.002	0.165
HPO_4^{2-}, $H_2PO_4^-$	-	0.01

Most ions are more abundant in seawater than they are in blood, with some notable exceptions. There is far more hydrogen carbonate ion (HCO_3^-) in blood than in seawater; indeed, it is the third most common ion in blood. This difference is significant because the HCO_3^- ion and some related species [CO_3^{2-}, $CO_2(aq)$] have an important role in controlling the acid-base properties of blood. Although there is a negligible amount of the two hydrogen phosphate ions (HPO_4^{2-} and $H_2PO_4^-$) in seawater, there is a small amount in blood, where these ions affect acid-base properties. Another notable difference is that blood has a negligible amount of the sulfate ion (SO_4^{2-}), but this ion is present in seawater. Gold is present in seawater—but only a tiny amount. A current estimate of the amount of gold is about 1 part per every 1×10^{13} parts of seawater, which makes the extraction of gold from seawater unfeasible. However, it does mean that there are about 1.4×10^{14} g of gold in the world's oceans!



Figure **5.1.1**: Gold in the Ocean. There are approximately 1.4×10^{14} g of gold in the oceans, but extracting it effectively is beyond current technologies. (CC SA-BY; Evgeni Dinev, Flickr.)

A solution is a *homogeneous mixture*: a mixture of two or more substances that are so intimately mixed that the mixture behaves in many ways like a single substance. Many chemical reactions occur when the reactants are dissolved in solution. In this chapter, we will introduce concepts that are applicable to solutions and the chemical reactions that occur in them.

5.1: Prelude to Solutions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **11.1: Prelude to Solutions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

5.2: Definitions

Learning Objective

- Learn some terminology involving solutions.
- Recognize which terminology is qualitative and which terminology is quantitative.
- Explain why certain substances dissolve in other substances.

The major component of a solution is called the **solvent.** The minor component of a solution is called the **solute**. "Major" and "minor" indicate which component has the greater or lesser presence by mass or moles, respectively. Sometimes this becomes confusing, especially when considering substances with very different molar masses. We will confine the discussion here to solutions for which the major component and the minor component are obvious.

Solutions exist for every possible phase of the solute and the solvent. Salt water, for example, is a solution of solid NaCl in liquid water; soda water is a solution of gaseous CO_2 in liquid water, while air is a solution of a gaseous solute (O_2) in a gaseous solvent (N_2). In all cases, however, the overall phase of the solution is the same phase as the solvent.

Example 5.2.1: Sugar Water

A solution is made by dissolving 1.00 g of sucrose ($C_{12}H_{22}O_{11}$) in 100.0 g of liquid water. Identify the solvent and solute in the resulting solution.

Solution

Either by mass or by moles, the obvious minor component is sucrose, so it is the solute. Water—the majority component—is the solvent. The fact that the resulting solution is the same phase as water also suggests that water is the solvent.

Exercise 5.2.1

A solution is made by dissolving 3.33 g of HCl(g) in 40.0 g of liquid methyl alcohol (CH_3OH). Identify the solvent and solute in the resulting solution.

Answer

solute: HCl(g); solvent: CH_3OH

One important concept of solutions is in defining how much solute is dissolved in a given amount of solvent. This concept is called **concentration**. Various words are used to describe the relative amounts of solute. **Dilute** describes a solution that has very little solute, while **concentrated** describes a solution that has a lot of solute. One problem is that these terms are qualitative; they describe more or less but not exactly how much.

In most cases, only a certain maximum amount of solute can be dissolved in a given amount of solvent. This maximum amount is called the **solubility** of the solute. It is usually expressed in terms of the amount of solute that can dissolve in 100 g of the solvent at a given temperature. Table **5.2.1** lists the solubilities of some simple ionic compounds. These solubilities vary widely: NaCl can dissolve up to 31.6 g per 100 g of H_2O, while AgCl can dissolve only 0.00019 g per 100 g of H_2O.

Table **5.2.1**: Solubilities of Some Ionic Compounds

Solute	Solubility (g per 100 g of H_2O at 25°C)
AgCl	0.00019
$CaCO_3$	0.0006
KBr	70.7
NaCl	36.1
$NaNO_3$	94.6

When the maximum amount of solute has been dissolved in a given amount of solvent, we say that the solution is **saturated** with solute. When less than the maximum amount of solute is dissolved in a given amount of solute, the solution is **unsaturated**. These terms are also qualitative terms because each solute has its own solubility. A solution of 0.00019 g of AgCl per 100 g of H_2O may be saturated, but with so little solute dissolved, it is also rather dilute. A solution of 36.1 g of NaCl in 100 g of H_2O is also saturated but rather concentrated. Ideally, we need more precise ways of specifying the amount of solute in a solution. We will introduce such ways in Section 11.3.

In some circumstances, it is possible to dissolve more than the maximum amount of a solute in a solution. Usually, this happens by heating the solvent, dissolving more solute than would normally dissolve at regular temperatures, and letting the solution cool down slowly and carefully. Such solutions are called **supersaturated** solutions and are not stable; given an opportunity (such as dropping a crystal of solute in the solution), the excess solute will precipitate from the solution.

It should be obvious that some solutes dissolve in certain solvents but not others. NaCl, for example, dissolves in water but not in vegetable oil. Beeswax dissolves in liquid hexane but not water. What is it that makes a solute soluble in some solvents but not others?

The answer is intermolecular interactions. The intermolecular interactions include London dispersion forces, dipole-dipole interactions, and hydrogen bonding (as described in Chapter 10). From experimental studies, it has been determined that if molecules of a solute experience the same intermolecular forces that the solvent does, the solute will likely dissolve in that solvent. So, NaCl—a very polar substance because it is composed of ions—dissolves in water, which is very polar, but not in oil, which is generally nonpolar. Nonpolar wax dissolves in nonpolar hexane but not in polar water. This concept leads to the general rule that "like dissolves like" for predicting whether a solute is soluble in a given solvent. However, this is a general rule, not an absolute statement, so it must be applied with care.

✔ Example 5.2.2: Polar and Nonpolar Solvents

Would I_2 be more soluble in CCl_4 or H_2O? Explain your answer.

Solution

I_2 is nonpolar. Of the two solvents, CCl_4 is nonpolar and H_2O is polar, so I_2 would be expected to be more soluble in CCl_4.

? Exercise 5.2.2

Would C_3H_7OH be more soluble in CCl_4 or H_2O? Explain your answer.

Answer

H_2O because both experience hydrogen bonding

Summary

- Solutions are composed of a solvent (major component) and a solute (minor component).
- Concentration is the expression of the amount of solute in a given amount of solvent and can be described by several qualitative terms.
- Solubility is a specific amount of solute that can dissolve in a given amount of solvent.

- "Like dissolves like" is a useful rule for deciding if a solute will be soluble in a solvent.

5.2: Definitions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **11.2: Definitions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

5.3: The Mole

Learning Objectives

- Describe the unit *mole*.
- Relate the mole quantity of substance to its mass.

So far, we have been talking about chemical substances in terms of individual atoms and molecules. Yet we do not typically deal with substances an atom or a molecule at a time; we work with millions, billions, and trillions of atoms and molecules at a time. We need a way to deal with macroscopic, rather than microscopic, amounts of matter. We need a unit of amount that relates quantities of substances on a scale that we can interact with.

Chemistry uses a unit called mole. A **mole** (mol) is a number of things equal to the number of atoms in exactly 12 g of carbon-12. Experimental measurements have determined that this number is very large:

$$1 \text{ mol} = 6.02214179 \times 10^{23} \text{ things}$$

Understand that a mole means a specific number of things, just like a dozen means a certain number of things—twelve, in the case of a dozen. But a mole is a much larger number of things. These things can be atoms, or molecules, or eggs; however, in chemistry, we usually use the mole to refer to the amounts of atoms or molecules. Although the number of things in a mole is known to eight decimal places, it is usually fine to use only two or three decimal places in calculations. The numerical value of things in a mole is often called *Avogadro's number* (N_A). *Avogadro's number* is also known as the *Avogadro constant*, after Amadeo Avogadro, an Italian chemist who first proposed its importance.

Example 5.3.1

How many molecules are present in 2.76 mol of H_2O? How many atoms is this?

Solution

The definition of a mole is an equality that can be used to construct a conversion factor. Also, because we know that there are three atoms in each molecule of H_2O, we can also determine the number of atoms in the sample.

$$2.76 \cancel{\text{ mol } H_2O} \times \frac{6.022 \times 10^{23} \text{ molecules } H_2O}{\cancel{\text{mol } H_2O}} = 1.66 \times 10^{24} \text{ molecules } H_2O$$

To determine the total number of atoms, we have

$$1.66 \times 10^{24} \cancel{\text{ molecules } H_2O} \times \frac{3 \text{ atoms}}{1 \text{ molecule}} = 4.99 \times 10^{24} \text{ atoms}$$

Exercise 5.3.1

How many molecules are present in 4.61×10^{-2} mol of O_2?

Answer

2.78×10^{22} molecules

How big is a mole? It is very large. Suppose you had a mole of dollar bills that need to be counted. If everyone on earth (about 6 billion people) counted one bill per second, it would take about 3.2 million years to count all the bills. A mole of sand would fill a cube about 32 km on a side. A mole of pennies stacked on top of each other would have about the same diameter as our galaxy, the Milky Way. A mole is a lot of things—but atoms and molecules are very tiny. One mole of carbon atoms would make a cube that is 1.74 cm on a side, small enough to carry in your pocket.

Why is the mole unit so important? It represents the link between the microscopic and the macroscopic, especially in terms of mass. *A mole of a substance has the same mass in grams as one unit (atom or molecules) has in atomic mass units*. The mole unit

allows us to express amounts of atoms and molecules in visible amounts that we can understand.

For example, we already know that, by definition, a mole of carbon has a mass of exactly 12 g. This means that exactly 12 g of C has 6.022×10^{23} atoms:

$$12 \text{ g C} = 6.022 \times 10^{23} \text{ atoms C}$$

We can use this equality as a conversion factor between the number of atoms of carbon and the number of grams of carbon. How many grams are there, say, in 1.50×10^{25} atoms of carbon? This is a one-step conversion:

$$1.50 \times 10^{25} \cancel{\text{ atoms C}} \times \frac{12.0000\, g\, C}{6.022 \times 10^{23} \cancel{\text{ atoms C}}} = 299\, g\, C$$

But it also goes beyond carbon. Previously we defined atomic and molecular masses as the number of atomic mass units per atom or molecule. Now we can do so in terms of grams. The atomic mass of an element is the number of grams in 1 mol of atoms of that element, while the molecular mass of a compound is the number of grams in 1 mol of molecules of that compound. Sometimes these masses are called **molar masses** to emphasize the fact that they are the mass for 1 mol of things. (The term *molar* is the adjective form of mole and has nothing to do with teeth.)

Here are some examples. The mass of a hydrogen atom is 1.0079 u; the mass of 1 mol of hydrogen atoms is 1.0079 g. Elemental hydrogen exists as a diatomic molecule, H_2. One molecule has a mass of 1.0079 + 1.0079 = 2.0158 u, while 1 mol H_2 has a mass of 2.0158 g. A molecule of H_2O has a mass of about 18.01 u; 1 mol H_2O has a mass of 18.01 g. A single unit of NaCl has a mass of 58.45 u; NaCl has a molar mass of 58.45 g. In each of these moles of substances, there are 6.022×10^{23} units: 6.022×10^{23} atoms of H, 6.022×10^{23} molecules of H_2 and H_2O, 6.022×10^{23} units of NaCl ions. These relationships give us plenty of opportunities to construct conversion factors for simple calculations.

Example 5.3.2: Sugar

What is the molar mass of sugar ($C_6H_{12}O_6$)?

Solution

To determine the molar mass, we simply add the atomic masses of the atoms in the molecular formula; but express the total in grams per mole, not atomic mass units. The masses of the atoms can be taken from the periodic table.

Solutions to Example 5.3.2

6 C = 6 × 12.011	= 72.066
12 H = 12 × 1.0079	= 12.0948
6 O = 6 × 15.999	= 95.994
TOTAL	= 180.155 g/mol

Per convention, the unit *grams per mole* is written as a fraction.

Exercise 5.3.2

What is the molar mass of $AgNO_3$?

Answer

169.87 g/mol

Knowing the molar mass of a substance, we can calculate the number of moles in a certain mass of a substance and vice versa, as these examples illustrate. The molar mass is used as the conversion factor.

Example 5.3.3

What is the mass of 3.56 mol of $HgCl_2$? The molar mass of $HgCl_2$ is 271.49 g/mol.

Solution

Use the molar mass as a conversion factor between moles and grams. Because we want to cancel the mole unit and introduce the gram unit, we can use the molar mass as given:

$$3.56 \cancel{mol\ HgCl_2} \times \frac{271.49\ g\ HgCl_2}{\cancel{mol\ HgCl_2}} = 967\ g\ HgCl_2$$

Exercise 5.3.3

What is the mass of 33.7 mol of $\mathbf{H_2O}$?

Answer

607 g

Example 5.3.4

How many moles of H_2O are present in 240.0 g of water (about the mass of a cup of water)?

Solution

Use the molar mass of H_2O as a conversion factor from mass to moles. The molar mass of water is

$$(1.0079 + 1.0079 + 15.999) = 18.015\ g/mol.$$

However, because we want to cancel the gram unit and introduce moles, we need to take the reciprocal of this quantity, or 1 mol/18.015 g:

$$240.0\ \cancel{g\ H_2O} \times \frac{1\ mol\ H_2O}{18.015\ \cancel{g\ H_2O}} = 13.32\ mol\ H_2O$$

Exercise 5.3.4

How many moles are present in 35.6 g of H_2SO_4 (molar mass = 98.08 g/mol)?

Answer

0.363 mol

Other conversion factors can be combined with the definition of mole—density, for example.

Example 5.3.5

The density of ethanol is 0.789 g/mL. How many moles are in 100.0 mL of ethanol? The molar mass of ethanol is 46.08 g/mol.

Solution

Here, we use density to convert from volume to mass and then use the molar mass to determine the number of moles.

$$100\ \cancel{ml}\ ethanol \times \frac{0.789\ g}{\cancel{ml}} \times \frac{1\ mol}{46.08\ \cancel{g}} = 1.71\ mol\ ethanol$$

Exercise 5.3.5

If the density of benzene, C_6H_6, is 0.879 g/mL, how many moles are present in 17.9 mL of benzene?

Answer

0.201 mol

Summary

The mole is a key unit in chemistry. The molar mass of a substance, in grams, is numerically equal to one atom's or molecule's mass in atomic mass units.

5.3: The Mole is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **5.3: The Mole** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

5.4: Quantitative Units of Concentration

Learning Objective

- Learn to determine specific concentrations with several common units.

Rather than qualitative terms (Section 11.2 - Definitions), we need quantitative ways to express the amount of solute in a solution; that is, we need specific units of concentration. In this section, we will introduce several common and useful units of concentration.

Molarity (M) is defined as the number of moles of solute divided by the number of liters of solution:

$$molarity = \frac{moles\ of\ solute}{liters\ of\ solution}$$

which can be simplified as

$$M = \frac{mol}{L},\ or\ mol/L$$

As with any mathematical equation, if you know any two quantities, you can calculate the third, unknown, quantity.

For example, suppose you have 0.500 L of solution that has 0.24 mol of NaOH dissolved in it. The concentration of the solution can be calculated as follows:

$$molarity = \frac{0.24\ mol\ NaOH}{0.500 L} = 0.48\ M\ NaOH$$

The concentration of the solution is 0.48 M, which is spoken as "zero point forty-eight molarity" or "zero point forty-eight molar." If the quantity of the solute is given in mass units, you must convert mass units to mole units before using the definition of molarity to calculate concentration. For example, what is the molar concentration of a solution of 22.4 g of HCl dissolved in 1.56 L? First, convert the mass of solute to moles using the molar mass of HCl (36.5 g/mol):

$$22.4\ \cancel{g HCl} \times \frac{1\ mol\ Hcl}{36.5\ \cancel{g HCl}} = 0.614\ M\ HCl$$

Now we can use the definition of molarity to determine a concentration:

$$M = \frac{0.614\ mol\ HCl}{1.56 L} = 0.394\ M$$

Example 5.4.1

What is the molarity of a solution made when 32.7 g of NaOH are dissolved to make 445 mL of solution?

Solution

To use the definition of molarity, both quantities must be converted to the proper units. First, convert the volume units from milliliters to liters:

$$445\ \cancel{mL} \times \frac{1\ L}{1000\ \cancel{mL}} = 0.445\ L$$

Now we convert the amount of solute to moles, using the molar mass of NaOH, which is 40.0 g/mol:

$$32.7\ \cancel{g NaOH} \times \frac{1\ mol\ NaOH}{40.0\ \cancel{g NaOH}} = 0.818\ mol\ NaOH$$

Now we can use the definition of molarity to determine the molar concentration:

$$M = \frac{0.818\ mol\ NaOH}{0.445 L} = 1.84\ M\ NaOH$$

Exercise 5.4.1

What is the molarity of a solution made when 66.2 g of $C_6H_{12}O_6$ are dissolved to make 235 mL of solution?

Answer

1.57 M

The definition of molarity can be used to determine the amount of solute or the volume of solution, if the other information is given. Example 4 illustrates this situation.

Example 5.4.2

How many moles of solute are present in 0.108 L of a 0.887 M NaCl solution?

Solution

We know the volume and the molarity; we can use the definition of molarity to mathematically solve for the amount in moles. Substituting the quantities into the definition of molarity:

$$0.887\,M = \frac{mol\ NaCl}{0.108 L}$$

We multiply the 0.108 L over to the other side of the equation and multiply the units together; "molarity × liters" equals moles, according to the definition of molarity. So

mol NaCl = (0.887 M)(0.108 L) = 0.0958 mol

Exercise 5.4.2

How many moles of solute are present in 225 mL of a 1.44 M $CaCl_2$ solution?

Answer

0.324 mol

If you need to determine volume, remember the rule that the unknown quantity must be by itself and in the numerator to determine the correct answer. Thus, rearrangement of the definition of molarity is required.

Example 5.4.3

What volume of a 2.33 M $NaNO_3$ solution is needed to obtain 0.222 mol of solute?

Solution

Using the definition of molarity, we have

$$2.33\,M = \frac{0.222 mol}{L}$$

To solve for the number of liters, we bring the 2.33 M over to the right into the denominator, and the number of liters over to the left in the numerator. We now have

$$L = \frac{0.222 mol}{2.33\,M}$$

Dividing, **the volume is 0.0953 L = 95.3 mL.**

? Exercise 5.4.3

What volume of a 0.570 M K_2SO_4 solution is needed to obtain 0.872 mol of solute?

Answer

1.53 L

A similar unit of concentration is **molality** (*m*), which is defined as the number of moles of solute per kilogram of solvent, not per liter of solution:

$$molality = \frac{moles\ solute}{kilograms\ solvent}$$

Mathematical manipulation of molality is the same as with molarity.

Another way to specify an amount is **percentage composition by mass** (or *mass percentage*, % m/m). It is defined as follows:

$$\%m/m = \frac{mass\ of\ solute}{mass\ of\ entire\ sample} \times 100\%$$

It is not uncommon to see this unit used on commercial products (see Figure **5.4.1** - Concentration in Commercial Applications).



Figure **5.4.1** Concentration in Commercial Applications © Thinkstock. *The percentage of urea in this package is 5% m/m, meaning that there are 5 g of urea per 100 g of product.*

✓ Example 5.4.4

What is the mass percentage of Fe in a piece of metal with 87.9 g of Fe in a 113 g sample?

Solution

Using the definition of mass percentage, we have

$$\%m/m = \frac{87.9\,g\,Fe}{113\,g\,sample} \times 100\% = 77.8\%\,Fe$$

Exercise 5.4.4

What is the mass percentage of H_2O_2 in a solution with 1.67 g of H_2O_2 in a 55.5 g sample?

Answer

3.01%

Related concentration units are **parts per thousand (ppth), parts per million (ppm)** and **parts per billion (ppb).** Parts per thousand is defined as follows:

$$ppth = \frac{mass\ of\ solute}{mass\ of\ sample} \times 1000$$

There are similar definitions for parts per million and parts per billion:

$$ppm = \frac{mass\ of\ solute}{mass\ of\ sample} \times 1,000,000$$

and

$$ppb = \frac{mass\ of\ solute}{mass\ of\ sample} \times 1,000,000,000$$

Each unit is used for progressively lower and lower concentrations. The two masses must be expressed in the same unit of mass, so conversions may be necessary.

Example 5.4.5

If there are 0.6 g of Pb present in 277 g of solution, what is the Pb concentration in parts per thousand?

Solution

Use the definition of parts per thousand to determine the concentration. Substituting

$$\frac{0.6g\ Pb}{277g\ solution} \times 1000 = 2.17\ ppth$$

Exercise 5.4.5

If there are 0.551 mg of As in 348 g of solution, what is the As concentration in ppm?

Answer

1.58 ppm

As with molarity and molality, algebraic rearrangements may be necessary to answer certain questions.

Example 5.4.6

The concentration of Cl^- ion in a sample of H_2O is 15.0 ppm. What mass of Cl^- ion is present in 240.0 mL of H_2O, which has a density of 1.00 g/mL?

Solution

First, use the density of H_2O to determine the mass of the sample:

$$240.0\ \cancel{mL} \times \frac{1.00\ g}{\cancel{mL}} = 240.0\ g$$

Now we can use the definition of ppm:

$$15.0\,ppm = \frac{mass\ of\ solute}{240.0\ g\ solution} \times 1,000,000$$

Rearranging to solve for the mass of solute,

$$mass\ solute = \frac{(15.0\,ppm)(240.0\ g\ solution)}{1,000,000} = 0.0036g = 3.6\,mg$$

Exercise 5.4.6

The concentration of Fe^{3+} ion in a sample of H_2O is 335.0 ppm. What mass of Fe^{3+} ion is present in 3,450 mL of H_2O, which has a density of 1.00 g/mL?

Answer

1.16 g

For ionic solutions, we need to differentiate between the concentration of the salt versus the concentration of each individual ion. Because the ions in ionic compounds go their own way when a compound is dissolved in a solution, the resulting concentration of the ion may be different from the concentration of the complete salt. For example, if 1 M NaCl were prepared, the solution could also be described as a solution of 1 M $Na^+(aq)$ and 1 M $Cl^-(aq)$ because there is one Na^+ ion and one Cl^- ion per formula unit of the salt. However, if the solution were 1 M $CaCl_2$, there are two $Cl^-(aq)$ ions for every formula unit dissolved, so the concentration of $Cl^-(aq)$ would be 2 M, not 1 M.

In addition, the total ion concentration is the sum of the individual ion concentrations. Thus for the 1 M NaCl, the total ion concentration is 2 M; for the 1 M $CaCl_2$, the total ion concentration is 3 M.

Key Takeaway

- Quantitative units of concentration include molarity, molality, mass percentage, parts per thousand (ppth), parts per million (ppm), and parts per billion (ppb).

5.4: Quantitative Units of Concentration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **11.3: Quantitative Units of Concentration** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

5.5: Dilutions and Concentrations

Learning Objective

- Learn how to dilute and concentrate solutions.

Often, a worker will need to change the concentration of a solution by changing the amount of solvent. **Dilution** is the addition of solvent, which decreases the concentration of the solute in the solution. **Concentration** is the removal of solvent, which increases the concentration of the solute in the solution. (Do not confuse the two uses of the word *concentration* here!)

In both dilution and concentration, the amount of solute stays the same. This gives us a way to calculate what the new solution volume must be for the desired concentration of solute. From the definition of molarity,

$$\text{molarity} = \frac{\text{moles of solute}}{\text{liters of solution}}$$

we can solve for the number of moles of solute:

moles of solute = (molarity)(liters of solution)

A simpler way of writing this is to use M to represent molarity and V to represent volume. So the equation becomes

moles of solute = MV

Because this quantity does not change before and after the change in concentration, the product MV must be the same before and after the concentration change. Using numbers to represent the initial and final conditions, we have

$$M_1V_1 = M_2V_2$$

as the **dilution equation**. The volumes must be expressed in the same units. Note that this equation gives only the initial and final conditions, not the amount of the change. The amount of change is determined by subtraction.

Example 5.5.1

If 25.0 mL of a 2.19 M solution are diluted to 72.8 mL, what is the final concentration?

Solution

It does not matter which set of conditions is labeled 1 or 2, as long as the conditions are paired together properly. Using the dilution equation, we have

$$(2.19\text{ M})(25.0\text{ mL}) = M_2(72.8\text{ mL})$$

Solving for the second concentration (noting that the milliliter units cancel),

$$M_2 = 0.752\text{ M}$$

The concentration of the solution has decreased. In going from 25.0 mL to 72.8 mL, 72.8 − 25.0 = 47.8 mL of solvent must be added.

Exercise 5.5.1

A 0.885 M solution of KBr whose initial volume is 76.5 mL has more water added until its concentration is 0.500 M. What is the new volume of the solution?

Answer

135.4 mL

Concentrating solutions involves removing solvent. Usually this is done by evaporating or boiling, assuming that the heat of boiling does not affect the solute. The dilution equation is used in these circumstances as well.

Chemistry is Everywhere: Preparing IV Solutions

In a hospital emergency room, a physician orders an intravenous (IV) delivery of 100 mL of 0.5% KCl for a patient suffering from hypokalemia (low potassium levels). Does an aide run to a supply cabinet and take out an IV bag containing this concentration of KCl?

Not likely. It is more probable that the aide must make the proper solution from an IV bag of sterile solution and a more concentrated, sterile solution, called a *stock solution*, of KCl. The aide is expected to use a syringe to draw up some stock solution and inject it into the waiting IV bag and dilute it to the proper concentration. Thus the aide must perform a dilution calculation.



Figure **5.5.1** Preparing IV Solution © Thinkstock. Medical personnel commonly must perform dilutions for IV solutions.

If the stock solution is 10.0% KCl and the final volume and concentration need to be 100 mL and 0.50%, respectively, then it is easy to calculate how much stock solution to use:

$$(10\%)V_1 = (0.50\%)(100\text{ mL})V_1 = 5\text{ mL}$$

Of course, the addition of the stock solution affects the total volume of the diluted solution, but the final concentration is likely close enough even for medical purposes.

Medical and pharmaceutical personnel are constantly dealing with dosages that require concentration measurements and dilutions. It is an important responsibility: calculating the *wrong* dose can be useless, harmful, or even fatal!

Summary

- Molarity and volume are used to determine dilutions and concentrations of solutions.

5.5: Dilutions and Concentrations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **11.4: Dilutions and Concentrations** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

5.6: Colligative Properties of Solutions

the solvent. However, some of the properties of solutions differ from pure solvents in measurable and predictable ways. The differences are proportional to the fraction that the particles occupy in the . These properties are called ; the word comes from the Greek word meaning "related to the number," implying that these properties are related to the number of particles, not their identities. unit. The of the i^{th} component in a , χ_i, is the number of moles of that component divided by the total number of moles in the sample: is always a number between 0 and 1 (inclusive) and has no units; it is just a number. fractions of all substances in a mixture equals 1. Thus the of C_6H_6 in Example 5.6.1 could be calculated by evaluating the definition of a second time, or—because there are only two substances in this particular mixture—we can subtract the of the $C_{10}H_8$ from 1 to get the of C_6H_6. unit has been introduced, the first colligative property can be considered. As mentioned in Chapter 10, all pure liquids have a characteristic in equilibrium with the liquid phase, the of which is dependent on temperature. Solutions, however, have a lower than the pure solvent has, and the amount of lowering is dependent on the fraction of particles, as long as the itself does not have a significant (the term is used to describe such solutes). This colligative property is called (or). The actual of the can be calculated as follows: }\] of the , χ_{solv} is the of the solvent particles, and P^*_{solv} is the of the pure solvent at that temperature (which is data that must be provided). Equation ??? is known as (the approximate pronunciation is). is rationalized by presuming that particles take positions at the surface in place of solvent particles, so not as many solvent particles can evaporate. are related to as expressed in . As a review, recall the definition of : } =\frac{\text{moles of }}{\text{kilograms of solvent}}\nonumber \] of a with a nonvolatile is depressed compared to that of the pure solvent, it requires a higher temperature for the 's to reach 1.00 atm (760). Recall that this is the definition of the : the temperature at which the of the liquid equals 1.00 atm. As such, the of the is higher than that of the pure solvent. This property is called . of the and is called the which is a characteristic of the solvent. Several constants (as well as boiling point temperatures) are listed in Table 5.6.1. . is higher than the boiling point of the pure solvent, but the opposite occurs with the freezing point. The freezing point of a is lower than the freezing point of the pure solvent. Think of this by assuming that particles interfere with solvent particles coming together to make a solid, so it takes a lower temperature to get the solvent particles to solidify. This is called . is similar to the equation for the : of the and is called the which is also a characteristic of the solvent only. Several constants (as well as freezing point temperatures) are listed in Table 5.6.2. is one colligative property that we use in everyday life. Many antifreezes used in automobile radiators use solutions that have a lower freezing point than normal so that automobile engines can operate at subfreezing temperatures. We also take advantage of when we sprinkle various compounds on ice to thaw it in the winter for safety (Figure 5.6.1). The compounds make solutions that have a lower freezing point, so rather than forming slippery ice, any ice is liquefied and runs off, leaving a safer pavement behind. is a thin membrane that will pass certain small but not others. A thin sheet of cellophane, for example, acts as a . Consider the system in Figure 5.6.2. of a is easy to calculate: of a , M is the of the , R is the constant, and T is the absolute temperature. This equation is reminiscent of the discussed previously. is important in biological systems because cell walls are semipermeable membranes. In particular, when a person is receiving intravenous (IV) fluids, the of the fluid needs to be approximately the same as blood serum; otherwise there may be negative consequences. Figure 5.6.3 shows three red blood cells: is also the reason you should not drink seawater if you're stranded in a lifeboat on an ocean; seawater has a higher than most of the fluids in your body. You drink the water, but ingesting it will pull water out of your cells as osmosis works to dilute the seawater. Ironically, your cells will die of thirst, and you will also die. (It is OK to drink the water if you are stranded on a body of freshwater, at least from an perspective.) is also thought to be important in getting water to the tops of tall trees, in addition to capillary action.

5.6: Colligative Properties of Solutions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

6: Chapter 6 - Survey of Chemical Reactions

6.1: Prelude to Chemical Reactions
6.2: The Chemical Equation
6.3: The Mole in Chemical Reactions
6.4: Stoichiometry
6.5: Limiting Reagents
6.6: Types of Chemical Reactions - Single and Double Displacement Reactions
6.7: Composition, Decomposition, and Combustion Reactions
6.8: Neutralization Reactions
6.9: Energy
6.10: Enthalpy and Chemical Reactions
6.11: Chemical Reaction Rate
6.12: Collision Theory
6.13: Activation Energy
6.14: Chemical Equilibrium
6.15: Equilibrium Constant
6.16: Le Châtelier's Principle

6: Chapter 6 - Survey of Chemical Reactions is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

6.1: Prelude to Chemical Reactions

The space shuttle—and any other rocket-based system—uses chemical reactions to propel itself into space and maneuver itself when it gets into orbit. The rockets that lift the orbiter are of two different types. The three main engines are powered by reacting liquid hydrogen with liquid oxygen to generate water. Then there are the two solid rocket boosters, which use a solid fuel mixture that contains mainly ammonium perchlorate and powdered aluminum. The chemical reaction between these substances produces aluminum oxide, water, nitrogen gas, and hydrogen chloride. Although the solid rocket boosters each have a significantly lower mass than the liquid oxygen and liquid hydrogen tanks, they provide over 80% of the lift needed to put the shuttle into orbit—all because of chemical reactions.



Figure **6.1.1** The Space Shuttle. (Public Domain; NASA).

6.1: Prelude to Chemical Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.1: Prelude to Chemical Reactions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.2: The Chemical Equation

Learning Objectives

- Define *chemical equation*.
- Identify the parts of a chemical equation.

A chemical reaction expresses a chemical change. For example, one chemical property of hydrogen is that it will react with oxygen to make water. We can write that as follows:

$$\textbf{hydrogen reacts with oxygen to make water}$$

We can represent this chemical change more succinctly as:

$$\textbf{hydrogen} + \textbf{oxygen} \rightarrow \textbf{water}$$

where the + sign means that the two substances interact chemically with each other and the → symbol implies that a chemical reaction takes place. But substances can also be represented by chemical formulas. Remembering that hydrogen and oxygen both exist as diatomic molecules, we can rewrite our chemical change as:

$$\mathrm{H_2 + O_2 \longrightarrow H_2O}$$

This is an example of a chemical equation, which is a concise way of representing a **chemical reaction**. The initial substances are called **reactants** and the final substances are called **products**.

Unfortunately, it is also an *incomplete* chemical equation. The law of conservation of matter says that matter cannot be created or destroyed. In chemical equations, the number of atoms of each element in the reactants must be the same as the number of atoms of each element in the products. If we count the number of hydrogen atoms in the reactants and products, we find two hydrogen atoms. But if we count the number of oxygen atoms in the reactants and products, we find that there are two oxygen atoms in the reactants, but only one oxygen atom in the products.

What can we do? Can we change the subscripts in the formula for water so that it has two oxygen atoms in it? No—you *cannot* change the formulas of individual substances, because the chemical formula for a given substance is characteristic of that substance. What you *can* do, however, is to change the number of molecules that react or are produced. We do this one element at a time, going from one side of the reaction to the other, changing the number of molecules of a substance until all elements have the same number of atoms on each side.

To accommodate the two oxygen atoms as reactants, let us assume that we have two water molecules as products:

$$\mathrm{H_2 + O_2 \longrightarrow 2\,H_2O}$$

The 2 in front of the formula for water is called a **coefficient**. Now there are the same number of oxygen atoms in the reactants as there are in the product. But in satisfying the need for the same number of oxygen atoms on both sides of the reaction, we have also changed the number of hydrogen atoms on the product side, so the number of hydrogen atoms is no longer equal. No problem, simply go back to the reactant side of the equation, and add a coefficient in front of the H_2. The coefficient that works is 2:

$$\mathrm{2\,H_2 + O_2 \longrightarrow 2\,H_2O}$$

There are now four hydrogen atoms in the reactants and also four atoms of hydrogen in the product. There are two oxygen atoms in the reactants and two atoms of oxygen in the product. The law of conservation of matter has been satisfied. When the reactants and products of a chemical equation have the same number of atoms of all elements present, we say that an equation is **balanced**. All proper chemical equations are balanced. If a substance does not have a coefficient written in front of it, it is assumed to be 1. Also, the convention is to use all whole numbers when balancing chemical equations. This sometimes makes us do a bit more "back and forth" work when balancing a chemical equation.

Example 6.2.1

Write and balance the chemical equation for each given chemical reaction.

1. Hydrogen and chlorine react to make HCl
2. Ethane, $\mathrm{C_2H_6}$, reacts with oxygen to make carbon dioxide and water.

Solution

1. Let us start by simply writing a chemical equation in terms of the formulas of the substances, remembering that both elemental hydrogen and chlorine are diatomic:

$$\mathrm{H_2 + Cl_2 \longrightarrow HCl}$$

There are two hydrogen atoms and two chlorine atoms in the reactants and one of each atom in the product. We can fix this by including the coefficient 2 on the product side:

$$\mathrm{H_2 + Cl_2 \longrightarrow 2\,HCl}$$

Now there are two hydrogen atoms and two chlorine atoms on both sides of the chemical equation, so it is balanced.

2. Start by writing the chemical equation in terms of the substances involved:

$$\mathrm{C_2H_6 + O_2 \longrightarrow CO_2 + H_2O}$$

We have two carbon atoms on the left, so we need two carbon dioxide molecules on the product side, so that each side has two carbon atoms; that element is balanced. We have six hydrogen atoms in the reactants, so we need six hydrogen atoms in the products. We can get this by having three water molecules:

$$\mathrm{C_2H_6 + O_2 \longrightarrow 2\,CO_2 + 3\,H_2O}$$

Now we have seven oxygen atoms in the products (four from the CO_2 and three from the H_2O). This means we need seven oxygen atoms in the reactants. However, because oxygen is a diatomic molecule, we can only get an even number of oxygen atoms at a time. We can achieve this by multiplying the other coefficients by 2:

$$\mathrm{2\,C_2H_6 + O_2 \longrightarrow 4\,CO_2 + 6\,H_2O}$$

By multiplying everything else by 2, we do not unbalance the other elements, and we now get an even number of oxygen atoms in the product—14. We can get 14 oxygen atoms on the reactant side by having 7 oxygen molecules:

$$\mathrm{2\,C_2H_6 + 7\,O_2 \longrightarrow 4\,CO_2 + 6\,H_2O}$$

As a check, recount everything to determine that each side has the same number of atoms of each element. This chemical equation is now balanced.

Exercise 6.2.1

Write and balance the chemical equation that represents nitrogen and hydrogen reacting to produce ammonia, NH_3.

Answer

\[\ce{N2 + 3H2 → 2NH3 \nonumber \nonumber \]

Many chemical equations also include phase labels for the substances: (s) for solid, (ℓ) for liquid, (g) for gas, and (aq) for aqueous (i.e., dissolved in water). Special conditions, such as temperature, may also be listed above the arrow. For example,

$$\mathrm{2\,NaHCO_3\,(s) \overset{200^\circ C}{\rightarrow} Na_2CO_3\,(s) + CO_2\,(g) + H_2O\,(l)}$$

Key Takeaways

- A chemical equation is a concise description of a chemical reaction.
- Proper chemical equations are balanced.

6.2: The Chemical Equation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.2: The Chemical Equation** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.3: The Mole in Chemical Reactions

Learning Objectives

- Balance a chemical equation in terms of moles.
- Use the balanced equation to construct conversion factors in terms of moles.
- Calculate moles of one substance from moles of another substance using a balanced chemical equation.

Consider this balanced chemical equation:

$$2\,H_2 + O_2 \longrightarrow 2\,H_2O$$

We interpret this as "two molecules of hydrogen react with one molecule of oxygen to make two molecules of water." The chemical equation is balanced as long as the coefficients are in the ratio 2:1:2. For instance, this chemical equation is also balanced:

$$100\,H_2 + 50\,O_2 \longrightarrow 100\,H_2O$$

This equation is not conventional—because convention says that we use the lowest ratio of coefficients—but it is balanced. So is this chemical equation:

$$5{,}000\,H_2 + 2{,}500\,O_2 \longrightarrow 5{,}000\,H_2O$$

Again, this is not conventional, but it is still balanced. Suppose we use a much larger number:

$$12.044 \times 10^{23} H_2 + 6.022 \times 10^{23} O_2 \rightarrow 12.044 \times 10^{23} H_2O$$

These coefficients are also in the ratio of 2:1:2. But these numbers are related to the number of things in a mole: the first and last numbers are two times Avogadro's number, while the second number is Avogadro's number. That means that the first and last numbers represent 2 mol, while the middle number is just 1 mol. Well, why not just use the number of moles in balancing the chemical equation?

$$2\,H_2 + O_2 \longrightarrow 2\,H_2O$$

is the same balanced chemical equation we started with! What this means is that chemical equations are not just balanced in terms of molecules; *they are also balanced in terms of moles*. We can just as easily read this chemical equation as "two moles of hydrogen react with one mole of oxygen to make two moles of water." All balanced chemical reactions are balanced in terms of moles.

Example 6.3.1

Interpret this balanced chemical equation in terms of moles.

$$P_4 + 5\,O_2 \longrightarrow P_4O_{10}$$

Solution

The coefficients represent the number of moles that react, not just molecules. We would speak of this equation as "one mole of molecular phosphorus reacts with five moles of elemental oxygen to make one mole of tetraphosphorus decoxide."

Exercise 6.3.1

Interpret this balanced chemical equation in terms of moles.

$$N_2 + 3\,H_2 \longrightarrow 2\,NH_3$$

Answer

One mole of elemental nitrogen reacts with three moles of elemental hydrogen to produce two moles of ammonia.

In Section 4.1, we stated that a chemical equation is simply a recipe for a chemical reaction. As such, chemical equations also give us equivalents—equivalents between the reactants and the products. However, now we understand that *these equivalents are expressed in terms of moles*. Consider the chemical equation

$$2\,H_2 + O_2 \longrightarrow 2\,H_2O$$

This chemical reaction gives us the following equivalents:

$$2 \text{ mol } H_2 \Leftrightarrow 1 \text{ mol } O_2 \Leftrightarrow 2 \text{ mol } H_2O$$

Any two of these quantities can be used to construct a conversion factor that lets us relate the number of moles of one substance to an equivalent number of moles of another substance. If, for example, we want to know how many moles of oxygen will react with 17.6 mol of hydrogen, we construct a conversion factor between 2 mol of H_2 and 1 mol of O_2 and use it to convert from moles of one substance to moles of another:

$$17.6\,\cancel{mol\,H_2} \times \frac{1\,mol\,O_2}{2\,\cancel{mol\,H_2}} = 8.80\,mol\,O_2$$

Note how the mol H_2 unit cancels, and mol O_2 is the new unit introduced. This is an example of a **mole-mole calculation,** when you start with moles of one substance and convert to moles of another substance by using the balanced chemical equation. The example may seem simple because the numbers are small, but numbers won't always be so simple!

Example 6.3.2

For the balanced chemical equation

$$2\,C_4H_{10}\,(g) + 13\,O_2 \longrightarrow 8\,CO_2\,(g) + 10\,H_2O(\ell)$$

if 154 mol of O_2 are reacted, how many moles of CO_2 are produced?

Solution

We are relating an amount of oxygen to an amount of carbon dioxide, so we need the equivalence between these two substances. According to the balanced chemical equation, the equivalence is

$$13 \text{ mol } O_2 \Leftrightarrow 8 \text{ mol } CO_2$$

We can use this equivalence to construct the proper conversion factor. We start with what we are given and apply the conversion factor:

$$154\,\cancel{mol\,O_2} \times \frac{8\,mol\,CO_2}{13\,\cancel{mol\,O_2}} = 94.8\,mol\,CO_2$$

The mol O_2 unit is in the denominator of the conversion factor so it cancels. Both the 8 and the 13 are exact numbers, so they do not contribute to the number of significant figures in the final answer.

Exercise 6.3.2

Using the above equation, how many moles of H_2O are produced when 154 mol of O_2 react?

Answer

118 mol

It is important to reiterate that balanced chemical equations are balanced in terms of *moles*. Not grams, kilograms, or liters—but moles. Any stoichiometry problem will likely need to work through the mole unit at some point, especially if you are working with a balanced chemical reaction.

Summary

Balanced chemical reactions are balanced in terms of moles. A balanced chemical reaction gives equivalents in moles that allow stoichiometry calculations to be performed.

6.3: The Mole in Chemical Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **5.4: The Mole in Chemical Reactions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.4: Stoichiometry

Learning Objectives

- Define *stoichiometry*.
- Relate quantities in a balanced chemical reaction on a molecular basis.

Consider a classic recipe for pound cake: 1 pound of eggs, 1 pound of butter, 1 pound of flour, and 1 pound of sugar. (That's why it's called "pound cake.") If you have 4 pounds of butter, how many pounds of sugar, flour, and eggs do you need? You would need 4 pounds each of sugar, flour, and eggs.

Now suppose you have 1.00 g H_2. If the chemical reaction follows the balanced chemical equation

$$2\,H_2\,(g) + O_2\,(g) \longrightarrow 2\,H_2O(\ell)$$

then what mass of oxygen do you need to make water?

Curiously, this chemical reaction question is very similar to the pound cake question. Both of them involve relating a quantity of one substance to a quantity of another substance or substances. The relating of one chemical substance to another using a balanced chemical reaction is called **stoichiometry**. Using stoichiometry is a fundamental skill in chemistry; it greatly broadens your ability to predict what will occur and, more importantly, how much is produced.

Let us consider a more complicated example. A recipe for pancakes calls for 2 cups (c) of pancake mix, 1 egg, and 1/2 c of milk. We can write this in the form of a chemical equation:

2 c mix + 1 egg + 1/2 c milk → 1 batch of pancakes

If you have 9 c of pancake mix, how many eggs and how much milk do you need? It might take a little bit of work, but eventually you will find you need 4½ eggs and 2¼ c milk.

How can we formalize this? We can make a conversion factor using our original recipe and use that conversion factor to convert from a quantity of one substance to a quantity of another substance. This is similar to the way we constructed a conversion factor between feet and yards in Chapter 2. Because one recipe's worth of pancakes requires 2 c of pancake mix, 1 egg, and 1/2 c of milk, we actually have the following mathematical relationships that relate these quantities:

2 c pancake mix ⇔ 1 egg ⇔ 1/2 c milk

where ⇔ is the mathematical symbol for "is equivalent to." This does not mean that 2 c of pancake mix equal 1 egg. However, *as far as this recipe is concerned*, these are the equivalent quantities needed for a single recipe of pancakes. So, any possible quantities of two or more ingredients must have the same numerical ratio as the ratios in the equivalence.

We can deal with these equivalences in the same way we deal with equalities in unit conversions: we can make conversion factors that essentially equal 1. For example, to determine how many eggs we need for 9 c of pancake mix, we construct the conversion factor:

$$\frac{1\,egg}{2\,c\,pancake\,mix}$$

This conversion factor is, in a strange way, equivalent to 1 because the recipe relates the two quantities. Starting with our initial quantity and multiplying by our conversion factor,

$$9\,\cancel{c\,pancake\,mix} \times \frac{1\,egg}{\cancel{2\,c\,pancake\,mix}} = 4.5\,eggs$$

Note how the units *cups pancake mix* canceled, leaving us with units of *eggs*. This is the formal, mathematical way of getting our amounts to mix with 9 c of pancake mix. We can use a similar conversion factor for the amount of milk:

$$9\,\cancel{c\,pancake\,mix} \times \frac{\frac{1}{2}\,c\,milk}{\cancel{2\,c\,pancake\,mix}} = 2.25\,c\,milk$$

Again, units cancel, and new units are introduced.

A balanced chemical equation is nothing more than *a recipe for a chemical reaction*. The difference is that a balanced chemical equation is written in terms of atoms and molecules, not cups, pounds, and eggs.

For example, consider the following chemical equation:

$$2\,H_2\,(g) + O_2\,(g) \longrightarrow 2\,H_2O(\ell)$$

We can interpret this as, literally, "two hydrogen molecules react with one oxygen molecule to make two water molecules." That interpretation leads us directly to some equivalencies, just as our pancake recipe did:

$$2H_2 \text{ molecules} \Leftrightarrow 1O_2 \text{ molecule} \Leftrightarrow 2H_2O \text{ molecules}$$

These equivalences allow us to construct conversion factors:

$$\frac{2\,molecules\,H_2}{1\,molecule\,O_2} \quad \frac{2\,molecules\,H_2}{2\,molecules\,H_2O} \quad \frac{1\,molecule\,H_2}{2\,molecules\,H_2O}$$

and so forth. These conversions can be used to relate quantities of one substance to quantities of another. For example, suppose we need to know how many molecules of oxygen are needed to react with 16 molecules of H_2. As we did with converting units, we start with our given quantity and use the appropriate conversion factor:

$$\cancel{16\,molecules\,H_2} \times \frac{1\,molecules\,O_2}{\cancel{2\,molecules\,H_2}} = 8\,molecules\,O_2$$

Note how the unit *molecules* H_2 cancels algebraically, just as any unit does in a conversion like this. The conversion factor came directly from the coefficients in the balanced chemical equation. This is another reason why a properly balanced chemical equation is important.

✔ Example 6.4.1

How many molecules of SO_3 are needed to react with 144 molecules of Fe_2O_3 given this balanced chemical equation?

$$Fe_2O_3 + 3\,SO_3 \longrightarrow Fe_2(SO_4)_3$$

Solution

We use the balanced chemical equation to construct a conversion factor between Fe_2O_3 and SO_3. The number of molecules of Fe_2O_3 goes on the bottom of our conversion factor so it cancels with our given amount, and the molecules of SO_3 go on the top. Thus, the appropriate conversion factor is

$$\frac{3\,molecules\,SO_3}{1\,molecule\,Fe_2O_3}$$

Starting with our given amount and applying the conversion factor, the result is

$$144\ \cancel{molecules}\ Fe_2O_3 \times \frac{3\,molecules\,SO_3}{1\ \cancel{molecule}\ Fe_2O_3} = 432\,molecules\,SO_3$$

We need 432 molecules of SO_3 to react with 144 molecules of Fe_2O_3.

? Exercise 6.4.1

How many molecules of H_2 are needed to react with 29 molecules of N_2 to make ammonia if the balanced chemical equation is:

$$N_2 + 3\,H_2 \longrightarrow 2\,NH_3?$$

Answer

87 molecules

Chemical equations also allow us to make conversions regarding the number of atoms in a chemical reaction, because a chemical formula lists the number of atoms of each element in a compound. The formula H_2O indicates that there are two hydrogen atoms and one oxygen atom in each molecule, and these relationships can be used to make conversion factors:

$$\frac{2\,atoms\,H}{1\,molecule\,H_2O} \quad \frac{1\,molecule\,H_2O}{1\,atom\,O}$$

Conversion factors like this can also be used in stoichiometry calculations.

Example 6.4.2

How many molecules of NH_3 can you make if you have 228 atoms of H_2?

Solution

From the formula, we know that one molecule of NH_3 has three H atoms. Use that fact as a conversion factor:

$$228\ \cancel{atoms\,H} \times \frac{1\,molecule\,NH_3}{3\ \cancel{atoms\,H}} = 76\,NH_3$$

Exercise 6.4.2

How many molecules of $Fe_2(SO_4)_3$ can you make from 777 atoms of S?

Answer

259 molecules

Summary

Quantities of substances can be related to each other using balanced chemical equations.

6.4: Stoichiometry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **5.2: Stoichiometry** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.5: Limiting Reagents

Learning Objectives

- Identify a limiting reagent from a set of reactants.
- Calculate how much product will be produced from the limiting reagent.
- Calculate how much reactant(s) remains when the reaction is complete.

In addition to the assumption that reactions proceed all the way to completion, one additional assumption we have made about chemical reactions is that all the reactants are present in the proper quantities to react to products; this is not always the case. In Figure **6.5.2** we are taking hydrogen atoms and oxygen atoms (left) to make water molecules (right). However, there are not enough oxygen atoms to use up all the hydrogen atoms. We run out of oxygen atoms and cannot make any more water molecules, so the process stops when we run out of oxygen atoms.

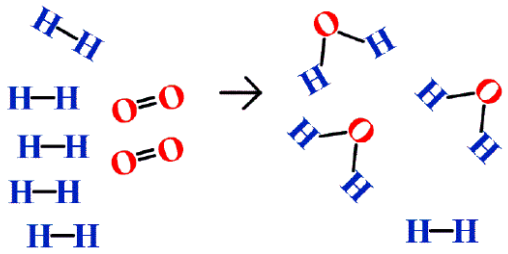


Figure **6.5.2**: Making Water. In this scenario for making water molecules, we run out of O atoms before we use up all the H atoms. Similar situations exist for many chemical reactions when one reactant runs out before the other.

A similar situation exists for many chemical reactions: you usually run out of one reactant before all of the other reactant has reacted. The reactant you run out of is called the limiting reagent; the other reactant or reactants are considered to be *in excess*. A crucial skill in evaluating the conditions of a chemical process is to determine which reactant is the limiting reagent and which is in excess.

The key to recognizing which reactant is the limiting reagent is based on a mole-mass or mass-mass calculation: whichever reactant gives the *lesser* amount of product is the limiting reagent. What we need to do is determine an amount of one product (either moles or mass), assuming all of each reactant reacts. Whichever reactant gives the least amount of that particular product is the limiting reagent. It does not matter which product we use, as long as we use the same one each time. It does not matter whether we determine the number of moles or grams of that product; however, we will see shortly that knowing the final mass of product can be useful.

For example, consider this reaction:

$$4As(s) + 3O_2(g) \rightarrow 2As_2O_3(s)$$

Suppose we start a reaction with 50.0 g of As and 50.0 g of O_2. Which one is the limiting reagent? We need to perform two mole-mass calculations, each assuming that each reactant reacts completely. Then we compare the amount of the product produced by each and determine which is less.

The calculations are as follows:

$$50.0\ \cancel{g\,As} \times \frac{1\ \cancel{mol\,As}}{74.92\ \cancel{g\,As}} \times \frac{2\,mol\,As_2O_3}{4\ \cancel{mol\,As}} = 0.334\,mol\,As_2O_3$$

$$50.0\ \cancel{g\,O_2} \times \frac{1\ \cancel{mol\,O_2}}{32.00\ \cancel{g\,O_2}} \times \frac{2\,mol\,As_2O_3}{3\ \cancel{mol\,O_2}} = 1.04\,mol\,As_2O_3$$

Comparing these two answers, it is clear that 0.334 mol of As_2O_3 is less than 1.04 mol of As_2O_3, so arsenic is the limiting reagent. If this reaction is performed under these initial conditions, the arsenic will run out before the oxygen runs out. We say that the oxygen is "in excess."

Identifying the limiting reagent, then, is straightforward. However, there are usually two associated questions: (1) what mass of product (or products) is then actually formed? and (2) what mass of what reactant is left over? The first question is straightforward to answer: simply perform a conversion from the number of moles of product formed to its mass, using its molar mass. For As_2O_3, the molar mass is 197.84 g/mol; knowing that we will form 0.334 mol of As_2O_3 under the given conditions, we will get

$$0.334\ \cancel{mol\ As_2O_3} \times \frac{197.84\ g\ As_2}{1\ \cancel{mol\ As_2O_3}} = 66.1\ g\ As_2O_3$$

The second question is somewhat more convoluted to answer. First, we must do a mass-mass calculation relating the limiting reagent (here, As) to the other reagent (O_2). Once we determine the mass of O_2 that reacted, we subtract that from the original amount to determine the amount left over. According to the mass-mass calculation,

$$50.0\ \cancel{g\ As} \times \frac{1\ \cancel{mol\ As}}{74.92\ \cancel{g\ As}} \times \frac{3\ \cancel{mol\ O_2}}{4\ \cancel{mol\ As}} \times \frac{32.00\ g\ O_2}{1\ \cancel{mol\ O_2}} = 16.0\ g\ O_2\ reacted$$

Because we reacted 16.0 g of our original O_2, we subtract that from the original amount, 50.0 g, to get the mass of O_2 remaining:

50.0 g O_2 − 16.0 g O_2 reacted = 34.0 g O_2 left over

You must remember to perform this final subtraction to determine the amount remaining; a common error is to report the 16.0 g as the amount remaining.

Example 6.5.1

A 5.00 g quantity of Rb is combined with 3.44 g of $MgCl_2$ according to this chemical reaction:

$$2Rb(s) + MgCl_2(s) \rightarrow Mg(s) + 2RbCl(s)$$

What mass of Mg is formed, and what mass of what reactant is left over?

Solution

Because the question asks what mass of magnesium is formed, we can perform two mass-mass calculations and determine which amount is less.

$$5.00\ \cancel{g\ Rb} \times \frac{1\ \cancel{mol\ Rb}}{85.47\ \cancel{g\ Rb}} \times \frac{1\ \cancel{mol\ Mg}}{2\ \cancel{mol\ Rb}} \times \frac{24.31\ g\ Mg}{1\ \cancel{mol\ Mg}} = 0.711\ g\ Mg$$

$$3.44\ \cancel{g\ MgCl_2} \times \frac{1\ \cancel{mol\ MgCl_2}}{95.21\ \cancel{g\ MgCl_2}} \times \frac{1\ \cancel{mol\ Mg}}{1\ \cancel{mol\ MgCl_2}} \times \frac{24.31\ g\ Mg}{1\ \cancel{mol\ Mg}} = 0.878\ g\ Mg$$

The 0.711 g of Mg is the lesser quantity, so the associated reactant—5.00 g of Rb—is the limiting reagent. To determine how much of the other reactant is left, we have to do one more mass-mass calculation to determine what mass of MgCl2 reacted with the 5.00 g of Rb, and then subtract the amount reacted from the original amount.

$$5.00\ \cancel{g\ Rb} \times \frac{1\ \cancel{mol\ Rb}}{85.47\ \cancel{g\ Rb}} \times \frac{1\ \cancel{mol\ MgCl_2}}{2\ \cancel{mol\ Rb}} \times \frac{95.21\ g\ Mg}{1\ \cancel{mol\ MgCl_2}} = 2.78\ g\ MgCl_2\ reacted$$

Because we started with 3.44 g of MgCl2, we have

3.44 g MgCl2 − 2.78 g MgCl2 reacted = 0.66 g MgCl2 left

Exercise 6.5.1

Given the initial amounts listed, what is the limiting reagent, and what is the mass of the leftover reagent?

$$\underbrace{22.7\ g}_{MgO(s)} + \underbrace{17.9\ g}_{H_2S} \rightarrow MgS(s) + H_2O(l)$$

Answer

H_2S is the limiting reagent; 1.5 g of MgO are left over.

Summary

The limiting reagent is the reactant that produces the least amount of product. Mass-mass calculations can determine how much product is produced and how much of the other reactants remain.

6.5: Limiting Reagents is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **5.7: Limiting Reagents** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.6: Types of Chemical Reactions - Single and Double Displacement Reactions

Learning Objectives

- Recognize chemical reactions as single-replacement reactions and double-replacement reactions.
- Use the periodic table, an activity series, or solubility rules to predict whether single-replacement reactions or double-replacement reactions will occur.

Up until now, we have presented chemical reactions as a topic, but we have not discussed how the products of a chemical reaction can be predicted. Here we will begin our study of certain types of chemical reactions that allow us to predict what the products of the reaction will be.

A **single-replacement reaction** is a chemical reaction in which one element is substituted for another element in a compound, generating a new element and a new compound as products. Presented below:

$$2\,HCl\,(aq) + Zn\,(s) \longrightarrow ZnCl_2\,(aq) + H_2\,(g)$$

is an example of a single-replacement reaction. The hydrogen atoms in **HCl** are replaced by **Zn** atoms, and in the process a new element—hydrogen—is formed. Another example of a single-replacement reaction is

$$2\,NaCl\,(aq) + F_2\,(g) \longrightarrow 2\,NaF\,(s) + Cl_2\,(g)$$

Here the negatively charged ion changes from chloride to fluoride. A typical characteristic of a single-replacement reaction is that there is one element as a reactant and another element as a product.

Not all proposed single-replacement reactions will occur between two given reactants. This is most easily demonstrated with fluorine, chlorine, bromine, and iodine. Collectively, these elements are called the *halogens* and are in the next-to-last column on the periodic table (Figure **6.6.1**). The elements on top of the column will replace the elements below them on the periodic table, but not the other way around. Thus, the reaction represented by

$$CaI_2\,(s) + Cl_2\,(g) \longrightarrow CaCl_2\,(s) + I_2\,(s)$$

will occur; but the reaction

$$CaF_2\,(s) + Br_2(\ell) \longrightarrow CaBr_2\,(s) + F_2\,(g)$$

will not, because bromine is below fluorine on the periodic table. This is just one of many ways the periodic table helps us to understand chemistry.

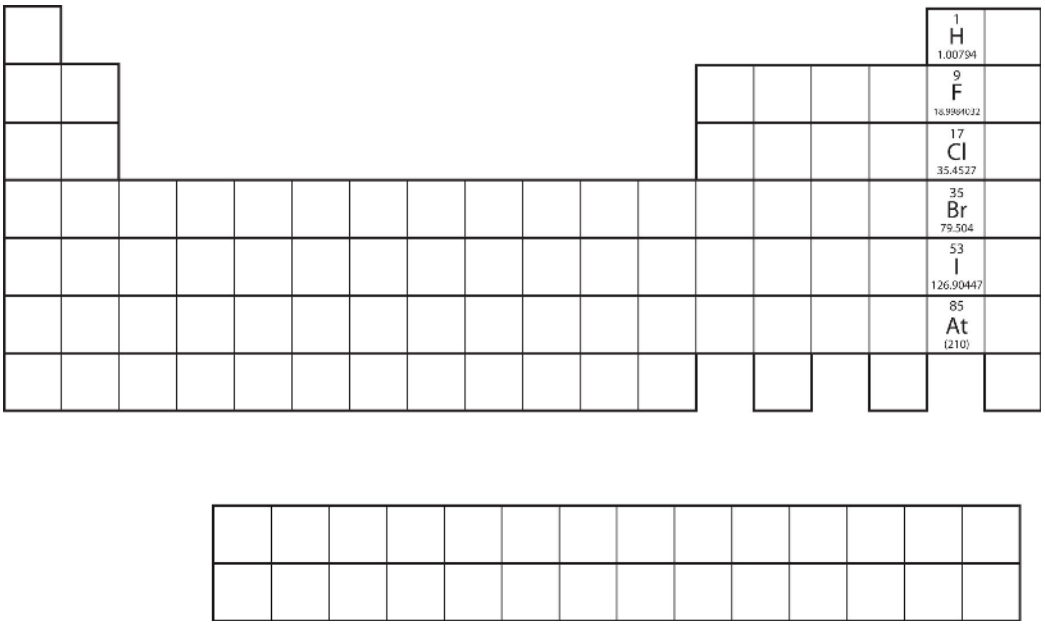


Figure **6.6.1**: Halogens on the Periodic Table. The halogens are the elements in the next-to-last column on the periodic table.

Example 6.6.1

Will a single-replacement reaction occur? If so, identify the products.

a. $MgCl_2 + I_2 \rightarrow ?$
b. $CaBr_2 + F_2 \rightarrow ?$

Solution

a. Because iodine is below chlorine on the periodic table, a single-replacement reaction will not occur.
b. Because fluorine is above bromine on the periodic table, a single-replacement reaction will occur, and the products of the reaction will be CaF_2 and Br_2.

Exercise 6.6.1

Will a single-replacement reaction occur? If so, identify the products.

$$\mathbf{FeI_2 + Cl_2 \longrightarrow}$$

Answer

Yes; $FeCl_2$ and I_2

Chemical reactivity trends are easy to predict when replacing anions in simple ionic compounds—simply use their relative positions on the periodic table. However, when replacing the cations, the trends are not as straightforward. This is partly because there are so many elements that can form cations; an element in one column on the periodic table may replace another element nearby, or it may not. A list called the **activity series** does the same thing the periodic table does for halogens: it lists the elements that will replace elements below them in single-replacement reactions. A simple activity series is shown below.

Activity Series for Cation Replacement in Single-Replacement Reactions

- Li
- K
- Ba
- Sr
- Ca

- Na
- Mg
- Al
- Mn
- Zn
- Cr
- Fe
- Ni
- Sn
- Pb
- H_2
- Cu
- Hg
- Ag
- Pd
- Pt
- Au

Using the activity series is similar to using the positions of the halogens on the periodic table. An element on top will replace an element below it in compounds undergoing a single-replacement reaction. Elements will not replace elements above them in compounds.

Example 6.6.2

Use the activity series to predict the products, if any, of each equation.

1. $FeCl_2 + Zn \rightarrow ?$
2. $HNO_3 + Au \rightarrow ?$

Solution

1. Because zinc is above iron in the activity series, it will replace iron in the compound. The products of this single-replacement reaction are $ZnCl_2$ and Fe.
2. Gold is below hydrogen in the activity series. As such, it will not replace hydrogen in a compound with the nitrate ion. No reaction is predicted.

Exercise 6.6.2

Use the activity series to predict the products, if any, of this equation.

$$AlPO_4 + Mg \longrightarrow$$

Answer

$Mg_3(PO_4)_2$ and Al

A **double-replacement reaction** occurs when parts of two ionic compounds are exchanged, making two new compounds. A characteristic of a double-replacement equation is that there are two compounds as reactants and two different compounds as products. An example is

$$CuCl_2\,(aq) + 2\,AgNO_3\,(aq) \longrightarrow Cu(NO_3)_2\,(aq) + 2\,AgCl\,(s)$$

There are two equivalent ways of considering a double-replacement equation: either the cations are swapped, or the anions are swapped. (You cannot swap both; you would end up with the same substances you started with.) Either perspective should allow you to predict the proper products, as long as you pair a cation with an anion, and not a cation with a cation or an anion with an anion.

Example 6.6.3

Predict the products of this double-replacement equation:

$$\mathrm{BaCl_2 + Na_2SO_4 \longrightarrow}$$

Solution

Thinking about the reaction as either switching the cations or switching the anions, we would expect the products to be $BaSO_4$ and NaCl.

