CONCEPTS OF PHYSICAL SCIENCE



Jamie MacArthur
Madera Community College

LibreTexts™

Concepts of Physical Science

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 01/12/2026

Prelude

Hello students, I hope that you enjoy this free online textbook and find it useful!

This text was made possible by many things: the existence of the Libretexts platform as an avenue through which to curate the necessary resources, the commitment by State Center Community College District to providing students with Open Education Resource (OER) materials, the time and efforts of the Madera Community College OER coordinator and committee to put this possibility in place, and the peer review team who dramatically reduced the number of errors in this text. Without each of these supports in place, this project would not have been possible.

There are many existing online free textbook options available for most courses, and hopefully you are able to find courses where these options are put into place. However, this particular textbook was a bit of a challenge in that there was no existing textbook created for the course. Many of the existing OER materials have historically been created by faculty in systems similar to the California University (CU) system. However, these universities don't often teach courses which cover the breadth of physical science in Science 1A, and so a free online textbook was not available for this course. In fact, they don't even have a category for it in the Libretexts library! (at least not yet, but there are many things that historically have fit outside of categories which are now finding categories that fit them, so perhaps one day there will be a category for this little textbook as well.)

Because of the category defying nature of this course, curation of materials for the textbook was a bit more of a challenge than it would have been for a course in a more highly populated field of study such as chemistry or physics. Existing Libretexts resources were curated from bookshelves in physics, chemistry, biology, environmental science, earth science, and psychology.

The most highly adopted materials were in the areas of physics and chemistry, with two sources used extensively: Introduction to Physics (Open Stax), and Introductory Chemistry (CK-12). Each of these texts would be a fine source for an introduction to physics or chemistry, respectively. However, neither of them is designed to cover the breadth of both disciplines within a single semester. And, of course, each text had its own unique stylistic approach. So the curation of this text involved both an adjustment in style to match these two texts in one, as well as a selection of a subsection of each text to highlight in order to adequately cover the breadth of material necessary.

This process was undertaken with the understanding that other instructors would use these materials (or even remix them for their own purposes). When it was difficult to decide which parts of the curated material to highlight, we often defaulted to including more instead of less. It is easier for another instructor to remove or ignore the areas they find less important than to add those materials to the text itself. As such, this text probably covers some material in more depth than is reasonable for coverage in a single semester course. This is particularly true in chapters 6-9, which cover classical physics. The expectation is that the instructor will pick and choose from the materials within these chapters. Hopefully your instructor is communicating which materials within a given chapter to focus on.

I hope that this text is a useful resource for you in understanding the fundamental concepts of physical science.

TABLE OF CONTENTS

Concepts of Physical Science

About Libretexts

Prelude

Licensing

1: Science Itself

- 1.1: Observing the World
- 1.2: Developing the Scientific Method
  - 1.2.1: Careful Observations- Documentation, Measurements, and Patterns
  - 1.2.2: Experiments
  - 1.2.3: A Scientific Method
- 1.3: Processes of Science
  - 1.3.1: Laws and Theories in Science
  - 1.3.2: The Scientific Process
- 1.4: The Limits of Science
  - 1.4.1: Other Ways of Knowing
  - 1.4.2: What Should Science be Allowed to do?
  - 1.4.3: Pseudoscience
- 1.5: Organization of Science
  - 1.5.1: Pure and Applied Science
  - 1.5.2: Fields of Science

2: Matter and Motion

- 2.1: Introduction to Motion
- 2.2: Modern Ideas about Motion
- 2.3: Direction
- 2.4: Vectors, Scalars, and Coordinate Systems
- 2.5: Speed and Velocity
- 2.6: Acceleration
- 2.7: Determining how to solve a motion equation problem
- 2.8: Motion Graphs
  - 2.8.1: Motion graphs of position, velocity, and speed versus time
  - 2.8.2: Motion graphs of acceleration, velocity, and position versus time.
  - 2.8.3: Predicting Motion graphs
- 2.9: Types of Acceleration
  - 2.9.1: Falling Objects
  - 2.9.2: Centripetal Acceleration
  - 2.9.3: Projectile Motion
- 2.10: Measuring the World
- 2.11: Development of Force Concept
- 2.12: Galileo and the Beginnings of Modern Science

3: Types of Motion and Their Causes

- 3.1: Introduction to Dynamics and Momentum.
- 3.2: Newton’s First Law of Motion- Inertia
- 3.3: Newton's Second Law of Motion- Force and Acceleration
- 3.4: Newton’s Third Law of Motion- Symmetry in Forces
- 3.5: Examples of Force
  - 3.5.1: Normal Force
  - 3.5.2: Tension
  - 3.5.3: Spring Force- Hooke’s Law
  - 3.5.4: Friction
  - 3.5.5: Centripetal Force
- 3.6: Summary of Newton's Laws
- 3.7: Newton’s Universal Law of Gravitation
- 3.8: Linear Momentum and Force
- 3.9: Conservation of Momentum
- 3.10: Rotation
- 3.11: Angular Momentum and Its Conservation

4: Fluid Mechanics and Waves

- 4.1: Introduction to Fluids and Waves
- 4.2: What Is a Fluid?
- 4.3: Density
- 4.4: Pressure
  - 4.4.1: Pressure Due to the Weight of Fluid
  - 4.4.2: Boyle's Law
- 4.5: Fluid Principles
  - 4.5.1: Archimedes’ Principle
  - 4.5.2: Pascal's Principle
  - 4.5.3: Bernoulli’s Principle
- 4.6: Wave Mechanics
  - 4.6.1: Period and Frequency in Oscillations
  - 4.6.2: Simple Harmonic Motion- A Special Periodic Motion
  - 4.6.3: Waves
- 4.7: Properties of Waves
  - 4.7.1: Transverse and Longitudinal Waves
  - 4.7.2: Forced Oscillations and Resonance
  - 4.7.3: Wave Interference- Standing Waves and Beats
  - 4.7.4: When Waves Sources Move- Doppler effects and Bow Wakes
- 4.8: Sound
  - 4.8.1: Speed of Sound, Frequency, and Wavelength

5: Work and Energy

- 5.1: Introduction to Work and Energy
- 5.2: Work- The Scientific Definition
- 5.3: Kinetic Energy and the Work-Energy Theorem
- 5.4: Potential Energy- Gravity and Springs

- 5.4.1: Gravitational Potential Energy
- 5.4.2: Spring Potential Energy
- 5.5: Conservative Forces, Potential Energy, and Conservation of Energy
- 5.6: Forms of Energy
- 5.7: Simple Machines
- 5.8: Power
- 5.9: Energy and Momentum

6: Heat and Energy

- 6.1: Introduction to Thermodynamics
- 6.2: Temperature
  - 6.2.1: Temperature and Thermometers
  - 6.2.2: The Molecular Basis for Temperature
- 6.3: Heat
- 6.4: Heat Transfer Methods
- 6.5: Temperature Change and Heat Capacity
- 6.6: The First Law of Thermodynamics
- 6.7: Heat Engines and the Second Law of Thermodynamics
  - 6.7.1: Efficiency and the Second Law of Thermodynamics
  - 6.7.2: Carnot's Perfect Heat Engine- The Second Law of Thermodynamics Restated
- 6.8: Entropy and the Second Law of Thermodynamics
  - 6.8.1: Statistical Interpretation of Entropy and the Second Law of Thermodynamics- The Underlying Explanation
  - 6.8.2: Entropy, Life, the Universe, and Everything
- 6.9: Energy Sources for Our Lives

7: Fundamentals and Applications of Electricity

- 7.1: Introduction to Electricity
- 7.2: Static Electricity and Charge- Conservation of Charge
- 7.3: Coulomb's Law
- 7.4: Representing Electric Fields
- 7.5: Electric Potential and Potential Energy
- 7.6: Conductors and Current
- 7.7: Ohm's Law
- 7.8: Circuits
- 7.9: Applications- Power and Safety
  - 7.9.1: Electric Power and Energy
  - 7.9.2: Electric Hazards and the Human Body