Exercise 6.6.3

Predict the products of this double-replacement equation:

$$\mathrm{KBr + AgNO_3 \longrightarrow}$$

Answer

KNO_3 and AgBr

Predicting whether a double-replacement reaction occurs is somewhat more difficult than predicting a single-replacement reaction. However, there is one type of double-replacement reaction that we can predict: the precipitation reaction. A **precipitation reaction** occurs when two ionic compounds are dissolved in water and form a new ionic compound that does not dissolve; this new compound falls out of solution as a solid **precipitate.** The formation of a solid precipitate is the driving force that makes the reaction proceed.

To judge whether double-replacement reactions will occur, we need to know what kinds of ionic compounds form precipitates. For this, we use **solubility rules**, which are general statements that predict which ionic compounds dissolve (are soluble) and which do not (are not soluble, or insoluble). Table **6.6.1** lists some general solubility rules. We need to consider each ionic compound (both the reactants and the possible products) in light of the solubility rules. If a compound is soluble, we use the (aq) label with it, indicating that it dissolves. If a compound is not soluble, we use the (s) label with it and assume that it will precipitate out of solution. If everything is soluble, then no reaction will be expected.

Table **6.6.1**: Some Useful Solubility Rules (soluble)

These compounds generally dissolve in water (are soluble):	Exceptions:
All compounds of Li^+, Na^+, K^+, Rb^+, Cs^+, and NH_4^+	None
All compounds of NO_3^- and $C_2H_3O_2^-$	None
Compounds of Cl^-, Br^-, I^-	Ag^+, Hg_2^{2+}, Pb^{2+}
Compounds of SO_4^{2}	Hg_2^{2+}, Pb^{2+}, Sr^{2+}, Ba^{2+}

Table **6.6.2**: Some Useful Solubility Rules (insoluble)

These compounds generally do not dissolve in water (are insoluble):	Exceptions:
Compounds of CO_3^{2-} and PO_4^{3-}	Compounds of Li^+, Na^+, K^+, Rb^+, Cs^+, and NH_4^+
Compounds of OH^-	Compounds of Li^+, Na^+, K^+, Rb^+, Cs^+, NH_4^+, Sr^{2+}, and Ba^{2+}

For example, consider the possible double-replacement reaction between Na_2SO_4 and $SrCl_2$. The solubility rules say that all ionic sodium compounds are soluble and all ionic chloride compounds are soluble, except for Ag^+, Hg_2^{2+}, and Pb^{2+}, which are not being considered here. Therefore, Na_2SO_4 and $SrCl_2$ are both soluble. The possible double-replacement reaction products are NaCl and $SrSO_4$. Are these soluble? NaCl is (by the same rule we just quoted), but what about $SrSO_4$? Compounds of the sulfate

ion are generally soluble, but Sr^{2+} is an exception: we expect it to be insoluble—a precipitate. Therefore, we expect a reaction to occur, and the balanced chemical equation would be:

$$Na_2SO_4\,(aq) + SrCl_2\,(aq) \longrightarrow 2\,NaCl\,(aq) + SrSO_4\,(s)$$

You would expect to see a visual change corresponding to $SrSO_4$ precipitating out of solution (Figure **6.6.2**).



Figure **6.6.2**: Double-Replacement Reactions. Some double-replacement reactions are obvious because you can see a solid precipitate coming out of solution. Source: Photo courtesy of Choij, http://commons.wikimedia.org/wiki/File:Copper_solution.jpg(opens in new window).

Example 6.6.4

Will a double-replacement reaction occur? If so, identify the products.

1. $Ca(NO_3)_2$ + KBr → ?
2. NaOH + $FeCl_2$ → ?

Solution

1. According to the solubility rules, both $Ca(NO_3)_2$ and KBr are soluble. Now we consider what the double-replacement products would be by switching the cations (or the anions)—namely, $CaBr_2$ and KNO_3. However, the solubility rules predict that these two substances would also be soluble, so no precipitate would form. Thus, we predict no reaction in this case.
2. According to the solubility rules, both NaOH and $FeCl_2$ are expected to be soluble. If we assume that a double-replacement reaction may occur, we need to consider the possible products, which would be NaCl and $Fe(OH)_2$. NaCl is soluble, but, according to the solubility rules, $Fe(OH)_2$ is not. Therefore, a reaction would occur, and $Fe(OH)_2$(s) would precipitate out of solution. The balanced chemical equation is

$$2\,NaOH\,(aq) + FeCl_2\,(aq) \longrightarrow 2\,NaCl\,(aq) + Fe(OH)_2\,(s)$$

Exercise 6.6.4

$$Sr(NO_3)_2 + KCl \longrightarrow$$

Answer

No reaction; all possible products are soluble.

Key Takeaways

- A single-replacement reaction replaces one element for another in a compound.
- The periodic table or an activity series can help predict whether single-replacement reactions occur.
- A double-replacement reaction exchanges the cations (or the anions) of two ionic compounds.
- A precipitation reaction is a double-replacement reaction in which one product is a solid precipitate.
- Solubility rules are used to predict whether some double-replacement reactions will occur.

6.6: Types of Chemical Reactions - Single and Double Displacement Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.3: Types of Chemical Reactions - Single and Double Replacement Reactions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.7: Composition, Decomposition, and Combustion Reactions

Learning Objectives

- Recognize composition, decomposition, and combustion reactions.
- Predict the products of a combustion reaction.

Three classifications of chemical reactions will be reviewed in this section. Predicting the products in some of them may be difficult, but the reactions are still easy to recognize.

A **composition reaction** (sometimes also called a *combination reaction* or a *synthesis reaction*) produces a single substance from multiple reactants. A single substance as a product is the key characteristic of the composition reaction. There may be a coefficient other than one for the substance, but if the reaction has only a single substance as a product, it can be called a composition reaction. In the reaction

$$2\,H_2\,(g) + O_2\,(g) \longrightarrow 2\,H_2O(\ell)$$

water is produced from hydrogen and oxygen. Although there are two molecules of water being produced, there is only one substance—water—as a product. So this is a composition reaction.

A **decomposition reaction** starts from a single substance and produces more than one substance; that is, it decomposes. The key characteristics of a decomposition reaction are: one substance as a reactant and more than one substance as the products. For example, in the decomposition of sodium hydrogen carbonate (also known as sodium bicarbonate):

$$2\,NaHCO_3\,(s) \longrightarrow Na_2CO_3\,(s) + CO_2\,(g) + H_2O(\ell)$$

sodium carbonate, carbon dioxide, and water are produced from the single substance sodium hydrogen carbonate.

Composition and decomposition reactions are difficult to predict; however, they should be easy to recognize.

Example 6.7.1: Identifying Reactions

Identify each equation as a composition reaction, a decomposition reaction, or neither.

a. $Fe_2O_3 + 3\,SO_3 \longrightarrow Fe_2(SO_4)_3$
b. $NaCl + AgNO_3 \longrightarrow AgCl + NaNO_3$
c. $(NH_4)_2Cr_2O_7 \longrightarrow Cr_2O_3 + 4\,H_2O + N_2$

Solution

a. In this equation, two substances combine to make a single substance. This is a composition reaction.
b. Two different substances react to make two new substances. This does not fit the definition of either a composition reaction or a decomposition reaction, so it is neither. In fact, you may recognize this as a double-replacement reaction.
c. A single substance reacts to make multiple substances. This is a decomposition reaction.

Exercise 6.7.1

Identify the equation as a composition reaction, a decomposition reaction, or neither.

$$C_3H_8 \longrightarrow C_3H_4 + 2\,H_2$$

Answer

decomposition

A **combustion reaction** occurs when a reactant combines with oxygen, many times from the atmosphere, to produce oxides of all other elements as products; any nitrogen in the reactant is converted to elemental nitrogen, N_2. Many reactants, called *fuels*,

contain mostly carbon and hydrogen atoms, reacting with oxygen to produce CO_2 and H_2O. For example, the balanced chemical equation for the combustion of methane, CH_4, is as follows:

$$\mathrm{CH_4 + 2\,O_2 \longrightarrow CO_2 + 2\,H_2O}$$

Kerosene can be approximated with the formula $\mathrm{C_{12}H_{26}}$, and its combustion equation is:

$$\mathrm{2\,C_{12}H_{26} + 37\,O_2 \longrightarrow 24\,CO\text{-}2 + 26\,H_2O}$$

Sometimes fuels contain oxygen atoms, which must be counted when balancing the chemical equation. One common fuel is ethanol, $\mathrm{C_2H_5OH}$, whose combustion equation is:

$$\mathrm{C_2H_5OH + 3\,O_2 \longrightarrow 2\,CO_2 + 3\,H_2O}$$

If nitrogen is present in the original fuel, it is converted to $\mathrm{N_2}$, not to a nitrogen-oxygen compound. Thus, for the combustion of the fuel dinitroethylene, whose formula is $\mathrm{C_2H_2N_2O_4}$, we have:

$$\mathrm{2\,C_2H_2N_2O_4 + O_2 \longrightarrow 4\,CO_2 + 2\,H_2O + 2\,N_2}$$

Example 6.7.2: Combustion Reactions

Complete and balance each combustion equation.

1. the combustion of propane ($\mathrm{C_3H_8}$)
2. the combustion of ammonia ($\mathrm{NH_3}$)

Solution

1. The products of the reaction are CO_2 and H_2O, so our unbalanced equation is

   $$\mathrm{C_3H_8 + O_2 \longrightarrow CO_2 + H_2O}$$

   Balancing (and you may have to go back and forth a few times to balance this), we get

   $$\mathrm{C_3H_8 + 5\,O_2 \longrightarrow 3\,CO_2 + 4\,H_2O}$$

2. The nitrogen atoms in ammonia will react to make N_2, while the hydrogen atoms will react with O_2 to make H_2O:

   $$\mathrm{NH_3 + O_2 \longrightarrow N_2 + H_2O}$$

   To balance this equation without fractions (which is the convention), we get

   $$\mathrm{4\,NH_3 + 3\,O_2 \longrightarrow 2\,N_2 + 6\,H_2O}$$

Exercise 6.7.2

Complete and balance the combustion equation for cyclopropanol (\ce{C3H6O}\)).

Answer

$$\mathrm{C_3H_6O + 4\,O_2 \longrightarrow 3\,CO_2 + 3\,H_2O}$$



Figure **6.7.1** A 20 lb (9.1 kg) steel propane cylinder.

Key Takeaways

- A composition reaction produces a single substance from multiple reactants.
- A decomposition reaction produces multiple products from a single reactant.
- Combustion reactions are the combination of some compound with oxygen to make oxides of the other elements as products (although nitrogen atoms react to make N_2).

6.7: Composition, Decomposition, and Combustion Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.5: Composition, Decomposition, and Combustion Reactions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.8: Neutralization Reactions

Learning Objectives

- Identify an acid and a base.
- Identify a neutralization reaction and predict its products.

In Chapter 3, we defined an acid as an ionic compound that contains H^+ as the cation. This is slightly incorrect, but until additional concepts were developed, a better definition needed to wait. Now we can redefine an acid: an **acid** is any compound that increases the amount of hydrogen ion (H^+) in an aqueous solution. The chemical opposite of an acid is a base. The equivalent definition of a base is that a **base** is a compound that increases the amount of hydroxide ion (OH^-) in an aqueous solution. These original definitions were proposed by Arrhenius (the same person who proposed ion dissociation) in 1884, so they are referred to as the **Arrhenius definition** of an acid and a base, respectively.

You may recognize that, based on the description of a hydrogen atom, an H^+ ion is a hydrogen atom that has lost its lone electron; that is, H^+ is simply a proton. Do we really have bare protons moving about in aqueous solution? No. What is more likely is that the H^+ ion has attached itself to one (or more) water molecule(s). To represent this chemically, we define the **hydronium ion** as H_3O^+, which represents an additional proton attached to a water molecule. We use the hydronium ion as the more logical way that a hydrogen ion appears in an aqueous solution, although in many chemical reactions H^+ and H_3O^+ are treated equivalently.

The reaction of an acid and a base is called a **neutralization reaction**. Although acids and bases have their own unique chemistries, the acid and base cancel each other's chemistry to produce a rather innocuous substance—water. In fact, the general reaction between an acid and a base is:

$$\text{acid} + \text{base} \longrightarrow \text{water} + \text{salt}$$

where the term **salt** is generally used to define any ionic compound (soluble or insoluble) that is formed from a reaction between an acid and a base. (In chemistry, the word *salt* refers to more than just table salt.) For example, the balanced chemical equation for the reaction between HCl(aq) and KOH(aq) is

$$\mathrm{HCl\,(aq) + KOH\,(aq) \longrightarrow H_2O(\ell) + KCl\,(aq)}$$

where the salt is KCl. By counting the number of atoms of each element, we find that only one water molecule is formed as a product. However, in the reaction between HCl(aq) and $Mg(OH)_2$(aq), additional molecules of HCl and H_2O are required to balance the chemical equation:

$$\mathrm{2\,HCl\,(aq) + Mg(OH)_2\,(aq) \longrightarrow 2\,H_2O(\ell) + MgCl_2\,(aq)}$$

Here, the salt is $MgCl_2$. (This is one of several reactions that take place when a type of antacid—a base—is used to treat stomach acid.)

Example 6.8.1

Write the neutralization reactions between each acid and base.

a. HNO_3(aq) and $Ba(OH)_2$(aq)
b. H_3PO_4(aq) and $Ca(OH)_2$(aq)

Solution

First, we will write the chemical equation with the formulas of the reactants and the expected products; then we will balance the equation.

a. The expected products are water and barium nitrate, so the initial chemical reaction is

$$\mathrm{HNO_3\,(aq) + Ba(OH)_2\,(aq) \longrightarrow H_2O(\ell) + Ba(NO_3)_2\,(aq)}$$

To balance the equation, we need to realize that there will be two H_2O molecules, so two HNO_3 molecules are required:

$$2\,HNO_3\,(aq) + Ba(OH)_2\,(aq) \longrightarrow 2\,H_2O(\ell) + Ba(NO_3)_2\,(aq)$$

This chemical equation is now balanced.

b. The expected products are water and calcium phosphate, so the initial chemical equation is

$$H_3PO_4\,(aq) + Ca(OH)_2\,(aq) \longrightarrow H_2O(\ell) + Ca_3(PO_4)_2\,(s)$$

According to the solubility rules, $Ca_3(PO_4)_2$ is insoluble, so it has an (s) phase label. To balance this equation, we need two phosphate ions and three calcium ions. We end up with six water molecules to balance the equation:

$$2\,H_3PO_4\,(aq) + 3\,Ca(OH)_2\,(aq) \longrightarrow 6\,H_2O(\ell) + Ca_3(PO_4)_2\,(s)$$

This chemical equation is now balanced.

? Exercise 6.8.1

Write the neutralization reaction between H_2SO_4(aq) and $Sr(OH)_2$(aq).

Answer

$$H_2SO_4\,(aq) + Sr(OH)_2\,(aq) \longrightarrow 2\,H_2O(\ell) + SrSO_4\,(aq)$$

Neutralization reactions are one type of chemical reaction that proceeds even if one reactant is not in the aqueous phase. For example, the chemical reaction between HCl(aq) and $Fe(OH)_3$(s) still proceeds according to the equation:

$$3\,HCl\,(aq) + Fe(OH)_3\,(s) \longrightarrow 3\,H_2O(\ell) + FeCl_3\,(aq)$$

even though $Fe(OH)_3$ is not soluble. When one realizes that $Fe(OH)_3$(s) is a component of rust, this explains why some cleaning solutions for rust stains contain acids—the neutralization reaction produces products that are soluble and wash away. (Washing with acids like HCl is one way to remove rust and rust stains, but HCl must be used with caution!)

Complete and net ionic reactions for neutralization reactions will depend on whether the reactants and products are soluble, even if the acid and base react. For example, in the reaction of HCl(aq) and NaOH(aq),

$$HCl\,(aq) + NaOH\,(aq) \longrightarrow H_2O(\ell) + NaCl\,(aq)$$

the complete ionic reaction is

$$H^+\,(aq) + Cl^-\,(aq) + Na^+\,(aq) + OH^-\,(aq) \longrightarrow H_2O(\ell) + Na^+\,(aq) + Cl^-\,(aq)$$

The Na^+(aq) and Cl^-(aq) ions are spectator ions, so we can remove them to have

$$H^+\,(aq) + OH^-\,(aq) \longrightarrow H_2O(\ell)$$

as the net ionic equation. If we wanted to write this in terms of the hydronium ion, H_3O^+(aq), we would write it as

$$H_3O^+\,(aq) + OH^-\,(aq) \longrightarrow 2\,H_2O(\ell)$$

With the exception of the introduction of an extra water molecule, these two net ionic equations are equivalent.

However, for the reaction between HCl(aq) and $Cr(OH)_2$(s), because chromium(II) hydroxide is insoluble, we cannot separate it into ions for the complete ionic equation:

$$2\,H^+\,(aq) + 2\,Cl^-\,(aq) + Cr(OH)_2\,(s) \longrightarrow 2\,H_2O(\ell) + Cr^{2+}\,(aq) + 2\,Cl^-\,(aq)$$

The chloride ions are the only spectator ions here, so the net ionic equation is

$$2\,H^+\,(aq) + Cr(OH)_2\,(s) \longrightarrow 2\,H_2O(\ell) + Cr^{2+}\,(aq)$$

✔ Example 6.8.2

Oxalic acid, $H_2C_2O_4$(s), and $Ca(OH)_2$(s) react very slowly. What is the net ionic equation between these two substances if the salt formed is insoluble? (The anion in oxalic acid is the oxalate ion, $C_2O_4^{2-}$.)

Solution

The products of the neutralization reaction will be water and calcium oxalate:

$H_2C_2O_4(s) + Ca(OH)_2(s) \rightarrow 2H_2O(\ell) + CaC_2O_4(s)$

Because nothing is dissolved, there are no substances to separate into ions, so the net ionic equation is the equation of the three solids and one liquid.

? Exercise 6.8.2

What is the net ionic equation between $HNO_3(aq)$ and $Ti(OH)_4(s)$?

Answer

$4H^+(aq) + Ti(OH)_4(s) \rightarrow 4H_2O(\ell) + Ti^{4+}(aq)$

Key Takeaways

- The Arrhenius definition of an acid is a substance that increases the amount of H^+ in an aqueous solution.
- The Arrhenius definition of a base is a substance that increases the amount of OH^- in an aqueous solution.
- Neutralization is the reaction of an acid and a base, which forms water and a salt.
- Net ionic equations for neutralization reactions may include solid acids, solid bases, solid salts, and water.

6.8: Neutralization Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.6: Neutralization Reactions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.9: Energy

Learning Objectives

- Define *energy*.
- Know the units of energy.
- Understand the law of conservation of energy.

Energy is the ability to do work. Think about it: when you have a lot of energy, you can do a lot of work; but if you're low on energy, you do not want to do much work. Work (w) itself is defined as a force (F) operating over a distance (Δx):

$$w = F \times \Delta x$$

In SI, force has units of newtons (N), while distance has units of meters. Therefore, work has units of N·m. This compound unit is redefined as a **Joule (J)**.

$$1 \text{ joule} = 1 \text{ newton·meter} 1 \text{ J} = 1 \text{ N·m}$$

Because energy is the ability to do work, energy is also measured in joules. This is the primary unit of energy we will use here.

How much is 1 J? It is enough to warm up about one-fourth of a gram of water by 1°C. It takes about 12,000 J to warm a cup of coffee from room temperature to 50°C. So a joule is not a lot of energy. It will not be uncommon to measure energies in thousands of joules, so the kilojoule (kJ) is a common unit of energy, with 1 kJ equal to 1,000 J.

An older—but still common—unit of energy is the *calorie*. The calorie (cal) was originally defined in terms of warming up a given quantity of water. The modern definition of calorie equates it to joules:

$$1 \text{ cal} = 4.184 \text{ J}$$

The calorie is used when considering nutrition. Energy contents of foods are often expressed in calories. However, the calorie unit used for foods is actually the kilocalorie (kcal). Most foods indicate this by spelling the word with a capital C—Calorie. Figure **6.9.1** - Calories on Food Labels, shows one example.



Figure **6.9.1** Calories on Food Labels © Thinkstock. *This label expresses the energy content of the food, but in Calories (which are actually kilocalories).*

Example 6.9.1

The label in Figure **6.9.1** states that the serving has 38 Cal. How many joules is this?

Solution

We recognize that with a capital C, the Calories unit is actually kilocalories. To determine the number of joules, we convert first from kilocalories to calories (using the definition of the *kilo-* prefix) and then from calories to joules (using the relationship between calories and joules). So

$$38 \cancel{\text{kcal}} \times \frac{1000 \cancel{\text{cal}}}{1 \cancel{\text{kcal}}} \times \frac{4.184 \text{ J}}{1 \cancel{\text{cal}}} = 160,000 \text{ J}$$

Exercise 6.9.1

A serving of breakfast cereal usually has 110 Cal. How many joules of energy is this?

Answer

460,000 J

In the study of energy, we use the term **system** to describe the part of the universe under study: a beaker, a flask, or a container whose contents are being observed and measured. An **isolated system** is a system that does not allow a transfer of energy or matter into or out of the system. A good approximation of an isolated system is a closed, insulated thermos-type bottle. The fact that the thermos-type bottle is closed keeps matter from moving in or out, and the fact that it is insulated keeps energy from moving in or out.

One of the fundamental ideas about the total energy of an isolated system is that it does not increase or decrease. When this happens to a quantity, we say that the quantity is *conserved*. The **law of conservation of energy** states that the total energy of an isolated system does not change. As a scientific law, this concept occupies the highest level of understanding we have about the natural universe.

Key Takeaways

- Energy is the ability to do work and uses the unit joule.
- The law of conservation of energy states that the total energy of an isolated system does not increase or decrease.

6.9: Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **7.2: Energy** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.10: Enthalpy and Chemical Reactions

Learning Objectives

- Define *enthalpy*.
- Properly express the enthalpy change of chemical reactions.
- Explain how enthalpy changes are measured experimentally.

Now that we have shown how energy, work, and heat are related, we are ready to consider energy changes in chemical reactions. A fundamental concept is that *every chemical reaction occurs with a concurrent change in energy*. Now we need to learn how to properly express these energy changes.

Our study of gases in Chapter 6, and our definition of work in Section 7.3, indicate that conditions like pressure, volume, and temperature affect the energy content of a system. What we need is a definition of energy that holds when some of these conditions are specified (somewhat similar to our definition of standard temperature and pressure in our study of gases). We define the **enthalpy change** (ΔH) as the heat of a process when pressure is held constant:

$$\Delta H \equiv q \text{ at constant pressure}$$

The letter H stands for "enthalpy," a kind of energy, while the Δ implies a change in the quantity. We will always be interested in the change in H, rather than the absolute value of H itself.

When a chemical reaction occurs, there is a characteristic change in enthalpy. The enthalpy change for a reaction is typically written after a balanced chemical equation and on the same line. For example, when two moles of hydrogen react with one mole of oxygen to make two moles of water, the characteristic enthalpy change is 570 kJ. We write the equation as

$$2\,H_2\,(g) + O_2\,(g) \longrightarrow 2\,H_2O(\ell)$$

with $\Delta H = -570\,kJ$

A chemical equation that includes an enthalpy change is called a **thermochemical equation**. A thermochemical equation is assumed to refer to the equation in molar quantities, which means it must be interpreted in terms of moles, not individual molecules.

Example 6.10.1

Write the thermochemical equation for the reaction of $PCl_3(g)$ with $Cl_2(g)$ to make $PCl_5(g)$, which has an enthalpy change of −88 kJ.

Solution

The thermochemical equation is

$$PCl_3(g) + Cl_2(g) \rightarrow PCl_5(g)\ \Delta H = -88\text{ kJ}$$

Exercise 6.10.1

Write the thermochemical equation for the reaction of $N_2(g)$ with $O_2(g)$ to make $2NO(g)$, which has an enthalpy change of 181 kJ.

Answer

$$N_2(g) + O_2(g) \rightarrow 2NO(g)\ \Delta H = 181\text{ kJ}$$

You may have noticed that the ΔH for a chemical reaction may be positive or negative. The number is assumed to be positive if it has no sign; a + sign can be added explicitly to avoid confusion. A chemical reaction that has a positive ΔH is said to be **endothermic**, while a chemical reaction that has a negative ΔH is said to be **exothermic**.

What does it mean if the ΔH of a process is positive? It means that the system in which the chemical reaction is occurring is gaining energy. If one considers the energy of a system as being represented as a height on a vertical energy plot, the enthalpy change that accompanies the reaction can be diagrammed as in part (a) of Figure **6.10.1** Reaction Energy: the energy of the reactants has some energy, and the system increases its energy as it goes to products. The products are higher on the vertical scale than the reactants. Endothermic, then, implies that the system *gains*, or absorbs, energy.

An opposite situation exists for an exothermic process, as shown in part (b) of Figure **6.10.1** - Reaction Energy. If the enthalpy change of a reaction is negative, the system is losing energy, so the products have less energy than the reactants, and the products are lower on the vertical energy scale than the reactants are. Exothermic, then, implies that the system *loses*, or gives off, energy.

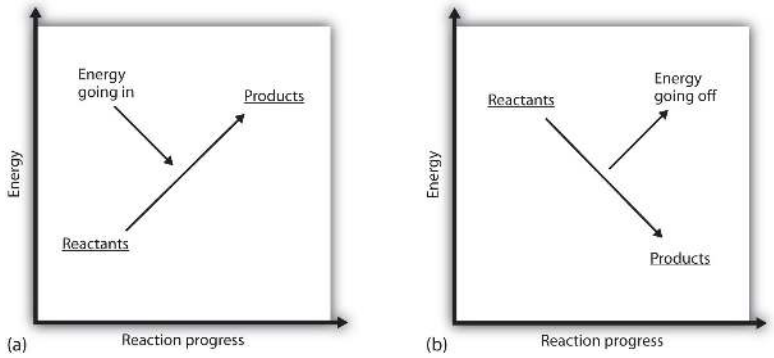


Figure **6.10.1** Reaction Energy (a) In an endothermic reaction, the energy of the system increases (i.e., moves higher on the vertical scale of energy). (b) In an exothermic reaction, the energy of the system decreases (i.e., moves lower on the vertical scale of energy).

Example 6.10.2

Consider this thermochemical equation.

$$2\,CO\,(g) + O_2\,(g) \longrightarrow 2\,CO_2\,(g)$$

with $\Delta H = -565$ kJ

Is it exothermic or endothermic? How much energy is given off or absorbed?

Solution

By definition, a chemical reaction that has a negative ΔH is exothermic, meaning that this much energy—in this case, 565 kJ—is given off by the reaction.

Exercise 6.10.2

Consider this thermochemical equation.

$$CO_2(g) + H_2(g) \rightarrow CO(g) + H_2O(g)\ \Delta H = 42\text{ kJ}$$

Is it exothermic or endothermic? How much energy is given off or absorbed?

Answer

Endothermic; 42 kJ are absorbed.

How are ΔH values measured experimentally? Actually, ΔH is not measured; q is measured. But the measurements are performed under conditions of constant pressure, so ΔH is equal to the q measured.

Experimentally, q is measured by taking advantage of the equation

$$q = mc\Delta T$$

We pre-measure the mass of the chemicals in a system. Then we let the chemical reaction occur and measure the change in temperature (ΔT) of the system. If we know the specific heat of the materials in the system (typically we do), we can calculate q. That value of q is numerically equal to the ΔH of the process, which we can scale up to a molar scale. The container in which the system resides is typically insulated, so any energy change goes into changing the temperature of the system, rather than being leaked from the system. The container is referred to as a **calorimeter**, and the process of measuring changes in enthalpy is called **calorimetry**.

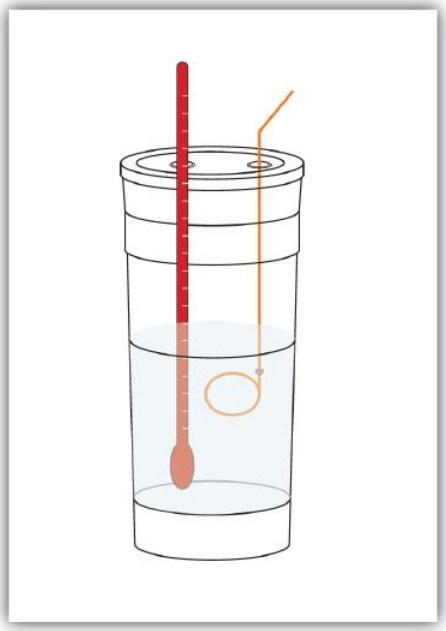


Figure **6.10.2** Calorimeters. A simple calorimeter can be constructed from some nested foam coffee cups, a cover, a thermometer, and a stirrer.

For example, suppose 4.0 g of NaOH, or 0.10 mol of NaOH, are dissolved to make 100.0 mL of aqueous solution; while 3.65 g of HCl, or 0.10 mol of HCl, are dissolved to make another 100.0 mL of aqueous solution. The two solutions are mixed in an insulated calorimeter, a thermometer is inserted, and the calorimeter is covered (See Figure **6.10.2** - Calorimeters for an example setup). The thermometer measures the temperature change as the following chemical reaction occurs:

$$\text{NaOH (aq)} + \text{HCl(aq)} \rightarrow \text{NaCl(aq)} + \text{H}_2\text{O}(\ell)$$

An observer notes that the temperature increases from 22.4°C to 29.1°C. Assuming that the heat capacities and densities of the solutions are the same as those of pure water, we now have the information we need to determine the enthalpy change of the chemical reaction. The total amount of solution is 200.0 mL, and with a density of 1.00 g/mL, we thus have 200.0 g of solution. Using the equation for q, we substitute for our experimental measurements and the specific heat of water (in Table **6.10.1** of Section 7.3).

$$q = (200.0 \cancel{g})\left(4.184 \frac{J}{\cancel{g} \cdot \cancel{^\circ C}}\right)(6.7 \cancel{^\circ C})$$

Solving for q, we get

$$q = 5600\ J \equiv \Delta H \text{ for the reaction}$$

The heat q is equal to the ΔH for the reaction because the chemical reaction occurs at constant pressure. However, the reaction is giving off this amount of energy, so the actual sign on ΔH is negative:

$$\Delta H = -5{,}600 \text{ J for the reaction}$$

Thus, we have the following thermochemical equation for the chemical reaction that occurred in the calorimeter:

$$\frac{1}{10}NaOH(aq) + \frac{1}{10}HCl(aq) \rightarrow \frac{1}{10}NaCl(aq) + \frac{1}{10}H_2O(l)\ \Delta H = -5600\ J$$

The 1/10 coefficients are present to remind us that we started with one-tenth of a mole of each reactant, so we make one-tenth of a mole of each product. Typically, however, we report thermochemical equations in terms of moles, not one-tenth of a mole. To scale up to molar quantities, we must multiply the coefficients by 10. However, when we do this, we get 10 times as much energy. Thus, we have

$$\text{NaOH (aq)} + \text{HCl(aq)} \rightarrow \text{NaCl(aq)} + \text{H}_2\text{O}(\ell)\ \Delta H = -56{,}000\ \text{J}$$

The ΔH can be converted into kJ units, so our final thermochemical equation is

$$\text{NaOH (aq)} + \text{HCl(aq)} \rightarrow \text{NaCl(aq)} + \text{H}_2\text{O}(\ell)\ \Delta H = -56\ \text{kJ}$$

We have just taken our experimental data from calorimetry and determined the enthalpy change of a chemical reaction. Similar measurements on other chemical reactions can determine the ΔH values of any chemical reaction you want to study.

Example 6.10.3

A 100 mL solution of 0.25 mol of $Ca^{2+}(aq)$ was mixed with 0.50 mol of $F^-(aq)$ ions, and CaF_2 was precipitated:

$$Ca^{2+}(aq) + 2F^-(aq) \rightarrow CaF_2(s)$$

The temperature of the solution increased by 10.5°C. What was the enthalpy change for the chemical reaction? What was the enthalpy change for the production of 1 mol of CaF_2? Assume that the solution has the same density and specific heat as water.

Solution

Because we are given ΔT directly, we can determine the heat of the reaction, which is equal to ΔH:

$$q = (100\ \cancel{g})\left(4.184\frac{J}{\cancel{g}\cdot\cancel{^\circ C}}\right)(10.5\ \cancel{^\circ C})$$

Solving for q, we get

$$q = 4{,}400\ \text{J}$$

Therefore, $\Delta H = -4{,}400$ J.

According to the stoichiometry of the reaction, exactly 0.25 mol of CaF_2 will form, so this quantity of heat is for 0.25 mol. For 1 mol of CaF_2, we need to scale up the heat by a factor of four:

$$q = 4{,}400\ \text{J} \times 4 = 17{,}600\ \text{J for 1 mol } CaF_2$$

On a molar basis, the change in enthalpy is

$$\Delta H = -17{,}600\ \text{J} = -17.6\ \text{kJ}$$

Exercise 6.10.3

In a calorimeter at constant pressure, 0.10 mol of $CH_4(g)$ and 0.20 mol of $O_2(g)$ are reacted.

$$CH_4(g) + 2O_2(g) \rightarrow CO_2(g) + 2H_2O(\ell)$$

The reaction warms 750.0 g of H_2O by 28.4°C. What is ΔH for the reaction on a molar scale?

Answer

−891 kJ

Key Takeaways

- Every chemical reaction occurs with a concurrent change in energy.
- The change in enthalpy equals heat at constant pressure.

- Enthalpy changes can be expressed by using thermochemical equations.
- Enthalpy changes are measured by using calorimetry.

6.10: Enthalpy and Chemical Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **7.4: Enthalpy and Chemical Reactions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/beginning-chemistry.

6.11: Chemical Reaction Rate

Drag racing is a sport that involves two cars starting from a dead stop, and driving as fast as they can down a quarter-mile strip. At the end of the strip are timers that determine both elapsed time (how long it took for the cars to cover the quarter mile) and top speed (how fast they were going as they went through the timer chute). Both pieces of data are important. One car may accelerate faster and get ahead that way, while the other car may be slower off the line, but can get up to a higher top speed at the end of the run.

Chemical Reaction Rate

Chemical reactions vary widely in terms of the speed with which they occur. Some reactions occur very quickly. If a lit match is brought in contact with lighter fluid or another flammable liquid, it erupts into flames instantly and burns fast. Other reactions occur very slowly. A container of milk in the refrigerator will be good to drink for weeks before it begins to turn sour. Millions of years were required for dead plants under Earth's surface to accumulate and eventually turn into fossil fuels such as coal and oil.

Chemists need to be concerned with the rates at which chemical reactions occur. Rate is another word for speed. If a sprinter takes $11.0\ \text{s}$ to run a $100\ \text{m}$ dash, his rate is given by the distance traveled divided by the time.

$$\text{speed} = \frac{\text{distance}}{\text{time}} = \frac{100\ \text{m}}{11.0\ \text{s}} = 9.09\ \text{m/s}$$

The sprinter's average running rate for the race is $9.09\ \text{m/s}$. We say that it is his average rate because he did not run at that speed for the entire race. At the very beginning of the race, while coming from a standstill, his rate must be slower until he is able to get up to his top speed. His top speed must then be greater than $9.09\ \text{m/s}$ so that, taken over the entire race, the average ends up at $9.09\ \text{m/s}$.



Figure **6.11.1**: Runner. (CC BY-NC; CK-12)

Chemical reactions can't be measured in units of meters per second, as that would not make any sense. A **reaction rate** is the change in concentration of a reactant or product with time. Suppose that a simple reaction were to take place in which a $1.00\ \text{M}$ aqueous solution of substance A was converted to substance B.

$$\text{A}\,(aq) \rightarrow \text{B}\,(aq)$$

Suppose that after 20.0 seconds, the concentration of A had dropped from $1.00\ \text{M}$ to $0.72\ \text{M}$ as A was slowly being converted to B. We can express the rate of this reaction as the change in concentration of A divided by time.

$$\text{rate} = -\frac{\Delta\,[\text{A}]}{\Delta t} = -\frac{[\text{A}]_{\text{final}} - [\text{A}]_{\text{initial}}}{\Delta t}$$

A bracket around a symbol or formula means the concentration in molarity of that substance. The change in concentration of A is its final concentration minus its initial concentration. Because the concentration of A is decreasing over time, the negative sign is used. Thus, the rate for the reaction is positive and the units are molarity per second or M/s.

$$\text{rate} = -\frac{0.72\ \text{M} - 1.00\ \text{M}}{20.0\ \text{s}} = -\frac{-0.28\ \text{M}}{20.0\ \text{s}} = 0.014\ \text{M/s}$$

The molarity of A decreases by an average rate of $0.014\ \text{M}$ every second. In summary, the rate of a chemical reaction is measured by the change in concentration over time for a reactant or product. The unit of measurement for a reaction rate is molarity per second (M/s).

Summary

- Chemists need to be concerned with the rates at which chemical reactions occur.
- The reaction rate indicates how fast the reaction proceeds.

6.11: Chemical Reaction Rate is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **18.1: Chemical Reaction Rate** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

6.12: Collision Theory

Car damage can be very expensive, especially if the driver who did the damage does not have car insurance. Many people have had the experience of backing up while parallel parking and hearing that "bump". Fortunately, cars are often not going fast enough to do any damage. Every once in a while, though, there is a rearrangement of the body parts of a car when it is hit with sufficient speed; then things need to be fixed.

Collision Theory

The behavior of the atoms, molecules, or ions that comprise the reactants is responsible for the rates of a given chemical reaction. **Collision theory** is a set of principles that states that the reacting particles can form products when they collide with one another, provided those collisions have enough kinetic energy and the correct orientation. Particles that lack the necessary kinetic energy may collide, but the particles will simply bounce off one another unchanged. The figure below illustrates the difference. In the first collision, the particles bounce off one another, and no rearrangement of atoms has occurred. The second collision occurs with greater kinetic energy, and so the bond between the two red atoms breaks. One red atom bonds with the other molecule as one product, while the single red atom is the other product. The first collision is called an **ineffective collision**, while the second collision is called an **effective collision**.

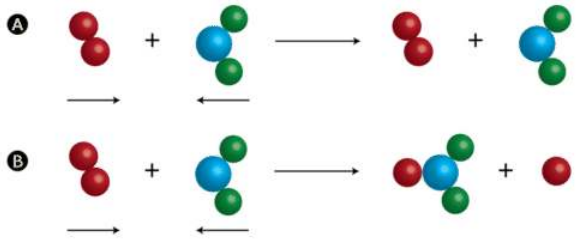


Figure **6.12.1**: An ineffective collision (A) is one that does not result in product formation. An effective collision (B) is one in which chemical bonds are broken and a product is formed. (CC BY-NC; CK-12)

Summary

- Collision theory explains how materials can collide and become new materials.
- Effective collisions result in product formation.
- Ineffective collisions do not result in product formation.

6.12: Collision Theory is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **18.2: Collision Theory** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

6.13: Activation Energy

The sight of fireworks cascading across the night sky is a hallmark of special occasions. These materials, invented hundreds of years ago, can be dangerous if not handled properly. The chemicals do not react until the fuse burns down and heat is applied to the system. Then, the rocket is launched and explodes high in the sky.

Activation Energy



Figure **6.13.1**: Calcium metal stored in an argon atmosphere. (CC BY-NC; CK-12)

Why do some chemical reactions occur readily while others require input of heat in order to take place? If we mix metallic sodium with water, a reaction occurs immediately, releasing a great deal of heat (enough to ignite the hydrogen gas that is formed). Group II metals, such as calcium, react at a much slower rate. Unlike the extremely vigorous reaction with sodium, the reaction with calcium is slow enough that we can trap the hydrogen gas released.

Supplying reactant particles with energy causes the bonds between the atoms to vibrate with a greater frequency. This increase in vibrational energy makes a chemical bond more likely to break, and a chemical reaction more likely to occur, when those particles collide with other particles. The **activation energy** for a reaction is the minimum energy that colliding particles must have in order to undergo a reaction. Some reactions occur readily at room temperature because the reacting particles already have the requisite activation energy at that temperature. Other reactions only occur when heated because the particles do not have enough energy unless an external source of heat provides the particles with more kinetic energy.

Summary

6.13: Activation Energy is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **18.3: Activation Energy** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

6.14: Chemical Equilibrium

A tug of war involves two teams at the ends of a rope. The goal is to pull the other team over a line in the middle. At first, there is a great deal of tension on the rope, but no apparent movement. A bystander might think that there is nothing happening. In reality, there is a great deal of tension on the rope as the two teams pull in opposite directions at the same time.

Chemical Equilibrium

Hydrogen and iodine gases react to form hydrogen iodide, according to the following reaction:

$$\mathrm{H_2}\,(g) + \mathrm{I_2}\,(g) \rightleftharpoons 2\mathrm{HI}\,(g)$$

$$\textbf{Forward reaction:}\quad \mathrm{H_2}\,(g) + \mathrm{I_2}\,(g) \rightarrow 2\mathrm{HI}\,(g)$$

$$\textbf{Reverse reaction:}\quad 2\mathrm{HI}\,(g) \rightarrow \mathrm{H_2}\,(g) + \mathrm{I_2}\,(g)$$

Initially, only the forward reaction occurs because no HI is present. As soon as some HI has formed, it begins to decompose back into $\mathrm{H_2}$ and $\mathrm{I_2}$. Gradually, the rate of the forward reaction decreases while the rate of the reverse reaction increases. Eventually, the rate of combination of $\mathrm{H_2}$ and $\mathrm{I_2}$ to produce HI becomes equal to the rate of decomposition of HI into $\mathrm{H_2}$ and $\mathrm{I_2}$. When the rates of the forward and reverse reactions have become equal to one another, the reaction has achieved a state of balance. **Chemical equilibrium** is the state of a system in which the rate of the forward reaction is equal to the rate of the reverse reaction.

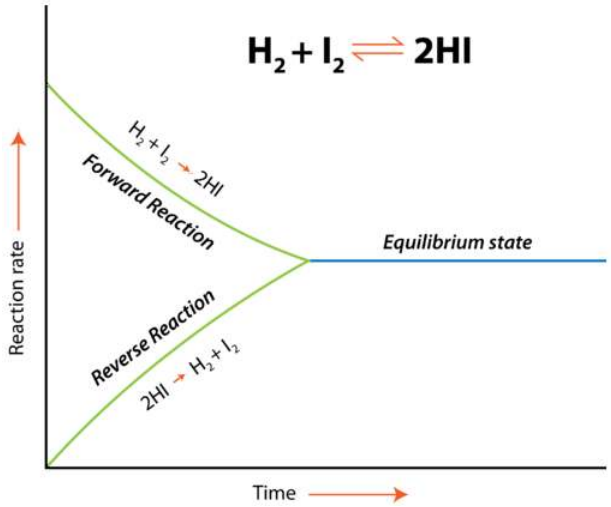


Figure **6.14.1**: Equilibrium in reaction: $\mathrm{H_2}\,(g) + \mathrm{I_2}\,(g) \rightleftharpoons 2\mathrm{HI}\,(g)$.

Chemical equilibrium can be attained whether the reaction begins with all reactants and no products, all products and no reactants, or some of both. The figure below shows changes in concentration of $\mathrm{H_2}$, $\mathrm{I_2}$, and HI for two different reactions. In the reaction depicted by the graph on the left (A), the reaction begins with only $\mathrm{H_2}$ and $\mathrm{I_2}$ present. There is no HI initially. As the reaction proceeds toward equilibrium, the concentrations of the $\mathrm{H_2}$ and $\mathrm{I_2}$ gradually decrease, while the concentration of the HI gradually increases. When the curve levels out and the concentrations all become constant, equilibrium has been reached. At equilibrium, concentrations of all substances are constant. In reaction B, the process begins with only HI and no $\mathrm{H_2}$ or $\mathrm{I_2}$. In this case, the concentration of HI gradually decreases while the concentrations of $\mathrm{H_2}$ and $\mathrm{I_2}$ gradually increase until equilibrium is again reached. Notice that in both cases, the relative position of equilibrium is the same, as shown by the relative concentrations of reactants and products. The concentration of HI at equilibrium is significantly higher than the concentrations of $\mathrm{H_2}$ and $\mathrm{I_2}$. This is true whether the reaction began with all reactants or all products. The position of equilibrium is a property of the particular reversible reaction and does not depend upon how equilibrium was achieved.

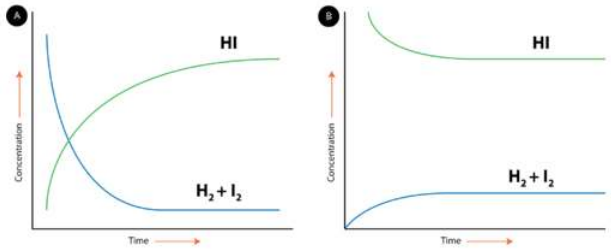


Figure **6.14.2**: Equilibrium between reactants and products.

Conditions for Equilibrium

It may be tempting to think that once equilibrium has been reached, the reaction stops. Chemical equilibrium is a dynamic process. The forward and reverse reactions continue to occur even after equilibrium has been reached. However, because the rates of the reactions are the same, there is no change in the relative concentrations of reactants and products for a reaction that is at equilibrium. The conditions and properties of a system at equilibrium are summarized below.

1. The system must be closed, meaning no substances can enter or leave the system.
2. Equilibrium is a dynamic process. Even though we don't necessarily see the reactions, both the forward and reverse reactions are taking place.
3. The rates of the forward and reverse reactions must be equal.
4. The amount of reactants and products do not have to be equal. However, after equilibrium is attained, the amounts of reactants and products will be constant.

The description of equilibrium in this concept refers primarily to equilibrium between reactants and products in a chemical reaction. Other types of equilibrium include phase equilibrium and solution equilibrium. A **phase equilibrium** occurs when a substance is in equilibrium between two states. For example, a stoppered flask of water attains equilibrium when the rate of evaporation is equal to the rate of condensation. A **solution equilibrium** occurs when a solid substance is in a saturated solution. At this point, the rate of dissolution is equal to the rate of recrystallization. Although these are all different types of transformations, most of the rules regarding equilibrium apply to any situation in which a process occurs reversibly.

Summary

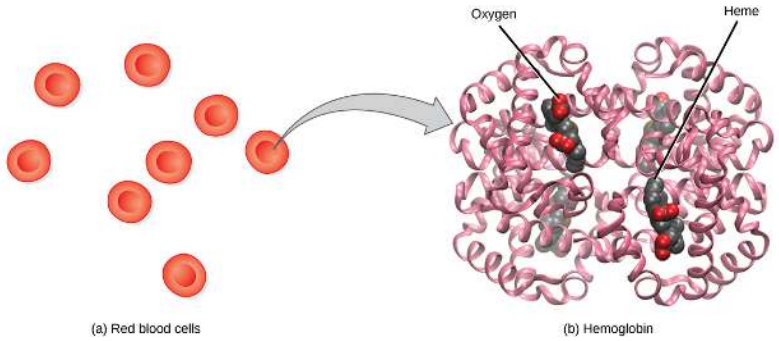
- Chemical equilibrium is the state of a system in which the rate of the forward reaction is equal to the rate of the reverse reaction.
- Conditions for chemical equilibrium are listed.
- A phase equilibrium occurs when a substance is in equilibrium between two states.
- A solution equilibrium occurs when a solid substance is in a saturated solution.

6.14: Chemical Equilibrium is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **19.2: Chemical Equilibrium** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

6.15: Equilibrium Constant

Red blood cells transport oxygen to tissues so that they can function. In the absence of oxygen, cells cannot carry out their biochemical responsibilities. Oxygen moves to the cells attached to hemoglobin, a protein found in red blood cells.



The protein inside (a) red blood cells that carries oxygen to cells and carbon dioxide to the lungs is (b) hemoglobin. Hemoglobin is made up of four symmetrical subunits and four heme groups. Iron associated with the heme binds oxygen. It is the iron in hemoglobin that gives blood its red color. (CC_BY Openstax)

In cases of carbon monoxide poisoning, $\mathbf{CO}$ binds much more strongly to the hemoglobin, blocking oxygen attachment and lowering the amount of oxygen reaching the cells. Treatment involves the patient breathing pure oxygen to displace the carbon monoxide. The equilibrium reaction shown below illustrates the shift toward the right when excess oxygen is added to the system:

$$\mathrm{Hb(CO)_4}\,(aq) + 4\mathrm{O_2}\,(g) \rightleftharpoons \mathrm{Hb(O_2)_4}\,(aq) + 4\mathrm{CO}\,(g)$$

Equilibrium Constant

Consider the hypothetical reversible reaction in which reactants $\mathbf{A}$ and $\mathbf{B}$ react to form products $\mathbf{C}$ and $\mathbf{D}$. This equilibrium can be shown as demonstrated below, where the lower case letters represent the coefficients of each substance.

$$a\mathrm{A} + b\mathrm{B} \rightleftharpoons c\mathrm{C} + d\mathrm{D}$$

As we have established, the rates of the forward and reverse reactions are the same at equilibrium, and so the concentrations of all of the substances are constant. Since that is the case, it stands to reason that a ratio of the concentrations for any given reaction at equilibrium maintains a constant value. The **equilibrium constant** (K_{eq}) is the ratio of the mathematical product of the concentrations of the products of a reaction to the mathematical product of the concentrations of the reactants of the reaction. Each concentration is raised to the power of its coefficient in the balanced chemical equation. For the general reaction above, the equilibrium constant expression is written as follows:

$$K_{\mathrm{eq}} = \frac{[\mathrm{C}]^c\,[\mathrm{D}]^d}{[\mathrm{A}]^a\,[\mathrm{B}]^b}$$

The concentrations of each substance, indicated by the square brackets around the formula, are measured in molarity units ($\mathrm{mol/L}$).

The value of the equilibrium constant for any reaction is only determined by experiment. As detailed in the previous section, the position of equilibrium for a given reaction does not depend on the starting concentrations, and so the value of the equilibrium constant is truly constant. It does, however, depend on the temperature of the reaction. This is because equilibrium is defined as a condition resulting from the rates of forward and reverse reactions being equal. If the temperature changes, the corresponding change in those reaction rates will alter the equilibrium constant. For any reaction in which a K_{eq} is given, the temperature should be specified.

Summary

6.15: Equilibrium Constant is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **19.3: Equilibrium Constant** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

https://chem.libretexts.org/@go/page/200146

6.16: Le Châtelier's Principle

There are some who enjoy going up in an airplane, strapping on a parachute, and diving out the door to free-fall before opening the chute, and dropping to the ground. This stressful activity (so they say) relieves the stress of everyday life. The release of adrenaline caused by this stressful behavior is said to promote a mood enhancement that helps you deal better with other stresses in your daily life.

Le Chatelier's Principle

Chemical equilibrium was studied by French chemist Henri Le Chatelier (1850-1936) and his description of how a system responds to a stress on equilibrium has become known as **Le Chatelier's principle:** When a chemical system that is at equilibrium is disturbed by a stress, the system will respond in order to relieve the stress. Stresses to a chemical system involve changes in the concentration of reactants or products, changes in the temperature of the system, or changes in the pressure of the system. We will discuss each of these stresses separately. The change to the equilibrium position in every case is either a favoring of the forward reaction or a favoring of the reverse reaction. When the forward reaction is favored, the concentrations of products increase, while the concentrations of reactants decrease. When the reverse reaction is favored, the concentrations of products decrease, while the concentrations of reactants increase.

Le Chatelier's Principle

Table 6.16.1

Original Equilibrium	Favored Reaction	Result
$\mathbf{A} \rightleftharpoons \mathbf{B}$	Forward: $\mathbf{A} \rightarrow \mathbf{B}$	[A] decreases; [B] increases
$\mathbf{A} \rightleftharpoons \mathbf{B}$	Reverse: $\mathbf{A} \leftarrow \mathbf{B}$	[A] increases; [B] decreases



Figure 6.16.1: Henri Le Chatelier.

Summary

6.16: Le Châtelier's Principle is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **19.5: Le Châtelier's Principle** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

CHAPTER OVERVIEW

7: Chapter 7 - Survey of Kinematics and Newtonian Motion

7.1: Prelude to One-Dimensional Kinematics
7.2: Displacement
7.3: Vectors, Scalars, and Coordinate Systems
7.4: Time, Velocity, and Speed
7.5: Acceleration
7.6: Kinematics in Two Dimensions - An Introduction
7.7: Projectile Motion
7.8: Development of Force Concept
7.9: Newton’s First Law of Motion- Inertia
7.10: Newton’s Second Law of Motion- Concept of a System
7.11: Newton’s Third Law of Motion- Symmetry in Forces
7.12: Linear Momentum and Force
7.13: Impulse

7: Chapter 7 - Survey of Kinematics and Newtonian Motion is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

7.1: Prelude to One-Dimensional Kinematics

Objects are in motion everywhere we look. Everything from a tennis game to a space-probe flyby of the planet Neptune involves motion. When you are resting, your heart moves blood through your veins. And even in inanimate objects, there is continuous motion in the vibrations of atoms and molecules. Questions about motion are interesting in and of themselves: *How long will it take for a space probe to get to Mars? Where will a football land if it is thrown at a certain angle?* But an understanding of motion is also key to understanding other concepts in physics. An understanding of acceleration, for example, is crucial to the study of force.



Figure **7.1.1**:The motion of an American kestrel through the air can be described by the bird's displacement, speed, velocity, and acceleration. When it flies in a straight line without any change in direction, its motion is said to be one dimensional. (credit: Vince Maidens, Wikimedia Commons)

Our formal study of physics begins with kinematics which is defined as the *study of motion without considering its causes*. The word "kinematics" comes from a Greek term meaning motion and is related to other English words such as "cinema" (movies) and "kinesiology" (the study of human motion). In one-dimensional kinematics and Two-Dimensional Kinematics we will study only the *motion* of a football, for example, without worrying about what forces cause or change its motion. Such considerations come in other chapters. In this chapter, we examine the simplest type of motion—namely, motion along a straight line, or one-dimensional motion. In Two-Dimensional Kinematics, we apply concepts developed here to study motion along curved paths (two- and three-dimensional motion); for example, that of a car rounding a curve.

This page titled 7.1: Prelude to One-Dimensional Kinematics is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **2.0: Prelude to One-Dimensional Kinematics** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.2: Displacement

Learning Objectives

- Define position, displacement, distance, and distance traveled.
- Explain the relationship between position and displacement.
- Distinguish between displacement and distance traveled.
- Calculate displacement and distance given initial position, final position, and the path between the two.

These cyclists in Vietnam can be described by their position relative to buildings and a canal. Their motion can be described by their change in position, or displacement, in the frame of reference.



Figure **7.2.1**: These cyclists in Vietnam can be described by their position relative to buildings and a canal. Their motion can be described by their change in position, or displacement, in the frame of reference. (credit: Suzan Black, Fotopedia)

Position

In order to describe the motion of an object, you must first be able to describe its position—where it is at any particular time. More precisely, you need to specify its position relative to a convenient reference frame. Earth is often used as a reference frame, and we often describe the position of an object as it relates to stationary objects in that reference frame. For example, a rocket launch would be described in terms of the position of the rocket with respect to the Earth as a whole, while a professor's position could be described in terms of where she is in relation to the nearby white board. In other cases, we use reference frames that are not stationary but are in motion relative to the Earth. To describe the position of a person in an airplane, for example, we use the airplane, not the Earth, as the reference frame.

Displacement

If an object moves relative to a reference frame (for example, if a professor moves to the right relative to a white board or a passenger moves toward the rear of an airplane), then the object's position changes. This change in position is known as displacement. The word "displacement" implies that an object has moved, or has been displaced.

Definition: Displacement

Displacement is the change in position of an object:

$$\Delta x = x_f - x_0,$$

where Δx is displacement, x_f is the final position, and x_0 is the initial position.

In this text the upper case Greek letter size Δ(delta) always means "change in" whatever quantity follows it; thus, size Δx means change in position. Always solve for displacement by subtracting initial position size x_0 from final position x_f.

Note that the SI unit for displacement is the meter (m) (see Section on Physical Quantities and Units), but sometimes kilometers, miles, feet, and other units of length are used. Keep in mind that when units other than the meter are used in a problem, you may

need to convert them into meters to complete the calculation.

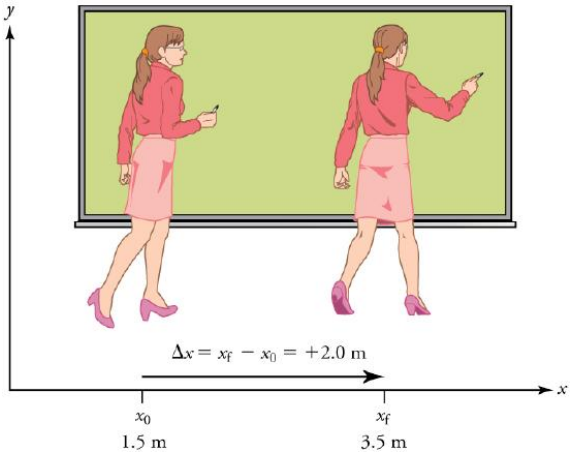


Figure 7.2.2: A professor paces left and right while lecturing. Her position relative to Earth is given by x. The +2 m displacement of the professor relative to Earth is represented by an arrow pointing to the right.

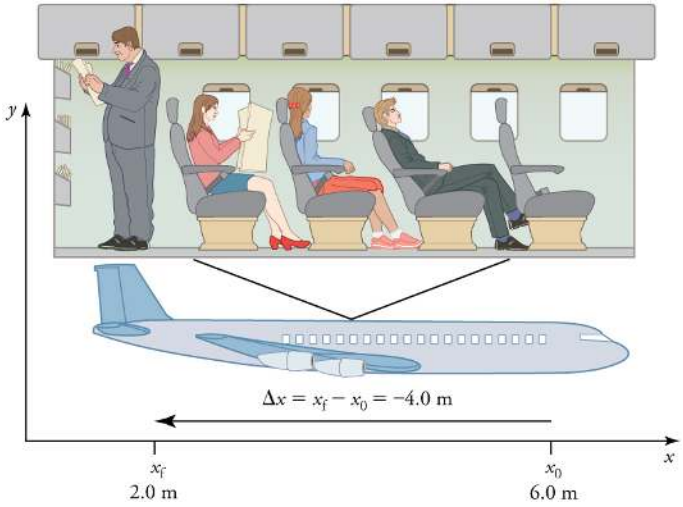


Figure 7.2.3: A passenger moves from his seat to the back of the plane. His location relative to the airplane is given by x. The -4 m displacement of the passenger relative to the plane is represented by an arrow toward the rear of the plane. Notice that the arrow representing his displacement is twice as long as the arrow representing the displacement of the professor (he moves twice as far)

Note that displacement has a direction as well as a magnitude. The professor's displacement in Figure 7.2.2 is 2.0 m to the right, and the airline passenger's displacement is 4.0 m toward the rear in Figure 7.2.3. In one-dimensional motion, direction can be specified with a plus or minus sign. When you begin a problem, you should select which direction is positive (usually that will be to the right or up, but you are free to select positive as being any direction). The professor's initial position is size x_0 = 1.5 m and her final position is x_f = 3.5 m. Thus her displacement is

$$\begin{aligned}\Delta x &= x_f - x_0 \\ &= 3.5\,m - 1.5\,m \\ &= +2.0\,m.\end{aligned}$$

In this coordinate system, motion to the right is positive, whereas motion to the left is negative. Similarly, the airplane passenger's initial position is $x_0 = 6.0m$ and his final position is $x_f = 2.0m$, so his displacement is

$$\begin{aligned}\Delta x &= x_f - x_0 \\ &= 2.0\,m - 6.0\,m \\ &= -4.0\,m.\end{aligned}$$

His displacement is negative because his motion is toward the rear of the plane, or in the negative size 12{x} {} direction in our coordinate system.

Distance

Although displacement is described in terms of direction, distance is not. **Distance** is defined to be the *magnitude or size of displacement between two positions*. Note that the distance between two positions is not the same as the distance traveled between them. **Distance traveled** is *the total length of the path traveled between two positions.* Distance has no direction and, thus, no sign. For example, the distance the professor walks is 2.0 m. The distance the airplane passenger walks is 4.0 m.

Misconception Alert: Distance Traveled vs. Magnitude of Displacement

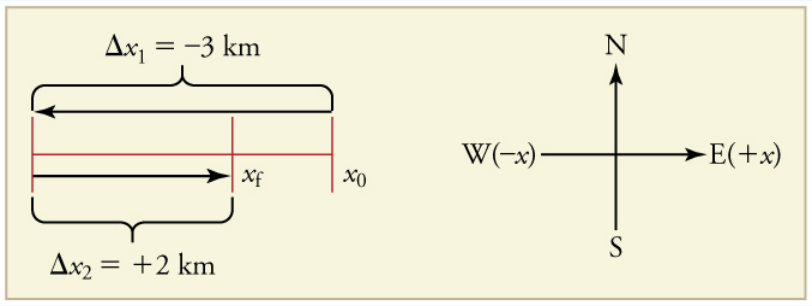
It is important to note that the distance traveled, however, can be greater than the magnitude of the displacement (by magnitude, we mean just the size of the displacement without regard to its direction; that is, just a number with a unit). For example, the professor could pace back and forth many times, perhaps walking a distance of 150 m during a lecture, yet still end up only 2.0 m to the right of her starting point. In this case her displacement would be +2.0 m, the magnitude of her displacement would be 2.0 m, but the distance she traveled would be 150 m. In kinematics we nearly always deal with displacement and magnitude of displacement, and almost never with distance traveled. One way to think about this is to assume you marked the start of the motion and the end of the motion. The displacement is simply the difference in the position of the two marks and is independent of the path taken in traveling between the two marks. The distance traveled, however, is the total length of the path taken between the two marks.

Exercise 7.2.1

A cyclist rides 3 km west and then turns around and rides 2 km east.

a. What is her displacement?
b. What distance does she ride?
c. What is the magnitude of her displacement?

Answer



Answer a

The rider's displacement is $\Delta x = x_f - x_0 = -1km$. (The displacement is negative because we take east to be positive and west to be negative.)

Answer b

The distance traveled is $3km + 2km = 5km$.

Answer c

The magnitude of the displacement is $1km$.

Summary

- Kinematics is the study of motion without considering its causes. In this chapter, it is limited to motion along a straight line, called one-dimensional motion.
- Displacement is the change in position of an object.
- In symbols, displacement Δx is defined to be

$$\Delta x = x_f - x_0,$$

where x_0 is the initial position and xf is the final position. In this text, the Greek letter Δ (delta) always means “change in” whatever quantity follows it. The SI unit for displacement is the meter (m). Displacement has a direction as well as a magnitude.

- When you start a problem, assign which direction will be positive.
- Distance is the magnitude of displacement between two positions.
- Distance traveled is the total length of the path traveled between two positions.

Glossary

kinematics

the study of motion without considering its causes

position

the location of an object at a particular time

displacement

the change in position of an object

distance

the magnitude of displacement between two positions

distance traveled

the total length of the path traveled between two positions

This page titled 7.2: Displacement is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **2.1: Displacement** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.3: Vectors, Scalars, and Coordinate Systems

Learning Objectives

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

- Define and distinguish between scalar and vector quantities.
- Assign a coordinate system for a scenario involving one-dimensional motion.

What is the difference between distance and displacement? Whereas displacement is defined by both direction and magnitude, distance is defined only by magnitude. Displacement is an example of a vector quantity. Distance is an example of a scalar quantity. A vector is any quantity with both *magnitude and direction*. Other examples of vectors include a velocity of 90 km/h east and a force of 500 newtons straight down.



Figure **7.3.1**: The motion of this Eclipse Concept jet can be described in terms of the distance it has traveled (a scalar quantity) or its displacement in a specific direction (a vector quantity). In order to specify the direction of motion, its displacement must be described based on a coordinate system. In this case, it may be convenient to choose motion toward the left as positive motion (it is the forward direction for the plane), although in many cases, the x -coordinate runs from left to right, with motion to the right as positive and motion to the left as negative. (credit: Armchair Aviator, Flickr)

The direction of a vector in one-dimensional motion is given simply by a plus (+) or minus (−) sign. Vectors are represented graphically by arrows. An arrow used to represent a vector has a length proportional to the vector's magnitude (e.g., the larger the magnitude, the longer the length of the vector) and points in the same direction as the vector.

Some physical quantities, like distance, either have no direction or none is specified. A scalar is any quantity that has a magnitude, but no direction. For example, a 20ºC temperature, the 250 kilocalories (250 Calories) of energy in a candy bar, a 90 km/h speed limit, a person's 1.8 m height, and a distance of 2.0 m are all scalars—quantities with no specified direction. Note, however, that a scalar can be negative, such as a −20ºC temperature. In this case, the minus sign indicates a point on a scale rather than a direction. Scalars are never represented by arrows.