8: The Amazing Connections Between Magnets and Light Switches

- 8.1: Introduction to Magnetism
- 8.2: Magnets
- 8.3: Natural Magnetism
- 8.4: The Source of Magnetism
- 8.5: Magnetic Fields and Magnetic Field Lines
- 8.6: The relationship between electric current, magnetic fields, and magnetic force
- 8.7: Force from Electricity and Magnetism- Motors and Meters
- 8.8: Faraday's Law- Electricity from Magnetism and Force

- 8.8.1: Applications of Electromagnetic Induction
- 8.9: Magnetism from Electricity
- 8.10: Maxwell’s Equations- Summarizing the Connections Between Electricity and Magnetism
- 8.11: Sources of Voltage and the Electrification of America
  - 8.11.1: Batteries
  - 8.11.2: Induced Voltage and Electrical Generators
  - 8.11.3: Alternating Current versus Direct Current
  - 8.11.4: Transformers

9: Electromagnetic Radiation and Optics

- 9.1: Introduction to Light
- 9.2: Maxwell’s Equations- Electromagnetic Waves Predicted and Observed
- 9.3: Producing Electromagnetic Radiation
- 9.4: The Electromagnetic Spectrum- an Overview
- 9.5: The Electromagnetic Spectrum- Application Notes
- 9.6: Interaction of Light with Matter
  - 9.6.1: Reflection
  - 9.6.2: Refraction
  - 9.6.3: Dispersion- The Rainbow and Prisms
- 9.7: Our Extraordinary World of Colors
- 9.8: Lenses, Mirrors, and Polarization
  - 9.8.1: Image Formation by Lenses
  - 9.8.2: Image Formation by Mirrors
  - 9.8.3: Polarization
- 9.9: The Greenhouse Effect and Climate Change

10: Elements Are Made of Atoms

- 10.1: Introduction to Atoms and Elements
- 10.2: Modern Understandings of Elements
- 10.3: The Development of Atomic Theory
  - 10.3.1: Law of Conservation of Mass
  - 10.3.2: Law of Definite Proportions
  - 10.3.3: Law of Multiple Proportions
  - 10.3.4: Dalton's Atomic Theory
- 10.4: Organizing the Elements
  - 10.4.1: Mendeleev's Periodic Table
  - 10.4.2: Periodic Law
  - 10.4.3: Modern Periodic Table- Periods and Groups
- 10.5: Understanding the Periodic Table
  - 10.5.1: Metals, Nonmetals, and Metalloids.
  - 10.5.2: Families of the Periodic Table
- 10.6: Inside the Atom
- 10.7: Electrons and Light
- 10.8: Quantum Mechanics
- 10.9: Electron Shells

11: Chemical Bonds

- 11.1: Chemical Bond
- 11.2: Elements, molecules, and compounds
- 11.3: The Octet Rule
- 11.4: Ions and Ion Formation
- 11.5: Ionic Bonding
- 11.6: Metallic Bonding
- 11.7: Covalent Bonds and Molecules
- 11.8: Bond Polarity
- 11.9: Intermolecular Forces
- 11.10: Naming Compounds

12: Physical Properties

- 12.1: Introduction to Physical Properties
- 12.2: Physical and Chemical
- 12.3: Classifying Matter
- 12.4: The Physical States of Matter
- 12.5: Transitions between Phases
- 12.6: Energy and Phase Transitions
- 12.7: Moles- Counting Atoms
  - 12.7.1: Moles and Avogrado's Number
  - 12.7.2: Moles and Mass
- 12.8: Solutions and Solubility
- 12.9: Concentrations of Solutions
- 12.10: Soap- an Application of Physical Properties