Coordinate Systems for One-Dimensional Motion

In order to describe the direction of a vector quantity, you must designate a coordinate system within the reference frame. For one-dimensional motion, this is a simple coordinate system consisting of a one-dimensional coordinate line. In general, when describing horizontal motion, motion to the right is usually considered positive, and motion to the left is considered negative. With vertical motion, motion up is usually positive and motion down is negative. In some cases, however, as with the jet in Figure \PageIndex{2}\), it can be more convenient to switch the positive and negative directions. For example, if you are analyzing the motion of falling objects, it can be useful to define downwards as the positive direction. If people in a race are running to the left, it is useful to define left as the positive direction. It does not matter as long as the system is clear and consistent. Once you assign a positive direction and start solving a problem, you cannot change it.

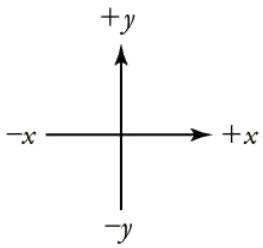


Figure **7.3.2**: It is usually convenient to consider motion upward or to the right as positive (+) and motion downward or to the left as negative (−) .

Exercise 7.3.1

A person's speed can stay the same as he or she rounds a corner and changes direction. Given this information, is speed a scalar or a vector quantity? Explain.

Answer

Speed is a scalar quantity. It does not change at all with direction changes; therefore, it has magnitude only. If it were a vector quantity, it would change as direction changes (even if its magnitude remained constant).

Summary

- A vector is any quantity that has magnitude and direction.
- A scalar is any quantity that has magnitude but no direction.
- Displacement and velocity are vectors, whereas distance and speed are scalars.
- In one-dimensional motion, direction is specified by a plus or minus sign to signify left or right, up or down, and the like.

Glossary

scalar

a quantity that is described by magnitude, but not direction

vector

a quantity that is described by both magnitude and direction

This page titled 7.3: Vectors, Scalars, and Coordinate Systems is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **2.2: Vectors, Scalars, and Coordinate Systems** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.4: Time, Velocity, and Speed

Learning Objectives

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

- Explain the relationships between instantaneous velocity, average velocity, instantaneous speed, average speed, displacement, and time.
- Calculate velocity and speed given initial position, initial time, final position, and final time.
- Derive a graph of velocity vs. time given a graph of position vs. time.
- Interpret a graph of velocity vs. time.

There is more to motion than distance and displacement. Questions such as, “How long does a foot race take?” and “What was the runner’s speed?” cannot be answered without an understanding of other concepts. In this section we add definitions of time, velocity, and speed to expand our description of motion.



Figure 7.4.1: The motion of these racing snails can be described by their speeds and their velocities. (credit: tobitasflickr, Flickr)

Time

As discussed in Physical Quantities and Units, the most fundamental physical quantities are defined by how they are measured. This is the case with time. Every measurement of time involves measuring a change in some physical quantity. It may be a number on a digital clock, a heartbeat, or the position of the Sun in the sky. In physics, the definition of **time** is simple—time is *change*, or the interval over which change occurs. It is impossible to know that time has passed unless something changes.

The amount of time or change is calibrated by comparison with a standard. The SI unit for time is the second, abbreviated s. We might, for example, observe that a certain pendulum makes one full swing every 0.75 s. We could then use the pendulum to measure time by counting its swings or, of course, by connecting the pendulum to a clock mechanism that registers time on a dial. This allows us to not only measure the amount of time, but also to determine a sequence of events.

How does time relate to motion? We are usually interested in elapsed time for a particular motion, such as how long it takes an airplane passenger to get from his seat to the back of the plane. To find elapsed time, we note the time at the beginning and end of the motion and subtract the two. For example, a lecture may start at 11:00 A.M. and end at 11:50 A.M., so that the elapsed time would be 50 min. **Elapsed time** Δt is the difference between the ending time and beginning time,

$$\Delta t = t_f - t_0,$$

where Δt is the change in time or elapsed time, t_f is the time at the end of the motion, and t_0 is the time at the beginning of the motion. (As usual, the delta symbol, Δ, means the change in the quantity that follows it.)

Life is simpler if the beginning time t_0 is taken to be zero, as when we use a stopwatch. If we were using a stopwatch, it would simply read zero at the start of the lecture and 50 min at the end. If $t_0 = 0$, then

$$\Delta t = t_f \equiv t. \tag{7.4.1}$$

In this text, for simplicity's sake,

- motion starts at time equal to zero ($t_0 = 0$)
- the symbol t is used for elapsed time unless otherwise specified ($\Delta t = t_f \equiv t$)

Velocity

Your notion of velocity is probably the same as its scientific definition. You know that if you have a large displacement in a small amount of time you have a large velocity, and that velocity has units of distance divided by time, such as miles per hour or kilometers per hour.

Definition: AVERAGE VELOCITY

Average velocity is *displacement (change in position) divided by the time of travel*,

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0}. \tag{7.4.2}$$

where $\bar{v}$ is the average (indicated by the bar over the v) velocity, Δx is the change in position (or displacement), and x_f and x_0 are the final and beginning positions at times t_f and t_0, respectively. If the starting time t_0 is taken to be zero, then the average velocity is simply

$$\bar{v} = \frac{\Delta x}{t}. \tag{7.4.3}$$

Notice that this definition indicates that *velocity is a vector because displacement is a vector*. It has both magnitude and direction. The SI unit for velocity is meters per second or m/s, but many other units, such as km/h, mi/h (also written as mph), and cm/s, are in common use. Suppose, for example, an airplane passenger took 5 seconds to move −4 m (the negative sign indicates that displacement is toward the back of the plane). His average velocity would be

$$\bar{v} = \frac{\Delta x}{t} = \frac{-4\,m}{5\,s} = -0.8\,m/s. \tag{7.4.4}$$

The minus sign indicates the average velocity is also toward the rear of the plane.

The average velocity of an object does not tell us anything about what happens to it between the starting point and ending point, however. For example, we cannot tell from average velocity whether the airplane passenger stops momentarily or backs up before he goes to the back of the plane. To get more details, we must consider smaller segments of the trip over smaller time intervals.

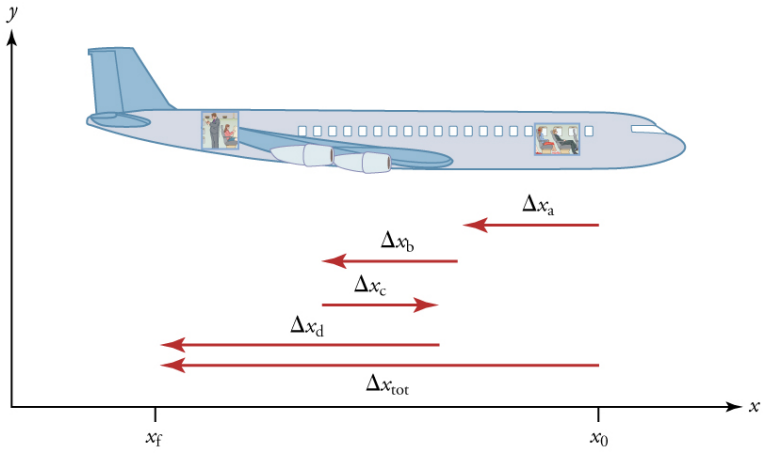


Figure **7.4.2**: A more detailed record of an airplane passenger heading toward the back of the plane, showing smaller segments of his trip.

The smaller the time intervals considered in a motion, the more detailed the information. When we carry this process to its logical conclusion, we are left with an infinitesimally small interval. Over such an interval, the average velocity becomes the *instantaneous velocity* or the *velocity at a specific instant*. A car's speedometer, for example, shows the magnitude (but not the direction) of the instantaneous velocity of the car. (Police give tickets based on instantaneous velocity, but when calculating how long it will take to

get from one place to another on a road trip, you need to use average velocity.) **Instantaneous velocity** v is the average velocity at a specific instant in time (or over an infinitesimally small time interval).

Mathematically, finding instantaneous velocity, v, at a precise instant t can involve taking a limit, a calculus operation beyond the scope of this text. However, under many circumstances, we can find precise values for instantaneous velocity without calculus.

Speed

In everyday language, most people use the terms “speed” and “velocity” interchangeably. In physics, however, they do not have the same meaning and they are distinct concepts. One major difference is that speed has no direction. Thus *speed is a scalar*. Just as we need to distinguish between instantaneous velocity and average velocity, we also need to distinguish between instantaneous speed and average speed.

Instantaneous speed is the magnitude of instantaneous velocity. For example, suppose the airplane passenger at one instant had an instantaneous velocity of −3.0 m/s (the minus meaning toward the rear of the plane). At that same time his instantaneous speed was 3.0 m/s. Or suppose that at one time during a shopping trip your instantaneous velocity is 40 km/h due north. Your instantaneous speed at that instant would be 40 km/h—the same magnitude but without a direction. Average speed, however, is very different from average velocity. **Average speed** is the distance traveled divided by elapsed time.

We have noted that distance traveled can be greater than displacement. So average speed can be greater than average velocity, which is displacement divided by time. For example, if you drive to a store and return home in half an hour, and your car’s odometer shows the total distance traveled was 6 km, then your average speed was 12 km/h. Your average velocity, however, was zero, because your displacement for the round trip is zero. (Displacement is change in position and, thus, is zero for a round trip.) Thus average speed is *not* simply the magnitude of average velocity.

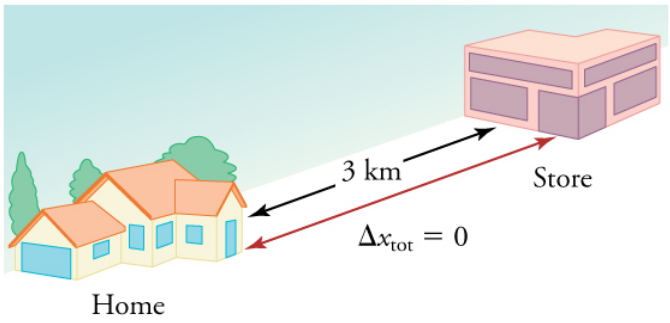


Figure **7.4.3**: During a 30-minute round trip to the store, the total distance traveled is 6 km. The average speed is 12 km/h. The displacement for the round trip is zero, since there was no net change in position. Thus the average velocity is zero.

Another way of visualizing the motion of an object is to use a graph. A plot of position or of velocity as a function of time can be very useful. For example, for this trip to the store, the position, velocity, and speed-vs.-time graphs are displayed in Figure **7.4.4**. (Note that these graphs depict a very simplified **model** of the trip. We are assuming that speed is constant during the trip, which is unrealistic given that we’ll probably stop at the store. But for simplicity’s sake, we will model it with no stops or changes in speed. We are also assuming that the route between the store and the house is a perfectly straight line.)

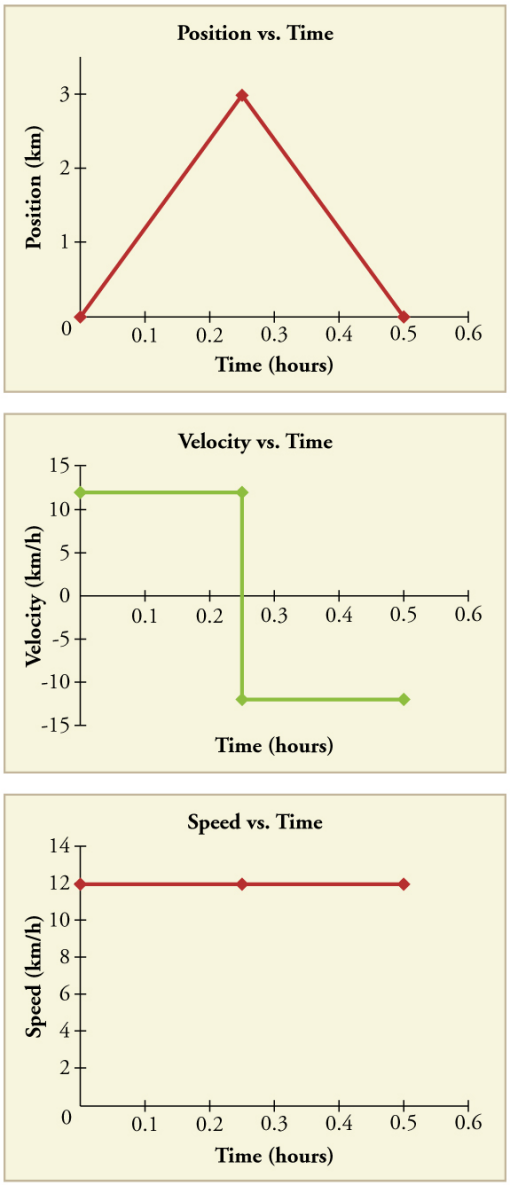


Figure **7.4.4**: Position vs. time, velocity vs. time, and speed vs. time on a trip. Note that the velocity for the return trip is negative.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION - GETTING A SENSE OF SPEED

If you have spent much time driving, you probably have a good sense of speeds between about 10 and 70 miles per hour. But what are these in meters per second? What do we mean when we say that something is moving at 10 m/s? To get a better sense of what these values really mean, do some observations and calculations on your own:

- calculate typical car speeds in meters per second
- estimate jogging and walking speed by timing yourself; convert the measurements into both m/s and mi/h
- determine the speed of an ant, snail, or falling leaf

Exercise 7.4.1

A commuter train travels from Baltimore to Washington, DC, and back in 1 hour and 45 minutes. The distance between the two stations is approximately 40 miles. What is

a. the average velocity of the train, and
b. the average speed of the train in m/s?

Answer

(a) The average velocity of the train is zero because $x_f = x_0$; the train ends up at the same place it starts.

(b) The average speed of the train is calculated below. Note that the train travels 40 miles one way and 40 miles back, for a total distance of 80 miles.

$$\frac{distance}{time} = \frac{80 miles}{105 minutes}$$

$$\frac{80 miles}{105 minutes} \times \frac{5280 feet}{1 mile} \times \frac{1 meter}{3.28 feet} \times \frac{1 minute}{60 seconds} = 20 m/s$$

Summary

- Time is measured in terms of change, and its SI unit is the second (s). Elapsed time for an event is

$$\Delta t = t_f - t_0,$$

  where t_f is the final time and t_0 is the initial time. The initial time is often taken to be zero, as if measured with a stopwatch; the elapsed time is then just t.
- Average velocity $\bar{v}$ is defined as displacement divided by the travel time. In symbols, average velocity is

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0}.$$

- The SI unit for velocity is m/s.
- Velocity is a vector and thus has a direction.
- Instantaneous velocity v is the velocity at a specific instant or the average velocity for an infinitesimal interval.
- Instantaneous speed is the magnitude of the instantaneous velocity.
- Instantaneous speed is a scalar quantity, as it has no direction specified.
- Average speed is the total distance traveled divided by the elapsed time. (Average speed is *not* the magnitude of the average velocity.) Speed is a scalar quantity; it has no direction associated with it.

Glossary

average speed
distance traveled divided by time during which motion occurs

average velocity
displacement divided by time over which displacement occurs

instantaneous velocity
velocity at a specific instant, or the average velocity over an infinitesimal time interval

instantaneous speed
magnitude of the instantaneous velocity

time
change, or the interval over which change occurs

model
simplified description that contains only those elements necessary to describe the physics of a physical situation

elapsed time
the difference between the ending time and beginning time

This page titled 7.4: Time, Velocity, and Speed is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **2.3: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.5: Acceleration

Learning Objectives

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

- Define and distinguish between instantaneous acceleration, average acceleration, and deceleration.
- Calculate acceleration given initial time, initial velocity, final time, and final velocity.

In everyday conversation, to accelerate means to speed up. The accelerator in a car can in fact cause it to speed up. The greater the **acceleration**, the greater the change in velocity over a given time. The formal definition of acceleration is consistent with these notions, but more inclusive.



Figure **7.5.1**: A plane decelerates, or slows down, as it comes in for landing in St. Maarten. Its acceleration is opposite in direction to its velocity. (credit: Steve Conry, Flickr)

Definition: Average Acceleration

The *average acceleration* is the rate at which velocity changes,

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0} \tag{7.5.1}$$

where $\bar{a}$ is average acceleration, v is velocity, and t is time. (The bar over the a means *average*.)

Because acceleration is velocity in m/s divided by time in s, the SI units for acceleration are m/s^2, meters per second squared or meters per second per second, which literally means by how many meters per second the velocity changes every second.

Recall that velocity is a vector—it has both magnitude and direction. This means that a change in velocity can be a change in magnitude (or speed), but it can also be a change in direction. For example, if a car turns a corner at constant speed, it is accelerating because its direction is changing. The quicker you turn, the greater the acceleration. So there is an acceleration when velocity changes either in magnitude (an increase or decrease in speed) or in direction, or both.

ACCELERATION AS A VECTOR

Acceleration is a vector in the same direction as the change in velocity, Δv. Since velocity is a vector, it can change either in magnitude or in direction. Acceleration is therefore a change in either speed or direction, or both.

Keep in mind that although acceleration is in the direction of the *change* in velocity, it is not always in the direction of *motion*. When an object slows down, its acceleration is opposite to the direction of its motion. This is known as **deceleration.**



Figure 7.5.2: A subway train in Sao Paulo, Brazil, decelerates as it comes into a station. It is accelerating in a direction opposite to its direction of motion. (credit: Yusuke Kawasaki, Flickr)

MISCONCEPTION ALERT: DECELERATION VS. NEGATIVE ACCELERATION

Deceleration always refers to acceleration in the direction opposite to the direction of the velocity. Deceleration always reduces speed. Negative acceleration, however, is acceleration in the negative direction in the chosen coordinate system. Negative acceleration may or may not be deceleration, and deceleration may or may not be considered negative acceleration. If acceleration has the same sign as the velocity, the object is speeding up. If acceleration has the opposite sign as the velocity, the object is slowing down. For example, consider Figure 7.5.2.

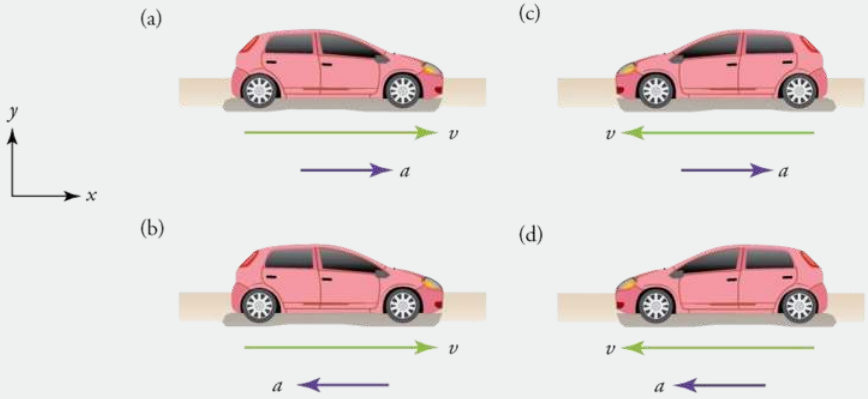


Figure 7.5.3: (a) This car is speeding up as it moves toward the right. It therefore has positive acceleration in our coordinate system. (b) This car is slowing down as it moves toward the right. Therefore, it has negative acceleration in our coordinate system, because its acceleration is toward the left. The car is also decelerating: the direction of its acceleration is opposite to its direction of motion. (c) This car is moving toward the left, but slowing down over time. Therefore, its acceleration is positive in our coordinate system because it is toward the right. However, the car is decelerating because its acceleration is opposite to its motion. (d) This car is speeding up as it moves toward the left. It has negative acceleration because it is accelerating toward the left. However, because its acceleration is in the same direction as its motion, it is speeding up (not decelerating).

Example 7.5.1: Calculating Acceleration: A Racehorse Leaves the Gate

A racehorse coming out of the gate accelerates from rest to a velocity of 15.0 m/s due west in 1.80 s. What is its average acceleration?



Figure 7.5.4: Two racehorses running toward the left. (credit: Jon Sullivan, PD Photo.org)

Strategy

First we draw a sketch and assign a coordinate system to the problem. This is a simple problem, but it always helps to visualize it. Notice that we assign east as positive and west as negative. Thus, in this case, we have negative velocity.

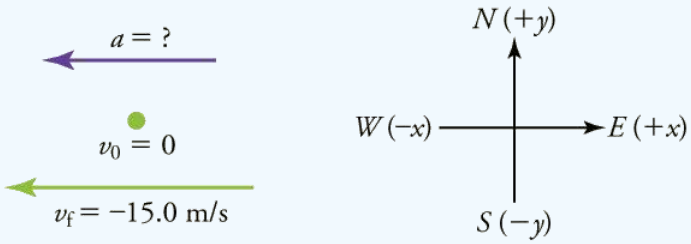


Figure 7.5.5.

We can solve this problem by identifying Δv and Δt from the given information and then calculating the average acceleration directly from the Equation ???:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}.$$

Solution

1. Identify the knowns. $v_0 = 0, v_f = -15.0 m/s$ (the negative sign indicates direction toward the west), $\Delta t = 1.80 s$.
2. Find the change in velocity. Since the horse is going from zero to $-15.0 m/s$, its change in velocity equals its final velocity:

$$\Delta v = v_f = -15.0\, m/s.$$

3. Plug in the known values (Δv and Δt) and solve for the unknown $\bar{a}$.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-15.0 m/s}{1.80 s} = -8.33 m/s^2.$$

Discussion

The negative sign for acceleration indicates that acceleration is toward the west. An acceleration of $8.33\ m/s^2$ due west means that the horse increases its velocity by 8.33 m/s due west each second, that is, 8.33 meters per second per second, which we write as $8.33\ m/s^2$. This is truly an average acceleration, because the ride is not smooth. We shall see later that an acceleration of this magnitude would require the rider to hang on with a force nearly equal to his weight.

Instantaneous Acceleration

Instantaneous acceleration a, or the acceleration at a specific instant in time, is obtained by the same process as discussed for instantaneous velocity in Time, Velocity, and Speed—that is, by considering an infinitesimally small interval of time. How do we find instantaneous acceleration using only algebra? The answer is that we choose an average acceleration that is representative of the motion. Figure 7.5.6 shows graphs of instantaneous acceleration versus time for two very different motions. In Figure 7.5.6a, the acceleration varies slightly and the average over the entire interval is nearly the same as the instantaneous acceleration at any time.

In this case, we should treat this motion as if it had a constant acceleration equal to the average (in this case about $1.8 m/s^2$). In Figure 7.5.6*b*, the acceleration varies drastically over time. In such situations it is best to consider smaller time intervals and choose an average acceleration for each. For example, we could consider motion over the time intervals from 0 to 1.0 s and from 1.0 to 3.0 s as separate motions with accelerations of $+3.0 m/s^2$ and $-2.0 m/s^2$, respectively.

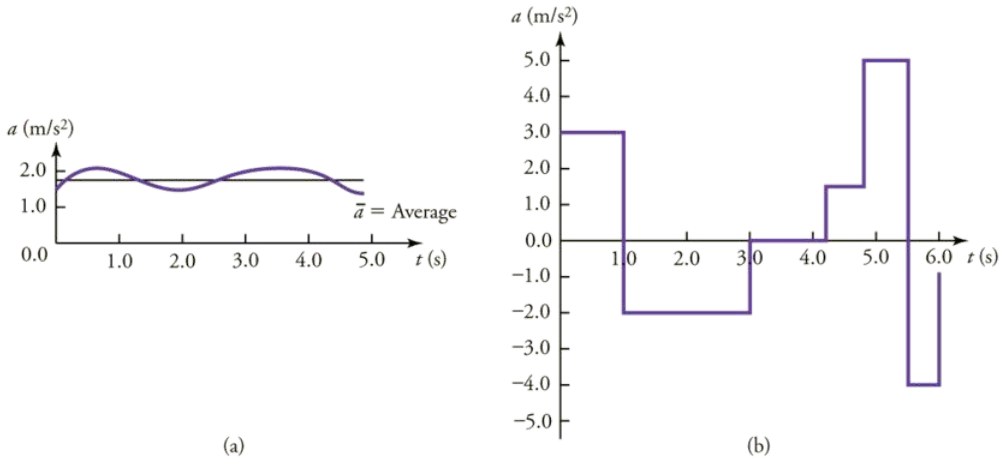


Figure 7.5.6: Graphs of instantaneous acceleration versus time for two different one-dimensional motions. (a) Here acceleration varies only slightly and is always in the same direction, since it is positive. The average over the interval is nearly the same as the acceleration at any given time. (b) Here the acceleration varies greatly, perhaps representing a package on a post office conveyor belt that is accelerated forward and backward as it bumps along. It is necessary to consider small time intervals (such as from 0 to 1.0 s) with constant or nearly constant acceleration in such a situation.

The next several examples consider the motion of the subway train shown in Figure 7.5.7. In Figure 7.5.7*a* the shuttle moves to the right, and in Figure 7.5.7*b* it moves to the left. The examples are designed to further illustrate aspects of motion and to illustrate some of the reasoning that goes into solving problems.

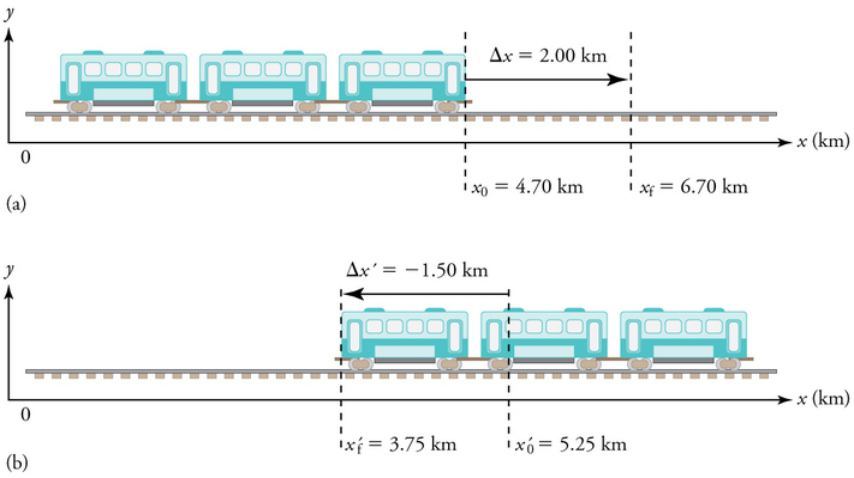


Figure 7.5.7: One-dimensional motion of a subway train considered in Examples 7.5.2 - 7.5.5. Here we have chosen the x -axis so that + means to the right and − means to the left for displacements, velocities, and accelerations. (a) The subway train moves to the right from x_0 to x_f. Its displacement Δx is +2.0 km. (b) The train moves to the left from x_0 to x_f. Its displacement Δx is $-1.5 km$. (Note that the prime symbol (′) is used simply to distinguish between displacement in the two different situations. The distances of travel and the size of the cars are on different scales to fit everything into the diagram.)

Example 7.5.2: Calculating Displacement - A Subway Train

What are the magnitude and sign of displacements for the motions of the subway train shown in parts (a) and (b) of Figure 7.5.7?

Strategy

A drawing with a coordinate system is already provided, so we don't need to make a sketch, but we should analyze it to make sure we understand what it is showing. Pay particular attention to the coordinate system. To find displacement, we use the equation $\Delta x = x_f - x_0$. This is straightforward since the initial and final positions are given.

Solution

1. Identify the knowns. In the figure we see that $x_f = 6.70\,km$ and $x_0 = 4.70\,km$ for part (a), and $x'_f = 3.75\,km$ and $x'_0 = 5.25\,km$ for part (b).
2. Solve for displacement in part (a).

$$\begin{aligned}\Delta x &= x_f - x_0 \\ &= 6.70\,km - 4.70\,km \\ &= +2.00\,km\end{aligned}$$

3. Solve for displacement in part (b).

$$\begin{aligned}\Delta x' &= x'_f - x'_0 \\ &= 3.75\,km - 5.25\,km \\ &= -1.50\,km\end{aligned}$$

Discussion

The direction of the motion in (a) is to the right and therefore its displacement has a positive sign, whereas motion in (b) is to the left and thus has a negative sign.

Example 7.5.3: Comparing Distance Traveled with Displacement - A Subway Train

What are the distances traveled for the motions shown in parts (a) and (b) of the subway train in Figure 7.5.7?

Strategy

To answer this question, think about the definitions of distance and distance traveled, and how they are related to displacement. Distance between two positions is defined to be the magnitude of displacement, which was found in Example 7.5.2. Distance traveled is the total length of the path traveled between the two positions (see Section on Displacement). In the case of the subway train shown in Figure 7.5.7, the distance traveled is the same as the distance between the initial and final positions of the train.

Solution

1. The displacement for part (a) was +2.00 km. Therefore, the distance between the initial and final positions was 2.00 km, and the distance traveled was 2.00 km.

2. The displacement for part (b) was $-1.5km$. Therefore, the distance between the initial and final positions was 1.50 km, and the distance traveled was 1.50 km.

Discussion

Distance is a scalar. It has magnitude but no sign to indicate direction.

Example 7.5.4: Calculating Acceleration: A Subway Train Speeding Up

Suppose the train in Figure 7.5.7a accelerates from rest to 30.0 km/h in the first 20.0 s of its motion. What is its average acceleration during that time interval?

Strategy

It is worth it at this point to make a simple sketch:

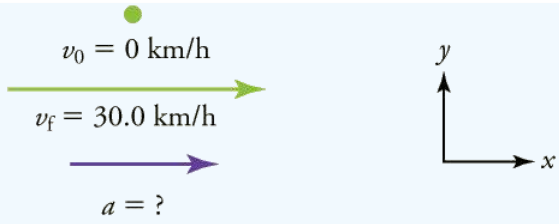


Figure 7.5.8: This problem involves three steps. First we must determine the change in velocity, then we must determine the change in time, and finally we use these values to calculate the acceleration.

Solution

1. Identify the knowns. $v_0 = 0$ (the trains starts at rest), $v_f = 30.0 km/h$, and $\Delta t = 20.0 s$.
2. Calculate Δv. Since the train starts from rest, its change in velocity is $\Delta v = +30.0 km/h$, where the plus sign means velocity to the right.
3. Plug in known values and solve for the unknown, $\bar{a}$.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+30.0 km/h}{20.0 s}$$

4. Since the units are mixed (we have both hours and seconds for time), we need to convert everything into SI units of meters and seconds. (See Physical Quantities and Units for more guidance.)

$$\bar{a} = \left(\frac{+30 km/h}{20.0 s}\right)\left(\frac{10^3 m}{1 km}\right)\left(\frac{1 h}{3600 s}\right) = 0.417 m/s^2$$

Discussion

The plus sign means that acceleration is to the right. This is reasonable because the train starts from rest and ends up with a velocity to the right (also positive). So acceleration is in the same direction as the change in velocity, as is always the case.

Example 7.5.5: Calculate Acceleration

A Subway Train Slowing Down: Now suppose that at the end of its trip, the train in Figure 7.5.7a slows to a stop from a speed of 30.0 km/h in 8.00 s. What is its average acceleration while stopping?

Strategy

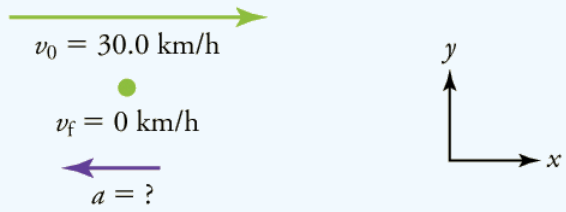


Figure 7.5.9:.

In this case, the train is decelerating and its acceleration is negative because it is toward the left. As in the previous example, we must find the change in velocity and the change in time and then solve for acceleration.

Solution

1. Identify the knowns. $v_0 = 30.0 km/h, v_f = 0 km/h$ (the train is stopped, so its velocity is 0), and $\Delta t = 8.00 s$.
2. Solve for the change in velocity, Δv.

$$\Delta v = v_f - v_0 = 0 - 30.0 km/h = -30.0 km/h$$

3. Plug in the knowns, Δv and Δt, and solve for $\bar{a}$.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-30.0 km/h}{8.00 s}$$

4. Convert the units to meters and seconds.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \left(\frac{-30.0 km/h}{8.00 s}\right)\left(\frac{10^3 m}{1 km}\right)\left(\frac{1 h}{3600 s}\right) = -1.04 m/s^2.$$

Discussion

The minus sign indicates that acceleration is to the left. This sign is reasonable because the train initially has a positive velocity in this problem, and a negative acceleration would oppose the motion. Again, acceleration is in the same direction as the change in velocity, which is negative here. This acceleration can be called a deceleration because it has a direction opposite to the velocity.

The graphs of position, velocity, and acceleration vs. time for the trains in Example **7.5.4** and **7.5.5** are displayed in Figure **7.5.10**. (We have taken the velocity to remain constant from 20 to 40 s, after which the train decelerates.)

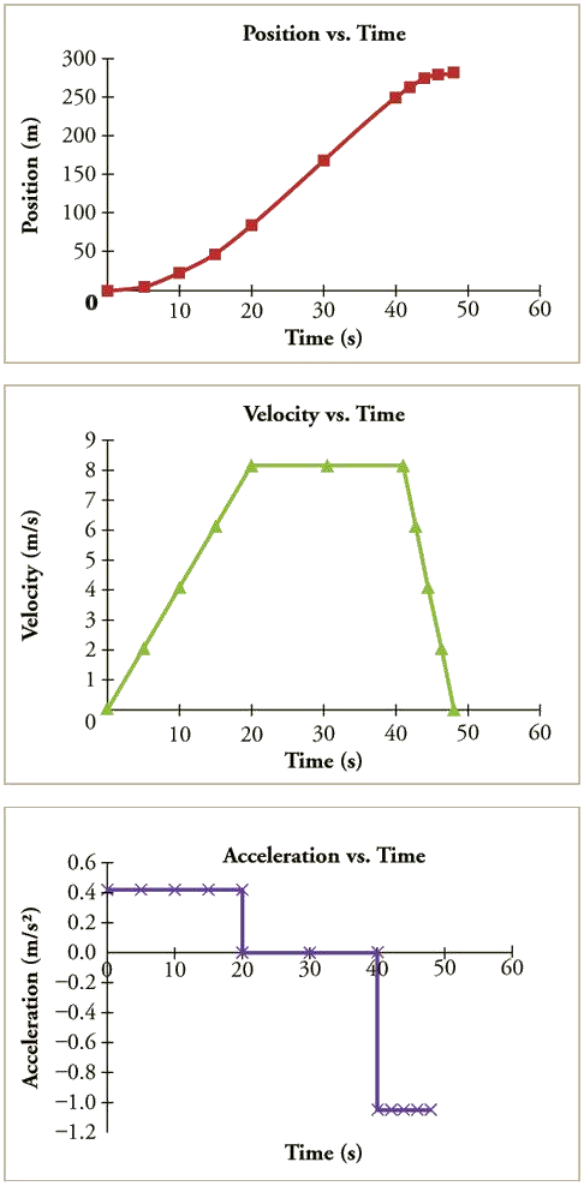


Figure **7.5.10**: (a) Position of the train over time. Notice that the train's position changes slowly at the beginning of the journey, then more and more quickly as it picks up speed. Its position then changes more slowly as it slows down at the end of the journey. In the middle of the journey, while the velocity remains constant, the position changes at a constant rate. (b) Velocity of the train over time. The train's velocity increases as it accelerates at the beginning of the journey. It remains the same in the middle of the journey (where there is no acceleration). It decreases as the train decelerates at the end of the journey. (c) The acceleration of the train over time. The train has positive acceleration as it speeds up at

the beginning of the journey. It has no acceleration as it travels at constant velocity in the middle of the journey. Its acceleration is negative as it slows down at the end of the journey.

Example 7.5.6: Calculating Average Velocity: The Subway Train

What is the average velocity of the train in part b of Example **7.5.2**, and shown again below, if it takes 5.00 min to make its trip?

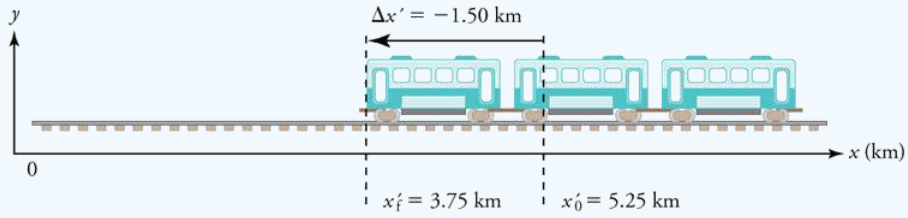


Figure **7.5.11**

Strategy

Average velocity is displacement divided by time. It will be negative here, since the train moves to the left and has a negative displacement.

Solution

1. Identify the knowns.

$$x'_f = 3.75km,\ x'_0 = 5.25km,\ \Delta t = 5.00min.$$

2. Determine displacement, $\Delta x'$. We found $\Delta x'$ to be $-1.5km$ in Example **7.5.7**.
3. Solve for average velocity.

$$\bar{v} = \frac{\Delta x'}{\Delta t} = \frac{-1.50km}{5.00min}$$

4. Convert units.

$$\bar{v} = \frac{\Delta x'}{\Delta t} = (\frac{-1.50km}{5.00min})(\frac{60min}{1}h) = -18.0km/h$$

Discussion

The negative velocity indicates motion to the left.

Example 7.5.7: Calculating Deceleration: The Subway Train

Finally, suppose the train in Figure **7.5.7** slows to a stop from a velocity of 20.0 km/h in 10.0 s. What is its average acceleration?

Strategy

Once again, let's draw a sketch:

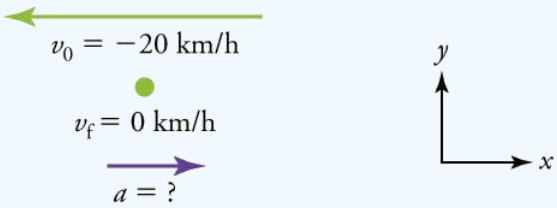


Figure **7.5.12**: As before, we must find the change in velocity and the change in time to calculate average acceleration.

Solution

1. Identify the knowns. $v_0 = -20\,km/h,\ v_f = 0\,km/h,\ \Delta t = 10.0\,s.$
2. Calculate Δv. The change in velocity here is actually positive, since

https://chem.libretexts.org/@go/page/200153

$$\Delta v = v_f - v_0 = 0 - (-20\,km/h) = +20\,km/h.$$

3. Solve for $\bar{a}$.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+20.0\,km/h}{10.0\,s}$$

4. Convert units.

$$\bar{a} = \left(\frac{+20.0\,km/h}{10.0\,s}\right)\left(\frac{10^3\,m}{1\,km}\right)\left(\frac{1\,h}{3600\,s}\right) = +0.556\,m/^s2$$

Discussion

The plus sign means that acceleration is to the right. This is reasonable because the train initially has a negative velocity (to the left) in this problem and a positive acceleration opposes the motion (and so it is to the right). Again, acceleration is in the same direction as the change in velocity, which is positive here. As in Example **7.5.5**, this acceleration can be called a deceleration since it is in the direction opposite to the velocity.

Sign and Direction

Perhaps the most important thing to note about these examples is the signs of the answers. In our chosen coordinate system, plus means the quantity is to the right and minus means it is to the left. This is easy to imagine for displacement and velocity. But it is a little less obvious for acceleration. Most people interpret negative acceleration as the slowing of an object. This was not the case in Example **7.5.5**, where a positive acceleration slowed a negative velocity. The crucial distinction was that the acceleration was in the opposite direction from the velocity. In fact, a negative acceleration will increase a negative velocity. For example, the train moving to the left in Figure Figure **7.5.11** is sped up by an acceleration to the left. In that case, both v and a are negative. The plus and minus signs give the directions of the accelerations. If acceleration has the same sign as the velocity, the object is speeding up. If acceleration has the opposite sign as the velocity, the object is slowing down.

Exercise 7.5.1

An airplane lands on a runway traveling east. Describe its acceleration.

Answer

If we take east to be positive, then the airplane has negative acceleration, as it is accelerating toward the west. It is also decelerating: its acceleration is opposite in direction to its velocity.

PHET EXPLORATIONS: MOVING MAN SIMULATION

Learn about position, velocity, and acceleration graphs with the PhET Moving Man simulation. Move the little man back and forth with the mouse and plot his motion. Set the position, velocity, or acceleration and let the simulation move the man for you.

Summary

- Acceleration is the rate at which velocity changes. In symbols, **average acceleration** $\bar{a}$ is $\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}$.
- The SI unit for acceleration is 2 .
- Acceleration is a vector, and thus has a both a magnitude and direction.
- Acceleration can be caused by either a change in the magnitude or the direction of the velocity.
- Instantaneous acceleration is the acceleration at a specific instant in time.
- Deceleration is an acceleration with a direction opposite to that of the velocity.

Glossary

acceleration

the rate of change in velocity; the change in velocity over time

average acceleration

the change in velocity divided by the time over which it changes

instantaneous acceleration

acceleration at a specific point in time

deceleration

acceleration in the direction opposite to velocity; acceleration that results in a decrease in velocity

---

This page titled 7.5: Acceleration is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **2.4: Acceleration** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.6: Kinematics in Two Dimensions - An Introduction

Learning Objectives

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

- Observe that motion in two dimensions consists of horizontal and vertical components.
- Understand the independence of horizontal and vertical vectors in two-dimensional motion.



Figure 7.6.1: Walkers and drivers in a city like New York are rarely able to travel in straight lines to reach their destinations. Instead, they must follow roads and sidewalks, making two-dimensional, zigzagged paths. (credit: Margaret W. Carruthers)

Two-Dimensional Motion: Walking in a City

Suppose you want to walk from one point to another in a city with uniform square blocks, as pictured in Figure 7.6.2.

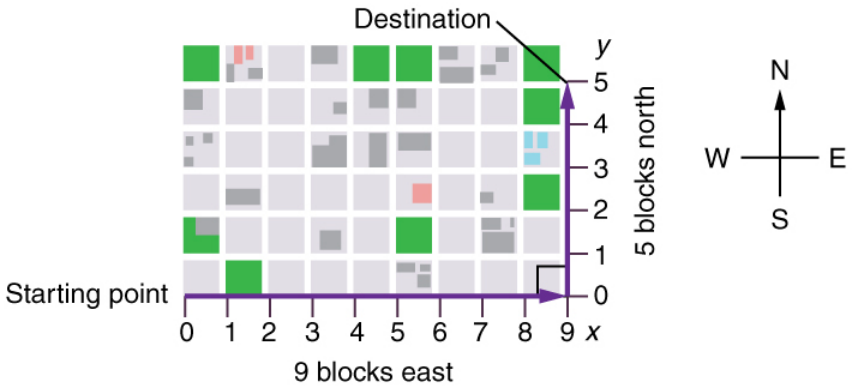


Figure 7.6.2: A pedestrian walks a two-dimensional path between two points in a city. In this scene, all blocks are square and are the same size.

The straight-line path that a helicopter might fly is blocked to you as a pedestrian, and so you are forced to take a two-dimensional path, such as the one shown. You walk 14 blocks in all, 9 east followed by 5 north. What is the straight-line distance?

An old adage states that the shortest distance between two points is a straight line. The two legs of the trip and the straight-line path form a right triangle, and so the Pythagorean theorem,

$$a^2 + b^2 = c^2 \tag{7.6.1}$$

can be used to find the straight-line distance.

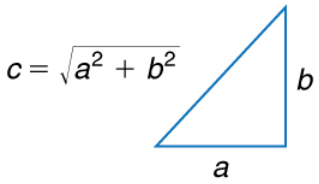


Figure **7.6.3**: The Pythagorean theorem relates the length of the legs of a right triangle, labeled a and b, with the hypotenuse, labeled c. The relationship is given by: $a^2 + b^2 = c^2$. This can be rewritten, solving for c : $c = \sqrt{a^2 + b^2}$.

The hypotenuse of the triangle is the straight-line path, and so in this case its length in units of city blocks is $\sqrt{(9blocks)^2 + (5blocks)^2} = 10.3blocks$, considerably shorter than the 14 blocks you walked. (Note that we are using three significant figures in the answer. Although it appears that "9" and "5" have only one significant digit, they are discrete numbers. In this case "9 blocks" is the same as "9.0 or 9.00 blocks." We have decided to use three significant figures in the answer in order to show the result more precisely.)

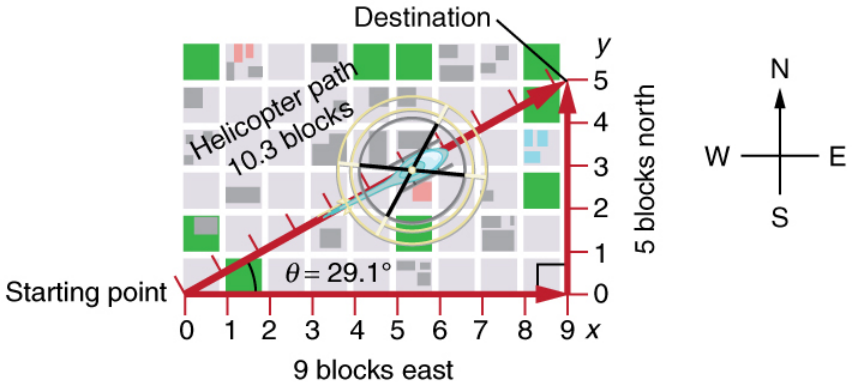


Figure **7.6.4**: The straight-line path followed by a helicopter between the two points is shorter than the 14 blocks walked by the pedestrian. All blocks are square and the same size.

The fact that the straight-line distance (10.3 blocks) in Figure is less than the total distance walked (14 blocks) is one example of a general characteristic of vectors. (Recall that vectors are quantities that have both magnitude and direction.)

As for one-dimensional kinematics, we use arrows to represent vectors. The length of the arrow is proportional to the vector's magnitude. The arrow's length is indicated by hash marks in Figure and Figure. The arrow points in the same direction as the vector. For two-dimensional motion, the path of an object can be represented with three vectors: one vector shows the straight-line path between the initial and final points of the motion, one vector shows the horizontal component of the motion, and one vector shows the vertical component of the motion. The horizontal and vertical components of the motion add together to give the straight-line path. For example, observe the three vectors in Figure. The first represents a 9-block displacement east. The second represents a 5-block displacement north. These vectors are added to give the third vector, with a 10.3-block total displacement. The third vector is the straight-line path between the two points. Note that in this example, the vectors that we are adding are perpendicular to each other and thus form a right triangle. This means that we can use the Pythagorean theorem to calculate the magnitude of the total displacement. (Note that we cannot use the Pythagorean theorem to add vectors that are not perpendicular. We will develop techniques for adding vectors having any direction, not just those perpendicular to one another, in Vector Addition and Subtraction: Graphical Methods and Vector Addition and Subtraction: Analytical Methods.)

The Independence of Perpendicular Motions

The person taking the path shown in Figure walks east and then north (two perpendicular directions). How far he or she walks east is only affected by his or her motion eastward. Similarly, how far he or she walks north is only affected by his or her motion northward.

INDEPENDENCE OF MOTION

The horizontal and vertical components of two-dimensional motion are independent of each other. Any motion in the horizontal direction does not affect motion in the vertical direction, and vice versa.

This is true in a simple scenario like that of walking in one direction first, followed by another. It is also true of more complicated motion involving movement in two directions at once. For example, let's compare the motions of two baseballs. One baseball is dropped from rest. At the same instant, another is thrown horizontally from the same height and follows a curved path. A stroboscope has captured the positions of the balls at fixed time intervals as they fall.

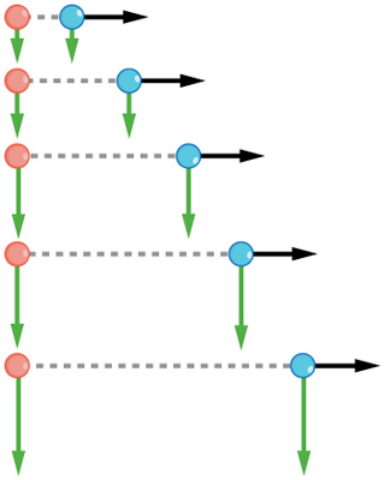


Figure 7.6.5: This shows the motions of two identical balls—one falls from rest, the other has an initial horizontal velocity. Each subsequent position is an equal time interval. Arrows represent horizontal and vertical velocities at each position. The ball on the right has an initial horizontal velocity, while the ball on the left has no horizontal velocity. Despite the difference in horizontal velocities, the vertical velocities and positions are identical for both balls. This shows that the vertical and horizontal motions are independent.

It is remarkable that for each flash of the strobe, the vertical positions of the two balls are the same. This similarity implies that the vertical motion is independent of whether or not the ball is moving horizontally. (Assuming no air resistance, the vertical motion of a falling object is influenced by gravity only, and not by any horizontal forces.) Careful examination of the ball thrown horizontally shows that it travels the same horizontal distance between flashes. This is due to the fact that there are no additional forces on the ball in the horizontal direction after it is thrown. This result means that the horizontal velocity is constant, and affected neither by vertical motion nor by gravity (which is vertical). Note that this case is true only for ideal conditions. In the real world, air resistance will affect the speed of the balls in both directions.

The two-dimensional curved path of the horizontally thrown ball is composed of two independent one-dimensional motions (horizontal and vertical). The key to analyzing such motion, called *projectile motion*, is to *resolve* (break) it into motions along perpendicular directions. Resolving two-dimensional motion into perpendicular components is possible because the components are independent. We shall see how to resolve vectors in Vector Addition and Subtraction: Graphical Methods and Vector Addition and Subtraction: Analytical Methods. We will find such techniques to be useful in many areas of physics.

PHET EXPLORATIONS: LADYBUG MOTION 2D

Learn about position, velocity and acceleration vectors. Move the ladybug by setting the position, velocity or acceleration, and see how the vectors change. Choose linear, circular or elliptical motion, and record and playback the motion to analyze the behavior.

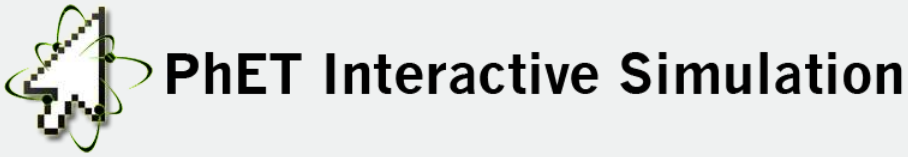


Figure 7.6.5: Ladybug Motion 2D

Summary

- The shortest path between any two points is a straight line. In two dimensions, this path can be represented by a vector with horizontal and vertical components.
- The horizontal and vertical components of a vector are independent of one another. Motion in the horizontal direction does not affect motion in the vertical direction, and vice versa.

Glossary

vector

a quantity that has both magnitude and direction; an arrow used to represent quantities with both magnitude and direction

---

This page titled 7.6: Kinematics in Two Dimensions - An Introduction is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **3.1: Kinematics in Two Dimensions - An Introduction** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.7: Projectile Motion

Learning Objectives

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

- Identify and explain the properties of a projectile, such as acceleration due to gravity, range, maximum height, and trajectory.
- Determine the location and velocity of a projectile at different points in its trajectory.
- Apply the principle of independence of motion to solve projectile motion problems.

Projectile motion is the **motion** of an object thrown or projected into the air, subject to only the acceleration of gravity. The object is called a **projectile**, and its path is called its **trajectory.** The motion of falling objects, as covered in Problem-Solving Basics for One-Dimensional Kinematics, is a simple one-dimensional type of projectile motion in which there is no horizontal movement. In this section, we consider two-dimensional projectile motion, such as that of a football or other object for which **air resistance** *is negligible*.

The most important fact to remember here is that motions along perpendicular axes are independent and thus can be analyzed separately. This fact was discussed in Kinematics in Two Dimensions: An Introduction, where vertical and horizontal motions were seen to be independent. The key to analyzing two-dimensional projectile motion is to break it into two motions, one along the horizontal axis and the other along the vertical. (This choice of axes is the most sensible, because acceleration due to gravity is vertical—thus, there will be no acceleration along the horizontal axis when air resistance is negligible.) As is customary, we call the horizontal axis the *x*-axis and the vertical axis the *y*-axis. Figure illustrates the notation for displacement, where $\boldsymbol{s}$ is defined to be the total displacement and $\boldsymbol{x}$ and $\boldsymbol{y}$ are its components along the horizontal and vertical axes, respectively. The magnitudes of these vectors are s, x, and y. (Note that in the last section we used the notation $\boldsymbol{A}$ to represent a vector with components $\boldsymbol{A_x}$ and $\boldsymbol{A_y}$. If we continued this format, we would call displacement $\boldsymbol{s}$ with components $\boldsymbol{s_x}$ and $\boldsymbol{s_y}$. However, to simplify the notation, we will simply represent the component vectors as $\boldsymbol{x}$ and $\boldsymbol{y}$.)

Of course, to describe motion we must deal with velocity and acceleration, as well as with displacement. We must find their components along the x- and *y*-axes, too. We will assume all forces except gravity (such as air resistance and friction, for example) are negligible. The components of acceleration are then very simple: $\boldsymbol{a_y = -g = -9.80 m/s^2}$(. (Note that this definition assumes that the upwards direction is defined as the positive direction. If you arrange the coordinate system instead such that the downwards direction is positive, then acceleration due to gravity takes a positive value.) Because gravity is vertical, $\boldsymbol{a_x = 0}$. Both accelerations are constant, so the kinematic equations can be used.

REVIEW OF KINEMATIC EQUATIONS (CONSTANT)

$$x = x_0 + \bar{v}t$$
$$\bar{v} = \frac{v_0 + v}{2}$$
$$v = v_0 + at$$
$$x = x_0 + v_0 t + \frac{1}{2}at^2$$
$$v^2 = v_0^2 + 2a(x - x_0).$$

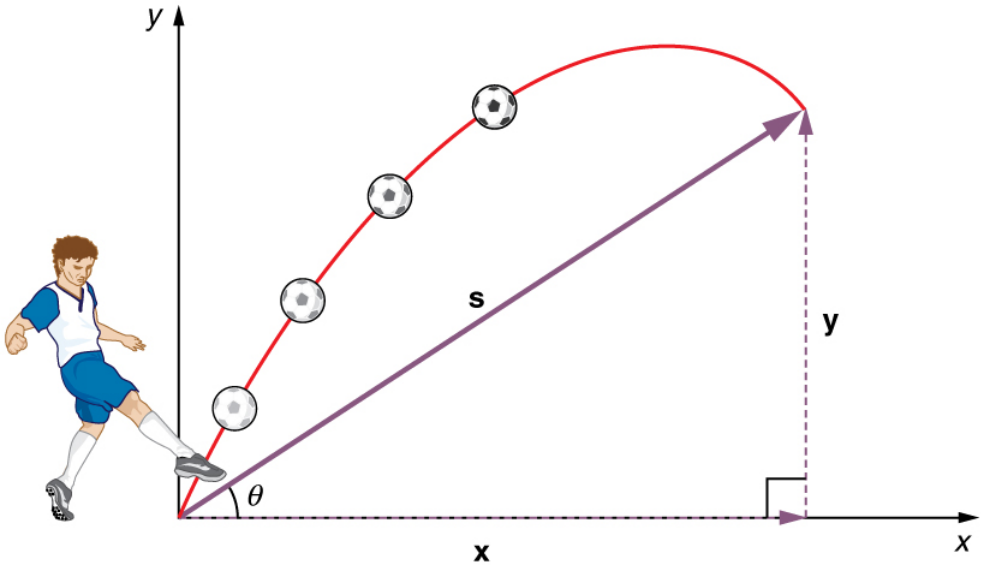


Figure 7.7.1: The total displacement s of a soccer ball at a point along its path. The vector s has components x and y along the horizontal and vertical axes. Its magnitude is s, and it makes an angle θ with the horizontal.

Given these assumptions, the following steps are then used to analyze projectile motion:

Step 1. *Resolve or break the motion into horizontal and vertical components along the x- and y-axes.* These axes are perpendicular, so $A_x = Acos\theta$ and $A_y = Asin\theta$ are used. The magnitude of the components of displacement s along these axes are x and y. The magnitudes of the components of the velocity v are $v_x = vcos\theta$ and $v_y = vsin\theta$, where v is the magnitude of the velocity and θ is its direction, as shown in Figure. Initial values are denoted with a subscript 0, as usual.

Step 2. *Treat the motion as two independent one-dimensional motions, one horizontal and the other vertical.* The kinematic equations for horizontal and vertical motion take the following forms:

$$\text{Horizontal Motion}(a_x = 0)$$
$$x = x_0 + v_x t$$
$$v_x = v_{0x} = v_x = \text{velocity is a constant.}$$
$$\text{Vertical Motion(assuming positive is up} a_y = -g = -9.80 m/s^2)$$
$$y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$$
$$v_y = v_{0y} - gt$$
$$y = y_0 + v_{0y}t - \frac{1}{2}gt^2$$
$$v_y^2 = v_{0y}^2 - 2g(y - y_0).$$

Step 3. *Solve for the unknowns in the two separate motions—one horizontal and one vertical.* Note that the only common variable between the motions is time *t*. The problem solving procedures here are the same as for one-dimensional **kinematics** and are illustrated in the solved examples below.

Step 4. *Recombine the two motions to find the total displacement s and velocity v.* Because the *x* - and *y* -motions are perpendicular, we determine these vectors by using the techniques outlined in the Vector Addition and Subtraction: Analytical Methods and employing $A = \sqrt{A_x^2 + A_y^2}$ and $\theta = tan^{-1}(A_y/A_x)$ in the following form, where θ is the direction of the displacement s and θ_v is the direction of the velocity v:

Total displacement and velocity

$$s = \sqrt{x^2 + y^2}$$
$$\theta = tan^{-1}(y/x)$$
$$v = \sqrt{v_x^2 + v_y^2}$$
$$\theta_v = tan^{-1}(v_y/v_x).$$

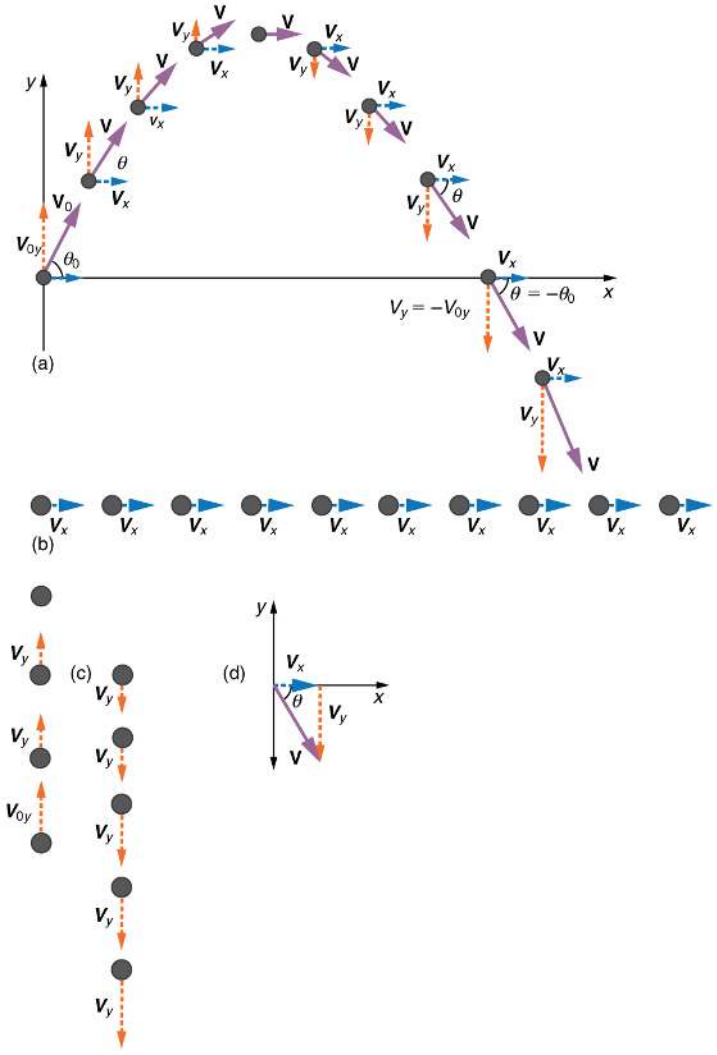


Figure **7.7.2**: (a) We analyze two-dimensional projectile motion by breaking it into two independent one-dimensional motions along the vertical and horizontal axes. (b) The horizontal motion is simple, because $a_x = 0$ and v_x is thus constant. (c) The velocity in the vertical direction begins to decrease as the object rises; at its highest point, the vertical velocity is zero. As the object falls towards the Earth again, the vertical velocity increases again in magnitude but points in the opposite direction to the initial vertical velocity. (d) The x - and y -motions are recombined to give the total velocity at any given point on the trajectory.

Example 7.7.1: A Fireworks Projectile Explodes High and Away

During a fireworks display, a shell is shot into the air with an initial speed of 70.0 m/s at an angle of **75.0°** above the horizontal, as illustrated in Figure. The fuse is timed to ignite the shell just as it reaches its highest point above the ground.

a. Calculate the height at which the shell explodes.
b. How much time passed between the launch of the shell and the explosion?
c. What is the horizontal displacement of the shell when it explodes?

Strategy

Because air resistance is negligible for the unexploded shell, the analysis method outlined above can be used. The motion can be broken into horizontal and vertical motions in which $a_x = 0$ and $a_y = -g$. We can then define x_0 and y_0 to be zero and solve for the desired quantities.

Solution for (a)

By “height” we mean the altitude or vertical position y above the starting point. The highest point in any trajectory, called the apex, is reached when $v_y = 0$. Since we know the initial and final velocities as well as the initial position, we

use the following equation to find y:

$$v_y^2 = v_{0y}^2 - 2g(y - y_0).$$

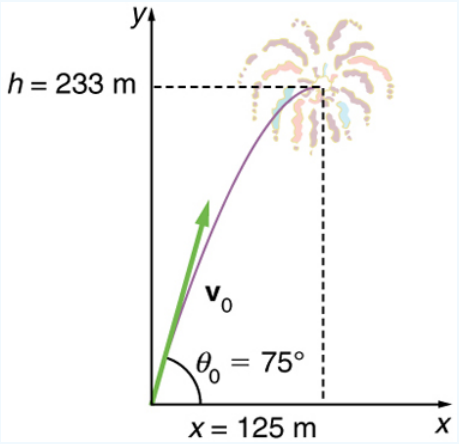


Figure 7.7.3: The trajectory of a fireworks shell. The fuse is set to explode the shell at the highest point in its trajectory, which is found to be at a height of 233 m and 125 m away horizontally.

Because y_0 and v_y are both zero, the equation simplifies to

$$0 = v_{0y}^2 - 2gy.$$

Solving for y gives

$$y = \frac{v_{0y}^2}{2g}.$$

Now we must find v_{0y}, the component of the initial velocity in the *y*-direction. It is given by $v_{0y} = v_0 sin\theta$, where v_{0y} is the initial velocity of 70.0 m/s, and $\theta_0 = 75.0^o$ is the initial angle. Thus,

$$v_{0y} = v_0 sin\theta_0 = (70.0 m/s)(sin75^o) = 67.6 m/s.$$

and y is

$$y = \frac{(67.6 m/s)^2}{2(9.80 m/s^2)},$$

so that

$$y = 233 m.$$

Discussion for (a)

Note that because up is positive, the initial velocity is positive, as is the maximum height, but the acceleration due to gravity is negative. Note also that the maximum height depends only on the vertical component of the initial velocity, so that any projectile with a 67.6 m/s initial vertical component of velocity will reach a maximum height of 233 m (neglecting air resistance). The numbers in this example are reasonable for large fireworks displays, the shells of which do reach such heights before exploding. In practice, air resistance is not completely negligible, and so the initial velocity would have to be somewhat larger than that given to reach the same height.

Solution for (b)

As in many physics problems, there is more than one way to solve for the time to the highest point. In this case, the easiest method is to use $y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$. Because y0 is zero, this equation reduces to simply

$$y = \frac{1}{2}(v_{0y} + v_y)t.$$

Note that the final vertical velocity, v_y, at the highest point is zero. Thus,

$$t = \frac{2y}{(v_{0y} + v_y)} = \frac{2(233 m)}{(67.6 m/s)} = 6.90 s.$$

Discussion for (b)

This time is also reasonable for large fireworks. When you are able to see the launch of fireworks, you will notice several seconds pass before the shell explodes. (Another way of finding the time is by using $y = y_0 + v_{0y}t - \frac{1}{2}gt^2$, and solving the quadratic equation for t.)