13: Chemical Reactions

- 13.1: Introduction to Chemical Changes
- 13.2: Aspects of a Chemical Change
- 13.3: Chemical Equations- Representing Chemical Changes
- 13.4: Balancing Equations
- 13.5: Energy and Chemical Reactions
- 13.6: Entropy and Chemical Reactions
- 13.7: Classifying Reactions
- 13.8: Acid Base Reactions
- 13.9: Redox Reactions- The Transfer of Electrons
- 13.10: Polymerization Reactions

14: Nuclear Radiation

- 14.1: Introduction to Radioactivity
- 14.2: The Atomic Nucleus
- 14.3: Discovery of Radioactivity
- 14.4: Nuclear Decay Processes
- 14.5: Nuclear Equations
- 14.6: Half-Life
- 14.7: What Holds the Nucleus Together
- 14.8: The Energy Within the Nucleus and all Matter
- 14.9: Nuclear Binding Energy

- 14.9.1: Nuclear Fission
- 14.9.2: Nuclear Fusion
- 14.10: Nuclear Radiation in Our Lives
- 14.11: Process of Science- Lise Meitner and Otto Hahn

Index

Glossary

Detailed Licensing

Detailed Licensing

Licensing

This book is licensed using a creative commons license, as detailed here. Much of the material has been curated from existing sources, and those details are attributed on each of the relevant pages.

CHAPTER OVERVIEW

1: Science Itself

1.1: Observing the World

1.2: Developing the Scientific Method

1.2.1: Careful Observations- Documentation, Measurements, and Patterns

1.2.2: Experiments

1.2.3: A Scientific Method

1.3: Processes of Science

1.3.1: Laws and Theories in Science

1.3.2: The Scientific Process

1.4: The Limits of Science

1.4.1: Other Ways of Knowing

1.4.2: What Should Science be Allowed to do?

1.4.3: Pseudoscience

1.5: Organization of Science

1.5.1: Pure and Applied Science

1.5.2: Fields of Science

This page titled 1: Science Itself is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.1: Observing the World

Learning Objectives

- Explain the importance of observation in science and some of the obstacles that have historical limited observations.

What is the role of observation in our lives? We can use our senses, not just our vision, but our hearing, smell, taste, and touch to learn more about the world. We might do this for simple enjoyment, or we might do it because we want a better understanding of how the world works. If we observe something because we want to understand the world, we engage in an act of science, even if we do not realize it yet.

Observations were not necessarily always held in such high regard, however. During the time of Greek philosophy, around 400 BCE, Plato suggested that we cannot truly trust our senses. Plato claimed that to truly understand the world we must apply reason and believe that we can only see the shadows of some more true reality. But reason without observation is merely conjecture. Plato's student Aristotle wrote that men have more teeth than women. This is a statement that is obviously not based on true observations, but on reasoning alone. (And faulty reasoning at that!) When our reasoning is not based on observations about the natural world it can go in incorrect, and sometimes dangerous, directions. Some credit Aristotle with the beginning of natural philosophy, the precursor to science. However, the approach did not give as much primacy to observations as we now understand them to have in science.

During the medieval period (1100-1700 CE) in Europe, scholasticism became a dominant mode of thought. Like Greek philosophy before it, this mode of thought was also based more on reasoning than on observations of nature. In this case, the reasoning was related to trying to reconcile these same ideas from Greek philosophy with religious texts and writings. Scholastics debated about things such as the bodily functions of angels among other ideas we would now find silly (possibly even another debate about teeth, in this case those of a horse, though it is unclear if this story is factual.) The scholastic tradition did give rise to universities in Europe, and it was from these universities that scientific thought would eventually emerge. However the tradition itself was about appealing to an authority and particular methods of debate rather than about the work of observation and inquiry that would later develop in science.

Throughout time and across many cultures, observations about the world have led to understanding how it works. Sometimes these observations led to useful information such as the knowledge of which plants to eat and which to avoid. Sometimes these observations were paired with ideas about what might have caused a natural effect, perhaps a god throwing around lightning bolts because he is upset about something. Observation itself is the first step of science. But for the field of science to fully develop as it is today, additional steps were also necessary.