Solution for (c)

Because air resistance is negligible, $a_x = 0$ and the horizontal velocity is constant, as discussed above. The horizontal displacement is horizontal velocity multiplied by time as given by $x = x_0 + v_x t$, where x_0 is equal to zero:

$$x = v_x t,$$

where v_x is the *x*-component of the velocity, which is given by $v_x = v_0 cos\theta_0$. Now,

$$v_x = v_0 cos\theta_0 = (70.0 m/s)(cos 75.0^\circ) = 18.1 m/s.$$

The time t for both motions is the same, and so x is

$$x = (18.1 m/s)(6.90 s) = 125 m.$$

Discussion for (c)

The horizontal motion is a constant velocity in the absence of air resistance. The horizontal displacement found here could be useful in keeping the fireworks fragments from falling on spectators. Once the shell explodes, air resistance has a major effect, and many fragments will land directly below.

In solving part (a) of the preceding example, the expression we found for y is valid for any projectile motion where air resistance is negligible. Call the maximum height $y = h$; then,

$$h = \frac{v_{0y}^2}{2g}.$$

This equation defines the *maximum height* of a projectile and depends only on the vertical component of the initial velocity.

DEFINING A COORDINATE SYSTEM

It is important to set up a coordinate system when analyzing projectile motion. One part of defining the coordinate system is to define an origin for the x and y positions. Often, it is convenient to choose the initial position of the object as the origin such that $x_0 = 0$ and $y_0 = 0$. It is also important to define the positive and negative directions in the x and y directions. Typically, we define the positive vertical direction as upwards, and the positive horizontal direction is usually the direction of the object's motion. When this is the case, the vertical acceleration, g, takes a negative value (since it is directed downwards towards the Earth). However, it is occasionally useful to define the coordinates differently. For example, if you are analyzing the motion of a ball thrown downwards from the top of a cliff, it may make sense to define the positive direction downwards since the motion of the ball is solely in the downwards direction. If this is the case, g takes a positive value.

Example 7.7.2: Calculating Projectile Motion: Hot Rock Projectile

Kilauea in Hawaii is the world's most continuously active volcano. Very active volcanoes characteristically eject red-hot rocks and lava rather than smoke and ash. Suppose a large rock is ejected from the volcano with a speed of 25.0 m/s and at an angle above the horizontal, as shown in Figure. The rock strikes the side of the volcano at an altitude 20.0 m lower than its starting point. (a) Calculate the time it takes the rock to follow this path. (b) What are the magnitude and direction of the rock's velocity at impact?

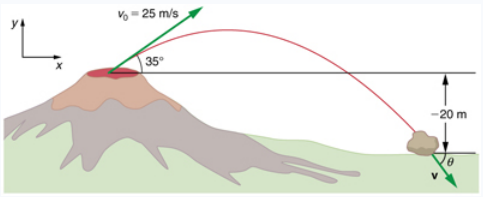


Figure 7.7.4: The trajectory of a rock ejected from the Kilauea volcano.

Strategy

Again, resolving this two-dimensional motion into two independent one-dimensional motions will allow us to solve for the desired quantities. The time a projectile is in the air is governed by its vertical motion alone. We will solve for t first. While the rock is rising and falling vertically, the horizontal motion continues at a constant velocity. This example asks for the final velocity. Thus, the vertical and horizontal results will be recombined to obtain v and θ_v at the final time t determined in the first part of the example.

Solution for (a)

While the rock is in the air, it rises and then falls to a final position 20.0 m lower than its starting altitude. We can find the time for this by using

$$y = y_0 + v_{0y}t - \frac{1}{2}gt^2.$$

If we take the initial position y_0 to be zero, then the final position is $y = -20.0m$. Now the initial vertical velocity is the vertical component of the initial velocity, found from $v_{0y} = v_0 sin\theta_0 = (25.0m/s)(sin35.0^o) = 14.3m/s$. Substituting known values yields

$$-20.0m = (14.3m/s)t - (4.90m/s^2)t^2.$$

Rearranging terms gives a quadratic equation in t:

$$(4.90m/s^2)t^2 - (14.3m/s)t - (20.0m) = 0.$$

This expression is a quadratic equation of the form $at^2 + bt + c = 0$, where the constants are $a = 4.90, b = -14.3$, and $c = -20.0$. Its solutions are given by the quadratic formula:

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This equation yields two solutions: $t = 3.96$ and $t = -1.03$. (It is left as an exercise for the reader to verify these solutions.) The time is $t = 3.96s$ or $-1.03s$. The negative value of time implies an event before the start of motion, and so we discard it. Thus,

$$t = 3.96s.$$

Discussion for (a)

The time for projectile motion is completely determined by the vertical motion. So any projectile that has an initial vertical velocity of 14.3 m/s and lands 20.0 m below its starting altitude will spend 3.96 s in the air.

Solution for (b)

From the information now in hand, we can find the final horizontal and vertical velocities v_x and v_y and combine them to find the total velocity v and the angle θ_0 it makes with the horizontal. Of course, v_x is constant so we can solve for it at any horizontal location. In this case, we chose the starting point since we know both the initial velocity and initial angle. Therefore:

$$v_x = v_0 cos\theta_0 = (25.0m/s)(cos35^o) = 20.5m/s.$$

The final vertical velocity is given by the following equation:

$$v_y = v_{0y} - gt,$$

where v_{0y} was found in part (a) to be **14.3 m/s**. Thus,

$$v_y = 14.3m/s - (9.80m/s^2)(3.96s)$$

so that

$$v_y = -24.5m/s.$$

To find the magnitude of the final velocity v we combine its perpendicular components, using the following equation:

$$v = \sqrt{v_x^2 + v_y^2} = \sqrt{(20.5m/s)^2 + (-24.5m/s)^2},$$

which gives

$$v = 31.9 m/s.$$

The direction θv is found from the equation:

$$\theta_v = tan^{-1}(v_y/v_x)$$

so that

$$\theta_v = tan^{-1}(-24.5/20.5) = tan^{-1}(-1.19).$$

Thus,

$$\theta_v = -50.1^o.$$

Discussion for (b)

The negative angle means that the velocity is **50.1°** below the horizontal. This result is consistent with the fact that the final vertical velocity is negative and hence downward—as you would expect because the final altitude is 20.0 m lower than the initial altitude. (See Figure.)

One of the most important things illustrated by projectile motion is that vertical and horizontal motions are independent of each other. Galileo was the first person to fully comprehend this characteristic. He used it to predict the range of a projectile. On level ground, we define **range** to be the horizontal distance traveled by a projectile. Galileo and many others were interested in the range of projectiles primarily for military purposes—such as aiming cannons. However, investigating the range of projectiles can shed light on other interesting phenomena, such as the orbits of satellites around the Earth. Let us consider projectile range further.

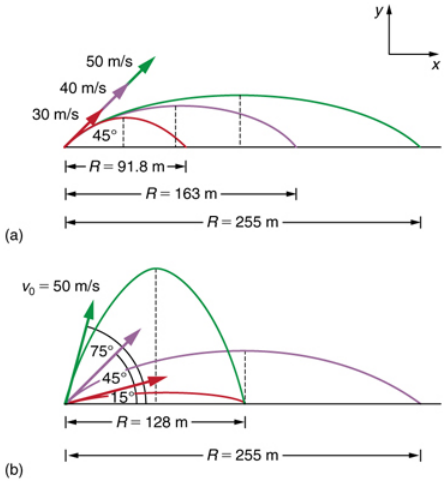


Figure **7.7.5**: Trajectories of projectiles on level ground. (a) The greater the initial speed v_0, the greater the range for a given initial angle. (b) The effect of initial angle θ_0 on the range of a projectile with a given initial speed. Note that the range is the same for 15° and 75°, although the maximum heights of those paths are different.

How does the initial velocity of a projectile affect its range? Obviously, the greater the initial speed v_0, the greater the range, as shown in Figure(a). The initial angle θ_0 also has a dramatic effect on the range, as illustrated in Figure(b). For a fixed initial speed, such as might be produced by a cannon, the maximum range is obtained with $\theta_0 = 45^o$. This is true only for conditions neglecting air resistance. If air resistance is considered, the maximum angle is approximately 38^o. Interestingly, for every initial angle except 45^o, there are two angles that give the same range—the sum of those angles is 90^o. The range also depends on the value of the acceleration of gravity g. The lunar astronaut Alan Shepherd was able to drive a golf ball a great distance on the Moon because gravity is weaker there. The range R of a projectile on level ground for which air resistance is negligible is given by

$$R = \frac{v_0^2 sin 2\theta_0}{g},$$

where v_0 is the initial speed and θ_0 is the initial angle relative to the horizontal. The proof of this equation is left as an end-of-chapter problem (hints are given), but it does fit the major features of projectile range as described.

When we speak of the range of a projectile on level ground, we assume that R is very small compared with the circumference of the Earth. If, however, the range is large, the Earth curves away below the projectile and acceleration of gravity changes direction

along the path. The range is larger than predicted by the range equation given above because the projectile has farther to fall than it would on level ground. (See Figure.) If the initial speed is great enough, the projectile goes into orbit. This is called exit velocity. This possibility was recognized centuries before it could be accomplished. When an object is in orbit, the Earth curves away from underneath the object at the same rate as it falls. The object thus falls continuously but never hits the surface. These and other aspects of orbital motion, such as the rotation of the Earth, will be covered analytically and in greater depth later in this text.

Once again we see that thinking about one topic, such as the range of a projectile, can lead us to others, such as the Earth orbits. In Addition of Velocities, we will examine the addition of velocities, which is another important aspect of two-dimensional kinematics and will also yield insights beyond the immediate topic.

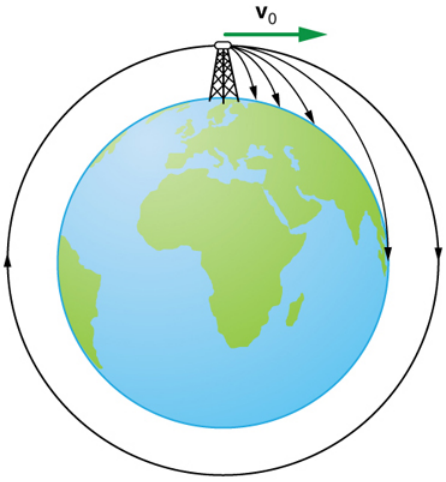


Figure **7.7.6**: Projectile to satellite. In each case shown here, a projectile is launched from a very high tower to avoid air resistance. With increasing initial speed, the range increases and becomes longer than it would be on level ground because the Earth curves away underneath its path. With a large enough initial speed, orbit is achieved.

PHET EXPLORATIONS: PROJECTILE MOTION

Blast a Buick out of a cannon! Learn about projectile motion by firing various objects. Set the angle, initial speed, and mass. Add air resistance. Make a game out of this simulation by trying to hit a target.

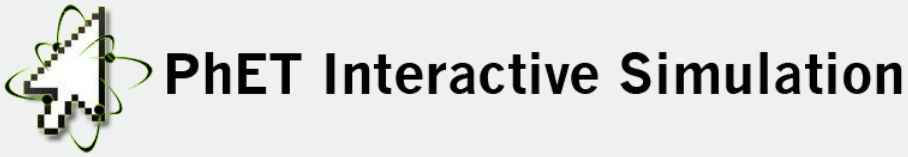


Figure **7.7.7**: Projectile Motion

Summary

- Projectile motion is the motion of an object through the air that is subject only to the acceleration of gravity.
- To solve projectile motion problems, perform the following steps:

1. Determine a coordinate system. Then, resolve the position and/or velocity of the object in the horizontal and vertical components. The components of position s are given by the quantities x and y, and the components of the velocity v are given by vx=vcosθ and vy=vsinθ, where v is the magnitude of the velocity and θ is its direction.

2. Analyze the motion of the projectile in the horizontal direction using the following equations:

$$\textbf{Horizontal motion}(a_x = 0)$$

$$x = x_0 + v_x t$$

$$v_x = v_{0x} = v_x = \textbf{velocity is a constant.}$$

3. Analyze the motion of the projectile in the vertical direction using the following equations:

$$\textbf{Vertical motion}(\textbf{Assuming positive direction is up;} a_y = -g = -9.80 m/s^2)$$

$$y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$$

$$v_y = v_{0y} - gt$$

$$y = y_0 + v_{0y}t - \frac{1}{2}gt^2$$

$$v_y^2 = v_{0y}^2 - 2g(y - y_0).$$

4. Recombine the horizontal and vertical components of location and/or velocity using the following equations:

$$s = \sqrt{x^2 + y^2}$$

$$\theta = tan^{-1}(y/x)$$

$$v = \sqrt{v_x^2 + v_y^2}$$

$$\theta_v = tan^{-1}(v_y/v_x).$$

- The maximum height h of a projectile launched with initial vertical velocity v_{0y} is given by

$$h = \frac{v_{0y}^2}{2g}.$$

- The maximum horizontal distance traveled by a projectile is called the **range**. The range R of a projectile on level ground launched at an angle θ_0 above the horizontal with initial speed v_0 is given by

$$R = \frac{v_0^2 sin2\theta_0}{g}.$$

Glossary

air resistance
: a frictional force that slows the motion of objects as they travel through the air; when solving basic physics problems, air resistance is assumed to be zero

kinematics
: the study of motion without regard to mass or force

motion
: displacement of an object as a function of time

projectile
: an object that travels through the air and experiences only acceleration due to gravity

projectile motion
: the motion of an object that is subject only to the acceleration of gravity

range
: the maximum horizontal distance that a projectile travels

trajectory
: the path of a projectile through the air

This page titled 7.7: Projectile Motion is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **3.4: Projectile Motion** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.8: 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. Our everyday experiences also give us a good idea of how multiple forces add. If two people push in different directions on a third person, as illustrated in Figure **7.8.1**, we might expect the total force to be in the direction shown. Since force is a vector, it adds just like other vectors, as illustrated in Figure **7.8.1***a* for two ice skaters. Forces, like other vectors, are represented by arrows and can be added using the familiar head-to-tail method or by trigonometric methods. These ideas were developed in Two-Dimensional Kinematics.

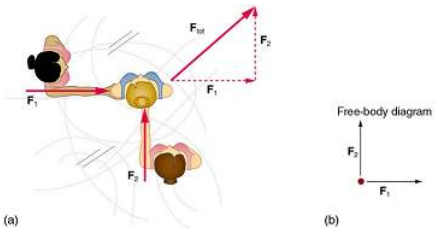


Figure **7.8.1**: Part (a) shows an overhead view of two ice skaters pushing on a third. Forces are vectors and add like other vectors, so the total force on the third skater is in the direction shown. In part (b), we see a free-body diagram representing the forces acting on the third skater.

Figure **7.8.1***b* is 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. (These forces are the only ones shown, because only external forces acting on the body affect its motion. We can ignore any internal forces within the body.) 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.

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 **7.8.2**, 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. (One that we will encounter in Magnetism is the magnetic force between two wires carrying electric current.) Some alternative definitions of force will be given later in this chapter.

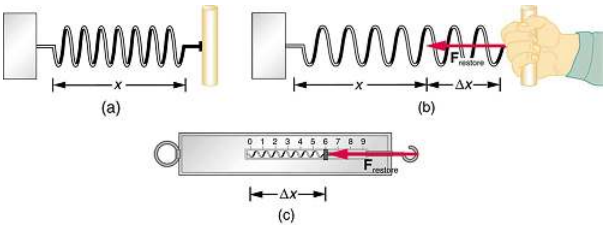


Figure **7.8.2**:The force exerted by a stretched spring can be used as a standard unit of force. (a) This spring has a length when undistorted. (b) When stretched a distance Δx the spring exerts a restoring force, $F_{restore}$ *which is reproducible. (c) A spring scale is one device that uses a spring 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. A **free-body diagram** is a drawing of all external forces acting 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

This page titled 7.8: Development of Force Concept is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **4.1: Development of Force Concept** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.9: 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 7.9.1

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

---

This page titled 7.9: Newton's First Law of Motion- Inertia is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **4.2: Newton's First Law of Motion - Inertia** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.10: 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 7.10.1*a* 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 7.10.1*a*, 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.

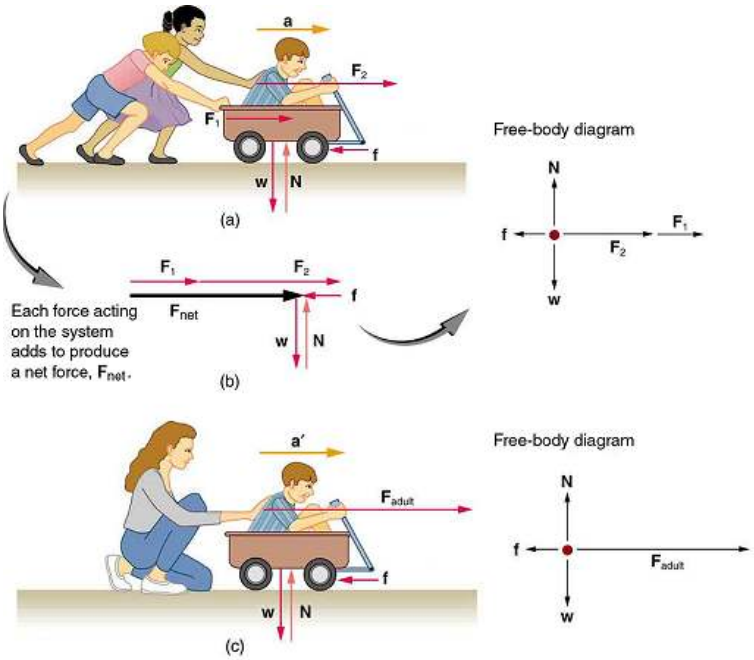


Figure 7.10.1: Different forces exerted on the same mass produce different accelerations. (a) Two children push a wagon with a child in it. Arrows representing all external forces are shown. The system of interest is the wagon and its rider. The weight w of the system and the support of the ground N 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 **7.10.1***b* 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

$$a \propto F_{net} \tag{7.10.1}$$

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

$$a \propto \frac{1}{m}, \tag{7.10.2}$$

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.

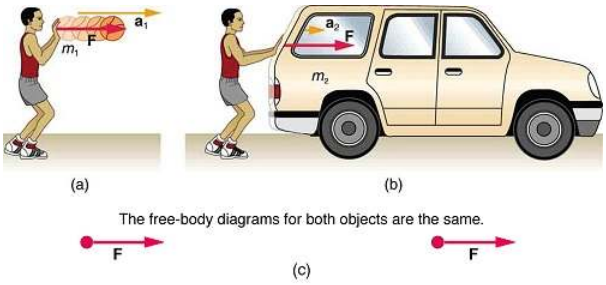


Figure **7.10.2**:The same force exerted on systems of different masses produces different accelerations. (a) A basketball player pushes on a basketball to make a pass. (The effect of gravity on the ball is ignored.) (b) The same player exerts an identical force on a stalled SUV and produces a far smaller 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

$$a = \frac{F_{net}}{m} \tag{7.10.3}$$

This is often written in the more familiar form

$$F_{net} = ma. \tag{7.10.4}$$

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

$$F_{net} = ma. \tag{7.10.5}$$

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 $1 m/s^2$ That is, since $F_{net} = ma$,

$$1N = 1kg \cdot s^2 \tag{7.10.6}$$

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:

$$w = mg \tag{7.10.7}$$

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:

$$w = mg = (1.0kg)(9.8m/s^2) = 9.8N. \tag{7.10.8}$$

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 7.10.1: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 7.10.3: 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

$$a = \frac{51\,N}{24\,kg} \tag{7.10.9}$$

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

$$a = \frac{51\,kg/s^2}{24\,kg} = 2.1\ m/s^2 \tag{7.10.10}$$

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 7.10.2: 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 $49m/s^2$ the mass of the system is 2100 kg, and the force of friction opposing the motion is known to be 650 N.

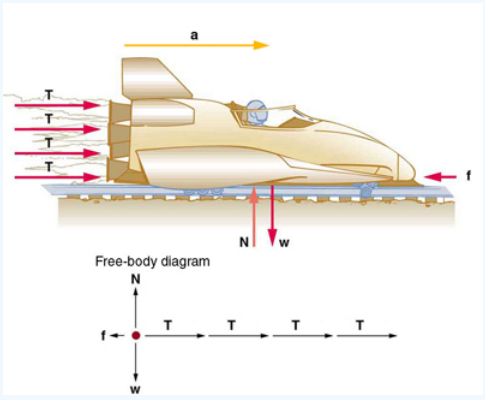


Figure 4.4.4. A sled experiences a rocket thrust that accelerates it to the right. Each rocket creates an identical thrust T. As in other situations where there is only horizontal acceleration, the vertical forces cancel. The ground exerts an upward force N on the system that is equal in magnitude and 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

$$F_{net} = ma. \tag{7.10.11}$$

, 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

$$F_{net} = 4T - f. \tag{7.10.12}$$

Substituting this into Newton's second law gives

$$F_{net} = ma = 4T - f. \tag{7.10.13}$$

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

$$4T = ma + f. \tag{7.10.14}$$

Substituting known values yields

$$4T = ma + f = (2100\,kg)(49\,m/s^2) + 650\,N \tag{7.10.15}$$

So the total thrust is

$$1 \times 10^5 N, \tag{7.10.16}$$

and the individual thrusts are

$$T = \frac{1 \times 10^5}{4} = 2.6 \times 10^4\,N \tag{7.10.17}$$

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:

$$w = mg. \tag{7.10.18}$$

- 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.

This page titled 7.10: Newton's Second Law of Motion- Concept of a System is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **4.3: Newton's Second Law of Motion- Concept of a System** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.11: 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 $\boldsymbol{F_{wall\ on\ feet}}$ is an external force on this system and affects its motion. The swimmer moves in the direction of $\boldsymbol{F_{wall\ on\ feet}}$. In contrast, the force $\boldsymbol{F_{feet\ on\ wall}}$ acts on the wall and not on our system of interest. Thus $\boldsymbol{F_{feet\ on\ wall}}$ does not directly affect the motion of the system and does not cancel $\boldsymbol{F_{wall\ on\ feet}}$. 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.

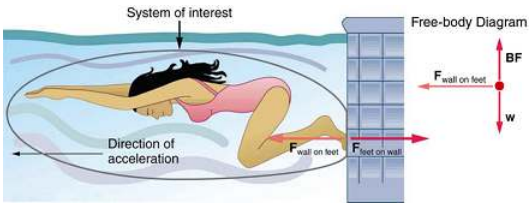


Figure **7.11.1**: When the swimmer exerts a force $\boldsymbol{F_{feet\ on\ wall}}$ on the wall, she accelerates in the direction opposite to that of her push. This means the net external force on her is in the direction opposite to $\boldsymbol{F_{feet\ on\ wall}}$ This opposition occurs because, in accordance with Newton's third law of motion, the wall exerts a force $\boldsymbol{F_{wall\ on\ feet}}$ 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 $\boldsymbol{F_{feet\ on\ wall}}$ does not act on this system (the swimmer) and, thus, does not cancel $\boldsymbol{F_{wall\ on\ feet}}$. Thus the free-body diagram shows only $\boldsymbol{F_{wall\ on\ feet}}$, $\boldsymbol{w}$, the gravitational force, and $\boldsymbol{BF}$, the buoyant force of the water supporting the swimmer's weight. The vertical forces $\boldsymbol{w}$ and $\boldsymbol{BF}$ 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 7.11.1: 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.

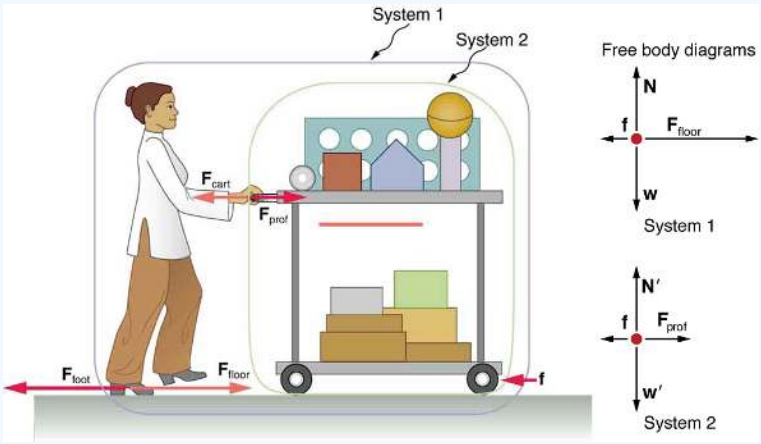


Figure 7.11.2: A professor pushes a cart of demonstration equipment. The lengths of the arrows are proportional to the magnitudes of the forces (except for f, since it is too small to draw to scale). Different questions are asked in each example; thus, the system of interest must be defined differently 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 7.11.2: 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

This page titled 7.11: Newton's Third Law of Motion- Symmetry in Forces is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **4.4: Newton's Third Law of Motion- Symmetry in Forces** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.12: 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:

$$p = mv \tag{7.12.1}$$

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 $\boldsymbol{p}$ is a vector having the same direction as the velocity $\boldsymbol{v}$. The SI unit for momentum is $\boldsymbol{kg \cdot m/s}$.

Example 7.12.1: 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, $\boldsymbol{p}$ (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 7.12.1, 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.

$$F_{net} = \frac{\Delta p}{\Delta t} \tag{7.12.2}$$

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

$$\Delta p = \Delta(mv) \tag{7.12.3}$$

If the mass of the system is constant, then

$$\Delta(mv) = m\Delta v. \tag{7.12.4}$$

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

$$F_{net} = \frac{\Delta p}{\Delta t} = \frac{m\Delta v}{\Delta t}. \tag{7.12.5}$$

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

$$F_{net} = ma \tag{7.12.6}$$

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 7.12.2: 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

This page titled 7.12: Linear Momentum and Force is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **8.1: Linear Momentum and Force** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

7.13: 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

$$\Delta p = F_{net}\Delta t, \tag{7.13.1}$$

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.

$$\Delta p = F_{net}\Delta t \tag{7.13.2}$$

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.

Example 7.13.1: Calculating Magnitudes of Impulses: Two Billiard Balls Striking a Rigid Wall

Two identical billiard balls strike a rigid wall with the same speed, and are reflected without any change of speed. The first ball strikes perpendicular to the wall. The second ball strikes the wall at an angle of 30^o from the perpendicular, and bounces off at an angle of 30^o from perpendicular to the wall.

a. Determine the direction of the force on the wall due to each ball.
b. Calculate the ratio of the magnitudes of impulses on the two balls by the wall.

Strategy for (a)

In order to determine the force on the wall, consider the force on the ball due to the wall using Newton's second law and then apply Newton's third law to determine the direction. Assume the $x-$axis to be normal to the wall and to be positive in the

initial direction of motion. Choose the y–axis to be along the wall in the plane of the second ball's motion. The momentum direction and the velocity direction are the same.

Strategy for (a)

In order to determine the force on the wall, consider the force on the ball due to the wall using Newton's second law and then apply Newton's third law to determine the direction. Assume the x–axis to be normal to the wall and to be positive in the initial direction of motion. Choose the y–axis to be along the wall in the plane of the second ball's motion. The momentum direction and the velocity direction are the same.

Solution for (a)

The first ball bounces directly into the wall and exerts a force on it in the +x direction. Therefore the wall exerts a force on the ball in the -y direction. The second ball continues with the same momentum component in the y– direction, but reverses its x-component of momentum, as seen by sketching a diagram of the angles involved and keeping in mind the proportionality between velocity and momentum.

These changes mean the change in momentum for both balls is in the $-x$ direction, so the force of the wall on each ball is along the $-x$ direction.

Strategy for (b)

Calculate the change in momentum for each ball, which is equal to the impulse imparted to the ball.

Solution for (b)

Let μ be the speed of each ball before and after collision with the wall, and m the mass of each ball. Choose the x–axis and y–axis as previously described, and consider the change in momentum of the first ball which strikes perpendicular to the wall.

$$p_{xi} = m\mu;\ p_{yi} = 0 \tag{7.13.3}$$

$$p_{xf} = -m\mu;\ p_{yf} = 0 \tag{7.13.4}$$

Impulse is the change in momentum vector. Therefore the x–component of impulse is equal to -$2m\mu$ and the y–component of impulse is equal to zero.

Now consider the change in momentum of the second ball.

$$p_{xi} = m/mu\,cos30^o;\ p_{yi} = -m\mu\,30^o \tag{7.13.5}$$

$$p_{xf} = -m/mu\,cos30^o;\ p_{yf} = -m\mu\,30^o \tag{7.13.6}$$

It should be noted here that while p_x changes sign after the collision, p_y does not. Therefore the -component of impulse is equal to $-2m\mu\,cos\ 30^o$ and the y–component of impulse is equal to zero.

The ratio of the magnitudes of the impulse imparted to the balls is

$$\frac{2m\mu}{2m\mu\,30^o} = \frac{2}{sqrt3} = 1.155. \tag{7.13.7}$$

Discussion

The direction of impulse and force is the same as in the case of (a); it is normal to the wall and along the negative x–direction. Making use of Newton's third law, the force on the wall due to each ball is normal to the wall along the positive x–direction.

Our definition of impulse includes an assumption that the force is constant over the time interval Δt.*Forces are usually not constant*. Forces vary considerably even during the brief time intervals considered. It is, however, possible to find an average effective force F_{eff} that produces the same result as the corresponding time-varying force. Figure shows a graph of what an actual force looks like as a function of time for a ball bouncing off the floor. The area under the curve has units of momentum and is equal to the impulse or change in momentum between times t_1 and t_2. That area is equal to the area inside the rectangle bounded by F_{eff}, t_1, and t_2 Thus the impulses and their effects are the same for both the actual and effective forces.

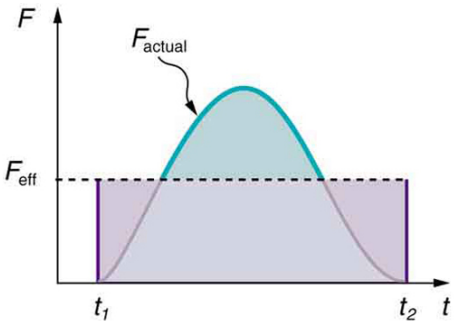


Figure 7.13.1: A graph of force versus time with time along the x—axis and force along the y—axis for an actual force and an equivalent effective force. The areas under the two curves are equal.

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?

MAKING CONNECTIONS: Constant Force anD Constant Acceleration

The assumption of a constant force in the definition of impulse is analogous to the assumption of a constant acceleration in kinematics. In both cases, nature is adequately described without the use of calculus.

Summary

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

$$\Delta p = F_{net}\Delta t. \tag{7.13.8}$$

- 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

---

This page titled 7.13: Impulse is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **8.2: Impulse** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

CHAPTER OVERVIEW

8: Chapter 8 - Energy

8.1: Prelude to Work, Energy, and Energy Resources
8.2: Work- The Scientific Definition
8.3: Kinetic Energy and the Work-Energy Theorem
8.4: Gravitational Potential Energy
8.5: Conservation of Energy
8.6: Nonconservative Forces
8.7: Power
8.8: Work, Energy, and Power in Humans

8: Chapter 8 - Energy is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.1: 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 **8.1.1**: 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.

This page titled 8.1: Prelude to Work, Energy, and Energy Resources is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **7.0: Prelude to Work, Energy, and Energy Resources** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

8.2: 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

$$W = \left|\vec{F}\right| \cos\theta \left|\vec{d}\right| \tag{8.2.2}$$

where W is work, d is the displacement of the system, and θ is the angle between the force vector $\vec{F}$ and the displacement vector $\vec{d}$, as in Figure 8.2.1. We can also write Equation 8.2.2 as

$$W = F\,d\cos\theta \tag{8.2.3}$$

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

$$W = F\,d\cos\theta \tag{8.2.4}$$

where W is work, F is the magnitude of the force on the system, d is the magnitude of the displacement of the system, and θ is the angle between the force vector F d the displacement vector d.

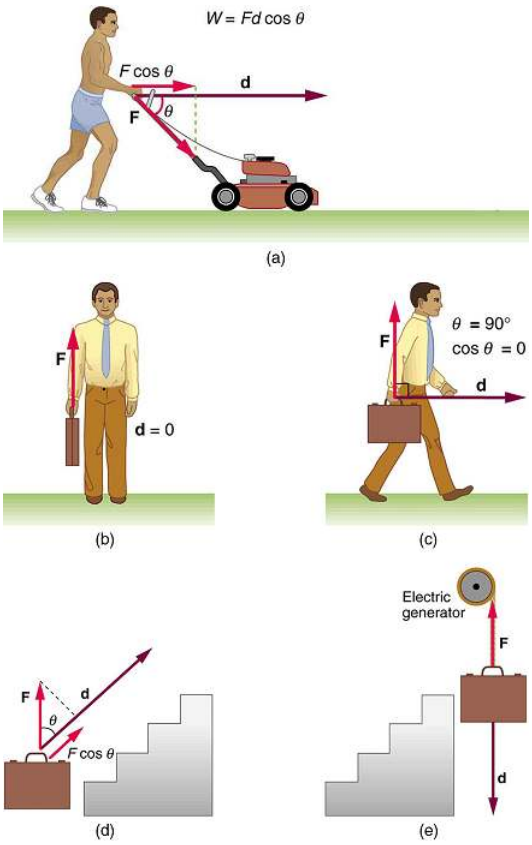


Figure 8.2.1: Examples of work. (a) The work done by the force F on this lawn mower is $Fd\cos\theta$. Note that $F\cos\theta$ is the component of the force in the direction of motion. (b) A person holding a briefcase does no work on it, because there is no motion. No energy is transferred to or from 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 8.2.1*b*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 8.2.1*c* does no work on it, because the force is perpendicular to the motion. That is, $\cos 90^o = 0$, so $W = 0$.

In contrast, when a force exerted on the system has a component in the direction of motion, such as in Figure 8.2.1*d*, work *is* done —energy is transferred to the briefcase. Finally, in Figure 8.2.1*e*, energy is transferred from the briefcase to a generator. There are two good ways to interpret this energy transfer. One interpretation is that the briefcase's weight does work on the generator, giving it energy. The other interpretation is that the generator does negative work on the briefcase, thus removing energy from it. The drawing shows the latter, with the force from the generator upward on the briefcase, and the displacement downward. This makes $\theta = 180^o$, and $\cos 180^o = -1$, therefore W is negative.

Calculating Work

Work and energy have the same units. From the definition of work, we see that those units are force times distance. Thus, in SI units, work and energy are measured in newton-meters. A newton-meter is given the special name joule (J), and $1\ J = 1\ N \cdot m = 1\ kg\ m^2/s^2$. One joule is not a large amount of energy; it would lift a small 100-gram apple a distance of about 1 meter.

Example 8.2.1: Calculating the Work You Do to Push a Lawn Mower Across a Large Lawn

How much work is done on the lawn mower by the person in Figure (a) if he exerts a constant force of 75.0 N at an angle 35^o below the horizontal and pushes the mower 25 m. on level ground? Convert the amount of work from joules to kilocalories and compare it with this person's average daily intake of **10,000 kJ** (about **2400 kcal**) of food energy. One *calorie* (1 cal) of heat is the amount required to warm 1 g of water by 1^oC and is equivalent to **4,184 J**, while one *food calorie* (**1 kcal**) is equivalent to **4,184 J**.

Strategy

We can solve this problem by substituting the given values into the definition of work done on a system, stated in the equation $W = Fd\cos\theta$. The force, angle, and displacement are given, so that only the work W is unknown.

Solution

The equation for the work is (Equation 8.2.3):

$$W = Fd\cos\theta$$

Substituting the known values gives

$$\begin{aligned} W &= (75\,N)(25.0\,m)(cos\,35^o) \\ &= 1536\,J \\ &= 1.54 \times 10^3\,J \end{aligned}$$

Converting the work in joules to kilocalories yields $W = (1536\ J)(1\ kcal/4184\ J) = 0.367 kcal$. The ratio of the work done to the daily consumption is

$$\frac{W}{2400\,kcal} = 1.53 \times 10^{-4}.$$

Discussion

This ratio is a tiny fraction of what the person consumes, but it is typical. Very little of the energy released in the consumption of food is used to do work. Even when we "work" all day long, less than 10% of our food energy intake is used to do work and more than 90% is converted to thermal energy or stored as chemical energy in fat.

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, times the cosine of the angle θ between them. In symbols,

$$W = Fd\cos\theta. \tag{8.2.5}$$

- 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

This page titled 8.2: Work- The Scientific Definition is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **7.1: Work- The Scientific Definition** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

8.3: 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 the lawn mower in [link](a) 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 [link](d) is stored in the briefcase-Earth system and can be recovered at any time, as shown in [link](e). 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.

Let us start by considering the total, or net, work done on a system. **Net work** is defined to be the sum of work done by all external forces—that is, net work is the work done by the net external force F_{net}. In equation form, this is $W_{net} = F_{net} d \cos\theta$, where θ is the angle between the force vector and the displacement vector. Figure (a) shows a graph of force versus displacement for the component of the force in the direction of the displacement—that is, an $F\cos\theta$ vs. d graph. In this case, $F\cos\theta$ is constant. You can see that the area under the graph is $F\cos\theta$, or the work done. Figure (b) shows a more general process where the force varies. The area under the curve is divided into strips, each having an average force $(F\cos\theta)_{i(ave)}$. The work done is e $(F\cos\theta)_{i(ave)} d_i$ for each strip, and the total work done is the sum of the W_i. Thus the total work done is the total area under the curve, a useful property to which we shall refer later.

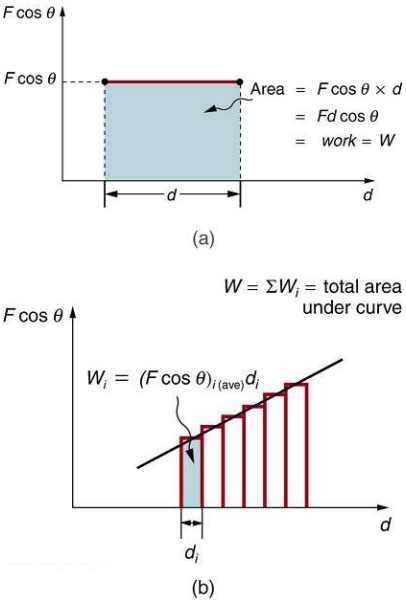


Figure 8.3.1: (a) A graph of $F\cos\theta$ vs. d when $F\cos\theta$ is constant. The area under the curve represents the work done by the force. (b) A graph of $F\cos\theta$ vs. d in which the force varies. The work done for each interval is the area of each strip; thus, the total area under the curve equals the total work done.

Net work will be simpler to examine if we consider a one-dimensional 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 in Figure.

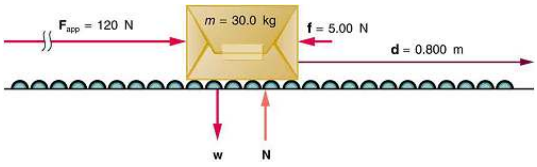


Figure 8.3.2: A package on a roller belt is pushed horizontally through a distance d.

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, so that $\theta = 0$ and $\cos\theta = 1$, and the net work is given by

$$W_{net} = F_{net}d. \tag{8.3.1}$$

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

$$W_{net} = mad. \tag{8.3.2}$$

To get a relationship between net work and the speed given to a system by the net force acting on it, we take $d = x - x_0$ and use the equation studied in Motion Equations for Constant Acceleration in One Dimension for the change in speed over a distance d if the acceleration has the constant value a, namely $v^2 = v_0^2 + 2ad$. (note that a appears in the expression for the net work). 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

$$W_{net} = m\left(\frac{v^2 - v_0^2}{2d}\right)d. \tag{8.3.3}$$

The d cancels, and we rearrange this to obtain

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{8.3.4}$$

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 $\frac{1}{2}mv^2$.

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{8.3.5}$$

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,

$$KE = \frac{1}{2}mv^2, \tag{8.3.6}$$

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 8.3.1: 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

$$KE = \frac{1}{2}mv^2. \tag{8.3.7}$$

Entering known values gives

$$KE = 0.5(30.0\,kg)(0.500\,m/s)^2, \tag{8.3.8}$$

which yields

$$KE = 3.75\,kg \cdot m^2/s^2 = 3.75\,J \tag{8.3.9}$$

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 8.3.2: Determining the Work to Accelerate a Package

Suppose that you push on the 30.0-kg package in Figure 7.03.2. with a constant force of 120 N through a distance of 0.800 m, and that the opposing friction force averages 5.00 N.

(a) Calculate the net work done on the package. (b) Solve the same problem as in part (a), this time by finding the work done by each force that contributes to the net force.

Strategy and Concept for (a)

This is a motion in one dimension problem, because the downward force (from the weight of the package) and the normal force have equal magnitude and opposite direction, so that they cancel in calculating the net force, while the applied force, friction, and the displacement are all horizontal. (See Figure 7.03.2.) As expected, the net work is the net force times distance.

Solution for (a)

The net force is the push force minus friction, or $F_{net} = 120\,N - 5.00\,N = 115\,N$. Thus the net work is

$$W_{net} = F_{net}d = (115\,N)(0.800\,m) \tag{8.3.10}$$

$$= 92.0\,N \cdot m = 92.0\,J \tag{8.3.11}$$

Discussion for (a)

This value is the net work done on the package. The person actually does more work than this, because friction opposes the motion. Friction does negative work and removes some of the energy the person expends and converts it to thermal energy. The net work equals the sum of the work done by each individual force.

Strategy and Concept for (b)

The forces acting on the package are gravity, the normal force, the force of friction, and the applied force. The normal force and force of gravity are each perpendicular to the displacement, and therefore do no work.

Solution for (b)

The applied force does work.

$$W_{app} = F_{app}d\,cos\,(0^o) = F_{app}d \tag{8.3.12}$$

$$= (120\,N)(0.800\,m) \tag{8.3.13}$$

$$= 96.0\,J \tag{8.3.14}$$

The friction force and displacement are in opposite directions, so that $\theta = 180^o$, and the work done by friction is

$$W_{fr} = F_{fr}d\,cos\,(180^o) \tag{8.3.15}$$

$$= -(5.00\,N)(0.800\,m) \tag{8.3.16}$$

$$= -4.00\,J \tag{8.3.17}$$

So the amounts of work done by gravity, by the normal force, by the applied force, and by friction are, respectively,

$$W_{gr} = 0, \tag{8.3.18}$$

$$W_N = 0, \tag{8.3.19}$$

$$W_{app} = 96.0\,J, \tag{8.3.20}$$

$$W_{fr} = -4\,J. \tag{8.3.21}$$

The total work done as the sum of the work done by each force is then seen to be

$$W_{total} = W_{gr} + W_N + W_{app} + W_{fr} = 92.0\,J. \tag{8.3.22}$$

Discussion for (b)

The calculated total work W_{total} as the sum of the work by each force agrees, as expected, with the work W_{net} done by the net force. The work done by a collection of forces acting on an object can be calculated by either approach.

Example 8.3.3: 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

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{8.3.23}$$

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

$$\frac{1}{2}mv^2 = W_{net} + \frac{1}{2}mv_0^2 \tag{8.3.24}$$

Thus,

$$\frac{1}{2}mv^2 = 92.0\,J + 3.75\,J = 95.75\,J. \tag{8.3.25}$$

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

$$v = \sqrt{\frac{2(95.75\,J)}{m}} = \sqrt{\frac{191.5\,kg \cdot m^2/s^2}{30.0\,kg}} \tag{8.3.26}$$

$$= 2.53\,m/s \tag{8.3.27}$$

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.

Example 8.3.4: Work and Energy Can Reveal Distance, Too

How far does the package in Figure 7.03.2. coast after the push, assuming friction remains constant? Use work and energy considerations.

Strategy

We know that once the person stops pushing, friction will bring the package to rest. In terms of energy, friction does negative work until it has removed all of the package's kinetic energy. The work done by friction is the force of friction times the distance traveled times the cosine of the angle between the friction force and displacement; hence, this gives us a way of finding the distance traveled after the person stops pushing.

Solution

The normal force and force of gravity cancel in calculating the net force. The horizontal friction force is then the net force, and it acts opposite to the displacement, so $\theta = 180^o$. To reduce the kinetic energy of the package to zero, the work W_{fr} by friction must be minus the kinetic energy that the package started with plus what the package accumulated due to the pushing. Thus $W_{fr} = -95.75\,J$. Furthermore, $W_{fr} = df'\cos\theta = -Fd'$, where d' is the distance it takes to stop. Thus,

$$d' = -\frac{W_{fr}}{f} = \frac{-95.75\,J}{5.00\,N}, \tag{8.3.28}$$

and so

$$d' = 19.2\,m \tag{8.3.29}$$

Discussion

> This is a reasonable distance for a package to coast on a relatively friction-free conveyor system. Note that the work done by friction is negative (the force is in the opposite direction of motion), so it removes the kinetic energy.

Some of the examples in this section can be solved without considering energy, but at the expense of missing out on gaining insights about what work and energy are doing in this situation. On the whole, solutions involving energy are generally shorter and easier than those using kinematics and dynamics alone.

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

---

This page titled 8.3: Kinetic Energy and the Work-Energy Theorem is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **7.2: Kinetic Energy and the Work-Energy Theorem** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

8.4: 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 8.4.2.) This shortcut makes it is easier to solve problems using energy (if possible) rather than explicitly using forces.

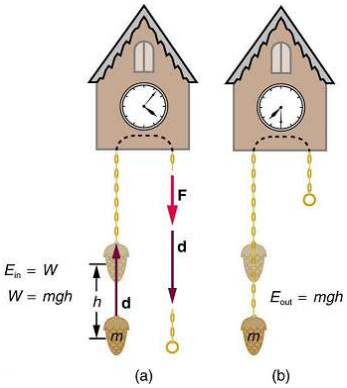


Figure 8.4.1: (a) The work done to lift the weight is stored in the mass-Earth system as gravitational potential energy. (b) As the weight moves downward, this gravitational potential energy is transferred to the cuckoo clock.

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

$$\Delta PE_g = mgh, \tag{8.4.1}$$

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

$$mgh = (0.500\,kg)(9.80\,m/s^2)(1.00\,m) \tag{8.4.2}$$

$$= 4.90\,kg \cdot m^2/s^2 = 4.90\,J. \tag{8.4.3}$$

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.

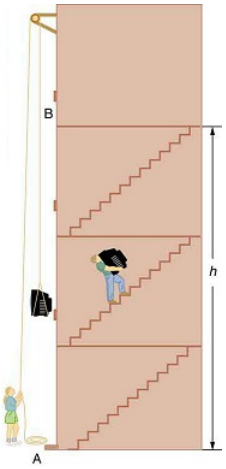


Figure 8.4.2: The change in gravitational potential energy (ΔPE_g) between points A and B is independent of the path $\Delta PE_g = mgh$ for any path between the two points. Gravity is one of a small class of forces where the work done by or against the force depends only on the starting and ending points, not on the path between them.

Example 8.4.1: The Force to Stop Falling

A 60.0-kg person jumps onto the floor from a height of 3.00 m. If he lands stiffly (with his knee joints compressing by 0.500 cm), calculate the force on the knee joints.

Strategy

This person's energy is brought to zero in this situation by the work done on him by the floor as he stops. The initial PE_g is transformed into KE as he falls. The work done by the floor reduces this kinetic energy to zero.

Solution

The work done on the person by the floor as he stops is given by

$$W = Fd\cos\theta = -Fd \tag{8.4.4}$$

with a minus sign because the displacement while stopping and the force from floor are in opposite directions ($\cos\theta = \cos 180^\circ = -1$.) The floor removes energy from the system, so it does negative work.

The kinetic energy the person has upon reaching the floor is the amount of potential energy lost by falling through height h:

$$KE = -\Delta PE_g = -mgh \tag{8.4.5}$$

The distance d that the person's knees bend is much smaller than the height h of the fall, so the additional change in gravitational potential energy during the knee bend is ignored. The work W done by the floor on the person stops the person and brings the person's kinetic energy to zero:

$$W = -KE = mgh \tag{8.4.6}$$

Combining this equation with the expression for W gives

$$-Fd = mgh. \tag{8.4.7}$$

Recalling that h is negative because the person fell *down*, the force on the knee joints is given by

$$F = -\frac{mgh}{d} = -\frac{(60.0\,kg)(9.80\,m/s^2)(-3.00\,m)}{5.00 \times 10^{-3}\,m} = 3.53 \times 10^5\,N. \tag{8.4.8}$$

Discussion

Such a large force (500 times more than the person's weight) over the short impact time is enough to break bones. A much better way to cushion the shock is by bending the legs or rolling on the ground, increasing the time over which the force acts. A bending motion of 0.5 m this way yields a force 100 times smaller than in the example. A kangaroo's hopping shows this method in action. The kangaroo is the only large animal to use hopping for locomotion, but the shock in hopping is cushioned by the bending of its hind legs in each jump.(See Figure.)



Figure 8.4.3: 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 8.4.2: 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?

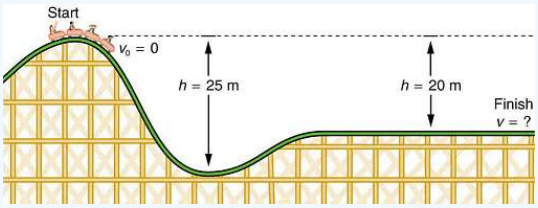


Figure 8.4.4: The speed of a roller coaster increases as gravity pulls it downhill and is greatest at its lowest point. Viewed in terms of energy, the roller-coaster-Earth system's gravitational potential energy is converted to kinetic energy. If work done by friction is negligible, all ΔPE_g is converted 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,

$$-\Delta PE_g = \Delta KE \tag{8.4.9}$$

becomes

$$mg\,|h| = \frac{1}{2}mv^2 \tag{8.4.10}$$

Solving for v we find that mass cancels and that

$$v = \sqrt{2g|h|}. \tag{8.4.11}$$

Substituting known values,

$$v = \sqrt{2(9.80\,m/s^2)(20.0\,m)} \tag{8.4.12}$$

$$= 19.8\,m/s \tag{8.4.13}$$

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,

$$mgh = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{8.4.14}$$

Rearranging gives

$$\frac{1}{2}mv^2 = mg\,|h| + \frac{1}{2}mv_0^2. \tag{8.4.15}$$

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

$$v = \sqrt{2g\,|h| + v_0^2}. \tag{8.4.16}$$

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

$$v = \sqrt{2(9.80\,m/s^2)(20.0\,m) + (5.00)^2} \tag{8.4.17}$$

$$= 20.4\,m/s. \tag{8.4.18}$$

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.

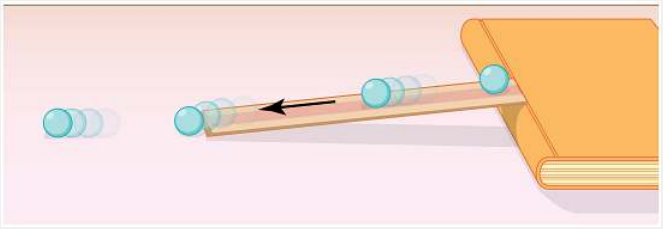


Figure 8.4.5: A marble rolls down a ruler, and its speed on the level surface is measured.

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

This page titled 8.4: Gravitational Potential Energy is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **7.3: Gravitational Potential Energy** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

8.5: 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

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f. \tag{8.5.2}$$

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

$$KE_i + PE_i = KE_f + PE_f. \tag{8.5.3}$$

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.

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f. \tag{8.5.4}$$

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 8.5.1: 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 g at 0.5 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

$$Efficiency\,(E_{ff}) = \frac{useful\ energy\ or\ work\ output}{total\ energy\ input} = \frac{W_{out}}{E_{in}}. \tag{8.5.5}$$

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.

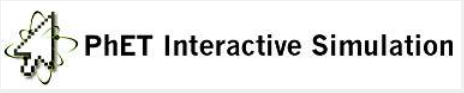


Figure 8.5.2: Masses and Springs

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

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f, \tag{8.5.6}$$

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

---

This page titled 8.5: Conservation of Energy is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

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

8.6: Nonconservative Forces

Learning Objectives

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

- Define nonconservative forces and explain how they affect mechanical energy.
- Show how the principle of conservation of energy can be applied by treating the conservative forces in terms of their potential energies and any nonconservative forces in terms of the work they do.

Nonconservative Forces and Friction

Forces are either conservative or nonconservative. A **nonconservative** force is one for which work depends on the path taken. Friction is a good example of a nonconservative force. As illustrated in Figure 8.6.1, work done against friction depends on the length of the path between the starting and ending points. Because of this dependence on path, there is no potential energy associated with nonconservative forces. An important characteristic is that the work done by a nonconservative force *adds or removes mechanical energy from a system*. **Friction**, for example, creates **thermal energy** that dissipates, removing energy from the system. Furthermore, even if the thermal energy is retained or captured, it cannot be fully converted back to work, so it is lost or not recoverable in that sense as well.

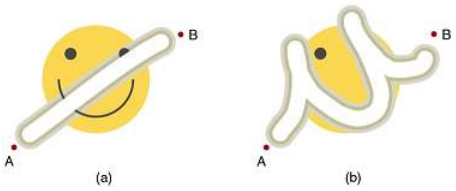


Figure 8.6.1: The amount of the happy face erased depends on the path taken by the eraser between points A and B, as does the work done against friction. Less work is done and less of the face is erased for the path in (a) than for the path in (b). The force here is friction, and most of the work goes into thermal energy that subsequently leaves the system (the happy face plus the eraser). The energy expended cannot be fully recovered.

How Nonconservative Forces Affect Mechanical Energy

Mechanical energy *may* not be conserved when nonconservative forces act. For example, when a car is brought to a stop by friction on level ground, it loses kinetic energy, which is dissipated as thermal energy, reducing its mechanical energy. Figure compares the effects of conservative and nonconservative forces. We often choose to understand simpler systems such as that described in Figure 8.6.2*a* first before studying more complicated systems as in Figure 8.6.2*b*.

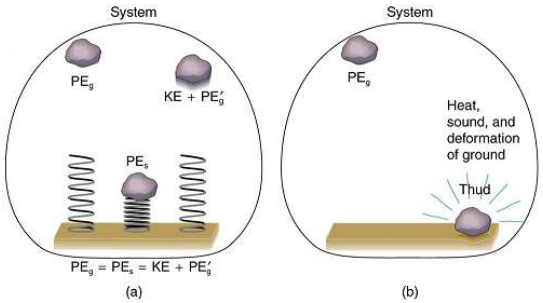


Figure 8.6.2: Comparison of the effects of conservative and nonconservative forces on the mechanical energy of a system. (a) A system with only conservative forces. When a rock is dropped onto a spring, its mechanical energy remains constant (neglecting air resistance) because the force in the spring is conservative. The spring can propel the rock back to its original height, where it once again has only potential energy due to gravity. (b) A system with nonconservative forces. When the same rock is dropped onto the ground, it is stopped by nonconservative forces that dissipate its mechanical energy as thermal energy, sound, and surface distortion. The rock has lost mechanical energy.

How the Work-Energy Theorem Applies

Now let us consider what form the work-energy theorem takes when both conservative and nonconservative forces act. We will see that the work done by nonconservative forces equals the change in the mechanical energy of a system. As noted in Kinetic Energy

and the Work-Energy Theorem, the work-energy theorem states that the net work on a system equals the change in its kinetic energy, or $W_{net} = \Delta KE$. The net work is the sum of the work by nonconservative forces plus the work by conservative forces. That is,

$$W_{net} = W_{nc} + W_c, \tag{8.6.1}$$

so that

$$W_{nc} + W_c = \Delta KE, \tag{8.6.2}$$

where W_{nc} is the total work done by all nonconservative forces and W_c is the total work done by all conservative forces.

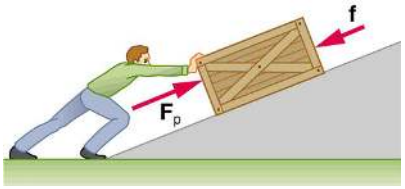


Figure 8.6.3: A person pushes a crate up a ramp, doing work on the crate. Friction and gravitational force (not shown) also do work on the crate; both forces oppose the person's push. As the crate is pushed up the ramp, it gains mechanical energy, implying that the work done by the person is greater than the work done by friction.

Consider Figure 8.6.3, in which a person pushes a crate up a ramp and is opposed by friction. As in the previous section, we note that work done by a conservative force comes from a loss of gravitational potential energy, so that $W_c = -\Delta PE$. Substituting this equation into the previous one and solving for W_{nc} gives

$$W_{nc} = \Delta KE + \Delta PE. \tag{8.6.3}$$

This equation means that the total mechanical energy ($KE + PE$) changes by exactly the amount of work done by nonconservative forces. In Figure 8.6.3, this is the work done by the person minus the work done by friction. So even if energy is not conserved for the system of interest (such as the crate), we know that an equal amount of work was done to cause the change in total mechanical energy. We rearrange Equation 8.6.4 to obtain

$$KE_i + PE_i + W_{nc} = KE_f + PE_f \tag{8.6.4}$$

This means that the amount of work done by nonconservative forces adds to the mechanical energy of a system. If W_{nc} is positive, then mechanical energy is increased, such as when the person pushes the crate up the ramp in Figure. If W_{nc} is negative, then mechanical energy is decreased, such as when the rock hits the ground in Figure 8.6.2*b*. If W_{nc} is negative, then mechanical energy is decreased, such as when the rock hits the ground in Figure 8.6.2*b*. If W_{nc} is zero, then mechanical energy is conserved, and nonconservative forces are balanced. For example, when you push a lawn mower at constant speed on level ground, your work done is removed by the work of friction, and the mower has a constant energy.

Applying Energy Conservation with Nonconservative Forces

When no change in potential energy occurs, applying Equation 8.6.4 amounts to applying the work-energy theorem by setting the change in kinetic energy to be equal to the net work done on the system, which in the most general case includes both conservative and nonconservative forces. But when seeking instead to find a change in total mechanical energy in situations that involve changes in both potential and kinetic energy, the previous equation $KE_i + PE_i + W_{nc} = KE_f + PE_f$ says that you can start by finding the change in mechanical energy that would have resulted from just the conservative forces, including the potential energy changes, and add to it the work done, with the proper sign, by any nonconservative forces involved.

Example 8.6.1: Calculating Distance Traveled: How Far a Baseball Player Slides

Consider the situation shown in Figure 8.6.4, where a baseball player slides to a stop on level ground. Using energy considerations, calculate the distance the 65.0-kg baseball player slides, given that his initial speed is 6.00 m/s and the force of friction against him is a constant 450 N.

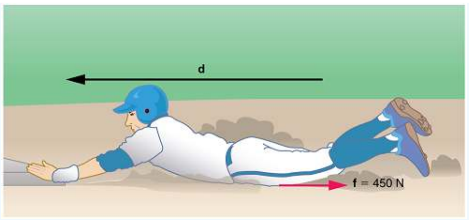


Figure 8.6.4: The baseball player slides to a stop in a distance d. In the process, friction removes the player's kinetic energy by doing an amount of work fd equal to the initial kinetic energy.

Strategy

Friction stops the player by converting his kinetic energy into other forms, including thermal energy. In terms of the work-energy theorem, the work done by friction, which is negative, is added to the initial kinetic energy to reduce it to zero. The work done by friction is negative, because f is in the opposite direction of the motion (that is, $\theta = 180^o$, and so $\cos\theta = -1$). Thus, $W_{nc} = -fd$. The equation simplifies to

$$\frac{1}{2}mv_i^2 - bfd = 0$$

or

$$fd = \frac{1}{2}mv_i^2.$$

This equation can now be solved for the distance d.

Solution

Solving the previous equation for d and substituting known values yields

$$\begin{aligned} d &= \frac{mv_i^2}{2f} \\ &= \frac{65.0\,kg)(6.00\,m/s)^2}{(2)(450\,N)} \\ &= 2.60\,m. \end{aligned}$$

Discussion

The most important point of this example is that the amount of nonconservative work equals the change in mechanical energy. For example, you must work harder to stop a truck, with its large mechanical energy, than to stop a mosquito.

Example 8.6.2: Calculating Distance Traveled: Sliding Up an Incline

Suppose that the player from Example 8.6.1 is running up a hill having a 5.00^o incline upward with a surface similar to that in the baseball stadium. The player slides with the same initial speed, and the frictional force is still 450 N. Determine how far he slides.

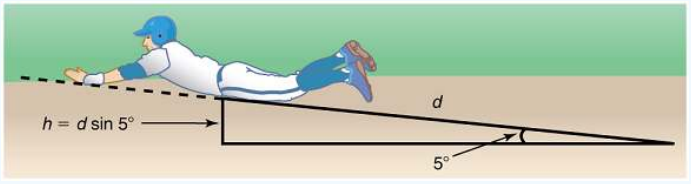


Figure 8.6.5: The same baseball player slides to a stop on a 5.00^o slope.

Strategy

In this case, the work done by the nonconservative friction force on the player reduces the mechanical energy he has from his kinetic energy at zero height, to the final mechanical energy he has by moving through distance d to reach height h along the hill, with $h = d \sin 5.00^o$. This is expressed by Equation 8.6.4

$$KE_i + PE_i + W_{nc} = KE_f + PE_f.$$

Solution

The work done by friction is again $W_{nc} = -fd$; initially the potential energy is $PE_i = mg \cdot 0 = 0$ and the kinetic energy is $KE_i = \frac{1}{2}mv^2$; the final energy contributions are $KE_f = 0$ for the kinetic energy and $PE_f = mgh = mgd\,sin\,\theta$ for the potential energy.

Substituting these values gives

$$\frac{1}{2}mv_i^2 + 0 + (-fd) = 0 + mgd\,\sin\,\theta.$$

Solve this for d to obtain

$$\begin{aligned} d &= \frac{\frac{1}{2}mv_i^2}{f + mg\,\sin\,\theta} \\ &= \frac{(0.5)(65.0\,kg)(6.00\,m/s)^2}{450\,N + (65.0\,kg)(9.80\,m/s^2)sin\,(5.00^o)} \\ &= 2.31\,m. \end{aligned}$$

Discussion

As might have been expected, the player slides a shorter distance by sliding uphill. Note that the problem could also have been solved in terms of the forces directly and the work energy theorem, instead of using the potential energy. This method would have required combining the normal force and force of gravity vectors, which no longer cancel each other because they point in different directions, and friction, to find the net force. You could then use the net force and the net work to find the distance d that reduces the kinetic energy to zero. By applying conservation of energy and using the potential energy instead, we need only consider the gravitational potential energy mgh, without combining and resolving force vectors. This simplifies the solution considerably.

Making Connections: Take-Home Investigation - Determining Friction from the Stopping Distance

This experiment involves the conversion of gravitational potential energy into thermal energy. Use the ruler, book, and marble from Take-Home Investigation—Converting Potential to Kinetic Energy. In addition, you will need a foam cup with a small hole in the side, as shown in Figure 8.6.6. From the 10-cm position on the ruler, let the marble roll into the cup positioned at the bottom of the ruler. Measure the distance d the cup moves before stopping. What forces caused it to stop? What happened to the kinetic energy of the marble at the bottom of the ruler? Next, place the marble at the 20-cm and the 30-cm positions and again measure the distance the cup moves after the marble enters it. Plot the distance the cup moves versus the initial marble position on the ruler. Is this relationship linear? With some simple assumptions, you can use these data to find the coefficient of kinetic friction μ_k of the cup on the table. The force of friction f on the cup is $\mu_k N$, where the normal force N is just the weight of the cup plus the marble. The normal force and force of gravity do no work because they are perpendicular to the displacement of the cup, which moves horizontally. The work done by friction is fd. You will need the mass of the marble as well to calculate its initial kinetic energy.

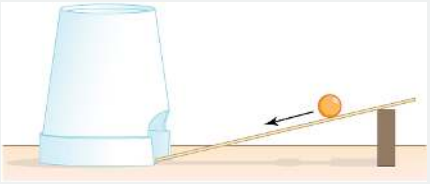


Figure 8.6.6: Rolling a marble down a ruler into a foam cup.

It is interesting to do the above experiment also with a steel marble (or ball bearing). Releasing it from the same positions on the ruler as you did with the glass marble, is the velocity of this steel marble the same as the velocity of the marble at the bottom of the ruler? Is the distance the cup moves proportional to the mass of the steel and glass marbles?

Phet Explorations: The Ramp

Explore forces, energy and work as you push household objects up and down a ramp. Lower and raise the ramp to see how the angle of inclination affects the parallel forces acting on the file cabinet. Graphs show forces, energy and work.



PhET Interactive Simulation

Figure 8.6.7: The Ramp

Summary

- A nonconservative force is one for which work depends on the path.
- Friction is an example of a nonconservative force that changes mechanical energy into thermal energy.
- Work W_{nc} done by a nonconservative force changes the mechanical energy of a system. In equation form, $W_{nc} = \Delta KE + \Delta PE$ or, equivalently, $KE_i + PE_i + W_{nc} = KE_f + PE_f$.
- When both conservative and nonconservative forces act, energy conservation can be applied and used to calculate motion in terms of the known potential energies of the conservative forces and the work done by nonconservative forces, instead of finding the net work from the net force, or having to directly apply Newton's laws

Glossary

nonconservative force

a force whose work depends on the path followed between the given initial and final configurations

friction

the force between surfaces that opposes one sliding on the other; friction changes mechanical energy into thermal energy

This page titled 8.6: Nonconservative Forces is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **7.5: Nonconservative Forces** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

8.7: Power

Learning Objectives

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

- Calculate power by calculating changes in energy over time.
- Examine power consumption and calculations of the cost of energy consumed.

What is Power?

Power—the word conjures up many images: a professional football player muscling aside his opponent, a dragster roaring away from the starting line, a volcano blowing its lava into the atmosphere, or a rocket blasting off, as in Figure.



Figure 8.7.1: This powerful rocket on the Space Shuttle Endeavor did work and consumed energy at a very high rate. (credit: NASA)

These images of power have in common the rapid performance of work, consistent with the scientific definition of power P as the rate at which work is done.

Power

Power is the rate at which work is done.

$$P = \frac{W}{t} \tag{8.7.1}$$

The SI unit for power is the watt W, where 1 watt equals 1 joule/second $(1\,W = 1\,J/s)$.

Because work is energy transfer, power is also the rate at which energy is expended. A 60-W light bulb, for example, expends 60 J of energy per second. Great power means a large amount of work or energy developed in a short time. For example, when a powerful car accelerates rapidly, it does a large amount of work and consumes a large amount of fuel in a short time.

Calculating Power from Energy

Example 8.7.1: Calculating the Power to Climb Stairs

What is the power output for a 60.0-kg woman who runs up a 3.00 m high flight of stairs in 3.50 s, starting from rest but having a final speed of 2.00 m/s? (See Figure.)

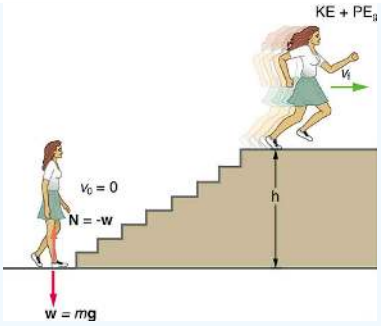


Figure 8.7.2:When this woman runs upstairs starting from rest, she converts the chemical energy originally from food into kinetic energy and gravitational potential energy. Her power output depends on how fast she does this.

Strategy and Concept

The work going into mechanical energy is $W = KE + PE$. At the bottom of the stairs, we take both KE and PE as initially zero; thus $W = KE_f + PE_g = \frac{1}{2}mv_f^2 + mgh$, where h is the vertical height of the stairs. Because all terms are given, we can calculate W and then divide it by time to get power.

Solution

Substituting the expression for W into the definition of power given in the previous equation, $P = W/t$ yields

$$W = \frac{W}{t} = \frac{\frac{1}{2}mv_f^2 + mgh}{t}. \tag{8.7.2}$$

Entering known values yields

$$P = \frac{0.5(60\,kg)(2.00m/s^2) + (60.0\,kg)(9.80m/s^2)(3.00\,m)}{3.50\,s} \tag{8.7.3}$$

$$= \frac{120\,J + 1764\,J}{3.50\,s} \tag{8.7.4}$$

$$= 538\,W. \tag{8.7.5}$$

Discussion

The woman does 1764 J of work to move up the stairs compared with only 120 J to increase her kinetic energy; thus, most of her power output is required for climbing rather than accelerating.

It is impressive that this woman's useful power output is slightly less than 1 horsepower ($1\,hp = 746\,W$). People can generate more than a horsepower with their leg muscles for short periods of time by rapidly converting available blood sugar and oxygen into work output. (A horse can put out 1 hp for hours on end.) Once oxygen is depleted, power output decreases and the person begins to breathe rapidly to obtain oxygen to metabolize more food—this is known as the *aerobic* stage of exercise. If the woman climbed the stairs slowly, then her power output would be much less, although the amount of work done would be the same.

Making Connections: Take-Home Investigation

—Measure Your Power Rating

- Determine your own power rating by measuring the time it takes you to climb a flight of stairs. We will ignore the gain in kinetic energy, as the above example showed that it was a small portion of the energy gain. Don't expect that your output will be more than about 0.5 hp.

Examples of Power

Examples of power are limited only by the imagination, because there are as many types as there are forms of work and energy. (See Table for some examples.) Sunlight reaching Earth's surface carries a maximum power of about 1.3 kilowatts per square meter kW/m^2 A tiny fraction of this is retained by Earth over the long term. Our consumption rate of fossil fuels is far greater than the

rate at which they are stored, so it is inevitable that they will be depleted. Power implies that energy is transferred, perhaps changing form. It is never possible to change one form completely into another without losing some of it as thermal energy. For example, a 60-W incandescent bulb converts only 5 W of electrical power to light, with 55 W dissipating into thermal energy. Furthermore, the typical electric power plant converts only 35 to 40% of its fuel into electricity. The remainder becomes a huge amount of thermal energy that must be dispersed as heat transfer, as rapidly as it is created. A coal-fired power plant may produce 1000 megawatts; 1 megawatt (MW) is 10^6 of electric power. But the power plant consumes chemical energy at a rate of about 2500 MW, creating heat transfer to the surroundings at a rate of 1500 MW. (See Figure.)



Figure 8.7.3: Tremendous amounts of electric power are generated by coal-fired power plants such as this one in China, but an even larger amount of power goes into heat transfer to the surroundings. The large cooling towers here are needed to transfer heat as rapidly as it is produced. The transfer of heat is not unique to coal plants but is an unavoidable consequence of generating electric power from any fuel—nuclear, coal, oil, natural gas, or the like. (credit: Kleinolive, Wikimedia Commons)

Object or Phenomenon	Power in Watts
Supernova (at peak)	5×10^{37}
Milky Way galaxy	10^{37}
Crab Nebula pulsar	10^{28}
The Sun	4×10^{26}
Volcanic eruption (maximum)	4×10^{15}
Lightning bolt	2×10^{12}
Nuclear power plant (total electric and heat transfer)	3×10^{9}
Aircraft carrier (total useful and heat transfer)	10^{8}
Dragster (total useful and heat transfer)	2×10^{6}
Car (total useful and heat transfer)	8×10^{4}
Football player (total useful and heat transfer)	5×10^{3}
Clothes dryer	4×10^{3}
Person at rest (all heat transfer)	100
Typical incandescent light bulb (total useful and heat transfer)	60
Heart, person at rest (total useful and heat transfer)	8
Electric clock	3
Pocket calculator	10^{-3}

Power and Energy Consumption

We usually have to pay for the energy we use. It is interesting and easy to estimate the cost of energy for an electrical appliance if its power consumption rate and time used are known. The higher the power consumption rate and the longer the appliance is used,

the greater the cost of that appliance. The power consumption rate is $P = \frac{W}{t} = \frac{E}{t}$, where E is the energy supplied by the electricity company. So the energy consumed over a time t is

$$E = Pt. \tag{8.7.6}$$

Electricity bills state the energy used in units of kilowatt-hours ($kW \cdot h$), which is the product of power in kilowatts and time in hours. This unit is convenient because electrical power consumption at the kilowatt level for hours at a time is typical.

Example 8.7.2: Calculating Energy Costs

What is the cost of running a 0.200-kW computer 6.00 h per day for 30.0 d if the cost of electricity is \$0.120 per $kW \cdot h$?

Strategy

Cost is based on energy consumed; thus, we must find E from $E = Pt$ and then calculate the cost. Because electrical energy is expressed in $kW \cdot h$ at the start of a problem such as this it is convenient to convert the units into kW and hours.

Solution

The energy consumed in $kW \cdot h$ is

$$E = Pt = (0.200\,kW)(6.00\,h/d)(30.0\,d) \tag{8.7.7}$$

$$= 36.0\,kW \cdot h, \tag{8.7.8}$$

and the cost is simply given by

$$cost = (36.0\,kW \cdot h)(\$0.120\,per\,kW \cdot h) = \$4.32\,per\,month. \tag{8.7.9}$$

Discussion

The cost of using the computer in this example is neither exorbitant nor negligible. It is clear that the cost is a combination of power and time. When both are high, such as for an air conditioner in the summer, the cost is high.

The motivation to save energy has become more compelling with its ever-increasing price. Armed with the knowledge that energy consumed is the product of power and time, you can estimate costs for yourself and make the necessary value judgments about where to save energy. Either power or time must be reduced. It is most cost-effective to limit the use of high-power devices that normally operate for long periods of time, such as water heaters and air conditioners. This would not include relatively high power devices like toasters, because they are on only a few minutes per day. It would also not include electric clocks, in spite of their 24-hour-per-day usage, because they are very low power devices. It is sometimes possible to use devices that have greater efficiencies—that is, devices that consume less power to accomplish the same task. One example is the compact fluorescent light bulb, which produces over four times more light per watt of power consumed than its incandescent cousin.

Modern civilization depends on energy, but current levels of energy consumption and production are not sustainable. The likelihood of a link between global warming and fossil fuel use (with its concomitant production of carbon dioxide), has made reduction in energy use as well as a shift to non-fossil fuels of the utmost importance. Even though energy in an isolated system is a conserved quantity, the final result of most energy transformations is waste heat transfer to the environment, which is no longer useful for doing work. As we will discuss in more detail in Thermodynamics, the potential for energy to produce useful work has been "degraded" in the energy transformation.

Summary

- Power is the rate at which work is done, or in equation form, for the average power P for work W done over a time t, $P = W/t$.
- The SI unit for power is the watt (W), where $1\,W = 1\,J/s$.
- The power of many devices such as electric motors is also often expressed in horsepower (hp), where $1\ hp = 746\,W$.

Glossary

power
the rate at which work is done

watt

(W) SI unit of power, with $1W = 1J/s$

horsepower

an older non-SI unit of power, with $1hp = 746W$

kilowatt-hour

($kW \cdot h$) unit used primarily for electrical energy provided by electric utility companies

---

This page titled 8.7: Power is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **7.7: Power** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

8.8: 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.

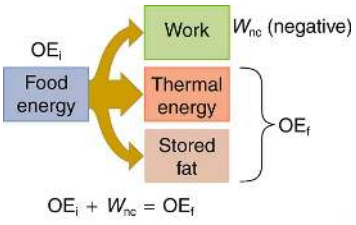


Figure 8.8.1: Energy consumed by humans is converted to work, thermal energy, and stored fat. By far the largest fraction goes to thermal energy, although the fraction varies depending on the type of physical activity.

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 8.8.1: 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

$$Time = \frac{energy}{\left(\frac{energy}{time}\right)} = \frac{1000\,kJ}{400\,W} = 2500\,s = 42\,min. \tag{8.8.1}$$

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

$$Fat\,loss = (1000\,kJ)\left(\frac{1\,g\,fat}{30\,kJ}\right) = 26\,g, \tag{8.8.2}$$

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



Figure 8.8.2: 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
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.

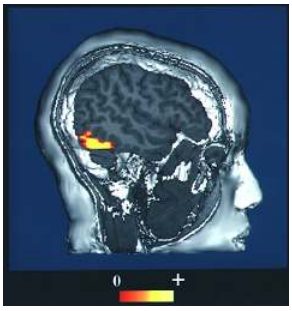


Figure **8.8.3**: This MRI scan shows an increased level of energy consumption in the vision center of the brain. Here, the patient was being asked to recognize faces. (credit: NIH via Wikimedia Commons)

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

This page titled 8.8: Work, Energy, and Power in Humans is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **7.8: Work, Energy, and Power in Humans** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

CHAPTER OVERVIEW

9: Chapter 9 - Waves

9.1: Free Waves
9.2: Sound
9.3: Hearing
9.4: Speed of Sound, Frequency, and Wavelength
9.5: Sound Intensity and Sound Level
9.6: Sound Interference and Resonance- Standing Waves in Air Columns
9.7: Images by Reflection
9.8: Refraction
9.9: Wave Optics

9: Chapter 9 - Waves is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

9.1: Free Waves

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

This page titled 9.1: Free Waves is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

- **7.1: Free Waves** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

9.2: Sound

Learning Objectives

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

- Define sound and hearing.
- Describe sound as a longitudinal wave.

Sound can be used as a familiar illustration of waves. Because hearing is one of our most important senses, it is interesting to see how the physical properties of sound correspond to our perceptions of it. **Hearing** is the perception of sound, just as vision is the perception of visible light. But sound has important applications beyond hearing. Ultrasound, for example, is not heard but can be employed to form medical images and is also used in treatment.



Figure **9.2.1**: This glass has been shattered by a high-intensity sound wave of the same frequency as the resonant frequency of the glass. While the sound is not visible, the effects of the sound prove its existence. (credit: ||read||, Flickr)

The physical phenomenon of **sound** is defined to be a disturbance of matter that is transmitted from its source outward. Sound is a wave. On the atomic scale, it is a disturbance of atoms that is far more ordered than their thermal motions. In many instances, sound is a periodic wave, and the atoms undergo simple harmonic motion. In this text, we shall explore such periodic sound waves.

A vibrating string produces a sound wave as illustrated in Figures **9.2.2**-**9.2.4**. As the string oscillates back and forth, it transfers energy to the air, mostly as thermal energy created by turbulence. But a small part of the string's energy goes into compressing and expanding the surrounding air, creating slightly higher and lower local pressures. These compressions (high pressure regions) and rarefactions (low pressure regions) move out as longitudinal pressure waves having the same frequency as the string—they are the disturbance that is a sound wave. (Sound waves in air and most fluids are longitudinal, because fluids have almost no shear strength. In solids, sound waves can be both transverse and longitudinal.) Figure **9.2.4** shows a graph of gauge pressure versus distance from the vibrating string.

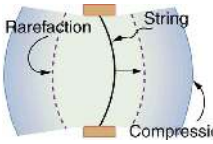


Figure **9.2.2**: A vibrating string moving to the right compresses the air in front of it and expands the air behind it.

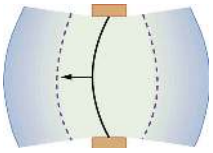
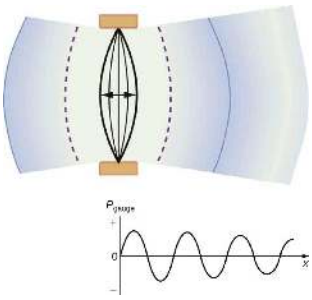


Figure **9.2.3**: As the string moves to the left, it creates another compression and rarefaction as the



ones on the right move away from the string. Figure **9.2.4**: After many vibrations, there are a series of compressions and rarefactions moving out from the string as a sound wave. The graph shows gauge pressure versus distance from the source. Pressures vary only slightly from atmospheric for ordinary sounds.

The amplitude of a sound wave decreases with distance from its source, because the energy of the wave is spread over a larger and larger area. But it is also absorbed by objects, such as the eardrum in Figure **9.2.5**, and converted to thermal energy by the viscosity of air. In addition, during each compression a little heat transfers to the air and during each rarefaction even less heat transfers from the air, so that the heat transfer reduces the organized disturbance into random thermal motions. (These processes can be viewed as a manifestation of the second law of thermodynamics presented in Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency.) Whether the heat transfer from compression to rarefaction is significant depends on how far apart they are—that is, it depends on wavelength. Wavelength, frequency, amplitude, and speed of propagation are important for sound, as they are for all waves.

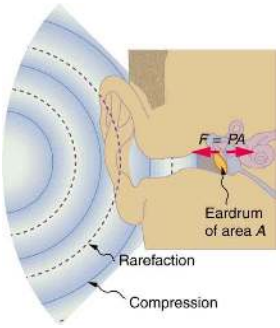


Figure **9.2.5**: Sound wave compressions and rarefactions travel up the ear canal and force the eardrum to vibrate. There is a net force on the eardrum, since the sound wave pressures differ from the atmospheric pressure found behind the eardrum. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person.

PHET EXPLORATIONS: WAVE INTERFERENCE

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

PhET Interactive Simulation

Figure **9.2.6**: Wave Interference

Summary

- Sound is a disturbance of matter that is transmitted from its source outward.
- Sound is one type of wave.
- Hearing is the perception of sound.

Glossary

sound

a disturbance of matter that is transmitted from its source outward

hearing

the perception of sound

This page titled 9.2: Sound is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **17.1: Sound** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

9.3: Hearing

Learning Objectives

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

- Define hearing, pitch, loudness, timbre, note, tone, phon, ultrasound, and infrasound.
- Compare loudness to frequency and intensity of a sound.
- Identify structures of the inner ear and explain how they relate to sound perception.

The human ear has a tremendous range and sensitivity. It can give us a wealth of simple information—such as pitch, loudness, and direction. And from its input we can detect musical quality and nuances of voiced emotion. How is our hearing related to the physical qualities of sound, and how does the hearing mechanism work?



Figure **9.3.1**: Hearing allows this vocalist, his band, and his fans to enjoy music. (credit: West Point Public Affairs, Flickr)

Hearing is the perception of sound. (Perception is commonly defined to be awareness through the senses, a typically circular definition of higher-level processes in living organisms.) Normal human hearing encompasses frequencies from 20 to 20,000 Hz, an impressive range. Sounds below 20 Hz are called **infrasound**, whereas those above 20,000 Hz are **ultrasound**. Neither is perceived by the ear, although infrasound can sometimes be felt as vibrations. When we do hear low-frequency vibrations, such as the sounds of a diving board, we hear the individual vibrations only because there are higher-frequency sounds in each. Other animals have hearing ranges different from that of humans. Dogs can hear sounds as high as 30,000 Hz, whereas bats and dolphins can hear up to 100,000-Hz sounds. You may have noticed that dogs respond to the sound of a dog whistle which produces sound out of the range of human hearing. Elephants are known to respond to frequencies below 20 Hz.

The perception of frequency is called **pitch**. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different frequency from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500.0 and 501.5 Hz are noticeably different. Pitch perception is directly related to frequency and is not greatly affected by other physical quantities such as intensity. Musical **notes** are particular sounds that can be produced by most instruments and in Western music have particular names. Combinations of notes constitute music. Some people can identify musical notes, such as A-sharp, C, or E-flat, just by listening to them. This uncommon ability is called perfect pitch. The ear is remarkably sensitive to low-intensity sounds. The lowest audible intensity or threshold is about $10^{-12}\ W/m^2$ or 0 dB. Sounds as much as 10^{12} more intense can be briefly tolerated. Very few measuring devices are capable of observations over a range of a trillion. The perception of intensity is called **loudness**. At a given frequency, it is possible to discern differences of about 1 dB, and a change of 3 dB is easily noticed. But loudness is not related to intensity alone. Frequency has a major effect on how loud a sound seems. The ear has its maximum sensitivity to frequencies in the range of 2000 to 5000 Hz, so that sounds in this range are perceived as being louder than, say, those at 500 or 10,000 Hz, even when they all have the same intensity. Sounds near the high-

and low-frequency extremes of the hearing range seem even less loud, because the ear is even less sensitive at those frequencies. Table **9.3.1** gives the dependence of certain human hearing perceptions on physical quantities.

Table **9.3.1**: Sound Perceptions

Perception	Physical quantity
Pitch	Frequency
Loudness	Intensity and Frequency
Timbre	Number and relative intensity of multiple frequencies. Subtle craftsmanship leads to non-linear effects and more detail.
Note	Basic unit of music with specific names, combined to generate tunes
Tone	Number and relative intensity of multiple frequencies.

When a violin plays middle C, there is no mistaking it for a piano playing the same note. The reason is that each instrument produces a distinctive set of frequencies and intensities. We call our perception of these combinations of frequencies and intensities **tone** quality, or more commonly the **timbre** of the sound. It is more difficult to correlate timbre perception to physical quantities than it is for loudness or pitch perception. Timbre is more subjective. Terms such as dull, brilliant, warm, cold, pure, and rich are employed to describe the timbre of a sound. So the consideration of timbre takes us into the realm of perceptual psychology, where higher-level processes in the brain are dominant. This is true for other perceptions of sound, such as music and noise. We shall not delve further into them; rather, we will concentrate on the question of loudness perception.

A unit called a **phon** is used to express loudness numerically. Phons differ from decibels because the phon is a unit of loudness perception, whereas the decibel is a unit of physical intensity. Figure **9.3.2** shows the relationship of loudness to intensity (or intensity level) and frequency for persons with normal hearing. The curved lines are equal-loudness curves. Each curve is labeled with its loudness in phons. Any sound along a given curve will be perceived as equally loud by the average person. The curves were determined by having large numbers of people compare the loudness of sounds at different frequencies and sound intensity levels. At a frequency of 1000 Hz, phons are taken to be numerically equal to decibels. The following example helps illustrate how to use the graph:

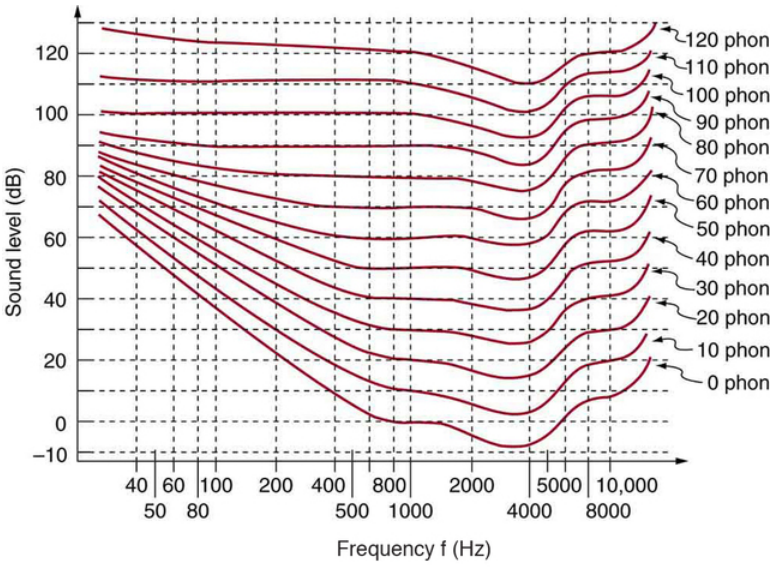


Figure **9.3.2**: The relationship of loudness in phons to intensity level (in decibels) and intensity (in watts per meter squared) for persons with normal hearing. The curved lines are equal-loudness curves—all sounds on a given curve are perceived as equally loud. Phons and decibels are defined to be the same at 1000 Hz.

Example 9.3.1: Measuring Loudness: Loudness Versus Intensity Level and Frequency

(a) What is the loudness in phons of a 100-Hz sound that has an intensity level of 80 dB? (b) What is the intensity level in decibels of a 4000-Hz sound having a loudness of 70 phons? (c) At what intensity level will an 8000-Hz sound have the same loudness as a 200-Hz sound at 60 dB?

Strategy for (a)

The graph in Figure **9.3.2** should be referenced in order to solve this example. To find the loudness of a given sound, you must know its frequency and intensity level and locate that point on the square grid, then interpolate between loudness curves to get the loudness in phons.

Solution for (a)

(1) Identify knowns:

- The square grid of the graph relating phons and decibels is a plot of intensity level versus frequency—both physical quantities.
- 100 Hz at 80 dB lies halfway between the curves marked 70 and 80 phons.

(2) Find the loudness: 75 phons.

Strategy for (b)

The graph in Figure **9.3.2** should be referenced in order to solve this example. To find the intensity level of a sound, you must have its frequency and loudness. Once that point is located, the intensity level can be determined from the vertical axis.

Solution for (b)

(1) Identify knowns:

- Values are given to be 4000 Hz at 70 phons.

(2) Follow the 70-phon curve until it reaches 4000 Hz. At that point, it is below the 70 dB line at about 67 dB.

(3) Find the intensity level:

67 dB

Strategy for (c)

The graph in Figure **9.3.2** should be referenced in order to solve this example.

Solution for (c)

(1) Locate the point for a 200 Hz and 60 dB sound.

(2) Find the loudness: This point lies just slightly above the 50-phon curve, and so its loudness is 51 phons.

(3) Look for the 51-phon level is at 8000 Hz: 63 dB.

Discussion

These answers, like all information extracted from Figure **9.3.2**, have uncertainties of several phons or several decibels, partly due to difficulties in interpolation, but mostly related to uncertainties in the equal-loudness curves.

Further examination of the graph in Figure **9.3.2** reveals some interesting facts about human hearing. First, sounds below the 0-phon curve are not perceived by most people. So, for example, a 60 Hz sound at 40 dB is inaudible. The 0-phon curve represents the threshold of normal hearing. We can hear some sounds at intensity levels below 0 dB. For example, a 3-dB, 5000-Hz sound is audible, because it lies above the 0-phon curve. The loudness curves all have dips in them between about 2000 and 5000 Hz. These dips mean the ear is most sensitive to frequencies in that range. For example, a 15-dB sound at 4000 Hz has a loudness of 20 phons, the same as a 20-dB sound at 1000 Hz. The curves rise at both extremes of the frequency range, indicating that a greater-intensity level sound is needed at those frequencies to be perceived to be as loud as at middle frequencies. For example, a sound at

10,000 Hz must have an intensity level of 30 dB to seem as loud as a 20 dB sound at 1000 Hz. Sounds above 120 phons are painful as well as damaging.

We do not often utilize our full range of hearing. This is particularly true for frequencies above 8000 Hz, which are rare in the environment and are unnecessary for understanding conversation or appreciating music. In fact, people who have lost the ability to hear such high frequencies are usually unaware of their loss until tested. The shaded region in Figure **9.3.3** is the frequency and intensity region where most conversational sounds fall. The curved lines indicate what effect hearing losses of 40 and 60 phons will have. A 40-phon hearing loss at all frequencies still allows a person to understand conversation, although it will seem very quiet. A person with a 60-phon loss at all frequencies will hear only the lowest frequencies and will not be able to understand speech unless it is much louder than normal. Even so, speech may seem indistinct, because higher frequencies are not as well perceived. The conversational speech region also has a gender component, in that female voices are usually characterized by higher frequencies. So the person with a 60-phon hearing impediment might have difficulty understanding the normal conversation of a woman.

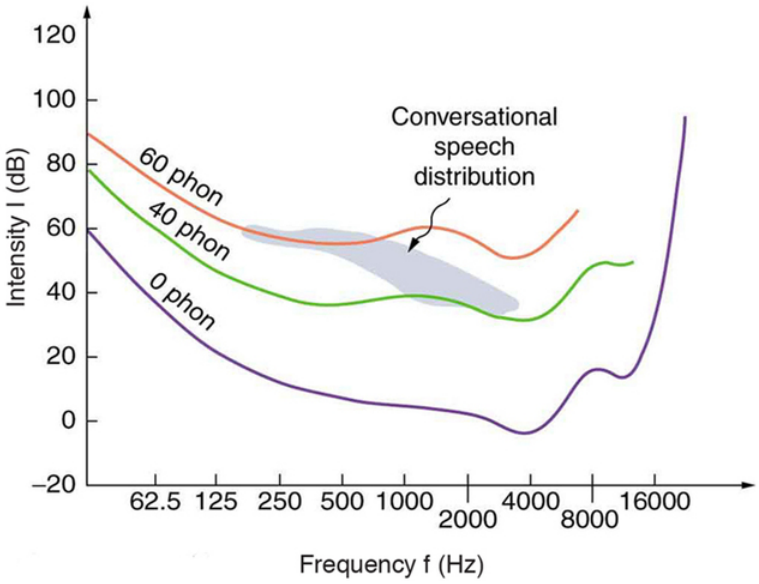


Figure **9.3.3**: The shaded region represents frequencies and intensity levels found in normal conversational speech. The 0-phon line represents the normal hearing threshold, while those at 40 and 60 represent thresholds for people with 40- and 60-phon hearing losses, respectively.

Hearing tests are performed over a range of frequencies, usually from 250 to 8000 Hz, and can be displayed graphically in an audiogram like that in Figure **9.3.4**. The hearing threshold is measured in dB *relative to the normal threshold*, so that normal hearing registers as 0 dB at all frequencies. Hearing loss caused by noise typically shows a dip near the 4000 Hz frequency, irrespective of the frequency that caused the loss and often affects both ears. The most common form of hearing loss comes with age and is called *presbycusis*—literally elder ear. Such loss is increasingly severe at higher frequencies, and interferes with music appreciation and speech recognition.

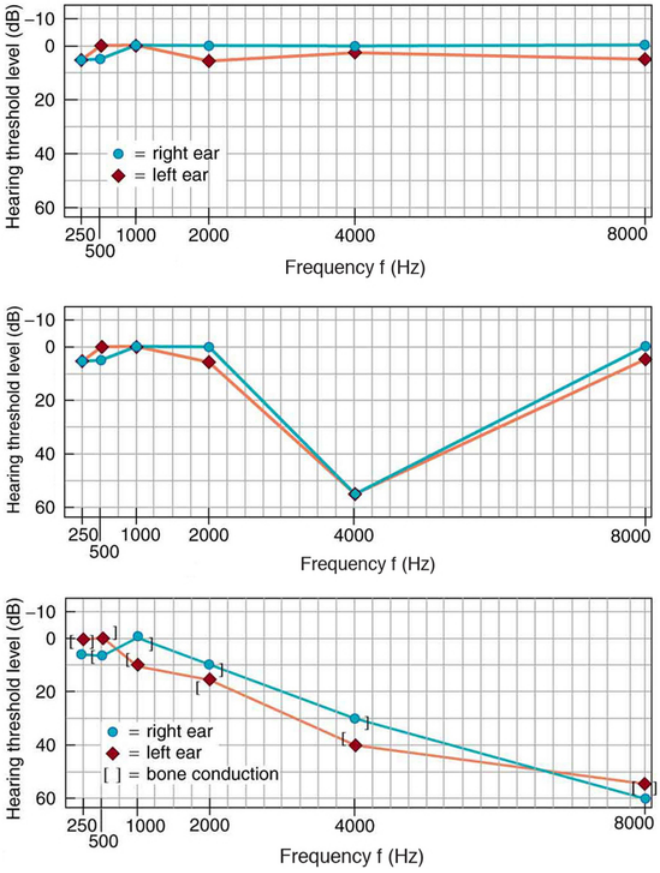


Figure **9.3.4**: Audiograms showing the threshold in intensity level versus frequency for three different individuals. Intensity level is measured relative to the normal threshold. The top left graph is that of a person with normal hearing. The graph to its right has a dip at 4000 Hz and is that of a child who suffered hearing loss due to a cap gun. The third graph is typical of presbycusis, the progressive loss of higher frequency hearing with age. Tests performed by bone conduction (brackets) can distinguish nerve damage from middle ear damage.

Hearing Mechanism

The hearing mechanism involves some interesting physics. The sound wave that impinges upon our ear is a pressure wave. The ear is a transducer that converts sound waves into electrical nerve impulses in a manner much more sophisticated than, but analogous to, a microphone. Figure **9.3.5** shows the gross anatomy of the ear with its division into three parts: the outer ear or ear canal; the middle ear, which runs from the eardrum to the cochlea; and the inner ear, which is the cochlea itself. The body part normally referred to as the ear is technically called the pinna.

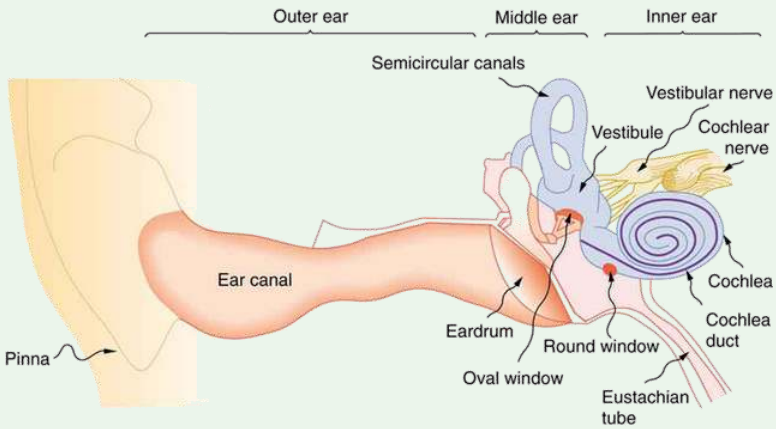


Figure **9.3.5**: The illustration shows the gross anatomy of the human ear.

https://chem.libretexts.org/@go/page/200174

The outer ear, or ear canal, carries sound to the recessed protected eardrum. The air column in the ear canal resonates and is partially responsible for the sensitivity of the ear to sounds in the 2000 to 5000 Hz range. The middle ear converts sound into mechanical vibrations and applies these vibrations to the cochlea. The lever system of the middle ear takes the force exerted on the eardrum by sound pressure variations, amplifies it and transmits it to the inner ear via the oval window, creating pressure waves in the cochlea approximately 40 times greater than those impinging on the eardrum (Figure **9.3.6**.) Two muscles in the middle ear (not shown) protect the inner ear from very intense sounds. They react to intense sound in a few milliseconds and reduce the force transmitted to the cochlea. This protective reaction can also be triggered by your own voice, so that humming while shooting a gun, for example, can reduce noise damage.

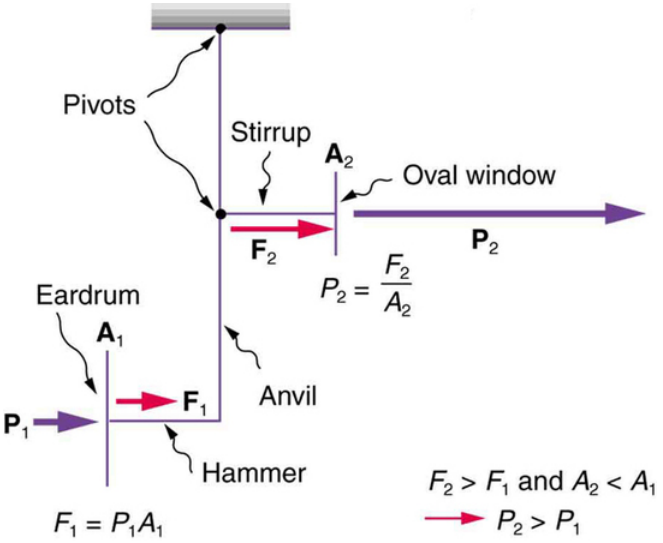


Figure **9.3.6**: This schematic shows the middle ear's system for converting sound pressure into force, increasing that force through a lever system, and applying the increased force to a small area of the cochlea, thereby creating a pressure about 40 times that in the original sound wave. A protective muscle reaction to intense sounds greatly reduces the mechanical advantage of the lever system.

Figure **9.3.7** shows the middle and inner ear in greater detail. Pressure waves moving through the cochlea cause the tectorial membrane to vibrate, rubbing cilia (called hair cells), which stimulate nerves that send electrical signals to the brain. The membrane resonates at different positions for different frequencies, with high frequencies stimulating nerves at the near end and low frequencies at the far end. The complete operation of the cochlea is still not understood, but several mechanisms for sending information to the brain are known to be involved. For sounds below about 1000 Hz, the nerves send signals at the same frequency as the sound. For frequencies greater than about 1000 Hz, the nerves signal frequency by position. There is a structure to the cilia, and there are connections between nerve cells that perform signal processing before information is sent to the brain. Intensity information is partly indicated by the number of nerve signals and by volleys of signals. The brain processes the cochlear nerve signals to provide additional information such as source direction (based on time and intensity comparisons of sounds from both ears). Higher-level processing produces many nuances, such as music appreciation.

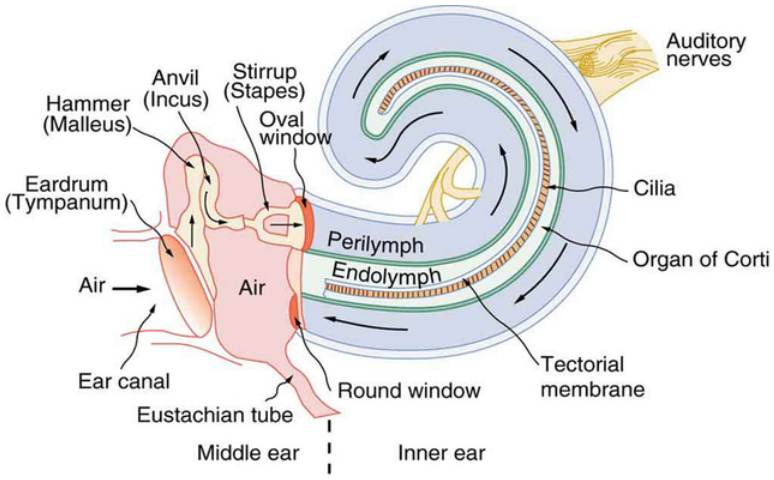


Figure **9.3.7**: The inner ear, or cochlea, is a coiled tube about 3 mm in diameter and 3 cm in length if uncoiled. When the oval window is forced inward, as shown, a pressure wave travels through the perilymph in the direction of the arrows, stimulating nerves at the base of cilia in the organ of Corti.

Hearing losses can occur because of problems in the middle or inner ear. Conductive losses in the middle ear can be partially overcome by sending sound vibrations to the cochlea through the skull. Hearing aids for this purpose usually press against the bone behind the ear, rather than simply amplifying the sound sent into the ear canal as many hearing aids do. Damage to the nerves in the cochlea is not repairable, but amplification can partially compensate. There is a risk that amplification will produce further damage. Another common failure in the cochlea is damage or loss of the cilia but with nerves remaining functional. Cochlear implants that stimulate the nerves directly are now available and widely accepted. Over 100,000 implants are in use, in about equal numbers of adults and children.

The cochlear implant was pioneered in Melbourne, Australia, by Graeme Clark in the 1970s for his deaf father. The implant consists of three external components and two internal components. The external components are a microphone for picking up sound and converting it into an electrical signal, a speech processor to select certain frequencies and a transmitter to transfer the signal to the internal components through electromagnetic induction. The internal components consist of a receiver/transmitter secured in the bone beneath the skin, which converts the signals into electric impulses and sends them through an internal cable to the cochlea and an array of about 24 electrodes wound through the cochlea. These electrodes in turn send the impulses directly into the brain. The electrodes basically emulate the cilia.

Exercise 9.3.1

Are ultrasound and infrasound imperceptible to all hearing organisms? Explain your answer.

Answer

No, the range of perceptible sound is based in the range of human hearing. Many other organisms perceive either infrasound or ultrasound.

Summary

- The range of audible frequencies is 20 to 20,000 Hz.
- Those sounds above 20,000 Hz are ultrasound, whereas those below 20 Hz are infrasound.
- The perception of frequency is pitch.
- The perception of intensity is loudness.
- Loudness has units of phons.

Glossary

loudness

the perception of sound intensity

timbre

number and relative intensity of multiple sound frequencies

note

basic unit of music with specific names, combined to generate tunes

tone

number and relative intensity of multiple sound frequencies

phon

the numerical unit of loudness

ultrasound

sounds above 20,000 Hz

infrasound

sounds below 20 Hz

This page titled 9.3: Hearing is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **17.6: Hearing** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

9.4: Speed of Sound, Frequency, and Wavelength

Learning Objectives

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

- Define pitch.
- Describe the relationship between the speed of sound, its frequency, and its wavelength.
- Describe the effects on the speed of sound as it travels through various media.
- Describe the effects of temperature on the speed of sound.

Sound, like all waves, travels at a certain speed and has the properties of frequency and wavelength. You can observe direct evidence of the speed of sound while watching a fireworks display. The flash of an explosion is seen well before its sound is heard, implying both that sound travels at a finite speed and that it is much slower than light. You can also directly sense the frequency of a sound. Perception of frequency is called pitch. The wavelength of sound is not directly sensed, but indirect evidence is found in the correlation of the size of musical instruments with their pitch. Small instruments, such as a piccolo, typically make high-pitch sounds, while large instruments, such as a tuba, typically make low-pitch sounds. High pitch means small wavelength, and the size of a musical instrument is directly related to the wavelengths of sound it produces. So a small instrument creates short-wavelength sounds. Similar arguments hold that a large instrument creates long-wavelength sounds.



Figure **9.4.1**: When a firework explodes, the light energy is perceived before the sound energy. Sound travels more slowly than light does. (credit: Dominic Alves, Flickr)

The relationship of the speed of sound, its frequency, and wavelength is the same as for all waves:

$$v_w = f\lambda, \tag{9.4.1}$$

where v_w is the speed of sound, f is its frequency, and λ is its wavelength. The wavelength of a sound is the distance between adjacent identical parts of a wave—for example, between adjacent compressions as illustrated in Figure **9.4.2**. The frequency is the same as that of the source and is the number of waves that pass a point per unit time.

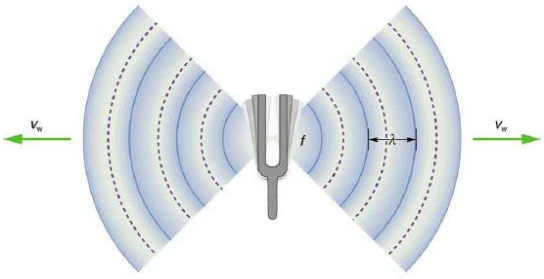


Figure **9.4.2**: A sound wave emanates from a source vibrating at a frequency f, propagates at v_w, and has a wavelength λ.

Table **9.4.1** makes it apparent that the speed of sound varies greatly in different media. The speed of sound in a medium is determined by a combination of the medium's rigidity (or compressibility in gases) and its density. The more rigid (or less compressible) the medium, the faster the speed of sound. For materials that have similar rigidities, sound will travel faster through the one with the lower density because the sound energy is more easily transferred from particle to particle. The speed of sound in air is low, because air is compressible. Because liquids and solids are relatively rigid and very difficult to compress, the speed of sound in such media is generally greater than in gases.

Table **9.4.1**: Speed of Sound in Various Media

Medium	$v_w (m/s)$
Gases at $0°C$	
Air	331
Carbon dioxide	259
Oxygen	316
Helium	965
Hydrogen	1290
Liquids at $20°C$	
Ethanol	1160
Mercury	1450
Water, fresh	1480
Sea water	1540
Human tissue	1540
Solids (longitudinal or bulk)	
Vulcanized rubber	54
Polyethylene	920
Marble	3810
Glass, Pyrex	5640
Lead	1960
Aluminum	5120
Steel	5960

Earthquakes, essentially sound waves in Earth's crust, are an interesting example of how the speed of sound depends on the rigidity of the medium. Earthquakes have both longitudinal and transverse components, and these travel at different speeds. The bulk modulus of granite is greater than its shear modulus. For that reason, the speed of longitudinal or pressure waves (P-waves) in earthquakes in granite is significantly higher than the speed of transverse or shear waves (S-waves). Both components of earthquakes travel slower in less rigid material, such as sediments. P-waves have speeds of 4 to 7 km/s, and S-waves correspondingly range in speed from 2 to 5 km/s, both being faster in more rigid material. The P-wave gets progressively farther ahead of the S-wave as they travel through Earth's crust. The time between the P- and S-waves is routinely used to determine the distance to their source, the epicenter of the earthquake.

The speed of sound is affected by temperature in a given medium. For air at sea level, the speed of sound is given by

$$v_w = (331\, m/s)\sqrt{\frac{T}{273\, K}}, \tag{9.4.2}$$

where the temperature (denoted as T) is in units of kelvin. The speed of sound in gases is related to the average speed of particles in the gas, v_{rms}, and that

$$v_{rms} = \sqrt{\frac{3\,kT}{m}}, \tag{9.4.3}$$

where k is the Boltzmann constant ($1.38 \times 10^{-23}\ J/K$) and m is the mass of each (identical) particle in the gas. So, it is reasonable that the speed of sound in air and other gases should depend on the square root of temperature. While not negligible, this is not a strong dependence. At $0°C$, the speed of sound is 331 m/s, whereas at $20°C$ it is 343 m/s, less than a 4% increase. Figure 9.4.3 shows a use of the speed of sound by a bat to sense distances. Echoes are also used in medical imaging.



Figure 9.4.3: A bat uses sound echoes to find its way about and to catch prey. The time for the echo to return is directly proportional to the distance.

One of the more important properties of sound is that its speed is nearly independent of frequency. This independence is certainly true in open air for sounds in the audible range of 20 to 20,000 Hz. If this independence were not true, you would certainly notice it for music played by a marching band in a football stadium, for example. Suppose that high-frequency sounds traveled faster—then the farther you were from the band, the more the sound from the low-pitch instruments would lag that from the high-pitch ones. But the music from all instruments arrives in cadence independent of distance, and so all frequencies must travel at nearly the same speed. Recall that

$$v_w = f\lambda. \tag{9.4.4}$$

In a given medium under fixed conditions, v_w is constant, so that there is a relationship between f and λ; the higher the frequency, the smaller the wavelength. See Figure 9.4.4 and consider the following example.

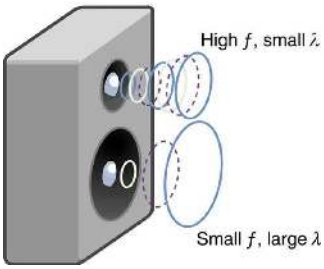


Figure 9.4.4: Because they travel at the same speed in a given medium, low-frequency sounds must have a greater wavelength than high-frequency sounds. Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, while the higher-frequency sounds are emitted by the small speaker, called a tweeter.

Example 9.4.1: Calculating Wavelengths: What Are the Wavelengths of Audible Sounds?

Calculate the wavelengths of sounds at the extremes of the audible range, 20 and 20,000 Hz, in $30.0°C$ air. (Assume that the frequency values are accurate to two significant figures.)

Strategy

To find wavelength from frequency, we can use $v_w = f\lambda$.

Solution

1. Identify knowns. The value for v_w, is given by

$$v_w = (331\ m/s)\sqrt{\frac{T}{273\ K}}.$$

2. Convert the temperature into kelvin and then enter the temperature into the equation

$$v_w = (331\ m/s)\sqrt{\frac{303\ K}{273\ K}} = 348.7\ m/s.$$

3. Solve the relationship between speed and wavelength for λ:

$$\lambda = \frac{v_w}{f}.$$

4. Enter the speed and the minimum frequency to give the maximum wavelength:

$$\lambda_{max} = \frac{348.7\,m/s}{20\,Hz} = 17\,m.$$

5. Enter the speed and the maximum frequency to give the minimum wavelength:

$$\lambda_{min} = \frac{348.7\,m/s}{20,000\,Hz} = 0.017\,m = 1.7\,cm.$$

Discussion

Because the product of f multiplied by λ equals a constant, the smaller f is, the larger λ must be, and vice versa.

The speed of sound can change when sound travels from one medium to another. However, the frequency usually remains the same because it is like a driven oscillation and has the frequency of the original source. If v_w changes and f remains the same, then the wavelength λ must change. That is, because $v_w = f\lambda$, the higher the speed of a sound, the greater its wavelength for a given frequency.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION - VOICE AS A SOUND WAVE

Suspend a sheet of paper so that the top edge of the paper is fixed and the bottom edge is free to move. You could tape the top edge of the paper to the edge of a table. Gently blow near the edge of the bottom of the sheet and note how the sheet moves. Speak softly and then louder such that the sounds hit the edge of the bottom of the paper, and note how the sheet moves. Explain the effects.

Exercise 9.4.1A

Imagine you observe two fireworks explode. You hear the explosion of one as soon as you see it. However, you see the other firework for several milliseconds before you hear the explosion. Explain why this is so.

Answer

Sound and light both travel at definite speeds. The speed of sound is slower than the speed of light. The first firework is probably very close by, so the speed difference is not noticeable. The second firework is farther away, so the light arrives at your eyes noticeably sooner than the sound wave arrives at your ears.

Exercise 9.4.1B

You observe two musical instruments that you cannot identify. One plays high-pitch sounds and the other plays low-pitch sounds. How could you determine which is which without hearing either of them play?

Answer

Compare their sizes. High-pitch instruments are generally smaller than low-pitch instruments because they generate a smaller wavelength.

Summary

- The relationship of the speed of sound v_w, its frequency f, and its wavelength λ is given by $v_w = f\lambda$, which is the same relationship given for all waves.
- In air, the speed of sound is related to air temperature T by $v_w = (331\,m/s)\sqrt{\frac{T}{273\,K}}$. v_w is the same for all frequencies and wavelengths.

Glossary

pitch

the perception of the frequency of a sound

This page titled 9.4: Speed of Sound, Frequency, and Wavelength is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **17.2: Speed of Sound, Frequency, and Wavelength** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

9.5: Sound Intensity and Sound Level

Learning Objectives

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

- Define intensity, sound intensity, and sound pressure level.
- Calculate sound intensity levels in decibels (dB).

In a quiet forest, you can sometimes hear a single leaf fall to the ground. After settling into bed, you may hear your blood pulsing through your ears. But when a passing motorist has his stereo turned up, you cannot even hear what the person next to you in your car is saying. We are all very familiar with the loudness of sounds and aware that they are related to how energetically the source is vibrating. In cartoons depicting a screaming person (or an animal making a loud noise), the cartoonist often shows an open mouth with a vibrating uvula, the hanging tissue at the back of the mouth, to suggest a loud sound coming from the throat Figure **9.5.1**. High noise exposure is hazardous to hearing, and it is common for musicians to have hearing losses that are sufficiently severe that they interfere with the musicians' abilities to perform. The relevant physical quantity is sound intensity, a concept that is valid for all sounds whether or not they are in the audible range.



Figure **9.5.1**: Noise on crowded roadways like this one in Delhi makes it hard to hear others unless they shout. (credit: Lingaraj G J, Flickr)

Intensity is defined to be the power per unit area carried by a wave. Power is the rate at which energy is transferred by the wave. In equation form, **intensity** I is

$$I = \frac{P}{A}, \tag{9.5.1}$$

where P is the power through an area A. The SI unit for I is W/m^2. The intensity of a sound wave is related to its amplitude squared by the following relationship:

$$I = \frac{(\Delta p)^2}{2\rho v_w}. \tag{9.5.2}$$

Here Δp is the pressure variation or pressure amplitude (half the difference between the maximum and minimum pressure in the sound wave) in units of pascals (Pa) or N/m^2. (We are using a lower case p for pressure to distinguish it from power, denoted by P above.) The energy (as kinetic energy $\frac{mv^2}{2}$) of an oscillating element of air due to a traveling sound wave is proportional to its amplitude squared. In this equation, ρ is the density of the material in which the sound wave travels, in units of kg/m^3, and v_w is the speed of sound in the medium, in units of m/s. The pressure variation is proportional to the amplitude of the oscillation, and so I varies as $(\Delta p)^2$ (Figure **9.5.2**). This relationship is consistent with the fact that the sound wave is produced by some vibration; the greater its pressure amplitude, the more the air is compressed in the sound it creates.

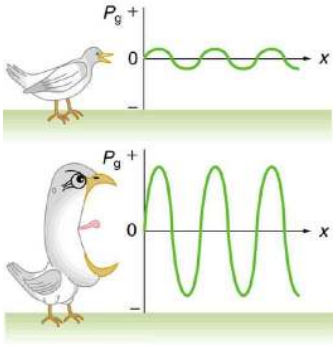


Figure 9.5.2: Graphs of the gauge pressures in two sound waves of different intensities. The more intense sound is produced by a source that has larger-amplitude oscillations and has greater pressure maxima and minima. Because pressures are higher in the greater-intensity sound, it can exert larger forces on the objects it encounters.

Sound intensity levels are quoted in decibels (dB) much more often than sound intensities in watts per meter squared. Decibels are the unit of choice in the scientific literature as well as in the popular media. The reasons for this choice of units are related to how we perceive sounds. How our ears perceive sound can be more accurately described by the logarithm of the intensity rather than directly to the intensity. The sound intensity level β in decibels of a sound having an intensity I in watts per meter squared is defined to be

$$\beta\,(dB) = 10\,\log_{10}\left(\frac{I}{I_0}\right), \tag{9.5.3}$$

where $I_0 = 10^{-12}\,W/m^2$ is a reference intensity. In particular, I_0 is the lowest or threshold intensity of sound a person with normal hearing can perceive at a frequency of 1000 Hz. Sound intensity level is not the same as intensity. Because β is defined in terms of a ratio, it is a unitless quantity telling you the *level* of the sound relative to a fixed standard ($10^{-12}\,W/m^2$, in this case). The units of decibels (dB) are used to indicate this ratio is multiplied by 10 in its definition. The bel, upon which the decibel is based, is named for Alexander Graham Bell, the inventor of the telephone.

Table 9.5.1: Sound Intensity Levels and Intensities

Sound intensity level β (dB)	Intensity $I(W?M^2)$	Example/effect
0	1×10^{-12}	Threshold of hearing at 1000 Hz
10	1×10^{-11}	Rustle of leaves
20	1×10^{-10}	Whisper at 1 m distance
30	1×10^{-9}	Quiet home
40	1×10^{-8}	Average home
50	1×10^{-7}	Average office, soft music
60	1×10^{-6}	Normal conversation
70	1×10^{-5}	Noisy office, busy traffic
80	1×10^{-4}	Loud radio, classroom lecture
90	1×10^{-3}	Inside a heavy truck; damage from prolonged exposure 1
100	1×10^{-2}	Noisy factory, siren at 30 m; damage from 8 h per day exposure
110	1×10^{-1}	Damage from 30 min per day exposure
120	1	Loud rock concert, pneumatic chipper at 2 m; threshold of pain
140	1×10^{2}	Jet airplane at 30 m; severe pain, damage in seconds
160	1×10^{4}	Bursting of eardrums

The decibel level of a sound having the threshold intensity of $10^{-12}\,W/m^2$ is $\beta = 0\,dB$, because $log_{10}1 = 0$. That is, the threshold of hearing is 0 decibels. Table **9.5.1** gives levels in decibels and intensities in watts per meter squared for some familiar sounds.

One of the more striking things about the intensities in Table **9.5.1** is that the intensity in watts per meter squared is quite small for most sounds. The ear is sensitive to as little as a trillionth of a watt per meter squared—even more impressive when you realize that the area of the eardrum is only about $1\,cm^2$, so that only 10^{-16} W falls on it at the threshold of hearing! Air molecules in a sound wave of this intensity vibrate over a distance of less than one molecular diameter, and the gauge pressures involved are less than 10^{-9} atm.

Another impressive feature of the sounds in Table **9.5.1** is their numerical range. Sound intensity varies by a factor of 10^{12} from threshold to a sound that causes damage in seconds. You are unaware of this tremendous range in sound intensity because how your ears respond can be described approximately as the logarithm of intensity. Thus, sound intensity levels in decibels fit your experience better than intensities in watts per meter squared. The decibel scale is also easier to relate to because most people are more accustomed to dealing with numbers such as 0, 53, or 120 than numbers such as 1.00×10^{-11}.

One more observation readily verified by examining Table **9.5.1** or using Equation 9.5.2 is that each factor of 10 in intensity corresponds to 10 dB. For example, a 90 dB sound compared with a 60 dB sound is 30 dB greater, or three factors of 10 (that is, 10^3 times) as intense. Another example is that if one sound is 10^7 as intense as another, it is 70 dB higher. See Table **9.5.2**.

Table **9.5.2**: Ratios of Intensities and Corresponding Differences in Sound Intensity Levels

I_2/I_1	β_2/β_1
2.0	3.0 dB
5.0	7.0 dB
10.0	10.0 dB

Example 9.5.1: Calculating Sound Intensity Levels: Sound Waves

Calculate the sound intensity level in decibels for a sound wave traveling in air at $0°C$ and having a pressure amplitude of 0.656 Pa.

Strategy

We are given Δp, so we can calculate I using the equation $I = (\Delta p)^2/(2\rho v_w)^2$. Using I, we can calculate β straight from its definition in $\beta\,(dB) = 10\,\log_{10}(I/I_0)$.

Solution

(1) Identify knowns:

Sound travels at 331 m/s in air at $0°C$.

Air has a density of $1.29\,kg/m^3$ at atmospheric pressure and $0°C$.

(2) Enter these values and the pressure amplitude into $I = (\Delta p)^2/(2\rho v_w)$:

$$I = \frac{(\Delta p)^2}{2\rho v_w}$$
$$= \frac{(0.656\,Pa)^2}{(1.29\,kg/m^3)(331\,m/s)}$$
$$= 5.04 \times 10^{-4}\,W/m^2$$

(3) Enter the value for I and the known value for I_0 into $\beta(dB = 10\,log_{10}(I/I_0)$. Calculate to find the sound intensity level in decibels:

$$10\,\log_{10}(5.04 \times 10^8) = 10(8.70)\,dB = 87\,dB.$$

Discussion

This 87 dB sound has an intensity five times as great as an 80 dB sound. So a factor of five in intensity corresponds to a difference of 7 dB in sound intensity level. This value is true for any intensities differing by a factor of five.

Example 9.5.2: Change Intensity Levels of a Sound: What Happens to the Decibel Level?

Show that if one sound is twice as intense as another, it has a sound level about 3 dB higher.

Strategy

You are given that the ratio of two intensities is 2 to 1, and are then asked to find the difference in their sound levels in decibels. You can solve this problem using of the properties of logarithms.

Solution

(1) Identify knowns:

The ratio of the two intensities is 2 to 1, or:

$$\frac{I_2}{I_1} = 2.00.$$

We wish to show that the difference in sound levels is about 3 dB. That is, we want to show:

$$\beta_2 - \beta_1 = 3\,dB. \tag{9.5.4}$$

Note that:

$$log_{10}b - log_{10}a = log_{10}\left(\frac{b}{a}\right).$$

(2) Use the definition of β to get:

$$\beta_2 - \beta_1 = 10\,log_{10}\left(\frac{I_2}{I_1}\right) = 10\,log_{10}2.00 = 10(0.301)\,dB. \tag{9.5.5}$$

Thus,

$$\beta_2 - \beta_1 = 3.01\,dB.$$

Discussion

This means that the two sound intensity levels differ by 3.01 dB, or about 3 dB, as advertised. Note that because only the ratio I_2/I_1 is given (and not the actual intensities), this result is true for any intensities that differ by a factor of two. For example, a 56.0 dB sound is twice as intense as a 53.0 dB sound, a 97.0 dB sound is half as intense as a 100 dB sound, and so on.

It should be noted at this point that there is another decibel scale in use, called the **sound pressure level**, based on the ratio of the pressure amplitude to a reference pressure. This scale is used particularly in applications where sound travels in water. It is beyond the scope of most introductory texts to treat this scale because it is not commonly used for sounds in air, but it is important to note that very different decibel levels may be encountered when sound pressure levels are quoted. For example, ocean noise pollution produced by ships may be as great as 200 dB expressed in the sound pressure level, where the more familiar sound intensity level we use here would be something under 140 dB for the same sound.

TAKE-HOME INVESTIGATION: FEELING SOUND

Find a CD player and a CD that has rock music. Place the player on a light table, insert the CD into the player, and start playing the CD. Place your hand gently on the table next to the speakers. Increase the volume and note the level when the table just begins to vibrate as the rock music plays. Increase the reading on the volume control until it doubles. What has happened to the vibrations?

Exercise 9.5.1

Describe how amplitude is related to the loudness of a sound.

Answer

Amplitude is directly proportional to the experience of loudness. As amplitude increases, loudness increases

Exercise 9.5.2

Identify common sounds at the levels of 10 dB, 50 dB, and 100 dB.

Answer

- 10 dB: Running fingers through your hair.
- 50 dB: Inside a quiet home with no television or radio.
- 100 dB: Take-off of a jet plane.

Summary

- Intensity is the same for a sound wave as was defined for all waves; it is $I = \frac{P}{A}$, where P is the power crossing area A. The SI unit for I is watts per meter squared. The intensity of a sound wave is also related to the pressure amplitude Δp $I = \frac{(\Delta p)^2}{2\rho v_w}$, where ρ is the density of the medium in which the sound wave travels and v_w is the speed of sound in the medium.

- Sound intensity level in units of decibels (dB) is $\beta(dB) = 10\log_{10}\left(\frac{I}{I_0}\right)$, where $I_0 = 10^{-12}\,W/m^2$ is the threshold intensity of hearing.

Footnotes

Several government agencies and health-related professional associations recommend that 85 dB not be exceeded for 8-hour daily exposures in the absence of hearing protection.

Glossary

intensity
: the power per unit area carried by a wave

sound intensity level
: a unitless quantity telling you the level of the sound relative to a fixed standard

sound pressure level
: the ratio of the pressure amplitude to a reference pressure

---

This page titled 9.5: Sound Intensity and Sound Level is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

- **17.3: Sound Intensity and Sound Level** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

9.6: Sound Interference and Resonance- Standing Waves in Air Columns

Learning Objectives

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

- Define antinode, node, fundamental, overtones, and harmonics.
- Identify instances of sound interference in everyday situations.
- Describe how sound interference occurring inside open and closed tubes changes the characteristics of the sound, and how this applies to sounds produced by musical instruments.
- Calculate the length of a tube using sound wave measurements.

Interference is the hallmark of waves, all of which exhibit constructive and destructive interference exactly analogous to that seen for water waves. In fact, one way to prove something "is a wave" is to observe interference effects. So, sound being a wave, we expect it to exhibit interference; we have already mentioned a few such effects, such as the beats from two similar notes played simultaneously.



Figure 9.6.1: Some types of headphones use the phenomena of constructive and destructive interference to cancel out outside noises. (credit: JVC America, Flickr)

Figure 9.6.2 shows a clever use of sound interference to cancel noise. Larger-scale applications of active noise reduction by destructive interference are contemplated for entire passenger compartments in commercial aircraft. To obtain destructive interference, a fast electronic analysis is performed, and a second sound is introduced with its maxima and minima exactly reversed from the incoming noise. Sound waves in fluids are pressure waves and consistent with Pascal's principle; pressures from two different sources add and subtract like simple numbers; that is, positive and negative gauge pressures add to a much smaller pressure, producing a lower-intensity sound. Although completely destructive interference is possible only under the simplest conditions, it is possible to reduce noise levels by 30 dB or more using this technique.

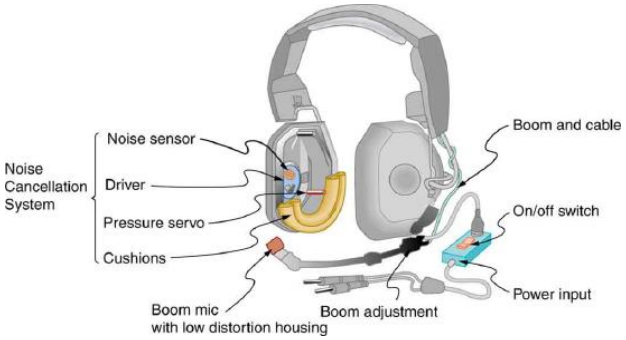


Figure 9.6.2: Headphones designed to cancel noise with destructive interference create a sound wave exactly opposite to the incoming sound. These headphones can be more effective than the simple passive attenuation used in most ear protection. Such headphones were used on the record-setting, around the world nonstop flight of the Voyager aircraft to protect the pilots' hearing from engine noise.

Where else can we observe sound interference? All sound resonances, such as in musical instruments, are due to constructive and destructive interference. Only the resonant frequencies interfere constructively to form standing waves, while others interfere destructively and are absent. From the toot made by blowing over a bottle, to the characteristic flavor of a violin's sounding box, to the recognizability of a great singer's voice, resonance and standing waves play a vital role.

Interference

Interference is such a fundamental aspect of waves that observing interference is proof that something is a wave. The wave nature of light was established by experiments showing interference. Similarly, when electrons scattered from crystals exhibited interference, their wave nature was confirmed to be exactly as predicted by symmetry with certain wave characteristics of light.

Suppose we hold a tuning fork near the end of a tube that is closed at the other end, as shown in Figure **9.6.3**, Figure **9.6.4**, Figure **9.6.5**, and Figure **9.6.6**. If the tuning fork has just the right frequency, the air column in the tube resonates loudly, but at most frequencies it vibrates very little. This observation just means that the air column has only certain natural frequencies. The figures show how a resonance at the lowest of these natural frequencies is formed. A disturbance travels down the tube at the speed of sound and bounces off the closed end. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes constructively with the continuing sound produced by the tuning fork. The incoming and reflected sounds form a standing wave in the tube as shown.