Section Summary

- Observation is the first step of science.

This page titled 1.1: Observing the World is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.2: Developing the Scientific Method

The first step of science is choosing to observe the world, but what comes next? There are several things that can be done: document results, repeat observations, make more careful observations, and eventually experiments. This process leads to something that is sometimes referred to as the scientific method. Within the subsections that follow, we will explore these ideas in greater detail.

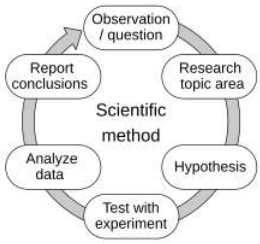


Figure 1.2.1: A schematic of a possible scientific method. These methods will be explored in more detail later in this chapter. Image created by Wikipedia user Efbrazil.

This page titled 1.2: Developing the Scientific Method is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.2.1: Careful Observations- Documentation, Measurements, and Patterns

Learning Objectives

- Explain the importance of documentation, measurement and pattern recognition in science.
- Correctly document measurements.
- Express patterns qualitatively, graphically, and mathematically.

It takes better observations to have a better understanding of the natural world. There are several ways that we can do this, many of which pre-date what we now refer to as science.

Documentation

It's one thing to say that observations do not matter, but it's quite another thing to say that maybe you made an observation but do not remember everything about it correctly. Careful documentation can aid in this process. Maybe you have an interest in starting a regular routine that you think will improve the quality of your life. Documenting the start of this routine and the quality of your life that you experience afterwards might be a way of learning whether the routine is helpful (and of course there are many apps for things like fitness, budgeting, time management, and cooking that have been created to help with this tracking process). The same process can be applied to any observation about the natural world. What time of year did you plant something, and did it grow the way you hoped it would? This information can be helpful to you the next year to make plans for planting again.

If enough people are able to document their observations of the world, this information can be compiled and shared for others to use. Note the difference between compiling observations and compiling ideas (as the scholastics did). If your goal is to understand the world that we live in, the written documentation of what someone else has observed can be quite helpful. Perhaps it was observed in a time or place that you don't have access to. Perhaps the records of someone else's plant growing process will help you to get the results that you want when you plant your own. This sort of documentation is one of the key aspects of science. Documentation is helpful in recognizing patterns in the natural world. As we shall soon see, this recognition of patterns is an important part of science.

Around 1400 CE, during the Ming dynasty, the Yongle Encyclopedia was created. This was one of the largest encyclopedias ever created, and included many subject areas beyond just science. The ability to share information in this way is one of the keys to scientific progress. Scientific knowledge is not built by the works of a single man (as one might assume based on the way Scientific Laws have often been named), but by the collective work of a community of diverse scholars.



Figure 1.2.1.1: A page from the Yongle Encyclopedia from the Chester Beatty Library. Public Domain

Measurements

Another way to learn more about the world is to make better observations. If you observe something as large, or tall, or tiny those are subjective observations which do not allow for finding patterns in nature. If you can quantify the observation it allows for an increased ability to recognize patterns. Maybe large things that are 8 feet tall follow different patterns than large things that are 10 feet tall. You will not know this if you are not able to measure (and then document) that difference.

People learned how to make repeatable measurements before documented history. This was helpful for economic reasons. If you wanted to sell and/or meal plan the food you had harvested, it was helpful to know how many ephas of grain you had and how many shekels of coin you could receive in return for them. The epha was a measure of volume and the shekel was a measure of mass. In fact, the process of measuring a harvest each year could be considered a scientific observation.

Measurements have become a very key part of science. As we make or report a measurement, it is important that we include both a number and a unit. Think of it like the difference between 100 dollars and 100 cents. If you do not know the unit, you do not actually know anything about the results of the measurement.

One of the things that has allowed science to progress so much is the ability to make better measurements. Sometimes new devices allow us to make better observations that unlock some understanding of the natural world that we did not have before.

> **Key Concept:** Every measurement must include both a number and a unit.