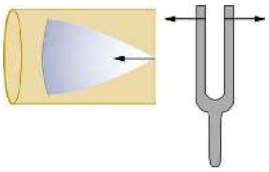


Figure **9.6.3**: Resonance of air in a tube closed at one end, caused by a tuning fork. A disturbance moves down the tube

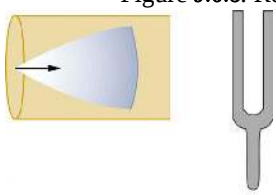


Figure **9.6.4**: Resonance of air in a tube closed at one end, caused by a tuning fork. The disturbance reflects from the closed end of the tube.

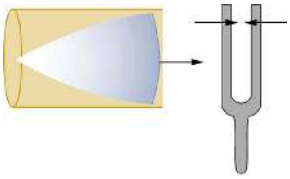


Figure **9.6.5**: Resonance of air in a tube closed at one end, caused by a tuning fork. If the length of the tube L is just right, the disturbance gets back to the tuning fork half a cycle later and interferes constructively with the continuing sound from the tuning fork. This interference forms a standing wave, and the air column resonates.

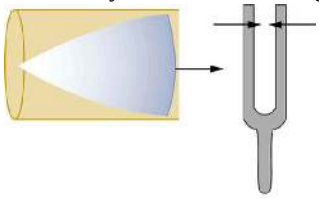


Figure **9.6.6**: Resonance of air in a tube closed at one end, caused by a tuning fork. A graph of air displacement along the length of the tube shows none at the closed end, where the motion is constrained, and a maximum at the open end. This standing wave has one-fourth of its wavelength in the tube, so that $\lambda = 4L$.

The standing wave formed in the tube has its maximum air displacement (an antinode) at the open end, where motion is unconstrained, and no displacement (a node) at the closed end, where air movement is halted. The distance from a node to an antinode is one-fourth of a wavelength, and this equals the length of the tube; thus, $\lambda = 4L$. This same resonance can be produced by a vibration introduced at or near the closed end of the tube, as shown in Figure **9.6.7**. It is best to consider this a natural vibration of the air column independently of how it is induced.

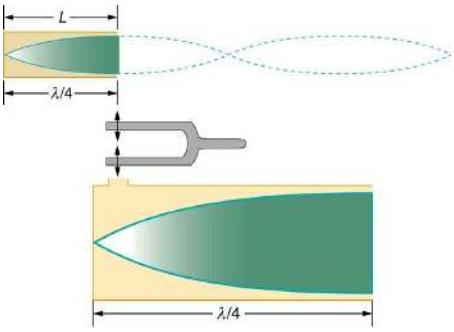


Figure 9.6.7: The same standing wave is created in the tube by a vibration introduced near its closed end.

Given that maximum air displacements are possible at the open end and none at the closed end, there are other, shorter wavelengths that can resonate in the tube, such as the one shown in Figure 9.6.8. Here the standing wave has three-fourths of its wavelength in the tube, or $L = (3/4)\lambda'$, so that $\lambda' = 4L/3$. Continuing this process reveals a whole series of shorter-wavelength and higher-frequency sounds that resonate in the tube. We use specific terms for the resonances in any system. The lowest resonant frequency is called the **fundamental**, while all higher resonant frequencies are called **overtones**. All resonant frequencies are integral multiples of the fundamental, and they are collectively called **harmonics**. The fundamental is the first harmonic, the first overtone is the second harmonic, and so on. Figure 9.6.9shows the fundamental and the first three overtones (the first four harmonics) in a tube closed at one end.

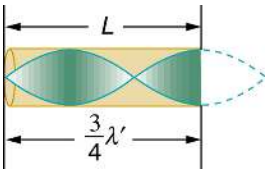


Figure 9.6.8: Another resonance for a tube closed at one end. This has maximum air displacements at the open end, and none at the closed end. The wavelength is shorter, with three-fourths λ' equaling the length of the tube, so that $\lambda' = 4L/3$. This higher-frequency vibration is the first overtone.

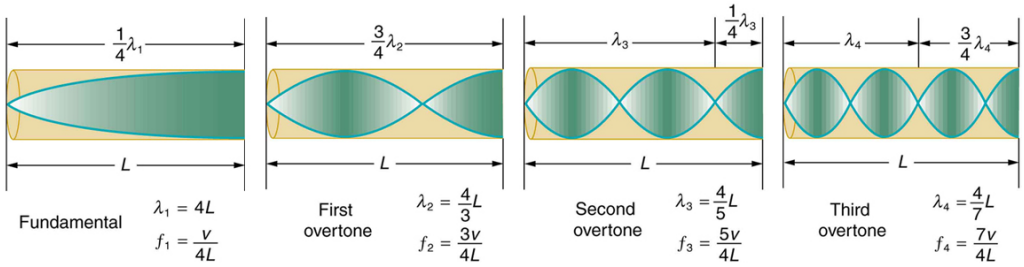


Figure 9.6.9: The fundamental and three lowest overtones for a tube closed at one end. All have maximum air displacements at the open end and none at the closed end.

The fundamental and overtones can be present simultaneously in a variety of combinations. For example, middle C on a trumpet has a sound distinctively different from middle C on a clarinet, both instruments being modified versions of a tube closed at one end. The fundamental frequency is the same (and usually the most intense), but the overtones and their mix of intensities are different and subject to shading by the musician. This mix is what gives various musical instruments (and human voices) their distinctive characteristics, whether they have air columns, strings, sounding boxes, or drumheads. In fact, much of our speech is determined by shaping the cavity formed by the throat and mouth and positioning the tongue to adjust the fundamental and combination of overtones. Simple resonant cavities can be made to resonate with the sound of the vowels, for example. (See Figure 9.6.10.) In boys, at puberty, the larynx grows and the shape of the resonant cavity changes giving rise to the difference in predominant frequencies in speech between men and women.

https://chem.libretexts.org/@go/page/200177

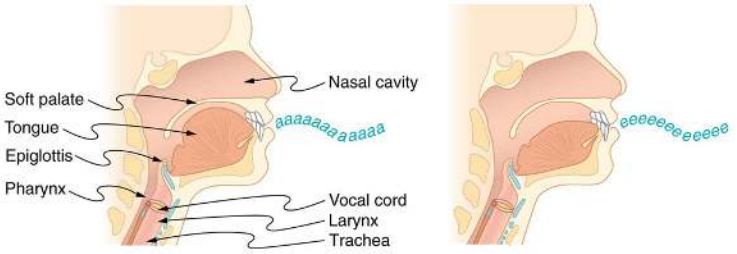


Figure 9.6.10: The throat and mouth form an air column closed at one end that resonates in response to vibrations in the voice box. The spectrum of overtones and their intensities vary with mouth shaping and tongue position to form different sounds. The voice box can be replaced with a mechanical vibrator, and understandable speech is still possible. Variations in basic shapes make different voices recognizable.

Now let us look for a pattern in the resonant frequencies for a simple tube that is closed at one end. The fundamental has $\lambda = 4L$, and frequency is related to wavelength and the speed of sound as given by:

$$v_w = f\lambda. \tag{9.6.1}$$

Solving for f in this equation gives

$$f = \frac{v_w}{\lambda} = \frac{v_w}{4L}, \tag{9.6.2}$$

where v_w is the speed of sound in air. Similarly, the first overtone has $\lambda' = 4L/3$ (see Figure 9.6.9), so that

$$f' = 3\frac{v_w}{4L} = 3f. \tag{9.6.3}$$

Because $f' = 3f$ we call the first overtone the third harmonic. Continuing this process, we see a pattern that can be generalized in a single expression. The resonant frequencies of a tube closed at one end are

$$f_n = n\frac{v_w}{4L},\ n = 1,\ 3,\ 5, \tag{9.6.4}$$

where f_1 is the fundamental, f_3 is the first overtone, and so on. It is interesting that the resonant frequencies depend on the speed of sound and, hence, on temperature. This dependence poses a noticeable problem for organs in old unheated cathedrals, and it is also the reason why musicians commonly bring their wind instruments to room temperature before playing them.

Example 9.6.1: Find the Length of a Tube with a 128 Hz Fundamental

a. What length should a tube closed at one end have on a day when the air temperature, is $22°C$, if its fundamental frequency is to be 128 Hz (C below middle C)?
b. What is the frequency of its fourth overtone?

Strategy

The length L can be found from the relationship in $f_n = n\frac{v_w}{4L}$, but we will first need to find the speed of sound v_w.

Solution for (a)

(1) Identify knowns:

- the fundamental frequency is 128 Hz
- the air temperature is $22°C$

(2) Use $f_n = n\frac{v_w}{4L}$ to find the fundamental frequency (n = 1).

$$f_1 = \frac{v_w}{4L} \tag{9.6.5}$$

(3) Solve this equation for length.

$$L = \frac{v_w}{4f_1} \tag{9.6.6}$$

(4) Find the speed of sound using $v_w = (331\ m/s)\sqrt{\frac{T}{273\ K}}$.

$$v_w = (331\,m/s)\sqrt{\frac{295\,K}{273\,K}} = 344\,m/s \tag{9.6.7}$$

(5) Enter the values of the speed of sound and frequency into the expression for L.

$$L = \frac{v_w}{4f_1} = \frac{344\,m/s}{4(128\,Hz)} = 0.672\,m \tag{9.6.8}$$

Discussion on (a)

Many wind instruments are modified tubes that have finger holes, valves, and other devices for changing the length of the resonating air column and hence, the frequency of the note played. Horns producing very low frequencies, such as tubas, require tubes so long that they are coiled into loops.

Solution for (b)

(1) Identify knowns:

- the first overtone has $n = 3$
- the second overtone has $n = 5$
- the third overtone has $n = 7$
- the fourth overtone has $n = 9$

(2) Enter the value for the fourth overtone into $f_n = n\frac{v_w}{4L}$.

$$f_9 = 9\frac{v_w}{4L} = 9f_1 = 1.15\,kHz \tag{9.6.9}$$

Discussion on (b)

Whether this overtone occurs in a simple tube or a musical instrument depends on how it is stimulated to vibrate and the details of its shape. The trombone, for example, does not produce its fundamental frequency and only makes overtones.

Another type of tube is one that is *open* at both ends. Examples are some organ pipes, flutes, and oboes. The resonances of tubes open at both ends can be analyzed in a very similar fashion to those for tubes closed at one end. The air columns in tubes open at both ends have maximum air displacements at both ends, as illustrated in Figure **9.6.11**. Standing waves form as shown.

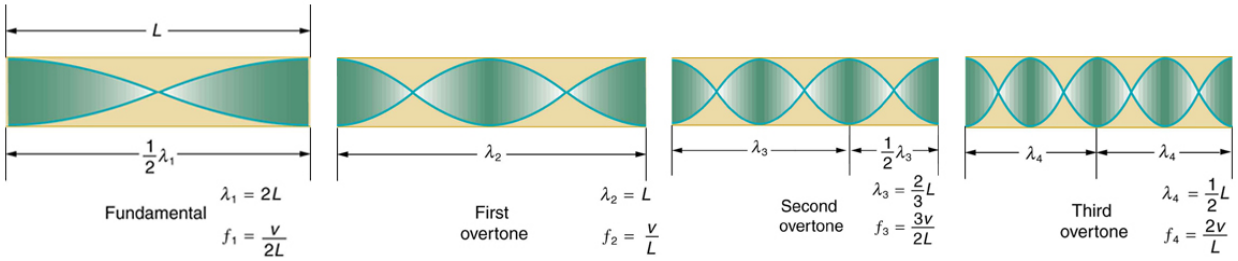


Figure **9.6.11**: The resonant frequencies of a tube open at both ends are shown, including the fundamental and the first three overtones. In all cases the maximum air displacements occur at both ends of the tube, giving it different natural frequencies than a tube closed at one end.

Based on the fact that a tube open at both ends has maximum air displacements at both ends, and using Figure **9.6.11** as a guide, we can see that the resonant frequencies of a tube open at both ends are:

$$f_n = n\frac{v_w}{2L},\ n = 1,\ 2,\ 3..., \tag{9.6.10}$$

where f_1 is fundamental, f_2 is the overtone, f_3 is the second overtone, and so on. Note that a tube open at both ends has a fundamental frequency twice what it would have if closed at one end. It also has a different spectrum of overtones than a tube closed at one end. So if you had two tubes with the same fundamental frequency but one was open at both ends and the other was closed at one end, they would sound different when played because they have different overtones. Middle C, for example, would sound richer played on an open tube, because it has even multiples of the fundamental as well as odd. A closed tube has only odd multiples.

REAL WORLD APPLICATIONS: RESONANCE IN EVERYDAY

Resonance occurs in many different systems, including strings, air columns, and atoms. Resonance is the driven or forced oscillation of a system at its natural frequency. At resonance, energy is transferred rapidly to the oscillating system, and the amplitude of its oscillations grows until the system can no longer be described by Hooke's law. An example of this is the distorted sound intentionally produced in certain types of rock music.

Wind instruments use resonance in air columns to amplify tones made by lips or vibrating reeds. Other instruments also use air resonance in clever ways to amplify sound. Figure **9.6.12**shows a violin and a guitar, both of which have sounding boxes but with different shapes, resulting in different overtone structures. The vibrating string creates a sound that resonates in the sounding box, greatly amplifying the sound and creating overtones that give the instrument its characteristic flavor. The more complex the shape of the sounding box, the greater its ability to resonate over a wide range of frequencies. The marimba, like the one shown in Figure **9.6.13**uses pots or gourds below the wooden slats to amplify their tones. The resonance of the pot can be adjusted by adding water.



Figure **9.6.12**: String instruments such as violins and guitars use resonance in their sounding boxes to amplify and enrich the sound created by their vibrating strings. The bridge and supports couple the string vibrations to the sounding boxes and air within. (credits: guitar, Feliciano Guimares, Fotopedia; violin, Steve Snodgrass, Flickr)



Figure **9.6.13**: Resonance has been used in musical instruments since prehistoric times. This marimba uses gourds as resonance chambers to amplify its sound. (credit: APC Events, Flickr)

We have emphasized sound applications in our discussions of resonance and standing waves, but these ideas apply to any system that has wave characteristics. Vibrating strings, for example, are actually resonating and have fundamentals and overtones similar to those for air columns. More subtle are the resonances in atoms due to the wave character of their electrons. Their orbitals can be viewed as standing waves, which have a fundamental (ground state) and overtones (excited states). It is fascinating that wave characteristics apply to such a wide range of physical systems.

Exercise **9.6.1**

Describe how noise-canceling headphones differ from standard headphones used to block outside sounds.

Answer

Regular headphones only block sound waves with a physical barrier. Noise-canceling headphones use destructive interference to reduce the loudness of outside sounds.

Exercise **9.6.2**

How is it possible to use a standing wave's node and antinode to determine the length of a closed-end tube?

Answer

When the tube resonates at its natural frequency, the wave's node is located at the closed end of the tube, and the antinode is located at the open end. The length of the tube is equal to one-fourth of the wavelength of this wave. Thus, if we know the

wavelength of the wave, we can determine the length of the tube.

PHET EXPLORATIONS: SOUND

This simulation lets you see sound waves. Adjust the frequency or volume and you can see and hear how the wave changes. Move the listener around and hear what she hears.

Summary

- Sound interference and resonance have the same properties as defined for all waves.
- In air columns, the lowest-frequency resonance is called the fundamental, whereas all higher resonant frequencies are called overtones. Collectively, they are called harmonics.
- The resonant frequencies of a tube closed at one end are: $f_n = n\frac{v_w}{4L}$, $n = 1, 3, 5...$, f_1 is fundamental and L is the length of the tube.
- The resonant frequencies of a tube open at both ends are: $f_n = n\frac{v_w}{2L}$, $n = 1, 2, 3...$

Glossary

antinode
point of maximum displacement

node
point of zero displacement

fundamental
the lowest-frequency resonance

overtones
all resonant frequencies higher than the fundamental

harmonics
the term used to refer collectively to the fundamental and its overtones

This page titled 9.6: Sound Interference and Resonance- Standing Waves in Air Columns is shared under a CC BY license and was authored, remixed, and/or curated by OpenStax.

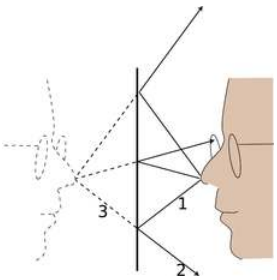
- **17.5: Sound Interference and Resonance- Standing Waves in Air Columns** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

9.7: Images by Reflection

Infants are always fascinated by the antics of the Baby in the Mirror. Now if you want to know something about mirror images that most people don't understand, try this. First bring this page closer and closer to your eyes, until you can no longer focus on it without straining. Then go in the bathroom and see how close you can get your face to the surface of the mirror before you can no longer easily focus on the image of your own eyes. You will find that the shortest comfortable eye-mirror distance is much less than the shortest comfortable eye-paper distance. This demonstrates that the image of your face in the mirror acts as if it had depth and existed in the space *behind* the mirror. If the image was like a flat picture in a book, then you wouldn't be able to focus on it from such a short distance.

In this chapter we will study the images formed by flat and curved mirrors on a qualitative, conceptual basis. Although this type of image is not as commonly encountered in everyday life as images formed by lenses, images formed by reflection are simpler to understand, so we discuss them first. In section 12.3 we will turn to a more mathematical treatment of images made by reflection. Surprisingly, the same equations can also be applied to lenses, which are the topic of section 12.4.

12.2.1 A virtual image



a / An image formed by a mirror.

We can understand a mirror image using a ray diagram. Figure a shows several light rays, 1, that originated by diffuse reflection at the person's nose. They bounce off the mirror, producing new rays, 2. To anyone whose eye is in the right position to get one of these rays, they appear to have come from a behind the mirror, 3, where they would have originated from a single point. This point is where the tip of the image-person's nose appears to be. A similar analysis applies to every other point on the person's face, so it looks as though there was an entire face behind the mirror. The customary way of describing the situation requires some explanation:

- **Customary description in physics:** There is an image of the face behind the mirror.
- **Translation:** The pattern of rays coming from the mirror is exactly the same as it would be if there were a face behind the mirror. Nothing is really behind the mirror.

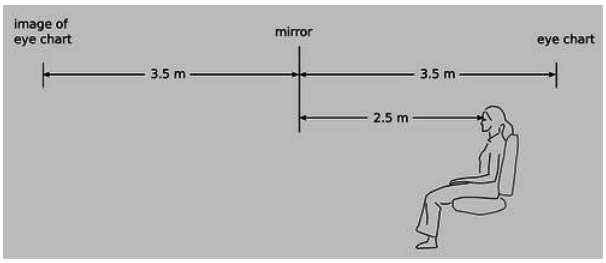
This is referred to as a *virtual* image, because the rays do not actually cross at the point behind the mirror. They only appear to have originated there.

self-check:

Imagine that the person in figure a moves his face down quite a bit --- a couple of feet in real life, or a few inches on this scale drawing. The mirror stays where it is. Draw a new ray diagram. Will there still be an image? If so, where is it visible from?

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

The geometry of specular reflection tells us that rays 1 and 2 are at equal angles to the normal (the imaginary perpendicular line piercing the mirror at the point of reflection). This means that ray 2's imaginary continuation, 3, forms the same angle with the mirror as ray 1. Since each ray of type 3 forms the same angles with the mirror as its partner of type 1, we see that the distance of the image from the mirror is the same as that of the actual face from the mirror, and it lies directly across from it. The image therefore appears to be the same size as the actual face.



b / Example 2.

Example 2: An eye exam

Figure b shows a typical setup in an optometrist's examination room. The patient's vision is supposed to be tested at a distance of 6 meters (20 feet in the U.S.), but this distance is larger than the amount of space available in the room. Therefore a mirror is used to create an image of the eye chart behind the wall.

Example 3: The Praxinoscope

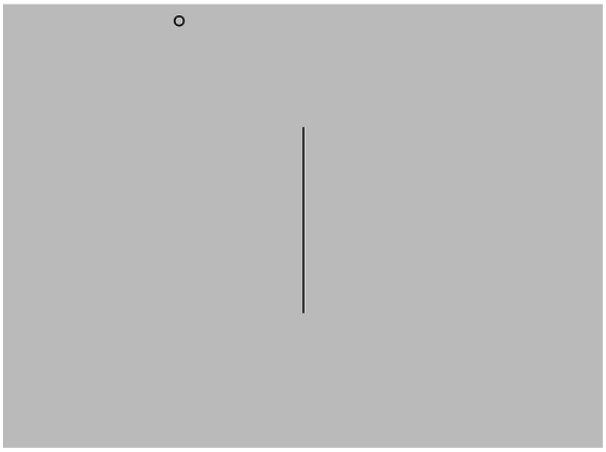


c / The praxinoscope.

Figure c shows an old-fashioned device called a praxinoscope, which displays an animated picture when spun. The removable strip of paper with the pictures printed on it has twice the radius of the inner circle made of flat mirrors, so each picture's virtual image is at the center. As the wheel spins, each picture's image is replaced by the next.

Discussion Question

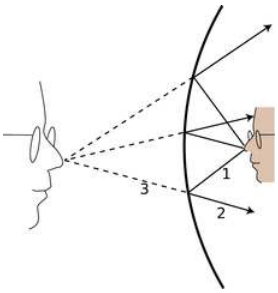
◊ The figure shows an object that is off to one side of a mirror. Draw a ray diagram. Is an image formed? If so, where is it, and from which directions would it be visible?



12.2.2 Curved mirrors

An image in a flat mirror is a pretechnological example: even animals can look at their reflections in a calm pond. We now pass to our first nontrivial example of the manipulation of an image by technology: an image in a curved mirror. Before we dive in, let's consider why this is an important example. If it was just a question of memorizing a bunch of facts about curved mirrors, then you

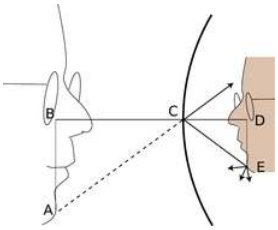
would rightly rebel against an effort to spoil the beauty of your liberally educated brain by force-feeding you technological trivia. The reason this is an important example is not that curved mirrors are so important in and of themselves, but that the results we derive for curved bowl-shaped mirrors turn out to be true for a large class of other optical devices, including mirrors that bulge outward rather than inward, and lenses as well. A microscope or a telescope is simply a combination of lenses or mirrors or both. What you're really learning about here is the basic building block of all optical devices from movie projectors to octopus eyes.



d / An image formed by a curved mirror.

Because the mirror in figure d is curved, it bends the rays back closer together than a flat mirror would: we describe it as *converging*. Note that the term refers to what it does to the light rays, not to the physical shape of the mirror's surface . (The surface itself would be described as *concave*. The term is not all that hard to remember, because the hollowed-out interior of the mirror is like a cave.) It is surprising but true that all the rays like 3 really do converge on a point, forming a good image. We will not prove this fact, but it is true for any mirror whose curvature is gentle enough and that is symmetric with respect to rotation about the perpendicular line passing through its center (not asymmetric like a potato chip). The old-fashioned method of making mirrors and lenses is by grinding them in grit by hand, and this automatically tends to produce an almost perfect spherical surface.

Bending a ray like 2 inward implies bending its imaginary continuation 3 outward, in the same way that raising one end of a seesaw causes the other end to go down. The image therefore forms deeper behind the mirror. This doesn't just show that there is extra distance between the image-nose and the mirror; it also implies that the image itself is bigger from front to back. It has been *magnified* in the front-to-back direction.



e / The image is magnified by the same factor in depth and in its other dimensions.

It is easy to prove that the same magnification also applies to the image's other dimensions. Consider a point like E in figure e. The trick is that out of all the rays diffusely reflected by E, we pick the one that happens to head for the mirror's center, C. The equal-angle property of specular reflection plus a little straightforward geometry easily leads us to the conclusion that triangles ABC and CDE are the same shape, with ABC being simply a scaled-up version of CDE. The magnification of depth equals the ratio BC/CD, and the up-down magnification is AB/DE. A repetition of the same proof shows that the magnification in the third dimension (out of the page) is also the same. This means that the image-head is simply a larger version of the real one, without any distortion. The scaling factor is called the magnification, M. The image in the figure is magnified by a factor $M = 1.9$.

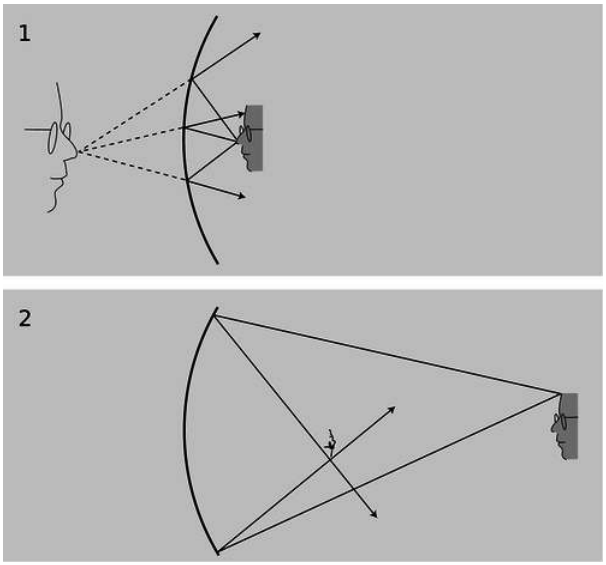


f / Increased magnification always comes at the expense of decreased field of view.

Note that we did not explicitly specify whether the mirror was a sphere, a paraboloid, or some other shape. However, we assumed that a focused image would be formed, which would not necessarily be true, for instance, for a mirror that was asymmetric or very deeply curved.

12.2.3 A real image

If we start by placing an object very close to the mirror, g/1, and then move it farther and farther away, the image at first behaves as we would expect from our everyday experience with flat mirrors, receding deeper and deeper behind the mirror. At a certain point, however, a dramatic change occurs. When the object is more than a certain distance from the mirror, g/2, the image appears upside-down and in *front* of the mirror.



g / 1. A virtual image. 2. A real image. As you'll verify in homework problem 12, the image is upside-down

Here's what's happened. The mirror bends light rays inward, but when the object is very close to it, as in g/1, the rays coming from a given point on the object are too strongly diverging (spreading) for the mirror to bring them back together. On reflection, the rays are still diverging, just not as strongly diverging. But when the object is sufficiently far away, g/2, the mirror is only intercepting the rays that came out in a narrow cone, and it is able to bend these enough so that they will reconverge.

Note that the rays shown in the figure, which both originated at the same point on the object, reunite when they cross. The point where they cross is the image of the point on the original object. This type of image is called a *real image*, in contradistinction to the virtual images we've studied before.

Definition: A real image is one where rays actually cross. A virtual image is a point from which rays only appear to have come.

The use of the word “real” is perhaps unfortunate. It sounds as though we are saying the image was an actual material object, which of course it is not.

The distinction between a real image and a virtual image is an important one, because a real image can be projected onto a screen or photographic film. If a piece of paper is inserted in figure g/2 at the location of the image, the image will be visible on the paper (provided the object is bright and the room is dark). Your eye uses a lens to make a real image on the retina.

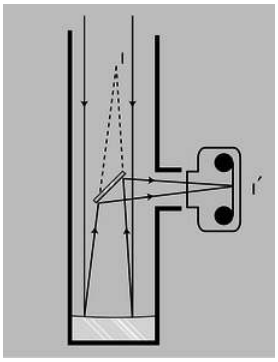
self-check:

Sketch another copy of the face in figure g/1, even farther from the mirror, and draw a ray diagram. What has happened to the location of the image?

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

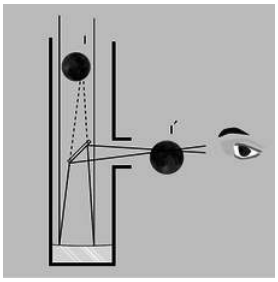
12.2.4 Images of images

If you are wearing glasses right now, then the light rays from the page are being manipulated first by your glasses and then by the lens of your eye. You might think that it would be extremely difficult to analyze this, but in fact it is quite easy. In any series of optical elements (mirrors or lenses or both), each element works on the rays furnished by the previous element in exactly the same manner as if the image formed by the previous element was an actual object.



h / A Newtonian telescope being used with a camera.

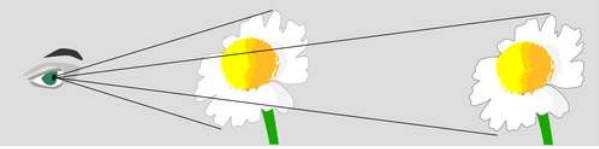
Figure h shows an example involving only mirrors. The Newtonian telescope, invented by Isaac Newton, consists of a large curved mirror, plus a second, flat mirror that brings the light out of the tube. (In very large telescopes, there may be enough room to put a camera or even a person inside the tube, in which case the second mirror is not needed.) The tube of the telescope is not vital; it is mainly a structural element, although it can also be helpful for blocking out stray light. The lens has been removed from the front of the camera body, and is not needed for this setup. Note that the two sample rays have been drawn parallel, because an astronomical telescope is used for viewing objects that are extremely far away. These two “parallel” lines actually meet at a certain point, say a crater on the moon, so they can't actually be perfectly parallel, but they are parallel for all practical purposes since we would have to follow them upward for a quarter of a million miles to get to the point where they intersect.



i / A Newtonian telescope being used for visual rather than photographic observing. In real life, an eyepiece lens is normally used for additional magnification, but this simpler setup will also work.

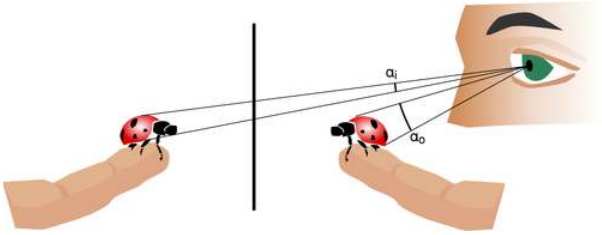
The large curved mirror by itself would form an image $\mathbf{I}$, but the small flat mirror creates an image of the image, $\mathbf{I}'$. The relationship between $\mathbf{I}$ and $\mathbf{I}'$ is exactly the same as it would be if $\mathbf{I}$ was an actual object rather than an image: $\mathbf{I}$ and $\mathbf{I}'$ are at equal distances from the plane of the mirror, and the line between them is perpendicular to the plane of the mirror.

One surprising wrinkle is that whereas a flat mirror used by itself forms a virtual image of an object that is real, here the mirror is forming a real image of virtual image I. This shows how pointless it would be to try to memorize lists of facts about what kinds of images are formed by various optical elements under various circumstances. You are better off simply drawing a ray diagram.



j / The angular size of the flower depends on its distance from the eye.

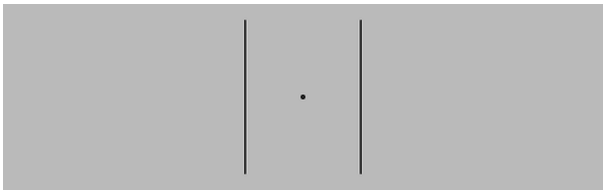
Although the main point here was to give an example of an image of an image, figure i also shows an interesting case where we need to make the distinction between *magnification* and *angular magnification*. If you are looking at the moon through this telescope, then the images I and I′ are much *smaller* than the actual moon. Otherwise, for example, image I would not fit inside the telescope! However, these images are very close to your eye compared to the actual moon. The small size of the image has been more than compensated for by the shorter distance. The important thing here is the amount of *angle* within your field of view that the image covers, and it is this angle that has been increased. The factor by which it is increased is called the *angular magnification*, M_a.



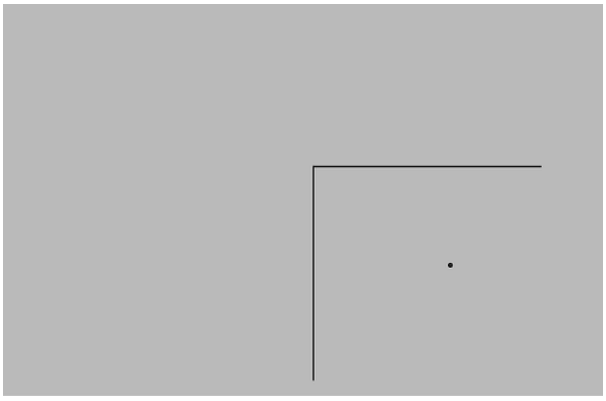
k / The person uses a mirror to get a view of both sides of the ladybug. Although the flat mirror has $M = 1$, it doesn't give an angular magnification of 1. The image is farther from the eye than the object, so the angular magnification $M_a = \alpha_i/\alpha_o$ is less than one.

Discussion Questions

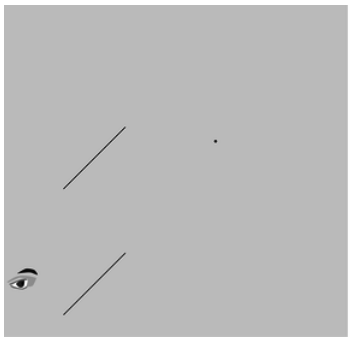
◊ Locate the images of you that will be formed if you stand between two parallel mirrors.



◊ Locate the images formed by two perpendicular mirrors, as in the figure. What happens if the mirrors are not perfectly perpendicular?



◊ Locate the images formed by the periscope.



Contributor

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

This page titled 9.7: Images by Reflection is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

- **13.2: Images by Reflection** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

9.8: Refraction

Economists normally consider free markets to be the natural way of judging the monetary value of something, but social scientists also use questionnaires to gauge the relative value of privileges, disadvantages, or possessions that cannot be bought or sold. They ask people to *imagine* that they could trade one thing for another and ask which they would choose. One interesting result is that the average light-skinned person in the U.S. would rather lose an arm than suffer the racist treatment routinely endured by African-Americans. Even more impressive is the value of sight. Many prospective parents can imagine without too much fear having a deaf child, but would have a far more difficult time coping with raising a blind one.

So great is the value attached to sight that some have imbued it with mystical aspects. Joan of Arc saw visions, and my college has a "vision statement." Christian fundamentalists who perceive a conflict between evolution and their religion have claimed that the eye is such a perfect device that it could never have arisen through a process as helter-skelter as evolution, or that it could not have evolved because half of an eye would be useless. In fact, the structure of an eye is fundamentally dictated by physics, and it has arisen separately by evolution somewhere between eight and 40 times, depending on which biologist you ask. We humans have a version of the eye that can be traced back to the evolution of a light-sensitive "eye spot" on the head of an ancient invertebrate. A sunken pit then developed so that the eye would only receive light from one direction, allowing the organism to tell where the light was coming from. (Modern flatworms have this type of eye.) The top of the pit then became partially covered, leaving a hole, for even greater directionality (as in the nautilus). At some point the cavity became filled with jelly, and this jelly finally became a lens, resulting in the general type of eye that we share with the bony fishes and other vertebrates. Far from being a perfect device, the vertebrate eye is marred by a serious design flaw due to the lack of planning or intelligent design in evolution: the nerve cells of the retina and the blood vessels that serve them are all in front of the light-sensitive cells, blocking part of the light. Squids and other molluscs, whose eyes evolved on a separate branch of the evolutionary tree, have a more sensible arrangement, with the light-sensitive cells out in front.

12.4.1 Refraction

The fundamental physical phenomenon at work in the eye is that when light crosses a boundary between two media (such as air and the eye's jelly), part of its energy is reflected, but part passes into the new medium. In the ray model of light, we describe the original ray as splitting into a reflected ray and a transmitted one (the one that gets through the boundary). Of course the reflected ray goes in a direction that is different from that of the original one, according to the rules of reflection we have already studied. More surprisingly --- and this is the crucial point for making your eye focus light --- the transmitted ray is bent somewhat as well. This bending phenomenon is called *refraction*. The origin of the word is the same as that of the word "fracture," i.e., the ray is bent or "broken." (Keep in mind, however, that light rays are not physical objects that can really be "broken.") Refraction occurs with all waves, not just light waves.

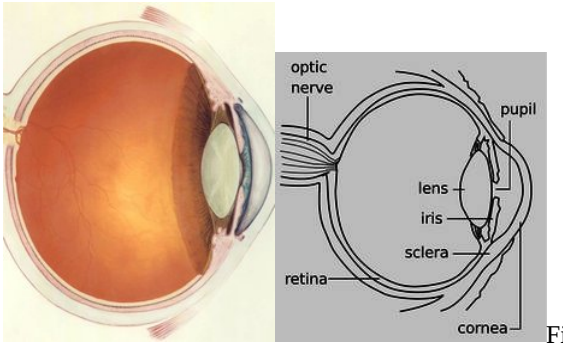
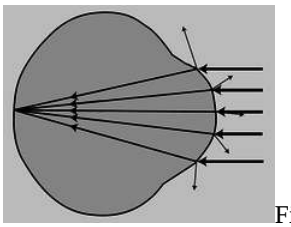


Figure 9.8.2: The anatomy of the eye.

The actual anatomy of the eye, Figure 9.8.2, is quite complex, but in essence it is very much like every other optical device based on refraction. The rays are bent when they pass through the front surface of the eye, Figure 9.8.3. Rays that enter farther from the central axis are bent more, with the result that an image is formed on the retina. There is only one slightly novel aspect of the situation. In most human-built optical devices, such as a movie projector, the light is bent as it passes into a lens, bent again as it reemerges, and then reaches a focus beyond the lens. In the eye, however, the "screen" is inside the eye, so the rays are only refracted once, on entering the jelly, and never emerge again.


Figure 9.8.3: A simplified optical diagram of the eye. Light rays are bent when they cross from the air into the eye. (A little of the incident rays' energy goes into the reflected rays rather than the ones transmitted into the eye.)

A common misconception is that the “lens” of the eye is what does the focusing. All the transparent parts of the eye are made of fairly similar stuff, so the dramatic change in medium is when a ray crosses from the air into the eye (at the outside surface of the cornea). This is where nearly all the refraction takes place. The lens medium differs only slightly in its optical properties from the rest of the eye, so very little refraction occurs as light enters and exits the lens. The lens, whose shape is adjusted by muscles attached to it, is only meant for fine-tuning the focus to form images of near or far objects.

Refractive Properties of Media

What are the rules governing refraction? The first thing to observe is that just as with reflection, the new, bent part of the ray lies in the same plane as the normal (perpendicular) and the incident ray, Figure 9.8.4.

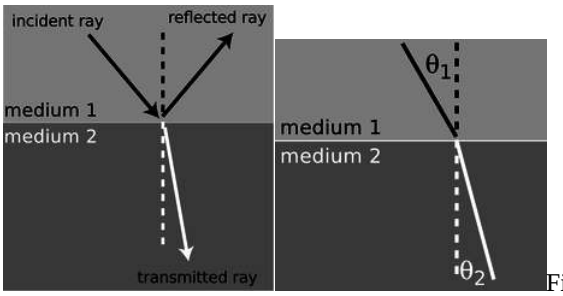


Figure 9.8.5: The angles θ_1 and θ_2 are related to each other, and also depend on the properties of the two media. Because refraction is time-reversal symmetric, there is no need to label the rays with arrowheads.

If you try shooting a beam of light at the boundary between two substances, say water and air, you'll find that regardless of the angle at which you send in the beam, the part of the beam in the water is always closer to the normal line, Figure 9.8.5.

It doesn't matter if the ray is entering the water or leaving, so refraction is symmetric with respect to time-reversal, Figure 9.8.6.

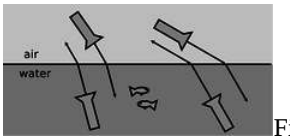


Figure 9.8.6: Refraction has time-reversal symmetry. Regardless of whether the light is going into or out of the water, the relationship between the two angles is the same, and the ray is closer to the normal while in the water.

If, instead of water and air, you try another combination of substances, say plastic and gasoline, again you'll find that the ray's angle with respect to the normal is consistently smaller in one and larger in the other. Also, we find that if substance A has rays closer to normal than in B, and B has rays closer to normal than in C, then A has rays closer to normal than C. This means that we can rank-order all materials according to their refractive properties. Isaac Newton did so, including in his list many amusing substances, such as “Danzig vitriol” and “a pseudo-topazius, being a natural, pellucid, brittle, hairy stone, of a yellow color.” Several general rules can be inferred from such a list:

- Vacuum lies at one end of the list. In refraction across the interface between vacuum and any other medium, the other medium has rays closer to the normal.
- Among gases, the ray gets closer to the normal if you increase the density of the gas by pressurizing it more.
- The refractive properties of liquid mixtures and solutions vary in a smooth and systematic manner as the proportions of the mixture are changed.
- Denser substances usually, but not always, have rays closer to the normal.

The second and third rules provide us with a method for measuring the density of an unknown sample of gas, or the concentration of a solution. The latter technique is very commonly used, and the CRC Handbook of Physics and Chemistry, for instance, contains

extensive tables of the refractive properties of sugar solutions, cat urine, and so on.

Snell's Law

The numerical rule governing refraction was discovered by Snell, who must have collected experimental data something like what is shown on this graph and then attempted by trial and error to find the right equation. The equation he came up with was

$$\frac{\sin\theta_1}{\sin\theta_2} = \text{constant}. \tag{9.8.1}$$

The value of the constant would depend on the combination of media used. For instance, any one of the data points in the graph would have sufficed to show that the constant was 1.3 for an air-water interface (taking air to be substance 1 and water to be substance 2).

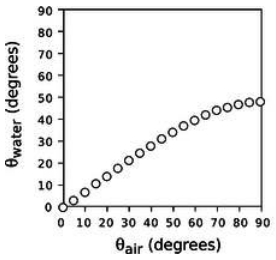


Figure 9.8.7: The relationship between the angles in refraction.

Snell further found that if media A and B gave a constant K_{AB} and media B and C gave a constant K_{BC}, then refraction at an interface between A and C would be described by a constant equal to the product, $K_{AC} = K_{AB}K_{BC}$. This is exactly what one would expect if the constant depended on the ratio of some number characterizing one medium to the number characteristic of the second medium. This number is called the *index of refraction* of the medium, written as n in equations. Since measuring the angles would only allow him to determine the *ratio* of the indices of refraction of two media, Snell had to pick some medium and define it as having $n = 1$. He chose to define vacuum as having $n = 1$. (The index of refraction of air at normal atmospheric pressure is 1.0003, so for most purposes it is a good approximation to assume that air has $n = 1$.) He also had to decide which way to define the ratio, and he chose to define it so that media with their rays closer to the normal would have larger indices of refraction. This had the advantage that denser media would typically have higher indices of refraction, and for this reason the index of refraction is also referred to as the optical density. Written in terms of indices of refraction, Snell's equation becomes

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}, \tag{9.8.2}$$

but rewriting it in the form

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{9.8.3}$$

[relationship between angles of rays at the interface between media with indices of refraction n_1 and n_2; angles are defined with respect to the normal] makes us less likely to get the 1's and 2's mixed up, so this the way most people remember Snell's law. A few indices of refraction are given in the back of the book.

? Exercise 9.8.1

1. What would the graph look like for two substances with the same index of refraction?
2. Based on the graph, when does refraction at an air-water interface change the direction of a ray most strongly?

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

✓ Example 9.8.1: Finding an angle using Snell's law

A submarine shines its searchlight up toward the surface of the water. What is the angle α shown in Figure 9.8.8?

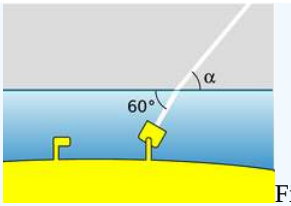


Figure 9.8.8: Example 10.

Solution

The tricky part is that Snell's law refers to the angles with *respect to the normal*. Forgetting this is a very common mistake. The beam is at an angle of 30° with respect to the normal in the water. Let's refer to the air as medium 1 and the water as 2. Solving Snell's law for θ_1, we find

$$\theta_1 = \sin^{-1}\left(\frac{n_2}{n_1}\sin\theta_2\right). \tag{9.8.4}$$

As mentioned above, air has an index of refraction very close to 1, and water's is about 1.3, so we find $\theta_1 = 40°$. The angle α is therefore 50°.

What neither Snell nor Newton knew was that there is a very simple interpretation of the index of refraction. This may come as a relief to the reader who is taken aback by the complex reasoning involving proportionalities that led to its definition. Later experiments showed that the index of refraction of a medium was inversely proportional to the speed of light in that medium. Since c is defined as the speed of light in vacuum, and $n = 1$ is defined as the index of refraction of vacuum, we have

$$n = \frac{c}{v}. \tag{9.8.5}$$

[n = medium's index of refraction, v = speed of light in that medium, c = speed of light in a vacuum]

Many textbooks start with this as the definition of the index of refraction, although that approach makes the quantity's name somewhat of a mystery, and leaves students wondering why c/v was used rather than v/c. It should also be noted that measuring angles of refraction is a far more practical method for determining n than direct measurement of the speed of light in the substance of interest.

A mechanical model of Snell's law

Why should refraction be related to the speed of light? The mechanical model shown in the figure may help to make this more plausible. Suppose medium 2 is thick, sticky mud, which slows down the car. The car's right wheel hits the mud first, causing the right side of the car to slow down. This will cause the car to turn to the right until is moves far enough forward for the left wheel to cross into the mud. After that, the two sides of the car will once again be moving at the same speed, and the car will go straight.

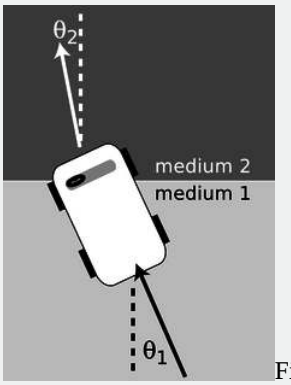


Figure 9.8.8: A mechanical model of refraction.

Of course, light isn't a car. Why should a beam of light have anything resembling a “left wheel” and “right wheel?” After all, the mechanical model would predict that a motorcycle would go straight, and a motorcycle seems like a better approximation to a ray of light than a car. The whole thing is just a model, not a description of physical reality.

A Derivation of Snell's law

However intuitively appealing the mechanical model may be, light is a wave, and we should be using wave models to describe refraction. In fact Snell's law can be derived quite simply from wave concepts. Figure **9.8.9** shows the refraction of a water wave. The water in the upper left part of the tank is shallower, so the speed of the waves is slower there, and their wavelengths is shorter. The reflected part of the wave is also very faintly visible.

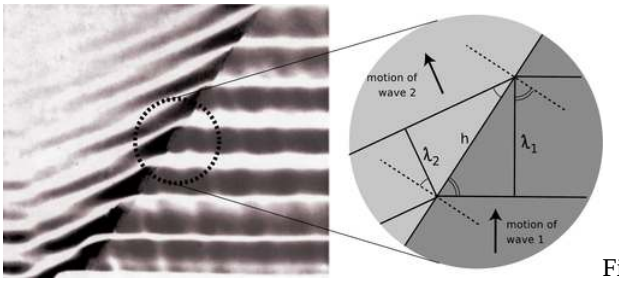


Figure **9.8.9**: A derivation of Snell's law.

In the close-up view on the right, the dashed lines are normals to the interface. The two marked angles on the right side are both equal to θ_1, and the two on the left to θ_2.

Trigonometry gives

$$\sin\theta_1 = \lambda_1/h \text{and}$$
$$\sin\theta_2 = \lambda_2/h.$$

Eliminating h by dividing the equations, we find

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{\lambda_1}{\lambda_2}. \tag{9.8.6}$$

The frequencies of the two waves must be equal or else they would get out of step, so by $v = f\lambda$ we know that their wavelengths are proportional to their velocities. Combining $\lambda \propto v$ with $v \propto 1/n$ gives $\lambda \propto 1/n$, so we find

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}, \tag{9.8.7}$$

which is one form of Snell's law.

Example 9.8.1: Ocean waves near and far from shore

Ocean waves are formed by winds, typically on the open sea, and the wavefronts are perpendicular to the direction of the wind that formed them. At the beach, however, you have undoubtedly observed that waves tend come in with their wavefronts very nearly (but not exactly) parallel to the shoreline. This is because the speed of water waves in shallow water depends on depth: the shallower the water, the slower the wave. Although the change from the fast-wave region to the slow-wave region is gradual rather than abrupt, there is still refraction, and the wave motion is nearly perpendicular to the normal in the slow region.

Color and Refraction

In general, the speed of light in a medium depends both on the medium and on the wavelength of the light. Another way of saying it is that a medium's index of refraction varies with wavelength. This is why a prism can be used to split up a beam of white light into a rainbow. Each wavelength of light is refracted through a different angle.

How much light is reflected, and how much is transmitted?

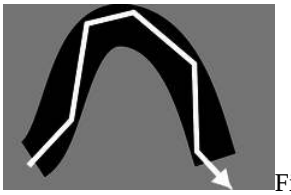
In section 6.2 we developed an equation for the percentage of the wave energy that is transmitted and the percentage reflected at a boundary between media. This was only done in the case of waves in one dimension, however, and rather than discuss the full three dimensional generalization it will be more useful to go into some qualitative observations about what happens. First, reflection happens only at the interface between two media, and two media with the same index of refraction act as if they were a single medium. Thus, at the interface between media with the same index of refraction, there is no reflection, and the ray keeps going

straight. Continuing this line of thought, it is not surprising that we observe very little reflection at an interface between media with similar indices of refraction.

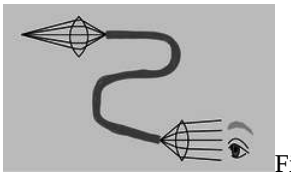
The next thing to note is that it is possible to have situations where no possible angle for the refracted ray can satisfy Snell's law. Solving Snell's law for θ_2, we find

$$\theta_2 = \sin^{-1}\left(\frac{n_1}{n_2}\sin\theta_1\right), \tag{9.8.8}$$

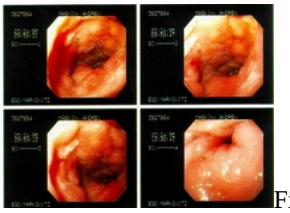
and if n_1 is greater than n_2, then there will be large values of θ_1 for which the quantity $(n_1/n_2)\sin\theta$ is greater than one, meaning that your calculator will flash an error message at you when you try to take the inverse sine. What can happen physically in such a situation? The answer is that all the light is reflected, so there is no refracted ray. This phenomenon is known as *total internal reflection*, and is used in the fiber-optic cables that nowadays carry almost all long-distance telephone calls.


Figure 9.8.10: Total internal reflection in a fiber-optic cable.

The electrical signals from your phone travel to a switching center, where they are converted from electricity into light. From there, the light is sent across the country in a thin transparent fiber. The light is aimed straight into the end of the fiber, and as long as the fiber never goes through any turns that are too sharp, the light will always encounter the edge of the fiber at an angle sufficiently oblique to give total internal reflection. If the fiber-optic cable is thick enough, one can see an image at one end of whatever the other end is pointed at.


Figure 9.8.11: A simplified drawing of a surgical endoscope. The first lens forms a real image at one end of a bundle of optical fibers. The light is transmitted through the bundle, and is finally magnified by the eyepiece.

Alternatively, a bundle of cables can be used, since a single thick cable is too hard to bend. This technique for seeing around corners is useful for making surgery less traumatic. Instead of cutting a person wide open, a surgeon can make a small “keyhole” incision and insert a bundle of fiber-optic cable (known as an endoscope) into the body.


Figure 9.8.12: Endoscopic images of a duodenal ulcer.

Since rays at sufficiently large angles with respect to the normal may be completely reflected, it is not surprising that the relative amount of reflection changes depending on the angle of incidence, and is greatest for large angles of incidence.

Discussion Questions

◊ What index of refraction should a fish have in order to be invisible to other fish?

◊ Does a surgeon using an endoscope need a source of light inside the body cavity? If so, how could this be done without inserting a light bulb through the incision?

◊ A denser sample of a gas has a higher index of refraction than a less dense sample (i.e., a sample under lower pressure), but why would it not make sense for the index of refraction of a gas to be proportional to density?

◊ The earth's atmosphere gets thinner and thinner as you go higher in altitude. If a ray of light comes from a star that is below the zenith, what will happen to it as it comes into the earth's atmosphere?

◊ Does total internal reflection occur when light in a denser medium encounters a less dense medium, or the other way around? Or can it occur in either case?

12.4.2 Lenses

Figures n/1 and n/2 show examples of lenses forming images. There is essentially nothing for you to learn about imaging with lenses that is truly new. You already know how to construct and use ray diagrams, and you know about real and virtual images. The concept of the focal length of a lens is the same as for a curved mirror. The equations for locating images and determining magnifications are of the same form. It's really just a question of flexing your mental muscles on a few examples. The following self-checks and discussion questions will get you started.

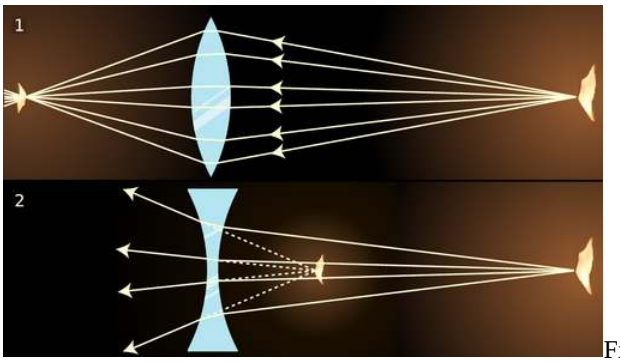


Figure **9.8.13**: 1. A converging lens forms an image of a candle flame. 2. A diverging lens.

? Exercise 9.8.1

1. In figures Figure **9.8.13;** and Figure **9.8.13;**, classify the images as real or virtual.
2. Glass has an index of refraction that is greater than that of air. Consider the topmost ray in Figure **9.8.13;**. Explain why the ray makes a slight left turn upon entering the lens, and another left turn when it exits.
3. If the flame in Figure **9.8.13;** were moved closer to the lens, what would happen to the location of the image?

Answer

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

Discussion Questions

◊ In figures n/1 and n/2, the front and back surfaces are parallel to each other at the center of the lens. What will happen to a ray that enters near the center, but not necessarily along the axis of the lens? Draw a BIG ray diagram, and show a ray that comes from off axis.

In discussion questions B-F, don't draw ultra-detailed ray diagrams as in A.

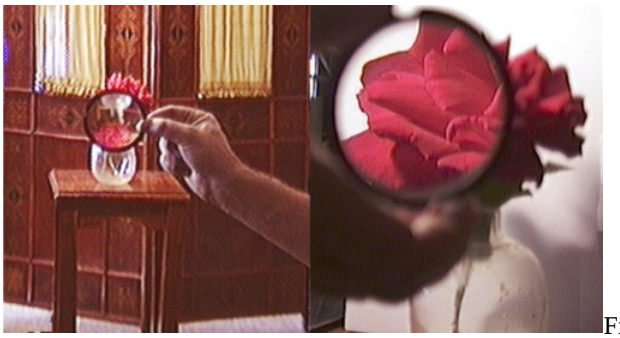
◊ Suppose you wanted to change the setup in figure n/1 so that the location of the actual flame in the figure would instead be occupied by an image of a flame. Where would you have to move the candle to achieve this? What about in n/2?

◊ There are three qualitatively different types of image formation that can occur with lenses, of which figures n/1 and n/2 exhaust only two. Figure out what the third possibility is. Which of the three possibilities can result in a magnification greater than one? Cf. problem 10, p. 797.

◊ Classify the examples shown in figure o according to the types of images delineated in discussion question C.

◊ In figures n/1 and n/2, the only rays drawn were those that happened to enter the lenses. Discuss this in relation to figure o.

◊ In the right-hand side of figure o, the image viewed through the lens is in focus, but the side of the rose that sticks out from behind the lens is not. Why?


Figure 9.8.14: Two images of a rose created by the same lens and recorded with the same camera.

The Lensmaker's Equation

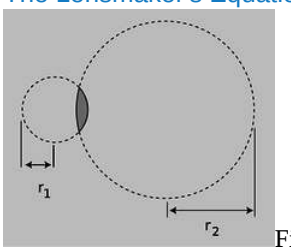


Figure 9.8.15: The radii of curvature appearing in the lensmaker's equation.

The focal length of a spherical mirror is simply $r/2$, but we cannot expect the focal length of a lens to be given by pure geometry, since it also depends on the index of refraction of the lens. Suppose we have a lens whose front and back surfaces are both spherical. (This is no great loss of generality, since any surface with a sufficiently shallow curvature can be approximated with a sphere.) Then if the lens is immersed in a medium with an index of refraction of 1, its focal length is given approximately by

$$f = \frac{1}{(n-1)\left|\frac{1}{r_1} \pm \frac{1}{r_2}\right|} \tag{9.8.9}$$

where n is the index of refraction and r_1 and r_2 are the radii of curvature of the two surfaces of the lens. Equation 9.8.9 is known as the *lensmaker's equation*. In my opinion it is not particularly worthy of memorization. The positive sign is used when both surfaces are curved outward or both are curved inward; otherwise a negative sign applies. The proof of this equation is left as an exercise to those readers who are sufficiently brave and motivated.

12.4.3 Dispersion

For most materials, we observe that the index of refraction depends slightly on wavelength, being highest at the blue end of the visible spectrum and lowest at the red. For example, white light disperses into a rainbow when it passes through a prism, q.

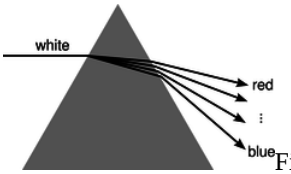


Figure 9.8.16: Dispersion of white light by a prism. White light is a mixture of all the wavelengths of the visible spectrum. Waves of different wavelengths undergo different amounts of refraction.

Even when the waves involved aren't light waves, and even when refraction isn't of interest, the dependence of wave speed on wavelength is referred to as dispersion. Dispersion inside spherical raindrops is responsible for the creation of rainbows in the sky, and in an optical instrument such as the eye or a camera it is responsible for a type of aberration called chromatic aberration (subsection 12.3.3 and problem 28). As we'll see in subsection 13.3.2, dispersion causes a wave that is not a pure sine wave to have its shape distorted as it travels, and also causes the speed at which energy and information are transported by the wave to be different from what one might expect from a naive calculation. The microscopic reasons for dispersion of light in matter are discussed in optional subsection 12.4.6.

The principle of least time for refraction

We have seen previously how the rules governing straight-line motion of light and reflection of light can be derived from the principle of least time. What about refraction? In the figure, it is indeed plausible that the bending of the ray serves to minimize the time required to get from a point A to point B. If the ray followed the unbent path shown with a dashed line, it would have to travel a longer distance in the medium in which its speed is slower. By bending the correct amount, it can reduce the distance it has to cover in the slower medium without going too far out of its way. It is true that Snell's law gives exactly the set of angles that minimizes the time required for light to get from one point to another. The proof of this fact is left as an exercise (problem 38, p. 802).

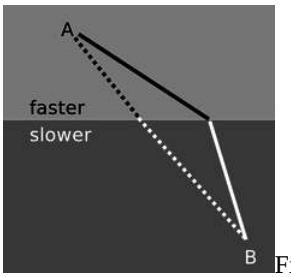


Figure 9.8.17: The principle of least time applied to refraction.

Microscopic description of refraction

Given that the speed of light is different in different media, we've seen two different explanations (on p. 774 and in subsection 12.4.5 above) of why refraction must occur. What we haven't yet explained is why the speed of light does depend on the medium.

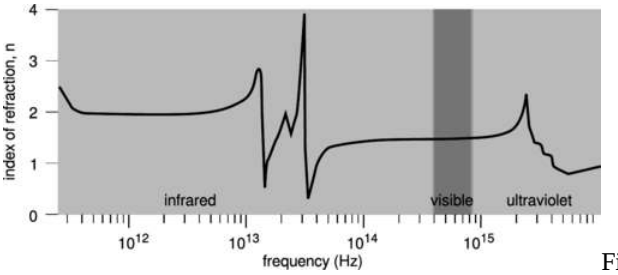


Figure 9.8.18: Index of refraction of silica glass, redrawn from Kitamura, Pilon, and Jonasz, Applied Optics 46 (2007) 8118, reprinted online at http://www.seas.ucla.edu/~pilon/Publ...s/AO2007-1.pdf.

A good clue as to what's going on comes from the figure s. The relatively minor variation of the index of refraction within the visible spectrum was misleading. At certain specific frequencies, n exhibits wild swings in the positive and negative directions. After each such swing, we reach a new, lower plateau on the graph. These frequencies are resonances. For example, the visible part of the spectrum lies on the left-hand tail of a resonance at about 2×10^{15} Hz, corresponding to the ultraviolet part of the spectrum. This resonance arises from the vibration of the electrons, which are bound to the nuclei as if by little springs. Because this resonance is narrow, the effect on visible-light frequencies is relatively small, but it is stronger at the blue end of the spectrum than at the red end. Near each resonance, not only does the index of refraction fluctuate wildly, but the glass becomes nearly opaque; this is because the vibration becomes very strong, causing energy to be dissipated as heat. The “staircase” effect is the same one visible in any resonance, e.g., figure k on p. 180: oscillators have a finite response for $f \ll f_0$, but the response approaches zero for $f \gg f_0$.

So far, we have a qualitative explanation of the frequency-variation of the loosely defined “strength” of the glass's effect on a light wave, but we haven't explained why the effect is observed as a change in speed, or why each resonance is an up-down swing rather than a single positive peak. To understand these effects in more detail, we need to consider the phase response of the oscillator. As shown in the bottom panel of figure j on p. 181, the phase response reverses itself as we pass through a resonance.

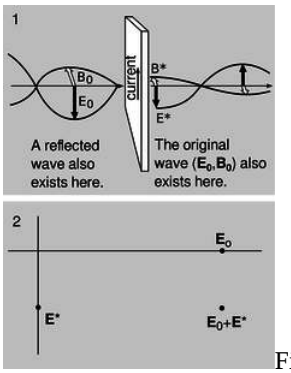


Figure 9.8.19: 1. A wave incident on a sheet of glass excites current in the glass, which produce a secondary wave. 2. The secondary wave superposes with the original wave, as represented in the complex-number representation introduced in subsection 10.5.7.

Suppose that a plane wave is normally incident on the left side of a thin sheet of glass, t/1, at $f \ll f_0$. The light wave observed on the right side consists of a superposition of the incident wave consisting of $\mathbf{E_0}$ and $\mathbf{B_0}$ with a secondary wave $\mathbf{E^*}$ and $\mathbf{B^*}$ generated by the oscillating charges in the glass. Since the frequency is far below resonance, the response $q\mathbf{x}$ of a vibrating charge q is in phase with the driving force $\mathbf{E_0}$. The current is the derivative of this quantity, and therefore 90 degrees ahead of it in phase. The magnetic field generated by a sheet of current has been analyzed in subsection 11.2.1, and the result, shown in figure e on p. 664, is just what we would expect from the right-hand rule. We find, t/1, that the secondary wave is 90 degrees ahead of the incident one in phase. The incident wave still exists on the right side of the sheet, but it is superposed with the secondary one. Their addition is shown in t/2 using the complex number representation introduced in subsection 10.5.7. The superposition of the two fields lags behind the incident wave, which is the effect we would expect if the wave had traveled more slowly through the glass.

In the case $f \gg 0$, the same analysis applies except that the phase of the secondary wave is reversed. The transmitted wave is advanced rather than retarded in phase. This explains the dip observed in figure s after each spike.

All of this is in accord with our understanding of relativity, ch. 7, in which we saw that the universal speed c was to be understood fundamentally as a conversion factor between the units used to measure time and space --- not as the speed of light. Since c isn't defined as the speed of light, it's of no fundamental importance whether light has a different speed in matter than it does in vacuum. In fact, the picture we've built up here is one in which all of our electromagnetic waves travel at c; propagation at some other speed is only what appears to happen because of the superposition of the $(\mathbf{E_0}, \mathbf{B_0})$ and $(\mathbf{E^*}, \mathbf{B^*})$ waves, both of which move at c.

But it is worrisome that at the frequencies where $n < 1$, the speed of the wave is greater than c. According to special relativity, information is never supposed to be transmitted at speeds greater than c, since this would produce situations in which a signal could be received before it was transmitted! This difficulty is resolved in subsection 13.3.2, where we show that there are two different velocities that can be defined for a wave in a dispersive medium, the phase velocity and the group velocity. The group velocity is the velocity at which information is transmitted, and it is always less than c.

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

This page titled 9.8: Refraction is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

- **13.4: Refraction** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

9.9: Wave Optics

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

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

12.5.1 Diffraction

Figure Figure **9.9.1*a*** shows a typical problem in wave optics, enacted with water waves. It may seem surprising that we don't get a simple pattern like Figure **9.9.1*b***, but the pattern would only be that simple if the wavelength was hundreds of times shorter than the distance between the gaps in the barrier and the widths of the gaps.

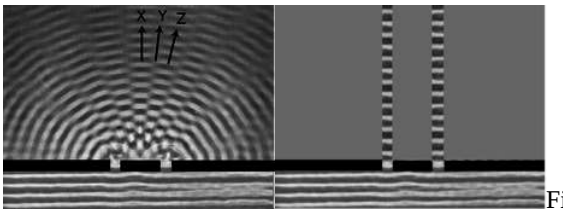


Figure **9.9.1**: a. In this view from overhead, a straight, sinusoidal water wave encounters a barrier with two gaps in it. Strong wave vibration occurs at angles X and Z, but there is none at all at angle Y. (The figure has been retouched from a real photo of water waves. In reality, the waves beyond the barrier would be much weaker than the ones before it, and they would therefore be difficult to see.) b / This doesn't happen.

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

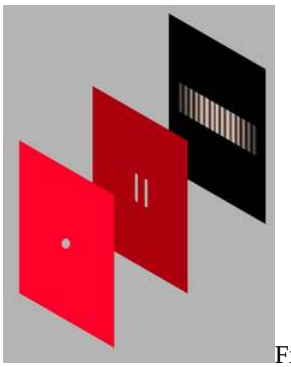
1. We restrict ourselves to cases in which a wave travels through a uniform medium, encounters a certain area in which the medium has different properties, and then emerges on the other side into a second uniform region.
2. We assume that the incoming wave is a nice tidy sine-wave pattern with wavefronts that are lines (or, in three dimensions, planes).
3. In Figure **9.9.1*a*** we can see that the wave pattern immediately beyond the barrier is rather complex, but farther on it sorts itself out into a set of wedges separated by gaps in which the water is still. We will restrict ourselves to studying the simpler wave patterns that occur farther away, so that the main question of interest is how intense the outgoing wave is at a given angle.

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

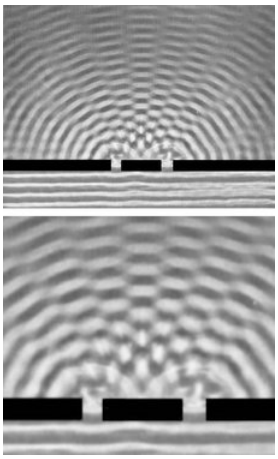
Diffraction can be used to find the structure of an unknown diffracting object: even if the object is too small to study with ordinary imaging, it may be possible to work backward from the diffraction pattern to learn about the object. The structure of a crystal, for example, can be determined from its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

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

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


Figure 9.9.2: A practical, low-tech setup for observing diffraction of light.

A second issue is that light from sources such as the sun or a lightbulb does not consist of a nice neat plane wave, except over very small regions of space. Different parts of the wave are out of step with each other, and the wave is referred to as *incoherent*. One way of dealing with this is shown in Figure 9.9.2. After filtering to select a certain wavelength of red light, we pass the light through a small pinhole. The region of the light that is intercepted by the pinhole is so small that one part of it is not out of step with another. Beyond the pinhole, light spreads out in a spherical wave; this is analogous to what happens when you speak into one end of a paper towel roll and the sound waves spread out in all directions from the other end. By the time the spherical wave gets to the double slit it has spread out and reduced its curvature, so that we can now think of it as a simple plane wave.



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

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

The parts of the final image on the screen in Figure 9.9.2 are called *diffraction fringes*. The center of each fringe is a point of maximum brightness, and halfway between two fringes is a minimum.

? Exercise 9.9.1: Discussion Question

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

12.5.2 Scaling of diffraction

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

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

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

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

12.5.3 The Correspondence Principle

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

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

? Exercise 9.9.1

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

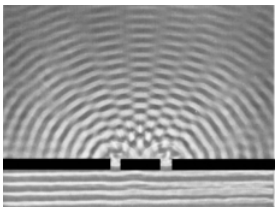
Answer

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

12.5.4 Huygens' Principle

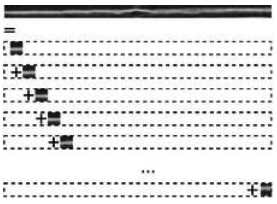


e / Christiaan Huygens (1629-1695).



f / Double-slit diffraction.

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



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



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



i / Adding up the ripples produces a new wavefront.

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

In double-slit diffraction the application of Huygens' principle is visually convincing: it is as though all the sets of ripples have been blocked except for two. It is a rather surprising mathematical fact, however, that Huygens' principle gives the right result in

the case of an unobstructed linear wave, h and i. A theoretically infinite number of circular wave patterns somehow conspire to add together and produce the simple linear wave motion with which we are familiar.

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

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

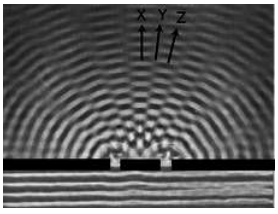
Almost all that was known about reflection and refraction of light could be interpreted equally well in terms of a particle model or a wave model, but Newton had one reason for strongly opposing Huygens' wave theory. Newton knew that waves exhibited diffraction, but diffraction of light is difficult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



j / Thomas Young

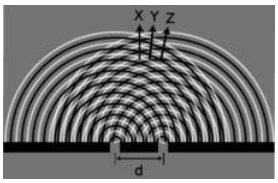
Thomas Young (1773-1829) was the person who finally, a hundred years later, did a careful search for wave interference effects with light and analyzed the results correctly. He observed double-slit diffraction of light as well as a variety of other diffraction effects, all of which showed that light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

12.5.5 Double-slit diffraction



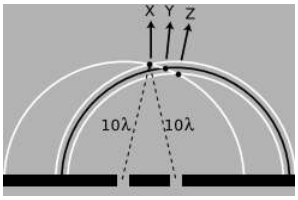
k / Double-slit diffraction.

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



l / Use of Huygens' principle.

If we assume that the width of the slits is small (on the order of the wavelength of the wave or less), then we can imagine only a single set of Huygens ripples spreading out from each one, l. White lines represent peaks, black ones troughs. The only dimension of the diffracting slits that has any effect on the geometric pattern of the overlapping ripples is then the center-to-center distance, d, between the slits.



m / Constructive interference along the center-line.

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

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

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

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

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

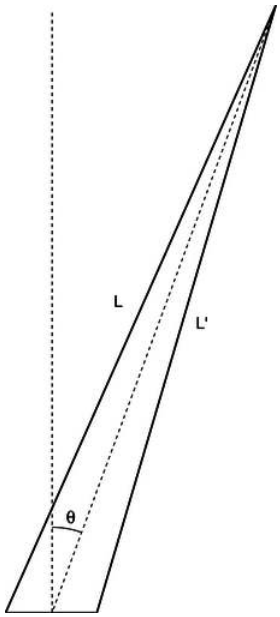
But the distance traveled does not have to be equal in order to get constructive interference. At the point along direction Z, one wave has gone nine wavelengths and the other ten. They are both at a positive extreme.

self-check:

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

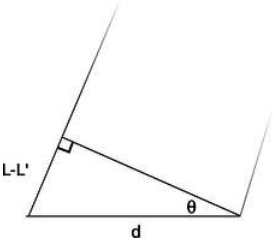
(answer in the back of the PDF version of the book)

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



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

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



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

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

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

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

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

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

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

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

[condition for a ma

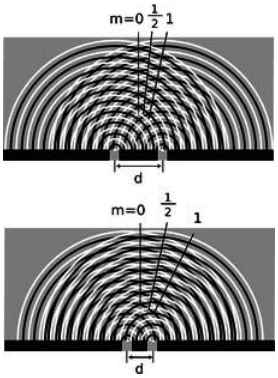
Similarly, the condition for a minimum is

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

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

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

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



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

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

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



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

Example 13: The correspondence principle

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

Example 14: Spacing of the fringes at small angles

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

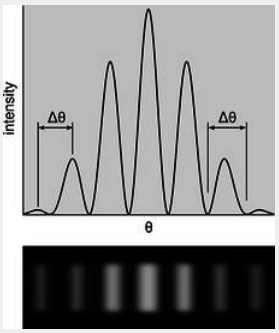
$$\frac{\lambda}{d} = \frac{\theta}{m},$$

which can be solved for θ to give

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

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

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



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

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

and similarly for any other neighboring pair of fringes.

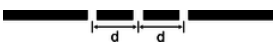
Although the equation $\lambda/d = \sin\theta/m$ is only valid for a double slit, it is can still be a guide to our thinking even if we are observing diffraction of light by a virus or a flea's leg: it is always true that

(1) large values of λ/d lead to a broad diffraction pattern, and

(2) diffraction patterns are repetitive.

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

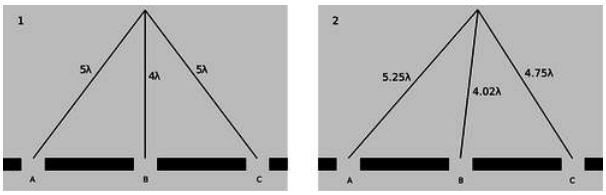
12.5.6 Repetition



s / A triple slit.

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

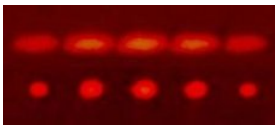
The answer is yes, as can be shown using figure u. For ease of visualization, I have violated our usual rule of only considering points very far from the diffracting object. The scale of the drawing is such that a wavelengths is one cm. In u/1, all three waves travel an integer number of wavelengths to reach the same point, so there is a bright central spot, as we would expect from our experience with the double slit. In figure u/2, we show the path lengths to a new point. This point is farther from slit A by a quarter of a wavelength, and correspondingly closer to slit C. The distance from slit B has hardly changed at all. Because the paths lengths traveled from slits A and C differ by half a wavelength, there will be perfect destructive interference between these two waves. There is still some uncanceled wave intensity because of slit B, but the amplitude will be three times less than in figure u/1, resulting in a factor of 9 decrease in brightness. Thus, by moving off to the right a little, we have gone from the bright central maximum to a point that is quite dark.



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

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

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



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

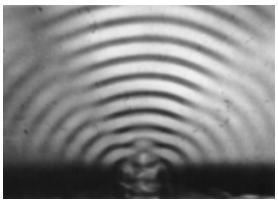
This is an example of a more general fact about diffraction: if some feature of the diffracting object is repeated, the locations of the maxima and minima are unchanged, but they become narrower.

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

12.5.7 Single-slit diffraction

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

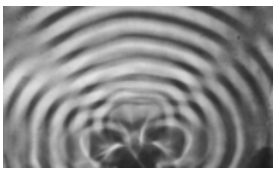
All of this, however, assumes that the slit is narrow compared to a wavelength of light. If, on the other hand, the slit is broader, there will indeed be interference among the sets of ripples spreading out from various points along the opening. Figure v shows an example with water waves, and figure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



v / Single-slit diffraction of water waves.



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



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

self-check:

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

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

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

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

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



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

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

sizes of the diffracted images depend on λ/a, the only way to improve the resolution of the images is to increase the diameter, a, of the tube. This is one of the main reasons (in addition to light-gathering power) why the best telescopes must be very large in diameter.



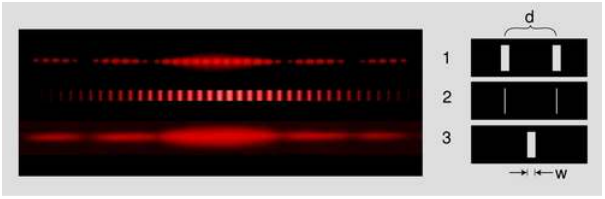
z / A radio telescope.

self-check:

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

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

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



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

Discussion Question

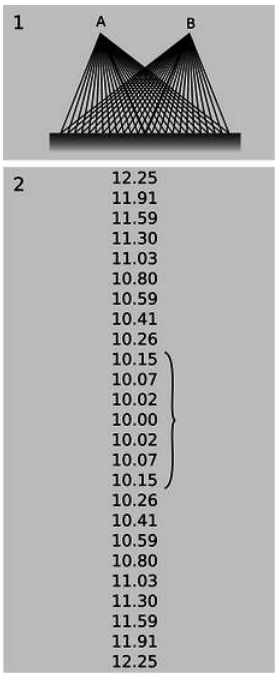
◊ Why is it optically impossible for bacteria to evolve eyes that use visible light to form images?

The principle of least time

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

Note how similar are the distances traveled by the group of 7 rays, indicated with a bracket, that come closest to obeying the principle of least time. If we think of each one as a wave, then all 7 are again nearly in phase at point B. However, the rays that are farther from satisfying the principle of least time show more rapidly changing distances; on reuniting at point B, their phases are a random jumble, and they will very nearly cancel each other out. Thus, almost none of the wave energy delivered to point B goes by these longer paths. Physically we find, for instance, that a wave pulse emitted at A is observed at B after a time interval

corresponding very nearly to the shortest possible path, and the pulse is not very “smeared out” when it gets there. The shorter the wavelength compared to the dimensions of the figure, the more accurate these approximate statements become.



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

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

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

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

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

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

This page titled 9.9: Wave Optics is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

- **13.5: Wave Optics** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

CHAPTER OVERVIEW

10: Chapter 10 - Electricty and Magnetism

10.1: Current and Voltage

10.2: Parallel and Series Circuits

10.3: More About the Magnetic Field

10.4: Induced Electric Fields

10: Chapter 10 - Electricty and Magnetism is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

10.1: Current and Voltage

9.1.1 Current

Unity of all types of electricity



a / *Gymnotus carapo*, a knifefish, uses electrical signals to sense its environment and to communicate with others of its species. (Greg DeGreef)Figure c/2 shows a setup that works. The battery pushes charge through the circuit, and recycles it over and over again. (We will have more to say later in this chapter about how batteries work.) This is called a *complete circuit*. Today, the electrical use of the word “circuit” is the only one that springs to mind for most people, but the original meaning was to travel around and make a round trip, as when a circuit court judge would ride around the boondocks, dispensing justice in each town on a certain date.

Note that an example like c/3 does not work. The wire will quickly begin acquiring a net charge, because it has no way to get rid of the charge flowing into it. The repulsion of this charge will make it more and more difficult to send any more charge in, and soon the electrical forces exerted by the battery will be canceled out completely. The whole process would be over so quickly that the filament would not even have enough time to get hot and glow. This is known as an *open circuit*. Exactly the same thing would happen if the complete circuit of figure c/2 was cut somewhere with a pair of scissors, and in fact that is essentially how an ordinary light switch works: by opening up a gap in the circuit.

The definition of electric current we have developed has the great virtue that it is easy to measure. In practical electrical work, one almost always measures current, not charge. The instrument used to measure current is called an *ammeter*. A simplified ammeter, c/4, simply consists of a coiled-wire magnet whose force twists an iron needle against the resistance of a spring. The greater the current, the greater the force. Although the construction of ammeters may differ, their use is always the same. We break into the path of the electric current and interpose the meter like a tollbooth on a road, c/5. There is still a complete circuit, and as far as the battery and bulb are concerned, the ammeter is just another segment of wire.

Does it matter where in the circuit we place the ammeter? Could we, for instance, have put it in the left side of the circuit instead of the right? Conservation of charge tells us that this can make no difference. Charge is not destroyed or “used up” by the lightbulb, so we will get the same current reading on either side of it. What is “used up” is energy stored in the battery, which is being converted into heat and light energy.

9.1.3 Voltage

The volt unit

Electrical circuits can be used for sending signals, storing information, or doing calculations, but their most common purpose by far is to manipulate energy, as in the battery-and-bulb example of the previous section. We know that lightbulbs are rated in units of watts, i.e., how many joules per second of energy they can convert into heat and light, but how would this relate to the flow of charge as measured in amperes? By way of analogy, suppose your friend, who didn't take physics, can't find any job better than pitching bales of hay. The number of calories he burns per hour will certainly depend on how many bales he pitches per minute, but it will also be proportional to how much mechanical work he has to do on each bale. If his job is to toss them up into a hayloft, he will get tired a lot more quickly than someone who merely tips bales off a loading dock into trucks. In metric units,

$$\frac{\text{joules}}{\text{second}} = \frac{\text{haybales}}{\text{second}} \times \frac{\text{joules}}{\text{haybale}}.$$

Similarly, the rate of energy transformation by a battery will not just depend on how many coulombs per second it pushes through a circuit but also on how much mechanical work it has to do on each coulomb of charge:

$$\frac{\text{joules}}{\text{second}} = \frac{\text{coulombs}}{\text{second}} \times \frac{\text{joules}}{\text{coulomb}}$$

or

$$\text{power} = \text{current} \times \text{work per unit charge.}$$

Units of joules per coulomb are abbreviated as *volts*, 1 V=1 J/C, named after the Italian physicist Alessandro Volta. Everyone knows that batteries are rated in units of volts, but the voltage concept is more general than that; it turns out that voltage is a property of every point in space. To gain more insight, let's think more carefully about what goes on in the battery and bulb circuit.

The voltage concept in general

To do work on a charged particle, the battery apparently must be exerting forces on it. How does it do this? Well, the only thing that can exert an electrical force on a charged particle is another charged particle. It's as though the haybales were pushing and pulling each other into the hayloft! This is potentially a horribly complicated situation. Even if we knew how much excess positive or negative charge there was at every point in the circuit (which realistically we don't) we would have to calculate zillions of forces using Coulomb's law, perform all the vector additions, and finally calculate how much work was being done on the charges as they moved along. To make things even more scary, there is more than one type of charged particle that moves: electrons are what move in the wires and the bulb's filament, but ions are the moving charge carriers inside the battery. Luckily, there are two ways in which we can simplify things:

The situation is unchanging. Unlike the imaginary setup in which we attempted to light a bulb using a rubber rod and a piece of fur, this circuit maintains itself in a steady state (after perhaps a microsecond-long period of settling down after the circuit is first assembled). The current is steady, and as charge flows out of any area of the circuit it is replaced by the same amount of charge flowing in. The amount of excess positive or negative charge in any part of the circuit therefore stays constant. Similarly, when we watch a river flowing, the water goes by but the river doesn't disappear.

Force depends only on position. Since the charge distribution is not changing, the total electrical force on a charged particle depends only on its own charge and on its location. If another charged particle of the same type visits the same location later on, it will feel exactly the same force.

The second observation tells us that there is nothing all that different about the experience of one charged particle as compared to another's. If we single out one particle to pay attention to, and figure out the amount of work done on it by electrical forces as it goes from point A to point B along a certain path, then this is the same amount of work that will be done on any other charged particles of the same type as it follows the same path. For the sake of visualization, let's think about the path that starts at one terminal of the battery, goes through the light bulb's filament, and ends at the other terminal. When an object experiences a force that depends only on its position (and when certain other, technical conditions are satisfied), we can define an electrical energy associated with the position of that object. The amount of work done on the particle by electrical forces as it moves from A to B equals the drop in electrical energy between A and B. This electrical energy is what is being converted into other forms of energy such as heat and light. We therefore define voltage in general as electrical energy per unit charge:

The difference in voltage between two points in space is defined as

$$\Delta V = \Delta U_{elec}/q,$$

where ΔU_{elec} is the change in the electrical energy of a particle with charge q as it moves from the initial point to the final point.

The amount of power dissipated (i.e., rate at which energy is transformed by the flow of electricity) is then given by the equation

$$P = I\Delta V.$$

Example 4: Energy stored in a battery

▹ The 1.2 V rechargeable battery in figure d is labeled 1800 milliamp-hours. What is the maximum amount of energy the battery can store?



d / Example 4.

▹ An ampere-hour is a unit of current multiplied by a unit of time. Current is charge per unit time, so an ampere-hour is in fact a funny unit of *charge*:

$$\begin{aligned}(1\text{ A})(1\text{ hour}) &= (1\text{ C/s})(3600\text{ s}) \\ &= 3600\text{ C}\end{aligned}$$

1800 milliamp-hours is therefore $1800 \times 10^{-3} \times 3600\text{ C} = 6.5 \times 10^{3}\text{ C}$. That's a huge number of charged particles, but the total loss of electrical energy will just be their total charge multiplied by the voltage difference across which they move:

$$\begin{aligned}\Delta U_{elec} &= q\Delta V \\ &= (6.5 \times 10^{3}\text{ C})(1.2\text{ V}) \\ &= 7.8\text{ kJ}\end{aligned}$$

Example 5: Units of volt-amps

▹ Doorbells are often rated in volt-amps. What does this combination of units mean?
▹ Current times voltage gives units of power, $P = I\Delta V$, so volt-amps are really just a nonstandard way of writing watts. They are telling you how much power the doorbell requires.

Example 6: Power dissipated by a battery and bulb

▹ If a 9.0-volt battery causes 1.0 A to flow through a lightbulb, how much power is dissipated?
▹ The voltage rating of a battery tells us what voltage difference ΔV it is designed to maintain between its terminals.

$$\begin{aligned}P &= I\ \Delta\text{V} \\ &= 9.0\text{ A} \cdot \text{V} \\ &= 9.0\ \frac{\text{C}}{\text{s}} \cdot \frac{\text{J}}{\text{C}} \\ &= 9.0\text{ J/s} \\ &= 9.0\text{ W}\end{aligned}$$

The only nontrivial thing in this problem was dealing with the units. One quickly gets used to translating common combinations like $\text{A} \cdot \text{V}$ into simpler terms.

Here are a few questions and answers about the voltage concept.

Question: OK, so what *is* voltage, really?
Answer: A device like a battery has positive and negative charges inside it that push other charges around the outside circuit. A higher-voltage battery has denser charges in it, which will do more work on each charged particle that moves through the outside circuit.

To use a gravitational analogy, we can put a paddlewheel at the bottom of either a tall waterfall or a short one, but a kg of water that falls through the greater gravitational energy difference will have more energy to give up to the paddlewheel at the bottom.

Question: Why do we define voltage as electrical energy divided by charge, instead of just defining it as electrical energy?
Answer: One answer is that it's the only definition that makes the equation $P = I\Delta V$ work. A more general answer is that we want to be able to define a voltage difference between any two points in space without having to know in advance how much charge the

particles moving between them will have. If you put a nine-volt battery on your tongue, then the charged particles that move across your tongue and give you that tingly sensation are not electrons but ions, which may have charges of $+e$, $-2e$, or practically anything. The manufacturer probably expected the battery to be used mostly in circuits with metal wires, where the charged particles that flowed would be electrons with charges of $-e$. If the ones flowing across your tongue happen to have charges of $-2e$, the electrical energy difference for them will be twice as much, but dividing by their charge of $-2e$ in the definition of voltage will still give a result of 9 V.

Question: Are there two separate roles for the charged particles in the circuit, a type that sits still and exerts the forces, and another that moves under the influence of those forces?
Answer: No. Every charged particle simultaneously plays both roles. Newton's third law says that any particle that has an electrical force acting on it must also be exerting an electrical force back on the other particle. There are no “designated movers” or “designated force-makers.”

Question: Why does the definition of voltage only refer to voltage *differences*?
Answer: It's perfectly OK to define voltage as $V = U_{elec}/q$. But recall that it is only *differences* in interaction energy, U, that have direct physical meaning in physics. Similarly, voltage differences are really more useful than absolute voltages. A voltmeter measures voltage differences, not absolute voltages.

Discussion Questions

◊ A roller coaster is sort of like an electric circuit, but it uses gravitational forces on the cars instead of electric ones. What would a high-voltage roller coaster be like? What would a high-current roller coaster be like?

◊ Criticize the following statements:

- “He touched the wire, and 10000 volts went through him.”
- “That battery has a charge of 9 volts.”
- “You used up the charge of the battery.”

◊ When you touch a 9-volt battery to your tongue, both positive and negative ions move through your saliva. Which ions go which way?

◊ I once touched a piece of physics apparatus that had been wired incorrectly, and got a several-thousand-volt voltage difference across my hand. I was not injured. For what possible reason would the shock have had insufficient power to hurt me?

9.1.4 Resistance

Resistance

So far we have simply presented it as an observed fact that a battery-and-bulb circuit quickly settles down to a steady flow, but why should it? Newton's second law, $a = F/m$, would seem to predict that the steady forces on the charged particles should make them whip around the circuit faster and faster. The answer is that as charged particles move through matter, there are always forces, analogous to frictional forces, that resist the motion. These forces need to be included in Newton's second law, which is really $a = F_{total}/m$, not $a = F/m$. If, by analogy, you push a crate across the floor at constant speed, i.e., with zero acceleration, the total force on it must be zero. After you get the crate going, the floor's frictional force is exactly canceling out your force. The chemical energy stored in your body is being transformed into heat in the crate and the floor, and no longer into an increase in the crate's kinetic energy. Similarly, the battery's internal chemical energy is converted into heat, not into perpetually increasing the charged particles' kinetic energy. Changing energy into heat may be a nuisance in some circuits, such as a computer chip, but it is vital in a lightbulb, which must get hot enough to glow. Whether we like it or not, this kind of heating effect is going to occur any time charged particles move through matter.

What determines the amount of heating? One flashlight bulb designed to work with a 9-volt battery might be labeled 1.0 watts, another 5.0. How does this work? Even without knowing the details of this type of friction at the atomic level, we can relate the heat dissipation to the amount of current that flows via the equation $P = I\Delta$V. If the two flashlight bulbs can have two different values of P when used with a battery that maintains the same ΔV, it must be that the 5.0-watt bulb allows five times more current to flow through it.

For many substances, including the tungsten from which lightbulb filaments are made, experiments show that the amount of current that will flow through it is directly proportional to the voltage difference placed across it. For an object made of such a substance, we define its electrical *resistance* as follows:

If an object inserted in a circuit displays a current flow which is proportional to the voltage difference across it, then we define its resistance as the constant ratio

$$R = \Delta V/I.$$

The units of resistance are volts/ampere, usually abbreviated as ohms, symbolized with the capital Greek letter omega, Ω.

Example 7: Resistance of a lightbulb

▹ A flashlight bulb powered by a 9-volt battery has a resistance of 10 Ω. How much current will it draw?
▹ Solving the definition of resistance for I, we find

$$\begin{aligned} I &= \Delta V/R \\ &= 0.9\ \mathrm{V}/\Omega \\ &= 0.9\ \mathrm{V}/(\mathrm{V}/\mathrm{A}) \\ &= 0.9\ \mathrm{A} \end{aligned}$$

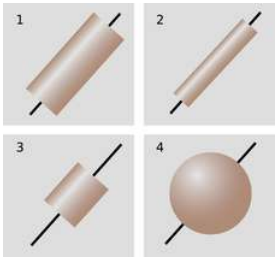


e / Georg Simon Ohm (1787-1854).

Ohm's law states that many substances, including many solids and some liquids, display this kind of behavior, at least for voltages that are not too large. The fact that Ohm's law is called a “law” should not be taken to mean that all materials obey it, or that it has the same fundamental importance as Newton's laws, for example. Materials are called *ohmic* or *nonohmic*, depending on whether they obey Ohm's law.

If objects of the same size and shape made from two different ohmic materials have different resistances, we can say that one material is more resistive than the other, or equivalently that it is less conductive. Materials, such as metals, that are very conductive are said to be good *conductors*. Those that are extremely poor conductors, for example wood or rubber, are classified as *insulators*. There is no sharp distinction between the two classes of materials. Some, such as silicon, lie midway between the two extremes, and are called semiconductors.

On an intuitive level, we can understand the idea of resistance by making the sounds “hhhhhh” and “ffffff.” To make air flow out of your mouth, you use your diaphragm to compress the air in your chest. The pressure difference between your chest and the air outside your mouth is analogous to a voltage difference. When you make the “h” sound, you form your mouth and throat in a way that allows air to flow easily. The large flow of air is like a large current. Dividing by a large current in the definition of resistance means that we get a small resistance. We say that the small resistance of your mouth and throat allows a large current to flow. When you make the “f” sound, you increase the resistance and cause a smaller current to flow.



f / Four objects made of the same substance have different resistances.

Note that although the resistance of an object depends on the substance it is made of, we cannot speak simply of the "resistance of gold" or the "resistance of wood." Figure f shows four examples of objects that have had wires attached at the ends as electrical connections. If they were made of the same substance, they would all nevertheless have different resistances because of their different sizes and shapes. A more detailed discussion will be more natural in the context of the following chapter, but it should not be too surprising that the resistance of f/2 will be greater than that of f/1 --- the image of water flowing through a pipe, however incorrect, gives us the right intuition. Object f/3 will have a smaller resistance than f/1 because the charged particles have less of it to get through.

Superconductors

All materials display some variation in resistance according to temperature (a fact that is used in thermostats to make a thermometer that can be easily interfaced to an electric circuit). More spectacularly, most metals have been found to exhibit a sudden change to *zero* resistance when cooled to a certain critical temperature. They are then said to be superconductors. Theoretically, superconductors should make a great many exciting devices possible, for example coiled-wire magnets that could be used to levitate trains. In practice, the critical temperatures of all metals are very low, and the resulting need for extreme refrigeration has made their use uneconomical except for such specialized applications as particle accelerators for physics research.



g / A superconducting segment of the ATLAS accelerator at Argonne National Laboratory near Chicago. It is used to accelerate beams of ions to a few percent of the speed of light for nuclear physics research. The shiny silver-colored surfaces are made of the element niobium, which is a superconductor at relatively high temperatures compared to other metals --- relatively high meaning the temperature of liquid helium! The beam of ions passes through the holes in the two small cylinders on the ends of the curved rods. Charge is shuffled back and forth between them at a frequency of 12 million cycles per second, so that they take turns being positive and negative. The positively charged beam consists of short spurts, each timed so that when it is in one of the segments it will be pulled forward by negative charge on the cylinder in front of it and pushed forward by the positively charged one behind. The huge currents involved would quickly melt any metal that was not superconducting, but in a superconductor they produce no heat at all.

But scientists have recently made the surprising discovery that certain ceramics are superconductors at less extreme temperatures. The technological barrier is now in finding practical methods for making wire out of these brittle materials. Wall Street is currently investing billions of dollars in developing superconducting devices for cellular phone relay stations based on these materials. In

2001, the city of Copenhagen replaced a short section of its electrical power trunks with superconducing cables, and they are now in operation and supplying power to customers.

There is currently no satisfactory theory of superconductivity in general, although superconductivity in metals is understood fairly well. Unfortunately I have yet to find a fundamental explanation of superconductivity in metals that works at the introductory level.

Example 8: Finding charge given current

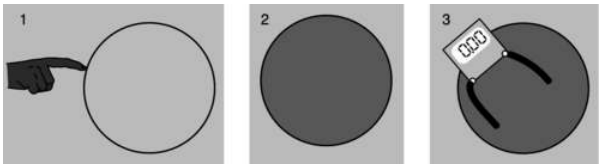
▹ In the segment of the ATLAS accelerator shown in figure g, the current flowing back and forth between the two cylinders is given by $I = a\cos bt$. What is the charge on one of the cylinders as a function of time? ▹ We are given the current and want to find the charge, i.e., we are given the derivative and we want to find the original function that would give that derivative. This means we need to integrate:

$$\begin{aligned} q &= \int I dt \\ &= \int a\cos bt dt \\ &= \frac{a}{b}\sin bt + q_0, \end{aligned}$$

where q_0 is a constant of integration.
We can interpret this in order to explain why a superconductor needs to be used. The constant b must be very large, since the current is supposed to oscillate back and forth millions of times a second. Looking at the final result, we see that if b is a very large number, and q is to be a significant amount of charge, then a must be a very large number as well. If a is numerically large, then the current must be very large, so it would heat the accelerator too much if it was flowing through an ordinary conductor.

Constant voltage throughout a conductor

The idea of a superconductor leads us to the question of how we should expect an object to behave if it is made of a very good conductor. Superconductors are an extreme case, but often a metal wire can be thought of as a perfect conductor, for example if the parts of the circuit other than the wire are made of much less conductive materials. What happens if R equals zero in the equation $R = \Delta V/I$? The result of dividing two numbers can only be zero if the number on top equals zero. This tells us that if we pick any two points in a perfect conductor, the voltage difference between them must be zero. In other words, the entire conductor must be at the same voltage.

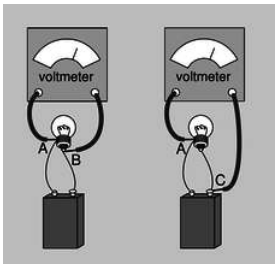


h / 1. The finger deposits charges on the solid, spherical, metal doorknob and is then withdrawn. 2. Almost instantaneously, the charges' mutual repulsion makes them redistribute themselves uniformly on the surface of the sphere. The only *excess* charge is on the surface; charges do exist in the atoms that form the interior of the sphere, but they are balanced. Charges on the interior feel zero total electrical force from the ones at the surface. Charges at the surface experience a net outward repulsion, but this is canceled out by the force that keep them from escaping into the air. 3. A voltmeter shows zero difference in voltage between any two points on the interior or surface of the sphere. If the voltage difference wasn't zero, then energy could be released by the flow of charge from one point to the other; this only happens before equilibrium is reached.

Constant voltage means that no work would be done on a charge as it moved from one point in the conductor to another. If zero work was done only along a certain path between two specific points, it might mean that positive work was done along part of the path and negative work along the rest, resulting in a cancellation. But there is no way that the work could come out to be zero for all possible paths unless the electrical force on a charge was in fact zero at every point. Suppose, for example, that you build up a static charge by scuffing your feet on a carpet, and then you deposit some of that charge onto a doorknob, which is a good conductor. How can all that charge be in the doorknob without creating any electrical force at any point inside it? The only possible answer is that the charge moves around until it has spread itself into just the right configuration so that the forces exerted by all the little bits of excess surface charge on any charged particle within the doorknob exactly cancel out.

We can explain this behavior if we assume that the charge placed on the doorknob eventually settles down into a stable equilibrium. Since the doorknob is a conductor, the charge is free to move through it. If it was free to move and any part of it did experience a nonzero total force from the rest of the charge, then it would move, and we would not have an equilibrium.

Excess charge placed on a conductor, once it reaches its equilibrium configuration, is entirely on the surface, not on the interior. This should be intuitively reasonable in figure h, for example, since the charges are all repelling each other. A proof is given in example 35 on p. 623.

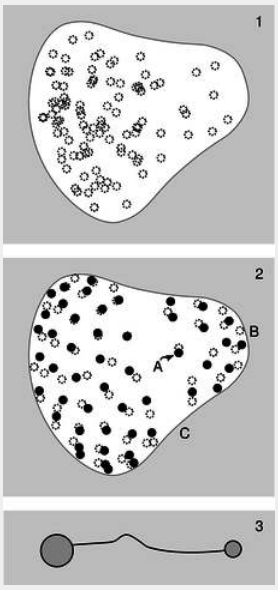


i / The voltmeter doesn't care which of these setups you use.

Since wires are good conductors, constancy of voltage throughout a conductor provides a convenient freedom in hooking up a voltmeter to a circuit. In figure i, points B and C are on the same piece of conducting wire, so $V_B = V_C$. Measuring $V_B - V_A$ gives the same result as measuring $V_C - V_A$.

Example 9: The lightning rod

Suppose you have a pear-shaped conductor like the one in figure j/1. Since the pear is a conductor, there are free charges everywhere inside it. Panels 1 and 2 of the figure show a computer simulation with 100 identical electric charges. In 1, the charges are released at random positions inside the pear. Repulsion causes them all to fly outward onto the surface and then settle down into an orderly but nonuniform pattern.



j / Example 9. In 1 and 2, charges that are visible on the front surface of the conductor are shown as solid dots; the others would have to be seen through the conductor, which we imagine is semi-transparent.

We might not have been able to guess the pattern in advance, but we can verify that some of its features make sense. For example, charge A has more neighbors on the right than on the left, which would tend to make it accelerate off to the left. But when we look at the picture as a whole, it appears reasonable that this is prevented by the larger number of more distant charges on its left than on its right.

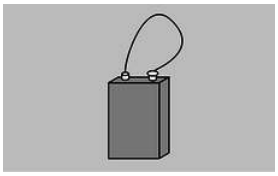
There also seems to be a pattern to the nonuniformity: the charges collect more densely in areas like B, where the surface is strongly curved, and less densely in flatter areas like C.

To understand the reason for this pattern, consider j/3. Two conducting spheres are connected by a conducting wire. Since the whole apparatus is conducting, it must all be at one voltage. As shown in problem 37 on p. 550, the density of charge is greater on the smaller sphere. This is an example of a more general fact observed in j/2, which is that the charge on a conductor packs itself more densely in areas that are more sharply curved.

Similar reasoning shows why Benjamin Franklin used a sharp tip when he invented the lightning rod. The charged stormclouds induce positive and negative charges to move to opposite ends of the rod. At the pointed upper end of the rod, the charge tends to concentrate at the point, and this charge attracts the lightning. The same effect can sometimes be seen when a scrap of aluminum foil is inadvertently put in a microwave oven. Modern experiments (Moore *et al.*, Journal of Applied Meteorology 39 (1999) 593) show that although a sharp tip is best at starting a spark, a more moderate curve, like the right-hand tip of the pear in this example, is better at successfully sustaining the spark for long enough to connect a discharge to the clouds.

Short circuits

So far we have been assuming a perfect conductor. What if it is a good conductor, but not a perfect one? Then we can solve for $\Delta V = IR$. An ordinary-sized current will make a very small result when we multiply it by the resistance of a good conductor such as a metal wire. The voltage throughout the wire will then be nearly constant. If, on the other hand, the current is extremely large, we can have a significant voltage difference. This is what happens in a *short-circuit:* a circuit in which a low-resistance pathway connects the two sides of a voltage source. Note that this is much more specific than the popular use of the term to indicate any electrical malfunction at all. If, for example, you short-circuit a 9-volt battery as shown in figure k, you will produce perhaps a thousand amperes of current, leading to a very large value of $P = I\Delta V$. The wire gets hot!



k / Short-circuiting a battery. Warning: you can burn yourself this way or start a fire! If you want to try this, try making the connection only very briefly, use a low-voltage battery, and avoid touching the battery or the wire, both of which will get hot.

self-check:

What would happen to the battery in this kind of short circuit?

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

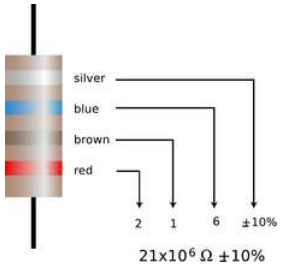
At this stage, most students have a hard time understanding why resistors would be used inside a radio or a computer. We obviously want a lightbulb or an electric stove to have a circuit element that resists the flow of electricity and heats up, but heating is undesirable in radios and computers. Without going too far afield, let's use a mechanical analogy to get a general idea of why a resistor would be used in a radio.

The main parts of a radio receiver are an antenna, a tuner for selecting the frequency, and an amplifier to strengthen the signal sufficiently to drive a speaker. The tuner resonates at the selected frequency, just as in the examples of mechanical resonance discussed in 3. The behavior of a mechanical resonator depends on three things: its inertia, its stiffness, and the amount of friction or damping. The first two parameters locate the peak of the resonance curve, while the damping determines the width of the resonance. In the radio tuner we have an electrically vibrating system that resonates at a particular frequency. Instead of a physical object moving back and forth, these vibrations consist of electrical currents that flow first in one direction and then in the other. In a mechanical system, damping means taking energy out of the vibration in the form of heat, and exactly the same idea applies to an electrical system: the resistor supplies the damping, and therefore controls the width of the resonance. If we set out to eliminate all resistance in the tuner circuit, by not building in a resistor and by somehow getting rid of all the inherent electrical resistance of the wires, we would have a useless radio. The tuner's resonance would be so narrow that we could never get close enough to the right frequency to bring in the station. The roles of inertia and stiffness are played by other circuit elements we have not discusses (a capacitor and a coil).

Resistors

Inside any electronic gadget you will see quite a few little circuit elements like the one shown below. These *resistors* are simply a cylinder of ohmic material with wires attached to the end.

color	meaning
black	0
brown	1
red	2
orange	3
yellow	4
green	5
blue	6
violet	7
gray	8
white	9
silver	±10%
gold	±5%



m / An example of a resistor with a color code.

Many electrical devices are based on electrical resistance and Ohm's law, even if they do not have little components in them that look like the usual resistor. The following are some examples.



n / The symbol used in schematics to represent a resistor.

Lightbulb

There is nothing special about a lightbulb filament --- you can easily make a lightbulb by cutting a narrow waist into a metallic gum wrapper and connecting the wrapper across the terminals of a 9-volt battery. The trouble is that it will instantly burn out.

Edison solved this technical challenge by encasing the filament in an evacuated bulb, which prevented burning, since burning requires oxygen.

Polygraph

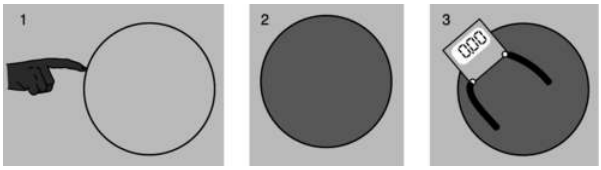
The polygraph, or “lie detector,” is really just a set of meters for recording physical measures of the subject's psychological stress, such as sweating and quickened heartbeat. The real-time sweat measurement works on the principle that dry skin is a good insulator, but sweaty skin is a conductor. Of course a truthful subject may become nervous simply because of the situation, and a practiced liar may not even break a sweat. The method's practitioners claim that they can tell the difference, but you should think twice before allowing yourself to be polygraph tested. Most U.S. courts exclude all polygraph evidence, but some employers attempt to screen out dishonest employees by polygraph testing job applicants, an abuse that ranks with such pseudoscience as handwriting analysis.

Fuse

A fuse is a device inserted in a circuit tollbooth-style in the same manner as an ammeter. It is simply a piece of wire made of metals having a relatively low melting point. If too much current passes through the fuse, it melts, opening the circuit. The purpose is to make sure that the building's wires do not carry so much current that they themselves will get hot enough to start a fire. Most modern houses use circuit breakers instead of fuses, although fuses are still common in cars and small devices. A circuit breaker is a switch operated by a coiled-wire magnet, which opens the circuit when enough current flows. The advantage is that once you turn off some of the appliances that were sucking up too much current, you can immediately flip the switch closed. In the days of fuses, one might get caught without a replacement fuse, or even be tempted to stuff aluminum foil in as a replacement, defeating the safety feature.

Voltmeter

A voltmeter is nothing more than an ammeter with an additional high-value resistor through which the current is also forced to flow. Ohm's law relates the current through the resistor is related directly to the voltage difference across it, so the meter can be calibrated in units of volts based on the known value of the resistor. The voltmeter's two probes are touched to the two locations in a circuit between which we wish to measure the voltage difference, o/2. Note how cumbersome this type of drawing is, and how difficult it can be to tell what is connected to what. This is why electrical drawing are usually shown in schematic form. Figure o/3 is a schematic representation of figure o/2.



o / 1. A simplified diagram of how a voltmeter works. 2. Measuring the voltage difference across a lightbulb. 3. The same setup drawn in schematic form. 4. The setup for measuring current is different.

The setups for measuring current and voltage are different. When we are measuring current, we are finding “how much stuff goes through,” so we place the ammeter where all the current is forced to go through it. Voltage, however, is not “stuff that goes through,” it is a measure of electrical energy. If an ammeter is like the meter that measures your water use, a voltmeter is like a measuring stick that tells you how high a waterfall is, so that you can determine how much energy will be released by each kilogram of falling water. We do not want to force the water to go through the measuring stick! The arrangement in figure o/3 is a ***parallel*** circuit: one in there are “forks in the road” where some of the current will flow one way and some will flow the other. Figure o/4 is said to be wired in *series:* all the current will visit all the circuit elements one after the other. We will deal with series and parallel circuits in more detail in the following chapter.

If you inserted a voltmeter incorrectly, in series with the bulb and battery, its large internal resistance would cut the current down so low that the bulb would go out. You would have severely disturbed the behavior of the circuit by trying to measure something about it.

Incorrectly placing an ammeter in parallel is likely to be even more disconcerting. The ammeter has nothing but wire inside it to provide resistance, so given the choice, most of the current will flow through it rather than through the bulb. So much current will flow through the ammeter, in fact, that there is a danger of burning out the battery or the meter or both! For this reason, most ammeters have fuses or circuit breakers inside. Some models will trip their circuit breakers and make an audible alarm in this situation, while others will simply blow a fuse and stop working until you replace it.

Discussion Questions

◊ In figure o/1, would it make any difference in the voltage measurement if we touched the voltmeter's probes to different points along the same segments of wire?

◊ Explain why it would be incorrect to define resistance as the amount of charge the resistor allows to flow.

9.1.5 Current-conducting properties of materials

Ohm's law has a remarkable property, which is that current will flow even in response to a voltage difference that is as small as we care to make it. In the analogy of pushing a crate across a floor, it is as though even a flea could slide the crate across the floor, albeit at some very low speed. The flea cannot do this because of static friction, which we can think of as an effect arising from the tendency of the microscopic bumps and valleys in the crate and floor to lock together. The fact that Ohm's law holds for nearly all solids has an interesting interpretation: at least some of the electrons are not “locked down” at all to any specific atom.

More generally we can ask how charge actually flows in various solids, liquids, and gases. This will lead us to the explanations of many interesting phenomena, including lightning, the bluish crust that builds up on the terminals of car batteries, and the need for electrolytes in sports drinks.

Solids

In atomic terms, the defining characteristic of a solid is that its atoms are packed together, and the nuclei cannot move very far from their equilibrium positions. It makes sense, then, that electrons, not ions, would be the charge carriers when currents flow in solids. This fact was established experimentally by Tolman and Stewart, in an experiment in which they spun a large coil of wire and then abruptly stopped it. They observed a current in the wire immediately after the coil was stopped, which indicated that charged particles that were not permanently locked to a specific atom had continued to move because of their own inertia, even after the material of the wire in general stopped. The direction of the current showed that it was negatively charged particles that kept moving. The current only lasted for an instant, however; as the negatively charged particles collected at the downstream end of the wire, farther particles were prevented joining them due to their electrical repulsion, as well as the attraction from the upstream end, which was left with a net positive charge. Tolman and Stewart were even able to determine the mass-to-charge ratio of the particles. We need not go into the details of the analysis here, but particles with high mass would be difficult to decelerate, leading to a stronger and longer pulse of current, while particles with high charge would feel stronger electrical forces decelerating them, which would cause a weaker and shorter pulse. The mass-to-charge ratio thus determined was consistent with the m/q of the electron to within the accuracy of the experiment, which essentially established that the particles were electrons.

The fact that only electrons carry current in solids, not ions, has many important implications. For one thing, it explains why wires don't fray or turn to dust after carrying current for a long time. Electrons are very small (perhaps even pointlike), and it is easy to imagine them passing between the cracks among the atoms without creating holes or fractures in the atomic framework. For those who know a little chemistry, it also explains why all the best conductors are on the left side of the periodic table. The elements in that area are the ones that have only a very loose hold on their outermost electrons.

Gases

The molecules in a gas spend most of their time separated from each other by significant distances, so it is not possible for them to conduct electricity the way solids do, by handing off electrons from atom to atom. It is therefore not surprising that gases are good insulators.

Gases are also usually nonohmic. As opposite charges build up on a stormcloud and the ground below, the voltage difference becomes greater and greater. Zero current flows, however, until finally the voltage reaches a certain threshold and we have an impressive example of what is known as a spark or electrical discharge. If air was ohmic, the current between the cloud and the ground would simply increase steadily as the voltage difference increased, rather than being zero until a threshold was reached.

This behavior can be explained as follows. At some point, the electrical forces on the air electrons and nuclei of the air molecules become so strong that electrons are ripped right off of some of the molecules. The electrons then accelerate toward either the cloud or the ground, whichever is positively charged, and the positive ions accelerate the opposite way. As these charge carriers accelerate, they strike and ionize other molecules, which produces a rapidly growing cascade.

Liquids

Molecules in a liquid are able to slide past each other, so ions as well as electrons can carry currents. Pure water is a poor conductor because the water molecules tend to hold onto their electrons strongly, and there are therefore not many electrons or ions available to move. Water can become quite a good conductor, however, with the addition of even a small amount of certain substances called electrolytes, which are typically salts. For example, if we add table salt, NaCl, to water, the NaCl molecules dissolve into Na^+ and Cl^- ions, which can then move and create currents. This is why electric currents can flow among the cells in our bodies: cellular fluid is quite salty. When we sweat, we lose not just water but electrolytes, so dehydration plays havoc with our cells' electrical systems. It is for this reason that electrolytes are included in sports drinks and formulas for rehydrating infants who have diarrhea.

Since current flow in liquids involves entire ions, it is not surprising that we can see physical evidence when it has occurred. For example, after a car battery has been in use for a while, the H_2SO_4 battery acid becomes depleted of hydrogen ions, which are the main charge carriers that complete the circuit on the inside of the battery. The leftover SO_4 then forms a visible blue crust on the battery posts.

Speed of currents and electrical signals

When I talk on the phone to my mother in law two thousand miles away, I do not notice any delay while the signal makes its way back and forth. Electrical signals therefore must travel very quickly, but how fast exactly? The answer is rather subtle. For the sake of concreteness, let's restrict ourselves to currents in metals, which consist of electrons.

The electrons themselves are only moving at speeds of perhaps a few thousand miles per hour, and their motion is mostly random thermal motion. This shows that the electrons in my phone cannot possibly be zipping back and forth between California and New York fast enough to carry the signals. Even if their thousand-mile-an-hour motion was organized rather than random, it would still take them many minutes to get there. Realistically, it will take the average electron even longer than that to make the trip. The current in the wire consists only of a slow overall drift, at a speed on the order of a few centimeters per second, superimposed on the more rapid random motion. We can compare this with the slow westward drift in the population of the U.S. If we could make a movie of the motion of all the people in the U.S. from outer space, and could watch it at high speed so that the people appeared to be scurrying around like ants, we would think that the motion was fairly random, and we would not immediately notice the westward drift. Only after many years would we realize that the number of people heading west over the Sierras had exceeded the number going east, so that California increased its share of the country's population.

So why are electrical signals so fast if the average drift speed of electrons is so slow? The answer is that a disturbance in an electrical system can move much more quickly than the charges themselves. It is as though we filled a pipe with golf balls and then inserted an extra ball at one end, causing a ball to fall out at the other end. The force propagated to the other end in a fraction of a second, but the balls themselves only traveled a few centimeters in that time.

Because the reality of current conduction is so complex, we often describe things using mental shortcuts that are technically incorrect. This is OK as long as we know that they are just shortcuts. For example, suppose the presidents of France and Russia shake hands, and the French politician has inadvertently picked up a positive electrical charge, which shocks the Russian. We may say that the excess positively charged particles in the French leader's body, which all repel each other, take the handshake as an opportunity to get farther apart by spreading out into two bodies rather than one. In reality, it would be a matter of minutes before the ions in one person's body could actually drift deep into the other's. What really happens is that throughout the body of the recipient of the shock there are already various positive and negative ions which are free to move. Even before the perpetrator's charged hand touches the victim's sweaty palm, the charges in the shocker's body begin to repel the positive ions and attract the negative ions in the other person. The split-second sensation of shock is caused by the sudden jumping of the victim's ions by distances of perhaps a micrometer, this effect occurring simultaneously throughout the whole body, although more violently in the hand and arm, which are closer to the other person.

Contributors

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

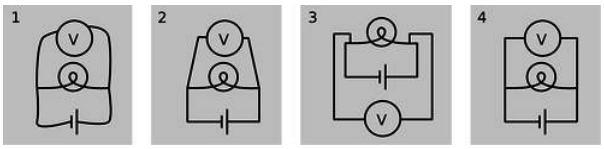
This page titled 10.1: Current and Voltage is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

10.2: Parallel and Series Circuits

In section 9.1, we limited ourselves to relatively simple circuits, essentially nothing more than a battery and a single lightbulb. The purpose of this chapter is to introduce you to more complex circuits, containing multiple resistors or voltage sources in series, in parallel, or both.

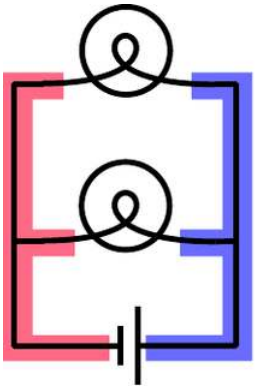
9.2.1 Schematics

I see a chess position; Kasparov sees an interesting Ruy Lopez variation. To the uninitiated a schematic may look as unintelligible as Mayan hieroglyphs, but even a little bit of eye training can go a long way toward making its meaning leap off the page. A schematic is a stylized and simplified drawing of a circuit. The purpose is to eliminate as many irrelevant features as possible, so that the relevant ones are easier to pick out.



a / 1. Wrong: The shapes of the wires are irrelevant. 2. Wrong: Right angles should be used. 3. Wrong: A simple pattern is made to look unfamiliar and complicated. 4. Right.

An example of an irrelevant feature is the physical shape, length, and diameter of a wire. In nearly all circuits, it is a good approximation to assume that the wires are perfect conductors, so that any piece of wire uninterrupted by other components has constant voltage throughout it. Changing the length of the wire, for instance, does not change this fact. (Of course if we used miles and miles of wire, as in a telephone line, the wire's resistance would start to add up, and its length would start to matter.) The shapes of the wires are likewise irrelevant, so we draw them with standardized, stylized shapes made only of vertical and horizontal lines with right-angle bends in them. This has the effect of making similar circuits look more alike and helping us to recognize familiar patterns, just as words in a newspaper are easier to recognize than handwritten ones. Figure a shows some examples of these concepts.



b / The two shaded areas shaped like the letter “E” are both regions of constant voltage.

The most important first step in learning to read schematics is to learn to recognize contiguous pieces of wire which must have constant voltage throughout. In figure b, for example, the two shaded E-shaped pieces of wire must each have constant voltage. This focuses our attention on two of the main unknowns we'd like to be able to predict: the voltage of the left-hand E and the voltage of the one on the right.

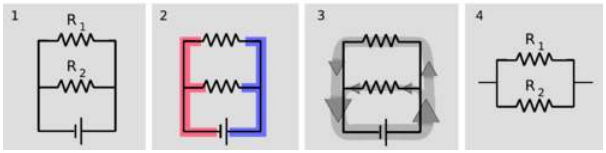
9.2.2 Parallel resistances and the junction rule

One of the simplest examples to analyze is the parallel resistance circuit, of which figure b was an example. In general we may have unequal resistances R_1 and R_2, as in c/1. Since there are only two constant-voltage areas in the circuit, c/2, all three components have the same voltage difference across them. A battery normally succeeds in maintaining the voltage differences

across itself for which it was designed, so the voltage drops ΔV_1 and ΔV_2 across the resistors must both equal the voltage of the battery:

$$\Delta V_1 = \Delta V_2 = \Delta V_{battery}.$$

Each resistance thus feels the same voltage difference as if it was the only one in the circuit, and Ohm's law tells us that the amount of current flowing through each one is also the same as it would have been in a one-resistor circuit. This is why household electrical circuits are wired in parallel. We want every appliance to work the same, regardless of whether other appliances are plugged in or unplugged, turned on or switched off. (The electric company doesn't use batteries of course, but our analysis would be the same for any device that maintains a constant voltage.)



c / 1. Two resistors in parallel. 2. There are two constant-voltage areas. 3. The current that comes out of the battery splits between the two resistors, and later reunites. 4. The two resistors in parallel can be treated as a single resistor with a smaller resistance value.

Of course the electric company can tell when we turn on every light in the house. How do they know? The answer is that we draw more current. Each resistance draws a certain amount of current, and the amount that has to be supplied is the sum of the two individual currents. The current is like a river that splits in half, c/3, and then reunites. The total current is

$$I_{total} = I_1 + I_2.$$

This is an example of a general fact called the junction rule:

In any circuit that is not storing or releasing charge, conservation of charge implies that the total current flowing out of any junction must be the same as the total flowing in.

Coming back to the analysis of our circuit, we apply Ohm's law to each resistance, resulting in

$$\begin{aligned} I_{total} &= \Delta V/R_1 + \Delta V/R_2 \\ &= \Delta V \left(\frac{1}{R_1} + \frac{1}{R_2} \right). \end{aligned}$$

As far as the electric company is concerned, your whole house is just one resistor with some resistance R, called the *equivalent resistance*. They would write Ohm's law as

$$I_{total} = \Delta V/R,$$

from which we can determine the equivalent resistance by comparison with the previous expression:

$$\begin{aligned} 1/R &= \frac{1}{R_1} + \frac{1}{R_2} \\ R &= \left(\frac{1}{R_1} + \frac{1}{R_2} \right)^{-1} \end{aligned}$$

[equivalent resistance of two resistors in parallel]

Two resistors in parallel, c/4, are equivalent to a single resistor with a value given by the above equation.

Example 10: Two lamps on the same household circuit

▷ You turn on two lamps that are on the same household circuit. Each one has a resistance of 1 ohm. What is the equivalent resistance, and how does the power dissipation compare with the case of a single lamp?

▷ The equivalent resistance of the two lamps in parallel is

$$\begin{aligned} R &= \left(\frac{1}{R_1}+\frac{1}{R_2}\right)^{-1} \\ &= \left(\frac{1}{1\ \Omega}+\frac{1}{1\ \Omega}\right)^{-1} \\ &= \left(1\ \Omega^{-1}+1\ \Omega^{-1}\right)^{-1} \\ &= \left(2\ \Omega^{-1}\right)^{-1} \\ &= 0.5\ \Omega \end{aligned}$$

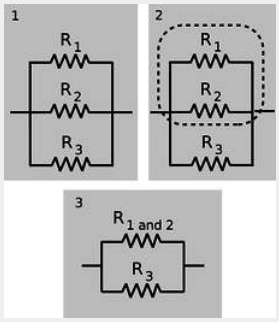
The voltage difference across the whole circuit is always the 110 V set by the electric company (it's alternating current, but that's irrelevant). The resistance of the whole circuit has been cut in half by turning on the second lamp, so a fixed amount of voltage will produce twice as much current. Twice the current flowing across the same voltage difference means twice as much power dissipation, which makes sense.

The cutting in half of the resistance surprises many students, since we are "adding more resistance" to the circuit by putting in the second lamp. Why does the equivalent resistance come out to be less than the resistance of a single lamp? This is a case where purely verbal reasoning can be misleading. A resistive circuit element, such as the filament of a lightbulb, is neither a perfect insulator nor a perfect conductor. Instead of analyzing this type of circuit in terms of "resistors," i.e., partial insulators, we could have spoken of "conductors." This example would then seem reasonable, since we "added more conductance," but one would then have the incorrect expectation about the case of resistors in series, discussed in the following section.

Perhaps a more productive way of thinking about it is to use mechanical intuition. By analogy, your nostrils resist the flow of air through them, but having two nostrils makes it twice as easy to breathe.

Example 11: Three resistors in parallel

▹ What happens if we have three or more resistors in parallel?



d / Three resistors in parallel.

▹ This is an important example, because the solution involves an important technique for understanding circuits: breaking them down into smaller parts and them simplifying those parts. In the circuit d/1, with three resistors in parallel, we can think of two of the resistors as forming a single big resistor, d/2, with equivalent resistance

$$R_{12} = \left(\frac{1}{R_1}+\frac{1}{R_2}\right)^{-1}.$$

We can then simplify the circuit as shown in d/3, so that it contains only two resistances. The equivalent resistance of the whole circuit is then given by

$$R_{123} = \left(\frac{1}{R_{12}}+\frac{1}{R_3}\right)^{-1}.$$

Substituting for R_{12} and simplifying, we find the result

$$R_{123} = \left(\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_3}\right)^{-1},$$

which you probably could have guessed. The interesting point here is the divide-and-conquer concept, not the mathematical result.

Example 12: An arbitrary number of identical resistors in parallel

▹ What is the resistance of N identical resistors in parallel?

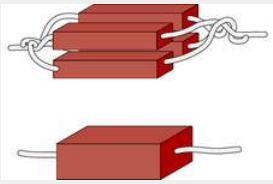
▹ Generalizing the results for two and three resistors, we have

$$R_N = \left(\frac{1}{R_1}+\frac{1}{R_2}+...\right)^{-1},$$

where "..." means that the sum includes all the resistors. If all the resistors are identical, this becomes

$$\begin{aligned} R_N &= \left(\frac{N}{R}\right)^{-1} \\ &= \frac{R}{N} \end{aligned}$$

Example 13: Dependence of resistance on cross-sectional area



e / Uniting four resistors in parallel is equivalent to making a single resistor with the same length but four times the cross-sectional area. The result is to make a resistor with one quarter the resistance.

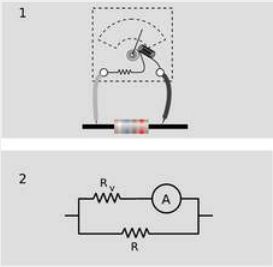
We have alluded briefly to the fact that an object's electrical resistance depends on its size and shape, but now we are ready to begin making more mathematical statements about it. As suggested by figure e, increasing a resistors's cross-sectional area is equivalent to adding more resistors in parallel, which will lead to an overall decrease in resistance. Any real resistor with straight, parallel sides can be sliced up into a large number of pieces, each with cross-sectional area of, say, 1 μm^2. The number, N, of such slices is proportional to the total cross-sectional area of the resistor, and by application of the result of the previous example we therefore find that the resistance of an object is inversely proportional to its cross-sectional area.



f / A fat pipe has less resistance than a skinny pipe.

An analogous relationship holds for water pipes, which is why high-flow trunk lines have to have large cross-sectional areas. To make lots of water (current) flow through a skinny pipe, we'd need an impractically large pressure (voltage) difference.

Example 14: Incorrect readings from a voltmeter



g / A voltmeter is really an ammeter with an internal resistor. When we measure the voltage difference across a resistor, 1, we are really constructing a parallel resistance circuit, 2.

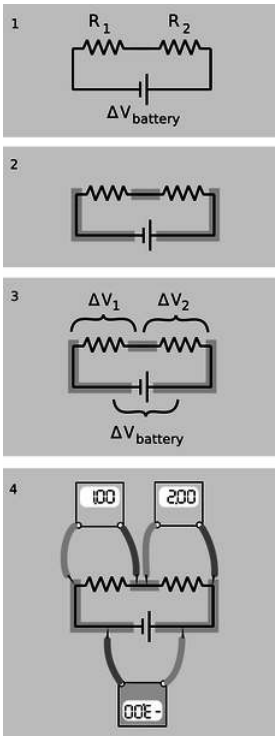
A voltmeter is really just an ammeter with an internal resistor, and we use a voltmeter in parallel with the thing that we're trying to measure the voltage difference across. This means that any time we measure the voltage drop across a resistor, we're essentially putting two resistors in parallel. The ammeter inside the voltmeter can be ignored for the purpose of analyzing what how current flows in the circuit, since it is essentially just some coiled-up wire with a very low resistance.

Now if we are carrying out this measurement on a resistor that is part of a larger circuit, we have changed the behavior of the circuit through our act of measuring. It is as though we had modified the circuit by replacing the resistance R with the smaller equivalent resistance of R and R_v in parallel. It is for this reason that voltmeters are built with the largest possible internal resistance. As a numerical example, if we use a voltmeter with an internal resistance of 1 $M\Omega$ to measure the voltage drop across a one-ohm resistor, the equivalent resistance is 0.999999 Ω, which is not different enough to make any difference. But if we tried to use the same voltmeter to measure the voltage drop across a $2 - M\Omega$ resistor, we would be reducing the resistance of that part of the circuit by a factor of three, which would produce a drastic change in the behavior of the whole circuit.

This is the reason why you can't use a voltmeter to measure the voltage difference between two different points in mid-air, or between the ends of a piece of wood. This is by no means a stupid thing to want to do, since the world around us is not a constant-voltage environment, the most extreme example being when an electrical storm is brewing. But it will not work with an ordinary voltmeter because the resistance of the air or the wood is many gigaohms. The effect of waving a pair of voltmeter probes around in the air is that we provide a reuniting path for the positive and negative charges that have been separated --- through the voltmeter itself, which is a good conductor compared to the air. This reduces to zero the voltage difference we were trying to measure.

In general, a voltmeter that has been set up with an open circuit (or a very large resistance) between its probes is said to be “floating.” An old-fashioned analog voltmeter of the type described here will read zero when left floating, the same as when it was sitting on the shelf. A floating digital voltmeter usually shows an error message.

9.2.3 Series resistances



h / 1. A battery drives current through two resistors in series. 2. There are three constant-voltage regions. 3. The three voltage differences are related. 4. If the meter crab-walks around the circuit without flipping over or crossing its legs, the resulting voltages have plus and minus signs that make them add up to zero.

The two basic circuit layouts are parallel and series, so a pair of resistors in series, h/1, is another of the most basic circuits we can make. By conservation of charge, all the current that flows through one resistor must also flow through the other (as well as through the battery):

$$I_1 = I_2.$$

The only way the information about the two resistance values is going to be useful is if we can apply Ohm's law, which will relate the resistance of each resistor to the current flowing through it and the voltage difference across it. Figure h/2 shows the three constant-voltage areas. Voltage differences are more physically significant than voltages, so we define symbols for the voltage differences across the two resistors in figure h/3.

We have three constant-voltage areas, with symbols for the difference in voltage between every possible pair of them. These three voltage differences must be related to each other. It is as though I tell you that Fred is a foot taller than Ginger, Ginger is a foot taller than Sally, and Fred is two feet taller than Sally. The information is redundant, and you really only needed two of the three pieces of data to infer the third. In the case of our voltage differences, we have

$$|\Delta V_1| + |\Delta V_2| = |\Delta V_{battery}|\,.$$

The absolute value signs are because of the ambiguity in how we define our voltage differences. If we reversed the two probes of the voltmeter, we would get a result with the opposite sign. Digital voltmeters will actually provide a minus sign on the screen if the wire connected to the "V" plug is lower in voltage than the one connected to the "COM" plug. Analog voltmeters pin the needle against a peg if you try to use them to measure negative voltages, so you have to fiddle to get the leads connected the right way, and then supply any necessary minus sign yourself.

Figure h/4 shows a standard way of taking care of the ambiguity in signs. For each of the three voltage measurements around the loop, we keep the same probe (the darker one) on the clockwise side. It is as though the voltmeter was sidling around the circuit like a crab, without ever "crossing its legs." With this convention, the relationship among the voltage drops becomes

$$\Delta V_1 + \Delta V_2 = -\Delta V_{battery},$$

or, in more symmetrical form,

$$\Delta V_1 + \Delta V_2 + \Delta V_{battery} = 0.$$

More generally, this is known as the loop rule for analyzing circuits:

Assuming the standard convention for plus and minus signs, the sum of the voltage drops around any closed loop in a circuit must be zero.

Looking for an exception to the loop rule would be like asking for a hike that would be downhill all the way and that would come back to its starting point!

For the circuit we set out to analyze, the equation

$$\Delta V_1 + \Delta V_2 + \Delta V_{battery} = 0$$

can now be rewritten by applying Ohm's law to each resistor:

$$I_1 R_1 + I_2 R_2 + \Delta V_{battery} = 0.$$

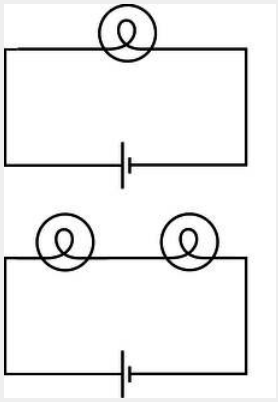
The currents are the same, so we can factor them out:

$$I\,(R_1 + R_2) + \Delta V_{battery} = 0,$$

and this is the same result we would have gotten if we had been analyzing a one-resistor circuit with resistance $R_1 + R_2$. Thus the equivalent resistance of resistors in series equals the sum of their resistances.

Example 15: Two lightbulbs in series

▹ If two identical lightbulbs are placed in series, how do their brightnesses compare with the brightness of a single bulb?



i / Example 15.

▹ Taken as a whole, the pair of bulbs act like a doubled resistance, so they will draw half as much current from the wall. Each bulb will be dimmer than a single bulb would have been.

The total power dissipated by the circuit is $I\Delta V$. The voltage drop across the whole circuit is the same as before, but the current is halved, so the two-bulb circuit draws half as much total power as the one-bulb circuit. Each bulb draws one-quarter of the normal power.

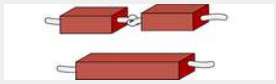
Roughly speaking, we might expect this to result in one quarter the light being produced by each bulb, but in reality lightbulbs waste quite a high percentage of their power in the form of heat and wavelengths of light that are not visible (infrared and ultraviolet). Less light will be produced, but it's hard to predict exactly how much less, since the efficiency of the bulbs will be changed by operating them under different conditions.

Example 16: More than two equal resistances in series

By straightforward application of the divide-and-conquer technique discussed in the previous section, we find that the equivalent resistance of N identical resistances R in series will be NR.

Example 17: Dependence of resistance on length

In the previous section, we proved that resistance is inversely proportional to cross-sectional area. By equivalent reason about resistances in series, we find that resistance is proportional to length. Analogously, it is harder to blow through a long straw than through a short one.



j / Doubling the length of a resistor is like putting two resistors in series. The resistance is doubled.

Putting the two arguments together, we find that the resistance of an object with straight, parallel sides is given by

$$R = (\text{constant}) \cdot L/A$$

The proportionality constant is called the resistivity, and it depends only on the substance of which the object is made. A resistivity measurement could be used, for instance, to help identify a sample of an unknown substance.

Example 18: Choice of high voltage for power lines

Thomas Edison got involved in a famous technological controversy over the voltage difference that should be used for electrical power lines. At this time, the public was unfamiliar with electricity, and easily scared by it. The president of the United States, for instance, refused to have electrical lighting in the White House when it first became commercially available because he considered it unsafe, preferring the known fire hazard of oil lamps to the mysterious dangers of electricity. Mainly as a way to overcome public fear, Edison believed that power should be transmitted using small voltages, and he publicized his opinion by giving demonstrations at which a dog was lured into position to be killed by a large voltage difference between two sheets of metal on the ground. (Edison's opponents also advocated alternating current rather than direct current, and AC is more dangerous than DC as well. As we will discuss later, AC can be easily stepped up and down to the desired voltage level using a device called a transformer.)

Now if we want to deliver a certain amount of power P_L to a load such as an electric lightbulb, we are constrained only by the equation $P_L = I\Delta V_L$. We can deliver any amount of power we wish, even with a low voltage, if we are willing to use large currents. Modern electrical distribution networks, however, use dangerously high voltage differences of tens of thousands of volts. Why did Edison lose the debate?

It boils down to money. The electric company must deliver the amount of power P_L desired by the customer through a transmission line whose resistance R_T is fixed by economics and geography. The same current flows through both the load and the transmission line, dissipating power usefully in the former and wastefully in the latter. The efficiency of the system is

$$\begin{aligned}\text{efficiency} &= \frac{\text{power paid for by the customer}}{\text{power paid for by the utility}} \\ &= \frac{P_L}{P_L + P_T} \\ &= \frac{1}{1 + P_T/P_L}\end{aligned}$$

Putting ourselves in the shoes of the electric company, we wish to get rid of the variable P_T, since it is something we control only indirectly by our choice of ΔV_T and I. Substituting $P_T = I\Delta V_T$, we find

$$\text{efficiency} = \frac{1}{1 + \frac{I\Delta V_T}{P_L}}$$

We assume the transmission line (but not necessarily the load) is ohmic, so substituting $\Delta V_T = IR_T$ gives

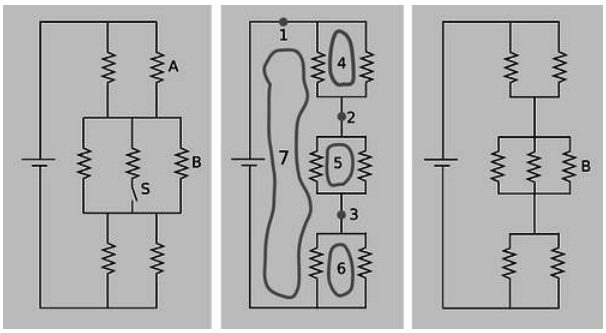
$$efficiency = \frac{1}{1 + \frac{I^2 R_T}{P_L}}$$

This quantity can clearly be maximized by making I as small as possible, since we will then be dividing by the smallest possible quantity on the bottom of the fraction. A low-current circuit can only deliver significant amounts of power if it uses high voltages, which is why electrical transmission systems use dangerous high voltages.

Example 19: Getting killed by your ammeter

As with a voltmeter, an ammeter can give erroneous readings if it is used in such a way that it changes the behavior the circuit. An ammeter is used in series, so if it is used to measure the current through a resistor, the resistor's value will effectively be changed to $R + R_a$, where R_a is the resistance of the ammeter. Ammeters are designed with very low resistances in order to make it unlikely that $R + R_a$ will be significantly different from R.

In fact, the real hazard is death, not a wrong reading! Virtually the only circuits whose resistances are significantly less than that of an ammeter are those designed to carry huge currents. An ammeter inserted in such a circuit can easily melt. When I was working at a laboratory funded by the Department of Energy, we got periodic bulletins from the DOE safety office about serious accidents at other sites, and they held a certain ghoulish fascination. One of these was about a DOE worker who was completely incinerated by the explosion created when he inserted an ordinary Radio Shack ammeter into a high-current circuit. Later estimates showed that the heat was probably so intense that the explosion was a ball of plasma --- a gas so hot that its atoms have been ionized.



k / Example 20.

Example 20: A complicated circuit

▹ All seven resistors in the left-hand panel of figure k are identical. Initially, the switch S is open as shown in the figure, and the current through resistor A is I_o. The switch is then closed. Find the current through resistor B, after the switch is closed, in terms of I_o.

▹ The second panel shows the circuit redrawn for simplicity, in the initial condition with the switch open. When the switch is open, no current can flow through the central resistor, so we may as well ignore it. I've also redrawn the junctions, without changing what's connected to what. This is the kind of mental rearranging that you'll eventually learn to do automatically from experience with analyzing circuits. The redrawn version makes it easier to see what's happening with the current. Charge is conserved, so any charge that flows past point 1 in the circuit must also flow past points 2 and 3. This would have been harder to reason about by applying the junction rule to the original version, which appears to have nine separate junctions.

In the new version, it's also clear that the circuit has a great deal of symmetry. We could flip over each parallel pair of identical resistors without changing what's connected to what, so that makes it clear that the voltage drops and currents must be equal for the members of each pair. We can also prove this by using the loop rule. The loop rule says that the two voltage drops in loop 4 must be equal, and similarly for loops 5 and 6. Since the resistors obey Ohm's law, equal voltage drops across them also imply equal currents. That means that when the current at point 1 comes to the top junction, exactly half of it goes through each resistor. Then the current reunites at 2, splits between the next pair, and so on. We conclude that each of the six resistors in the circuit experiences the same voltage drop and the same current. Applying the loop rule to loop 7, we find that the sum of the three voltage drops across the three left-hand resistors equals the battery's voltage, V, so each resistor in the circuit experiences a voltage drop $V/3$. Letting R stand for the resistance of one of the resistors, we find that the current through resistor B, which is the same as the currents through all the others, is given by $I_o = V/3R$.

We now pass to the case where the switch is closed, as shown in the third panel. The battery's voltage is the same as before, and each resistor's resistance is the same, so we can still use the same symbols V and R for them. It is no longer true, however, that each resistor feels a voltage drop $V/3$. The equivalent resistance of the whole circuit is $R/2 + R/3 + R/2 = 4R/3$, so the total current drawn from the battery is $3V/4R$. In the middle group of resistors, this current is split three ways, so the new current through B is $(1/3)(3V/4R) = V/4R = 3I_o/4$.

Interpreting this result, we see that it comes from two effects that partially cancel. Closing the switch reduces the equivalent resistance of the circuit by giving charge another way to flow, and increases the amount of current drawn from the battery. Resistor B, however, only gets a 1/3 share of this greater current, not 1/2. The second effect turns out to be bigger than the first, and therefore the current through resistor B is lessened over all.

Discussion Question

◊ We have stated the loop rule in a symmetric form where a series of voltage drops adds up to zero. To do this, we had to define a standard way of connecting the voltmeter to the circuit so that the plus and minus signs would come out right. Suppose we wish to restate the junction rule in a similar symmetric way, so that instead of equating the current coming in to the current going out, it simply states that a certain sum of currents at a junction adds up to zero. What standard way of inserting the ammeter would we have to use to make this work?

Contributors

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

This page titled 10.2: Parallel and Series Circuits is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

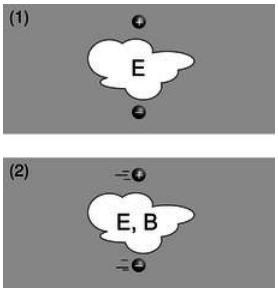
- **10.2: Parallel and Series Circuits** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

10.3: More About the Magnetic Field

11.1.1 Magnetic forces

In this chapter, I assume you know a few basic ideas about Einstein's theory of relativity, as described in sections 7.1 and 7.2. Unless your typical workday involves rocket ships or particle accelerators, all this relativity stuff might sound like a description of some bizarre futuristic world that is completely hypothetical. There is, however, a relativistic effect that occurs in everyday life, and it is obvious and dramatic: magnetism. Magnetism, as we discussed previously, is an interaction between a moving charge and another moving charge, as opposed to electric forces, which act between any pair of charges, regardless of their motion. Relativistic effects are weak for speeds that are small compared to the speed of light, and the average speed at which electrons drift through a wire is quite low (centimeters per second, typically), so how can relativity be behind an impressive effect like a car being lifted by an electromagnet hanging from a crane? The key is that matter is almost perfectly electrically neutral, and electric forces therefore cancel out almost perfectly. Magnetic forces really aren't very strong, but electric forces are even weaker.

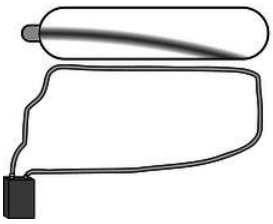
What about the word “relativity” in the name of the theory? It would seem problematic if moving charges interact differently than stationary charges, since motion is a matter of opinion, depending on your frame of reference. Magnetism, however, comes not to destroy relativity but to fulfill it. Magnetic interactions *must* exist according to the theory of relativity. To understand how this can be, consider how time and space behave in relativity. Observers in different frames of reference disagree about the lengths of measuring sticks and the speeds of clocks, but the laws of physics are valid and self-consistent in either frame of reference. Similarly, observers in different frames of reference disagree about what electric and magnetic fields and forces there are, but they agree about concrete physical events.



a / The pair of charged particles, as seen in two different frames of reference.

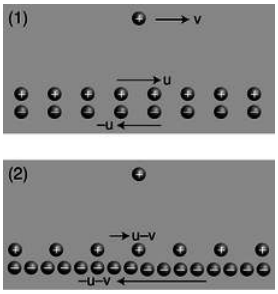
For instance, figure a/1 shows two particles, with opposite charges, which are not moving at a particular moment in time. An observer in this frame of reference says there are electric fields around the particles, and predicts that as time goes on, the particles will begin to accelerate towards one another, eventually colliding. A different observer, a/2, says the particles are moving. This observer also predicts that the particles will collide, but explains their motion in terms of both an electric field, $\mathbf{E}$, and a magnetic field, $\mathbf{B}$. As we'll see shortly, the magnetic field is *required* in order to maintain consistency between the predictions made in the two frames of reference.

To see how this really works out, we need to find a nice simple example that is easy to calculate. An example like figure a is *not* easy to handle, because in the second frame of reference, the moving charges create fields that change over time at any given location. Examples like figure b are easier, because there is a steady flow of charges, and all the fields stay the same over time.[1] What is remarkable about this demonstration is that there can be no electric fields acting on the electron beam at all, since the total charge density throughout the wire is zero. Unlike figure a/2, figure b is purely magnetic.



b / A large current is created by shorting across the leads of the battery. The moving charges in the wire attract the moving charges in the electron beam, causing the electrons to curve.

To see why this must occur based on relativity, we make the mathematically idealized model shown in figure c.



c / A charged particle and a current, seen in two different frames of reference. The second frame is moving at velocity v with respect to the first frame, so all the velocities have v subtracted from them. (As discussed in the main text, this is only approximately correct.)

The charge by itself is like one of the electrons in the vacuum tube beam of figure b, and a pair of moving, infinitely long line charges has been substituted for the wire. The electrons in a real wire are in rapid thermal motion, and the current is created only by a slow drift superimposed on this chaos. A second deviation from reality is that in the real experiment, the protons are at rest with respect to the tabletop, and it is the electrons that are in motion, but in c/1 we have the positive charges moving in one direction and the negative ones moving the other way. If we wanted to, we could construct a third frame of reference in which the positive charges were at rest, which would be more like the frame of reference fixed to the tabletop in the real demonstration. However, as we'll see shortly, frames c/1 and c/2 are designed so that they are particularly easy to analyze. It's important to note that even though the two line charges are moving in opposite directions, their currents don't cancel. A negative charge moving to the left makes a current that goes to the right, so in frame c/1, the total current is twice that contributed by either line charge.

Frame 1 is easy to analyze because the charge densities of the two line charges cancel out, and the electric field experienced by the lone charge is therefore zero:

$$\mathbf{E}_1 = 0$$

In frame 1, any force experienced by the lone charge must therefore be attributed solely to magnetism.

Frame 2 shows what we'd see if we were observing all this from a frame of reference moving along with the lone charge. Why don't the charge densities also cancel in this frame? Here's where the relativity comes in. Relativity tells us that moving objects appear contracted to an observer who is not moving along with them. Both line charges are in motion in both frames of reference, but in frame 1, the line charges were moving at equal speeds, so their contractions were equal, and their charge densities canceled out. In frame 2, however, their speeds are unequal. The positive charges are moving more slowly than in frame 1, so in frame 2 they are less contracted. The negative charges are moving more quickly, so their contraction is greater now. Since the charge densities don't cancel, there is an electric field in frame 2, which points into the wire, attracting the lone charge. Furthermore, the attraction felt by the lone charge must be purely electrical, since the lone charge is at rest in this frame of reference, and magnetic effects occur only between moving charges and other moving charges.[2]

To summarize, frame 1 displays a purely magnetic attraction, while in frame 2 it is purely electrical.

Now we can calculate the force in frame 2, and equating it to the force in frame 1, we can find out how much magnetic force occurs. To keep the math simple, and to keep from assuming too much about your knowledge of relativity, we're going to carry out this whole calculation in the approximation where all the speeds are fairly small compared to the speed of light.[3] For instance, if we find an expression such as $(v/c)^2 + (v/c)^4$, we will assume that the fourth-order term is negligible by comparison. This is known as a calculation "to leading order in v/c." In fact, I've already used the leading-order approximation twice without saying so! The first time I used it implicitly was in figure c, where I assumed that the velocities of the two line charges were $u - v$ and $-u - v$. Relativistic velocities don't just combine by simple addition and subtraction like this, but this is an effect we can ignore in the present approximation. The second sleight of hand occurred when I stated that we could equate the forces in the two frames of

reference. Force, like time and distance, is distorted relativistically when we change from one frame of reference to another. Again, however, this is an effect that we can ignore to the desired level of approximation.

Let $\pm\lambda$ be the charge per unit length of each line charge without relativistic contraction, i.e., in the frame moving with that line charge. Using the approximation $\gamma = (1 - v^2/c^2)^{-1/2} \approx 1 + v^2/2c^2$ for $v \ll c$, the total charge per unit length in frame 2 is

$$\begin{aligned}\lambda_{total,\,2} &\approx \lambda\left[1 + \frac{(u-v)^2}{2c^2}\right] - \lambda\left[1 + \frac{(-u-v)^2}{2c^2}\right] \\ &= \frac{-2\lambda uv}{c^2}.\end{aligned}$$

Let R be the distance from the line charge to the lone charge. Applying Gauss' law to a cylinder of radius R centered on the line charge, we find that the magnitude of the electric field experienced by the lone charge in frame 2 is

$$E = \frac{4k\lambda uv}{c^2 R},$$

and the force acting on the lone charge q is

$$F = \frac{4k\lambda quv}{c^2 R}.$$

In frame 1, the current is $I = 2\lambda_1 u$ (see homework problem 5), which we can approximate as $I = 2\lambda u$, since the current, unlike $\lambda_{total,\,2}$, doesn't vanish completely without the relativistic effect. The magnetic force on the lone charge q due to the current I is

$$F = \frac{2kIqv}{c^2 R}.$$

Discussion Question

◊ Resolve the following paradox concerning the argument given in this section. We would expect that at any given time, electrons in a solid would be associated with protons in a definite way. For simplicity, let's imagine that the solid is made out of hydrogen (which actually does become a metal under conditions of very high pressure). A hydrogen atom consists of a single proton and a single electron. Even if the electrons are moving and forming an electric current, we would imagine that this would be like a game of musical chairs, with the protons as chairs and the electrons as people. Each electron has a proton that is its “friend,” at least for the moment. This is the situation shown in figure c/1. How, then, can an observer in a different frame see the electrons and protons as not being paired up, as in c/2?

11.1.2 The magnetic field

Definition in terms of the force on a moving particle

With electricity, it turned out to be useful to define an electric field rather than always working in terms of electric forces. Likewise, we want to define a magnetic field, **B**. Let's look at the result of the preceding subsection for insight. The equation

$$F = \frac{2kIqv}{c^2 R}$$

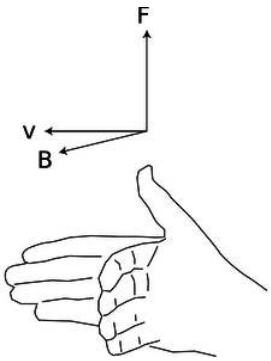
shows that when we put a moving charge near other moving charges, there is an extra magnetic force on it, in addition to any electric forces that may exist. Equations for electric forces always have a factor of k in front --- the Coulomb constant k is called the coupling constant for electric forces. Since magnetic effects are relativistic in origin, they end up having a factor of k/c^2 instead of just k. In a world where the speed of light was infinite, relativistic effects, including magnetism, would be absent, and the coupling constant for magnetism would be zero. A cute feature of the metric system is that we have $k/c^2 = 10^{-7}\ \mathrm{N \cdot s^2/C^2}$ exactly, as a matter of definition.

Naively, we could try to work by analogy with the electric field, and define the magnetic field as the magnetic force per unit charge. However, if we think of the lone charge in our example as the test charge, we'll find that this approach fails, because the force depends not just on the test particle's charge, but on its velocity, v, as well. Although we only carried out calculations for the case where the particle was moving parallel to the wire, in general this velocity is a vector, **v**, in three dimensions. We can also anticipate that the magnetic field will be a vector. The electric and gravitational fields are vectors, and we expect intuitively based on our experience with magnetic compasses that a magnetic field has a particular direction in space. Furthermore, reversing the

current I in our example would have reversed the force, which would only make sense if the magnetic field had a direction in space that could be reversed. Summarizing, we think there must be a magnetic field vector **B**, and the force on a test particle moving through a magnetic field is proportional both to the **B** vector and to the particle's own **v** vector. In other words, the magnetic force vector **F** is found by some sort of vector multiplication of the vectors **v** and **B**. As proved on page 912, however, there is only one physically useful way of defining such a multiplication, which is the cross product.

We therefore define the magnetic field vector, **B**, as the vector that determines the force on a charged particle according to the following rule:

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \text{[definition of the magnetic field]}$$



d / The right-hand relationship between the velocity of a positively charged particle, the magnetic field through which it is moving, and the magnetic force on it.

From this definition, we see that the magnetic field's units are $\mathrm{N \cdot s/C \cdot m}$, which are usually abbreviated as teslas, $\mathrm{1\ T = 1\ N \cdot s/C \cdot m}$. The definition implies a right-hand-rule relationship among the vectors, figure d, if the charge q is positive, and the opposite handedness if it is negative.



e / The unit of magnetic field, the tesla, is named after Serbian-American inventor Nikola Tesla.

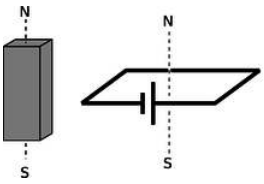
This is not just a definition but a bold prediction! Is it really true that for any point in space, we can always find a vector **B** that successfully predicts the force on any passing particle, regardless of its charge and velocity vector? Yes --- it's not obvious that it can be done, but experiments verify that it can. How? Well for example, the cross product of parallel vectors is zero, so we can try particles moving in various directions, and hunt for the direction that produces zero force; the **B** vector lies along that line, in either the same direction the particle was moving, or the opposite one. We can then go back to our data from one of the other cases, where the force was nonzero, and use it to choose between these two directions and find the magnitude of the **B** vector. We could then verify that this vector gave correct force predictions in a variety of other cases.

Even with this empirical reassurance, the meaning of this equation is not intuitively transparent, nor is it practical in most cases to measure a magnetic field this way. For these reasons, let's look at an alternative method of defining the magnetic field which, although not as fundamental or mathematically simple, may be more appealing.

Definition in terms of the torque on a dipole

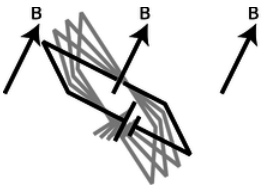
A compass needle in a magnetic field experiences a torque which tends to align it with the field. This is just like the behavior of an electric dipole in an electric field, so we consider the compass needle to be a *magnetic dipole*. In subsection 10.1.3 on page 567, we gave an alternative definition of the electric field in terms of the torque on an electric dipole.

To define the strength of a magnetic field, however, we need some way of defining the strength of a test dipole, i.e., we need a definition of the magnetic dipole moment. We could use an iron permanent magnet constructed according to certain specifications, but such an object is really an extremely complex system consisting of many iron atoms, only some of which are aligned with each other. A more fundamental standard dipole is a square current loop. This could be little resistive circuit consisting of a square of wire shorting across a battery, f.



f / A standard dipole made from a square loop of wire shorting across a battery. It acts very much like a bar magnet, but its strength is more easily quantified.

Applying $\mathbf{F} = \mathbf{v} \times \mathbf{B}$, we find that such a loop, when placed in a magnetic field, g, experiences a torque that tends to align plane so that its interior "face" points in a certain direction.



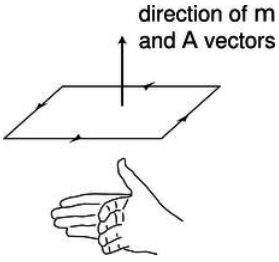
g / A dipole tends to align itself to the surrounding magnetic field.

Since the loop is symmetric, it doesn't care if we rotate it like a wheel without changing the plane in which it lies. It is this preferred facing direction that we will end up using as our alternative definition of the magnetic field.

If the loop is out of alignment with the field, the torque on it is proportional to the amount of current, and also to the interior area of the loop. The proportionality to current makes sense, since magnetic forces are interactions between moving charges, and current is a measure of the motion of charge. The proportionality to the loop's area is also not hard to understand, because increasing the length of the sides of the square increases both the amount of charge contained in this circular "river" and the amount of leverage supplied for making torque. Two separate physical reasons for a proportionality to length result in an overall proportionality to length squared, which is the same as the area of the loop. For these reasons, we define the magnetic dipole moment of a square current loop as

$$\mathbf{m} = I\mathbf{A},$$

where the direction of the vectors is defined as shown in figure h.



h / The **m** and **A** vectors.

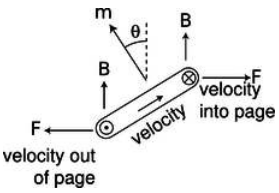
We can now give an alternative definition of the magnetic field:

The magnetic field vector, $\mathbf{B}$, at any location in space is defined by observing the torque exerted on a magnetic test dipole $\mathbf{m}_t$ consisting of a square current loop. The field's magnitude is

$$|\mathbf{B}| = \frac{\tau}{|\mathbf{m}_t| \sin\theta},$$

where θ is the angle between the dipole vector and the field. This is equivalent to the vector cross product $\boldsymbol{\tau} = \mathbf{m}_t \times \mathbf{B}$.

Let's show that this is consistent with the previous definition, using the geometry shown in figure i.



i / The torque on a current loop in a magnetic field. The current comes out of the page, goes across, goes back into the page, and then back across the other way in the hidden side of the loop.

The velocity vector that point in and out of the page are shown using the convention defined in figure j.

⊙ out of the page

⊗ into the page

j / A vector coming out of the page is shown with the tip of an arrowhead. A vector going into the page is represented using the tailfeathers of the arrow.

Let the mobile charge carriers in the wire have linear density λ, and let the sides of the loop have length h, so that we have $I = \lambda v$, and $m = h^2 \lambda v$. The only nonvanishing torque comes from the forces on the left and right sides. The currents in these sides are perpendicular to the field, so the magnitude of the cross product $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$ is simply $|\mathbf{F}| = qvB$. The torque supplied by each of these forces is $\mathbf{r} \times \mathbf{F}$, where the lever arm $\mathbf{r}$ has length $h/2$, and makes an angle θ with respect to the force vector. The magnitude of the total torque acting on the loop is therefore

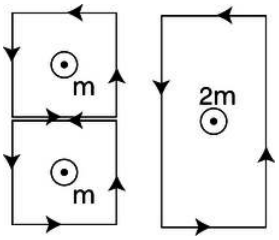
$$\begin{aligned} |\tau| &= 2\frac{h}{2}|\mathbf{F}| \sin\theta \\ &= h\, qvB \sin\theta, \end{aligned}$$

and substituting $q = \lambda h$ and $v = m/h^2\lambda$, we have

$$\begin{aligned} |\tau| &= h\, \lambda h \frac{m}{h^2 \lambda} B \sin\theta \\ &= mB \sin\theta, \end{aligned}$$

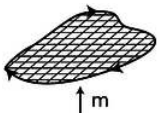
which is consistent with the second definition of the field.

It undoubtedly seems artificial to you that we have discussed dipoles only in the form of a square loop of current. A permanent magnet, for example, is made out of atomic dipoles, and atoms aren't square! However, it turns out that the shape doesn't matter. To see why this is so, consider the additive property of areas and dipole moments, shown in figure k.



k / Dipole vectors can be added.

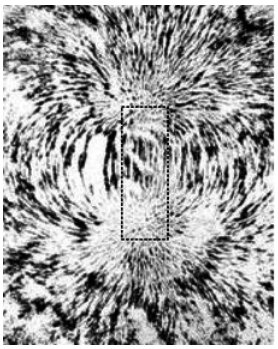
Each of the square dipoles has a dipole moment that points out of the page. When they are placed side by side, the currents in the adjoining sides cancel out, so they are equivalent to a single rectangular loop with twice the area. We can break down any irregular shape into little squares, as shown in figure l, so the dipole moment of any planar current loop can be calculated based on its area, regardless of its shape.



l / An irregular loop can be broken up into little squares.

Example 1: The magnetic dipole moment of an atom

Let's make an order-of-magnitude estimate of the magnetic dipole moment of an atom. A hydrogen atom is about 10^{-10} m in diameter, and the electron moves at speeds of about $10^{-2}c$. We don't know the shape of the orbit, and indeed it turns out that according to the principles of quantum mechanics, the electron doesn't even have a well-defined orbit, but if we're brave, we can still estimate the dipole moment using the cross-sectional area of the atom, which will be on the order of $(10^{-10}\ \mathrm{m})^2 = 10^{-20}\ \mathrm{m}^2$. The electron is a single particle, not a steady current, but again we throw caution to the winds, and estimate the current it creates as $e/\Delta t$, where Δt, the time for one orbit, can be estimated by dividing the size of the atom by the electron's velocity. (This is only a rough estimate, and we don't know the shape of the orbit, so it would be silly, for instance, to bother with multiplying the diameter by π based on our intuitive visualization of the electron as moving around the circumference of a circle.) The result for the dipole moment is $m \sim 10^{-23}\ \mathrm{A \cdot m^2}$.
Should we be impressed with how small this dipole moment is, or with how big it is, considering that it's being made by a single atom? Very large or very small numbers are never very interesting by themselves. To get a feeling for what they mean, we need to compare them to something else. An interesting comparison here is to think in terms of the total number of atoms in a typical object, which might be on the order of 10^{26} (Avogadro's number). Suppose we had this many atoms, with their moments all aligned. The total dipole moment would be on the order of $10^{3}\ \mathrm{A \cdot m^2}$, which is a pretty big number. To get a dipole moment this strong using human-scale devices, we'd have to send a thousand amps of current through a one-square meter loop of wire! The insight to be gained here is that, even in a permanent magnet, we must not have all the atoms perfectly aligned, because that would cause more spectacular magnetic effects than we really observe. Apparently, nearly all the atoms in such a magnet are oriented randomly, and do not contribute to the magnet's dipole moment.



m / The magnetic field pattern around a bar magnet is created by the superposition of the dipole fields of the individual iron atoms. Roughly speaking, it looks like the field of one big dipole, especially farther away from the magnet. Closer in, however, you can see a hint of the magnet's rectangular shape. The picture was made by placing iron filings on a piece of paper, and then bringing a magnet up underneath.

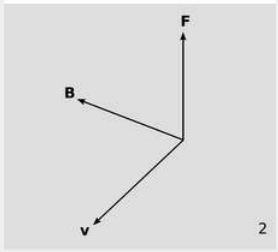
Discussion Questions

◊ The physical situation shown in figure c on page 648 was analyzed entirely in terms of forces. Now let's go back and think about it in terms of fields. The charge by itself up above the wire is like a test charge, being used to determine the magnetic and electric fields created by the wire. In figures c/1 and c/2, are there fields that are purely electric or purely magnetic? Are there fields that are a mixture of $\mathbf{E}$ and $\mathbf{B}$? How does this compare with the forces?

◊ Continuing the analysis begun in discussion question A, can we come up with a scenario involving some charged particles such that the fields are purely magnetic in one frame of reference but a mixture of $\mathbf{E}$ and $\mathbf{B}$ in another frame? How about an example where the fields are purely electric in one frame, but mixed in another? Or an example where the fields are purely electric in one frame, but purely magnetic in another?

Example 2: Magnetic levitation





n / Example 2.

In figure n, a small, disk-shaped permanent magnet is stuck on the side of a battery, and a wire is clasped loosely around the battery, shorting it. A large current flows through the wire. The electrons moving through the wire feel a force from the magnetic field made by the permanent magnet, and this force levitates the wire.

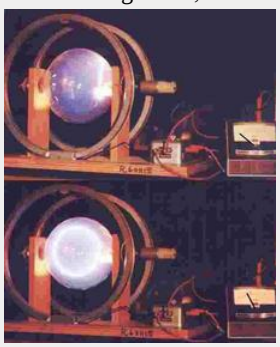
From the photo, it's possible to find the direction of the magnetic field made by the permanent magnet. The electrons in the copper wire are negatively charged, so they flow from the negative (flat) terminal of the battery to the positive terminal (the one with the bump, in front). As the electrons pass by the permanent magnet, we can imagine that they would experience a field either toward the magnet, or away from it, depending on which way the magnet was flipped when it was stuck onto the battery. By the right-hand rule (figure d on page 651), the field must be toward the battery.

Example 3: Nervous-system effects during an MRI scan<</th>

During an MRI scan of the head, the patient's nervous system is exposed to intense magnetic fields, and there are ions moving around in the nerves. The resulting forces on the ions can cause symptoms such as vertigo.

Example 4: A circular orbit

The magnetic force is always perpendicular to the motion of the particle, so it can never do any work, and a charged particle moving through a magnetic field does not experience any change in its kinetic energy: its velocity vector can change its direction, but not its magnitude. If the velocity vector is initially perpendicular to the field, then the curve of its motion will remain in the plane perpendicular to the field, so the magnitude of the magnetic force on it will stay the same. When an object experiences a force with constant magnitude, which is always perpendicular to the direction of its motion, the result is that it travels in a circle.



o / Magnetic forces cause a beam of electrons to move in a circle.

Figure o shows a beam of electrons in a spherical vacuum tube. In the top photo, the beam is emitted near the right side of the tube, and travels straight up. In the bottom photo, a magnetic field has been imposed by an electromagnet surrounding the vacuum tube; the ammeter on the right shows that the current through the electromagnet is now nonzero. We observe that the beam is bent into a circle.

self-check:

Infer the direction of the magnetic field. Don't forget that the beam is made of electrons, which are negatively charged!

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

Homework problem 12 is a quantitative analysis of circular orbits.

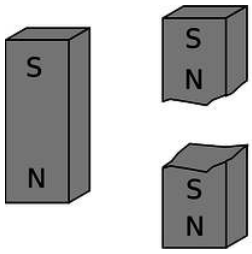
Example 5: A velocity filter

Suppose you see the electron beam in figure o, and you want to determine how fast the electrons are going. You certainly can't do it with a stopwatch! Physicists may also encounter situations where they have a beam of unknown charged particles, and they don't even know their charges. This happened, for instance, when alpha and beta radiation were discovered. One solution to this problem relies on the fact that the force experienced by a charged particle in an electric field, $\mathbf{F}_E = q\mathbf{E}$, is independent of its velocity, but the force due to a magnetic field, $\mathbf{F}_B = q\mathbf{v} \times \mathbf{B}$, isn't. One can send a beam of charged particles through a space containing both an electric and a magnetic field, setting up the fields so that the two forces will cancel out perfectly for a certain velocity. Note that since both forces are proportional to the charge of the particles, the cancellation is independent of charge. Such a *velocity filter* can be used either to determine the velocity of an unknown beam or particles, or to select from a beam of particles only those having velocities within a certain desired range. Homework problem 7 is an analysis of this application.

11.1.4 No magnetic monopoles

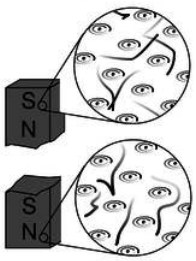
If you could play with a handful of electric dipoles and a handful of bar magnets, they would appear very similar. For instance, a pair of bar magnets wants to align themselves head-to-tail, and a pair of electric dipoles does the same thing. (It is unfortunately not that easy to make a permanent electric dipole that can be handled like this, since the charge tends to leak.)

You would eventually notice an important difference between the two types of objects, however. The electric dipoles can be broken apart to form isolated positive charges and negative charges. The two-ended device can be broken into parts that are not two-ended. But if you break a bar magnet in half, p, you will find that you have simply made two smaller two-ended objects.



p / You can't isolate the poles of a magnet by breaking it in half.

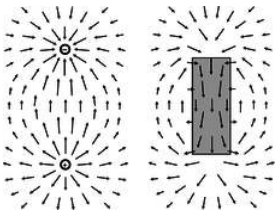
The reason for this behavior is not hard to divine from our microscopic picture of permanent iron magnets. An electric dipole has extra positive “stuff” concentrated in one end and extra negative in the other. The bar magnet, on the other hand, gets its magnetic properties not from an imbalance of magnetic “stuff” at the two ends but from the orientation of the rotation of its electrons. One end is the one from which we could look down the axis and see the electrons rotating clockwise, and the other is the one from which they would appear to go counterclockwise. There is no difference between the “stuff” in one end of the magnet and the other, q.



q / A magnetic dipole is made out of other dipoles, not out of monopoles.

Nobody has ever succeeded in isolating a single magnetic pole. In technical language, we say that magnetic *monopoles* not seem to exist. Electric monopoles *do* exist --- that's what charges are.

Electric and magnetic forces seem similar in many ways. Both act at a distance, both can be either attractive or repulsive, and both are intimately related to the property of matter called charge. (Recall that magnetism is an interaction between moving charges.) Physicists's aesthetic senses have been offended for a long time because this seeming symmetry is broken by the existence of electric monopoles and the absence of magnetic ones. Perhaps some exotic form of matter exists, composed of particles that are magnetic monopoles. If such particles could be found in cosmic rays or moon rocks, it would be evidence that the apparent asymmetry was only an asymmetry in the composition of the universe, not in the laws of physics. For these admittedly subjective reasons, there have been several searches for magnetic monopoles. Experiments have been performed, with negative results, to look for magnetic monopoles embedded in ordinary matter. Soviet physicists in the 1960's made exciting claims that they had created and detected magnetic monopoles in particle accelerators, but there was no success in attempts to reproduce the results there or at other accelerators. The most recent search for magnetic monopoles, done by reanalyzing data from the search for the top quark at Fermilab, turned up no candidates, which shows that either monopoles don't exist in nature or they are extremely massive and thus hard to create in accelerators.



r / Magnetic fields have no sources or sinks.

The nonexistence of magnetic monopoles means that unlike an electric field, a magnetic one, can never have sources or sinks. The magnetic field vectors lead in paths that loop back on themselves, without ever converging or diverging at a point, as in the fields

shown in figure r. Gauss' law for magnetism is therefore much simpler than Gauss' law for electric fields:

$$\Phi_B = \sum \mathbf{B}_j \cdot \mathbf{A}_j = 0$$

The magnetic flux through any closed surface is zero.

self-check:

Draw a Gaussian surface on the electric dipole field of figure r that has nonzero electric flux through it, and then draw a similar surface on the magnetic field pattern. What happens?

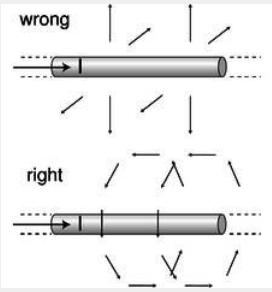
(answer in the back of the PDF version of the book)

Example 6: The field of a wire

▷ On page 650, we showed that a long, straight wire carrying current I exerts a magnetic force

$$F = \frac{2kIqv}{c^2R}$$

on a particle with charge q moving parallel to the wire with velocity v. What, then, is the magnetic field of the wire?



s / Example 6.

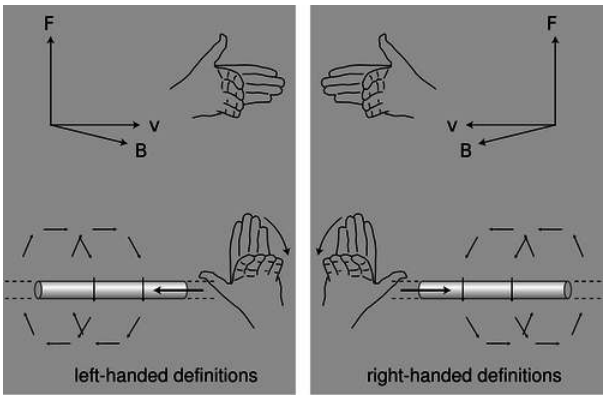
▷ Comparing the equation above to the first definition of the magnetic field, $\mathbf{F} = \mathbf{v} \times \mathbf{B}$, it appears that the magnetic field is one that falls off like $1/R$, where R is the distance from the wire. However, it's not so easy to determine the direction of the field vector. There are two other axes along which the particle could have been moving, and the brute-force method would be to carry out relativistic calculations for these cases as well. Although this would probably be enough information to determine the field, we don't want to do that much work.

Instead, let's consider what the possibilities are. The field can't be parallel to the wire, because a cross product vanishes when the two vectors are parallel, and yet we know from the case we analyzed that the force doesn't vanish when the particle is moving parallel to the wire. The other two possibilities that are consistent with the symmetry of the problem are shown in figure s. One is like a bottle brush, and the other is like a spool of thread. The bottle brush pattern, however, violates Gauss' law for magnetism. If we made a cylindrical Gaussian surface with its axis coinciding with the wire, the flux through it would *not* be zero. We therefore conclude that the spool-of-thread pattern is the correct one.[4] Since the particle in our example was moving perpendicular to the field, we have $|F| = |q|\,|v|\,|B|$, so

$$\begin{aligned} |B| &= \frac{|F|}{|q||v|} \\ &= \frac{2kI}{c^2R} \end{aligned}$$

11.1.5 Symmetry and handedness

Imagine that you establish radio contact with an alien on another planet. Neither of you even knows where the other one's planet is, and you aren't able to establish any landmarks that you both recognize. You manage to learn quite a bit of each other's languages, but you're stumped when you try to establish the definitions of left and right (or, equivalently, clockwise and counterclockwise). Is there any way to do it?



t / Left-handed and right-handed definitions.

If there was any way to do it without reference to external landmarks, then it would imply that the laws of physics themselves were asymmetric, which would be strange. Why should they distinguish left from right? The gravitational field pattern surrounding a star or planet looks the same in a mirror, and the same goes for electric fields. However, the magnetic field patterns shown in figure s seems to violate this principle. Could you use these patterns to explain left and right to the alien? No. If you look back at the definition of the magnetic field, it also contains a reference to handedness: the direction of the vector cross product. The aliens might have reversed their definition of the magnetic field, in which case their drawings of field patterns would look like mirror images of ours, as in the left panel of figure t.

Until the middle of the twentieth century, physicists assumed that any reasonable set of physical laws would have to have this kind of symmetry between left and right.



u / In this scene from Swan Lake, the the choreography has a symmetry with respect to left and right.

An asymmetry would be grotesque. Whatever their aesthetic feelings, they had to change their opinions about reality when experiments by C.S. Wu et al. showed that the weak nuclear force violates right-left symmetry!



v / C.S. Wu

It is still a mystery why right-left symmetry is observed so scrupulously in general, but is violated by one particular type of physical process.

Contributors

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

This page titled 10.3: More About the Magnetic Field is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

- **12.1: More About the Magnetic Field** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

10.4: Induced Electric Fields

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

This page titled 10.4: Induced Electric Fields is shared under a CC BY-SA license and was authored, remixed, and/or curated by Benjamin Crowell.

- **12.5: Induced Electric Fields** by Benjamin Crowell is licensed CC BY-SA 4.0. Original source: http://www.lightandmatter.com/cp/.

Index

A

acceleration
7.5: Acceleration
acid
6.8: Neutralization Reactions
activity series
6.6: Types of Chemical Reactions - Single and Double Displacement Reactions
air resistance
7.7: Projectile Motion
anions
3.5: Ions and Ionic Compounds
antinode
9.6: Sound Interference and Resonance- Standing Waves in Air Columns
Arrhenius definition
6.8: Neutralization Reactions
atomic mass
3.3: Masses of Atoms and Molecules
atomic mass unit
3.3: Masses of Atoms and Molecules
5.3: The Mole
atomic mass units
5.3: The Mole
atomic theory
3.2: Atomic Theory
atoms
3.2: Atomic Theory
average acceleration
7.5: Acceleration
average speed
7.4: Time, Velocity, and Speed
average velocity
7.4: Time, Velocity, and Speed
avogadros number
5.3: The Mole

B

balanced
6.2: The Chemical Equation
basal metabolic rate
8.8: Work, Energy, and Power in Humans
base
6.8: Neutralization Reactions
bel (units)
9.5: Sound Intensity and Sound Level
Boiling Point
5.6: Colligative Properties of Solutions
boiling point elevation
5.6: Colligative Properties of Solutions
boiling point elevation constant
5.6: Colligative Properties of Solutions

C

calorimeter
6.10: Enthalpy and Chemical Reactions
calorimetry
6.10: Enthalpy and Chemical Reactions
carbon tetrachloride
4.5: Nomenclature of Molecular Compounds
cations
3.5: Ions and Ionic Compounds
change in momentum
7.13: Impulse
chemical change
2.2: Basic Definitions
chemical energy
8.5: Conservation of Energy
chemical equation
6.2: The Chemical Equation
6.3: The Mole in Chemical Reactions
chemical properties
2.2: Basic Definitions
chemical reactions
6.3: The Mole in Chemical Reactions
6.10: Enthalpy and Chemical Reactions
coefficient
6.2: The Chemical Equation
colligative
5.6: Colligative Properties of Solutions
colligative properties
5.6: Colligative Properties of Solutions
combustion reaction
6.7: Composition, Decomposition, and Combustion Reactions
combustion reactions
6.7: Composition, Decomposition, and Combustion Reactions
composition
6.7: Composition, Decomposition, and Combustion Reactions
composition reaction
6.7: Composition, Decomposition, and Combustion Reactions
compound
2.2: Basic Definitions
concentrated
5.2: Definitions
concentration
5.2: Definitions
5.5: Dilutions and Concentrations
conversion factor
1.3: Converting Units
1.4: Other Units - Temperature and Density
5.3: The Mole
conversion factors
1.4: Other Units - Temperature and Density
6.3: The Mole in Chemical Reactions
coordinate systems
7.3: Vectors, Scalars, and Coordinate Systems
covalent
4.1: Compounds, Lewis Diagrams and Ionic Bonds
covalent bond
4.3: Covalent Bonding

D

deceleration
7.5: Acceleration
decibels (units)
9.5: Sound Intensity and Sound Level
decomposition
6.7: Composition, Decomposition, and Combustion Reactions
decomposition reaction
6.7: Composition, Decomposition, and Combustion Reactions
density
1.4: Other Units - Temperature and Density
diffraction fringes
9.9: Wave Optics
dilute
5.2: Definitions
dilution
5.5: Dilutions and Concentrations
dilution equation
5.5: Dilutions and Concentrations
dimensional analysis
1.3: Converting Units
dinitrogen tetroxide
4.5: Nomenclature of Molecular Compounds
dispersion force
4.7: Intermolecular Forces
dispersion forces
4.7: Intermolecular Forces
displacement
7.2: Displacement
7.3: Vectors, Scalars, and Coordinate Systems
distance
7.2: Displacement
7.3: Vectors, Scalars, and Coordinate Systems
distance traveled
7.2: Displacement
double bond
4.3: Covalent Bonding
dynamics
7.8: Development of Force Concept

E

efficiency
8.5: Conservation of Energy
elapsed time
7.4: Time, Velocity, and Speed
electrical energy
8.5: Conservation of Energy
electron
3.2: Atomic Theory
electrostatic potential map
4.6: Electronegativity and the Polar Covalent Bond
element
2.2: Basic Definitions
endothermic
6.10: Enthalpy and Chemical Reactions
energy
6.9: Energy
8.2: Work- The Scientific Definition
enthalpy
6.10: Enthalpy and Chemical Reactions

enthalpy change
6.10: Enthalpy and Chemical Reactions
exact number
1.3: Converting Units
exothermic
6.10: Enthalpy and Chemical Reactions
external force
7.8: Development of Force Concept

F
force
7.8: Development of Force Concept
Forces
7.10: Newton's Second Law of Motion- Concept of a System
freezing point depression
5.6: Colligative Properties of Solutions
freezing point depression constant
5.6: Colligative Properties of Solutions
frequency
9.4: Speed of Sound, Frequency, and Wavelength
friction
8.6: Nonconservative Forces

G
gimli flier
1.3: Converting Units
gravitational potential energy
8.4: Gravitational Potential Energy

H
harmonics
9.6: Sound Interference and Resonance- Standing Waves in Air Columns
hearing
9.2: Sound
heterogeneous mixture
2.2: Basic Definitions
homogeneous mixture
2.2: Basic Definitions
horsepower
8.7: Power
Huygens's principle
9.9: Wave Optics
hydrogen bonding
4.7: Intermolecular Forces
hypothesis
1.1: Chemistry as a Science

I
impulse
7.13: Impulse
index of refraction
9.8: Refraction
inertia
7.9: Newton's First Law of Motion- Inertia
infrasound
9.3: Hearing
instantaneous acceleration
7.5: Acceleration
instantaneous velocity
7.4: Time, Velocity, and Speed
intensity
9.5: Sound Intensity and Sound Level
intermolecular forces
4.7: Intermolecular Forces
ionic
4.1: Compounds, Lewis Diagrams and Ionic Bonds
ionic compound
3.5: Ions and Ionic Compounds
ionic compounds
3.5: Ions and Ionic Compounds
4.2: Nomenclature of Ionic Compounds
Ionic Formulas
3.5: Ions and Ionic Compounds
ions
3.5: Ions and Ionic Compounds
isolated system
6.9: Energy
IUPAC
3.4: Organization of the Elements - The Periodic Table

J
joule
6.9: Energy
8.2: Work- The Scientific Definition

K
kinematics
7.2: Displacement
7.7: Projectile Motion
kinetic energy
8.3: Kinetic Energy and the Work-Energy Theorem

L
law of conservation of energy
6.9: Energy
8.5: Conservation of Energy
law of conservation of matter
6.2: The Chemical Equation
Law of inertia
7.9: Newton's First Law of Motion- Inertia
limiting reagent
6.5: Limiting Reagents
Limiting Reagents
6.5: Limiting Reagents
linear momentum
7.12: Linear Momentum and Force
lithium cation
4.1: Compounds, Lewis Diagrams and Ionic Bonds
loudness
9.3: Hearing

M
mass
3.3: Masses of Atoms and Molecules
5.3: The Mole
7.9: Newton's First Law of Motion- Inertia
mass percentage
5.4: Quantitative Units of Concentration
matter
2.2: Basic Definitions
3.2: Atomic Theory
Melting Point
5.6: Colligative Properties of Solutions
metabolic rate
8.8: Work, Energy, and Power in Humans
metal
2.2: Basic Definitions
mixtures
2.2: Basic Definitions
model
7.4: Time, Velocity, and Speed
mol
6.3: The Mole in Chemical Reactions
molality
5.4: Quantitative Units of Concentration
molar masses
5.3: The Mole
molarity
5.4: Quantitative Units of Concentration
5.5: Dilutions and Concentrations
mole
5.3: The Mole
6.3: The Mole in Chemical Reactions
mole fraction
5.6: Colligative Properties of Solutions
molecular compounds
4.5: Nomenclature of Molecular Compounds
molecular mass
3.3: Masses of Atoms and Molecules
molecular surface
4.6: Electronegativity and the Polar Covalent Bond
molecule
4.3: Covalent Bonding
molecules
3.3: Masses of Atoms and Molecules
moles of solute
5.5: Dilutions and Concentrations
motion
7.7: Projectile Motion

N
net work
8.3: Kinetic Energy and the Work-Energy Theorem
neutralization reaction
6.8: Neutralization Reactions
neutron
3.2: Atomic Theory
Newton's third law of motion
7.11: Newton's Third Law of Motion- Symmetry in Forces
nitrogen monoxide
4.5: Nomenclature of Molecular Compounds
node
9.6: Sound Interference and Resonance- Standing Waves in Air Columns
nomenclature
4.2: Nomenclature of Ionic Compounds
4.5: Nomenclature of Molecular Compounds
nonconservative force
8.6: Nonconservative Forces

nonmetal
2.2: Basic Definitions
note
9.3: Hearing
nuclear energy
8.5: Conservation of Energy

O
octet rule
4.1: Compounds, Lewis Diagrams and Ionic Bonds
4.4: Exceptions to the Octet Rule
osmosis
5.6: Colligative Properties of Solutions
osmotic pressure
5.6: Colligative Properties of Solutions
overtones
9.6: Sound Interference and Resonance- Standing Waves in Air Columns

P
percentage composition by mass
5.4: Quantitative Units of Concentration
periodic table
3.4: Organization of the Elements - The Periodic Table
phase
2.2: Basic Definitions
4.7: Intermolecular Forces
phase to
4.7: Intermolecular Forces
phon (unit)
9.3: Hearing
physical change
2.2: Basic Definitions
physical properties
2.2: Basic Definitions
pitch
9.4: Speed of Sound, Frequency, and Wavelength
polyatomic ions
3.5: Ions and Ionic Compounds
position
7.2: Displacement
power
8.7: Power
precipitate
6.6: Types of Chemical Reactions - Single and Double Displacement Reactions
precipitation reaction
6.6: Types of Chemical Reactions - Single and Double Displacement Reactions
prefixes
1.2: Expressing Units
products
6.2: The Chemical Equation
projectile
7.7: Projectile Motion
Projectile motion
7.7: Projectile Motion
proton
3.2: Atomic Theory

Q
qualitative
5.2: Definitions
quantitative
5.2: Definitions

R
radiant energy
8.5: Conservation of Energy
range
7.7: Projectile Motion
Raoult's law
5.6: Colligative Properties of Solutions
reactants
6.2: The Chemical Equation

S
salt
6.8: Neutralization Reactions
scalar
7.3: Vectors, Scalars, and Coordinate Systems
scientific method
1.1: Chemistry as a Science
second law of motion
7.12: Linear Momentum and Force
semimetals
2.2: Basic Definitions
semipermeable membrane
5.6: Colligative Properties of Solutions
SI units
1.2: Expressing Units
Snell’s law of refraction
9.8: Refraction
solubility
5.2: Definitions
solubility rules
6.6: Types of Chemical Reactions - Single and Double Displacement Reactions
solute
5.2: Definitions
solutions
5.6: Colligative Properties of Solutions
solvent
5.2: Definitions
sound
9.2: Sound
sound intensity level
9.5: Sound Intensity and Sound Level
sound pressure level
9.5: Sound Intensity and Sound Level
speed of sound
9.4: Speed of Sound, Frequency, and Wavelength
state
2.2: Basic Definitions
stock solution
5.5: Dilutions and Concentrations
stoichiometry
6.4: Stoichiometry
sublimation
4.7: Intermolecular Forces
Substance
2.2: Basic Definitions
substances
4.7: Intermolecular Forces
sulfur dioxide
4.5: Nomenclature of Molecular Compounds
supersaturated
5.2: Definitions
system
6.9: Energy
7.10: Newton’s Second Law of Motion- Concept of a System

T
temperature
1.4: Other Units - Temperature and Density
thermal energy
8.5: Conservation of Energy
thermochemical equation
6.10: Enthalpy and Chemical Reactions
thrust
7.11: Newton’s Third Law of Motion- Symmetry in Forces
timbre
9.3: Hearing
time
7.4: Time, Velocity, and Speed
trajectory
7.7: Projectile Motion
triple covalent bond
4.3: Covalent Bonding

U
u
3.3: Masses of Atoms and Molecules
Ultrasound
9.3: Hearing
units
1.2: Expressing Units
unsaturated
5.2: Definitions
useful work
8.8: Work, Energy, and Power in Humans

V
valence expansion
4.4: Exceptions to the Octet Rule
valence shell
4.1: Compounds, Lewis Diagrams and Ionic Bonds
vapor pressure
5.6: Colligative Properties of Solutions
vapor pressure depression
5.6: Colligative Properties of Solutions
vapor pressure lowering
5.6: Colligative Properties of Solutions
vector
7.3: Vectors, Scalars, and Coordinate Systems
7.6: Kinematics in Two Dimensions - An Introduction

W

watt
8.7: Power

wavelength
9.4: Speed of Sound, Frequency, and Wavelength

weighted average
3.3: Masses of Atoms and Molecules

work
8.2: Work- The Scientific Definition

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Elementary Concepts in Physics and Chemistry

Webpages: 95

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 28.4% (27 pages)
- CC BY 4.0: 27.4% (26 pages)
- Undeclared: 21.1% (20 pages)
- CC BY-SA 4.0: 16.8% (16 pages)
- CC BY-NC 4.0: 6.3% (6 pages)

By Page

- Elementary Concepts in Physics and Chemistry - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Chapter 1 - Descriptions in the Physical Sciences - *Undeclared*
    - 1.1: Chemistry as a Science - *CC BY-NC-SA 4.0*
    - 1.2: Expressing Units - *CC BY-NC-SA 4.0*
    - 1.3: Converting Units - *CC BY-NC-SA 4.0*
    - 1.4: Other Units - Temperature and Density - *CC BY-NC-SA 4.0*
    - 1.5: 0.1 Introduction and Review - *CC BY-SA 4.0*
  - 2: Chapter 2 - Introduction to Chemistry - *Undeclared*
    - 2.1: Prelude to Chemistry - *CC BY-NC-SA 4.0*
    - 2.2: Basic Definitions - *CC BY-NC-SA 4.0*
  - 3: Chapter 3 - Atoms and Elemental Properties - *Undeclared*
    - 3.1: Prelude to Atoms, Molecules, and Ions - *CC BY-NC-SA 4.0*
    - 3.2: Atomic Theory - *CC BY-NC-SA 4.0*
    - 3.3: Masses of Atoms and Molecules - *CC BY-NC-SA 4.0*
    - 3.4: Organization of the Elements - The Periodic Table - *CC BY-SA 4.0*
    - 3.5: Ions and Ionic Compounds - *CC BY-NC-SA 4.0*
  - 4: Chapter 4 - Molecules, Bonding, and Forces - *Undeclared*
    - 4.1: Compounds, Lewis Diagrams and Ionic Bonds - *CC BY-SA 4.0*
    - 4.2: Nomenclature of Ionic Compounds - *CC BY-SA 4.0*
    - 4.3: Covalent Bonding - *CC BY-SA 4.0*
    - 4.4: Exceptions to the Octet Rule - *CC BY-SA 4.0*
    - 4.5: Nomenclature of Molecular Compounds - *CC BY-SA 4.0*
    - 4.6: Electronegativity and the Polar Covalent Bond - *CC BY-SA 4.0*
    - 4.7: Intermolecular Forces - *CC BY-NC-SA 4.0*
  - 5: Chapter 5 - Solution Chemistry - *Undeclared*
    - 5.1: Prelude to Solutions - *CC BY-NC-SA 4.0*
    - 5.2: Definitions - *CC BY-NC-SA 4.0*
    - 5.3: The Mole - *CC BY-NC-SA 4.0*
    - 5.4: Quantitative Units of Concentration - *CC BY-NC-SA 4.0*
    - 5.5: Dilutions and Concentrations - *CC BY-NC-SA 4.0*
    - 5.6: Colligative Properties of Solutions - *CC BY-NC-SA 4.0*
  - 6: Chapter 6 - Survey of Chemical Reactions - *Undeclared*
    - 6.1: Prelude to Chemical Reactions - *CC BY-NC-SA 4.0*
    - 6.2: The Chemical Equation - *CC BY-NC-SA 4.0*
    - 6.3: The Mole in Chemical Reactions - *CC BY-NC-SA 4.0*
    - 6.4: Stoichiometry - *CC BY-NC-SA 4.0*
    - 6.5: Limiting Reagents - *CC BY-NC-SA 4.0*
    - 6.6: Types of Chemical Reactions - Single and Double Displacement Reactions - *CC BY-NC-SA 4.0*

- 6.7: Composition, Decomposition, and Combustion Reactions - *CC BY-NC-SA 4.0*
- 6.8: Neutralization Reactions - *CC BY-NC-SA 4.0*
- 6.9: Energy - *CC BY-NC-SA 4.0*
- 6.10: Enthalpy and Chemical Reactions - *CC BY-NC-SA 4.0*
- 6.11: Chemical Reaction Rate - *CC BY-NC 4.0*
- 6.12: Collision Theory - *CC BY-NC 4.0*
- 6.13: Activation Energy - *CC BY-NC 4.0*
- 6.14: Chemical Equilibrium - *CC BY-NC 4.0*
- 6.15: Equilibrium Constant - *CC BY-NC 4.0*
- 6.16: Le Châtelier's Principle - *CC BY-NC 4.0*

7: Chapter 7 - Survey of Kinematics and Newtonian Motion - *Undeclared*

- 7.1: Prelude to One-Dimensional Kinematics - *CC BY 4.0*
- 7.2: Displacement - *CC BY 4.0*
- 7.3: Vectors, Scalars, and Coordinate Systems - *CC BY 4.0*
- 7.4: Time, Velocity, and Speed - *CC BY 4.0*
- 7.5: Acceleration - *CC BY 4.0*
- 7.6: Kinematics in Two Dimensions - An Introduction - *CC BY 4.0*
- 7.7: Projectile Motion - *CC BY 4.0*
- 7.8: Development of Force Concept - *CC BY 4.0*
- 7.9: Newton's First Law of Motion- Inertia - *CC BY 4.0*
- 7.10: Newton's Second Law of Motion- Concept of a System - *CC BY 4.0*
- 7.11: Newton's Third Law of Motion- Symmetry in Forces - *CC BY 4.0*
- 7.12: Linear Momentum and Force - *CC BY 4.0*
- 7.13: Impulse - *CC BY 4.0*

8: Chapter 8 - Energy - *Undeclared*

- 8.1: Prelude to Work, Energy, and Energy Resources - *CC BY 4.0*
- 8.2: Work- The Scientific Definition - *CC BY 4.0*
- 8.3: Kinetic Energy and the Work-Energy Theorem - *CC BY 4.0*
- 8.4: Gravitational Potential Energy - *CC BY 4.0*
- 8.5: Conservation of Energy - *CC BY 4.0*
- 8.6: Nonconservative Forces - *CC BY 4.0*
- 8.7: Power - *CC BY 4.0*
- 8.8: Work, Energy, and Power in Humans - *CC BY 4.0*

9: Chapter 9 - Waves - *Undeclared*

- 9.1: Free Waves - *CC BY-SA 4.0*
- 9.2: Sound - *CC BY 4.0*
- 9.3: Hearing - *CC BY 4.0*
- 9.4: Speed of Sound, Frequency, and Wavelength - *CC BY 4.0*
- 9.5: Sound Intensity and Sound Level - *CC BY 4.0*
- 9.6: Sound Interference and Resonance- Standing Waves in Air Columns - *CC BY 4.0*
- 9.7: Images by Reflection - *CC BY-SA 4.0*
- 9.8: Refraction - *CC BY-SA 4.0*
- 9.9: Wave Optics - *CC BY-SA 4.0*

10: Chapter 10 - Electricty and Magnetism - *Undeclared*

- 10.1: Current and Voltage - *CC BY-SA 4.0*
- 10.2: Parallel and Series Circuits - *CC BY-SA 4.0*
- 10.3: More About the Magnetic Field - *CC BY-SA 4.0*
- 10.4: Induced Electric Fields - *CC BY-SA 4.0*

Back Matter - *Undeclared*

- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*