Patterns

As more observations are documented, and the quality of those observations is improved by the use of measurements, patterns might begin to emerge. These patterns are often the key to understanding the next step in the process of science. As we begin to notice patterns, we can then use our understanding of those patterns to make predictions about what might happen next.

Some of the earliest scientists lived before recorded history and they used this sort of pattern recognition of the natural world for survival. Keeping track of the patterns of both weather and astronomy helped ancient nomadic people decide when it was time to move towards geographies that had better resources for them. By around 700 BCE, Babylonian astronomers had developed an observation based predictive model for the movement of the planets. This may have been the very first scientific revolution.



Figure 1.2.1.2: A Babylonian tablet from 164 BCE recording an observation of what has since been called Halley's comet (British Museum, public domain)

We have many skills to aid with pattern recognition in science that we will introduce throughout this text. There are often trends where two aspects of an object change together. For example, as a child gets older they get taller. One way to measure this trend would be to mark their height on a wall on each of their birthdays. You would likely find that children grow at variable rates (hence the term growth spurt). In this case, a purely qualitative description is sufficient, a sentence such as “children grow as they get older.”

But for some patterns, the trend is more pronounced. For example, if you were to measure the mass and volume of a single aluminum pellet and then repeat this process for 5, 10, and 20 aluminum pellets you would definitely see a trend: as the mass goes up, so does the volume. But you might notice this trend is not quite as variable as that for growing children. At this point we could make a line graph of the data to see what kind of variability exists. The graph would show a straight line connecting the data points. We call this sort of relationship a direct relationship. We can express it as

$$m = kV$$

where m is the mass, V is the volume, and k is some constant that relates these two quantities together. (It turns out the constant k is a property called density, and this is a relationship we will explore in more detail later in this text.)

This practice of changing some aspect of nature and measuring another aspect of it is a key part of scientific experiments that we will explore in more detail in the next subsection. It is also a very powerful tool. If you wanted a certain mass of aluminum pellets but were only able to measure the volume of those pellets, you could use the direct relationship developed here in order to

determine what volume of pellets to use. These sorts of relationships in the field of science are what has allowed us to create the high tech society in which we now live.

Section Summary

- Documentation is a useful aspect of the process of science, and access to information previously documented by others can be just as useful.
- More accurate observations can occur via measurements, and new measuring devices continually expand the scope of science.
- Every measurement contains both a number and a unit.
- Pattern recognition is very important in science, and leads to important descriptions of the relationship between measurements.
- In a direct relationship, two aspects of nature increase together which can be shown graphically or in an equation.

This page titled 1.2.1: Careful Observations- Documentation, Measurements, and Patterns is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.2.2: Experiments

Learning Objectives

- Explain the key aspects of an experiment.

Careful experiments such as the one we discussed with the aluminum pellets are a hallmark of science. But science was not always done in this way. Two researchers in the 1600s proposed new ideas for exploring the natural world that turned the focus of science towards experimentation (also referred to as empiricism).

Francis Bacon (1561-1626) advocated for inductive reasoning in contrast to the deductive reasoning that was practiced by the scholastics at the time. Deductive reasoning was based on ideas, whereas inductive reasoning was based on observations. The challenge for Bacon was that if there was some flaw in the ideas that were used to perform the deduction, it would carry over into what was deduced (as we have already seen in this text). Bacon advocated for making several observations and from those observations a cause might be observed to emerge. This was the initial the basis for what we might now call a controlled experiment. In a controlled experiment, we change one of the variables and see what might happen while leaving other variables unexamined.

The year after Bacon died, Robert Boyle was born (1627-1691). There were still many scholastics at the universities who wanted to understand nature via deductive reasoning from ancient texts, but Boyle continued the focus on inductive reasoning and experimentation. Boyle expanded on these ideas to include the need to replicate an experiment, and even to record experiments that failed to achieve the expected result. Boyle's innovations were owing to both his curiosity and his considerable wealth, which allowed him to conduct such costly experiments over and over again. (Imagine how much more science could have been done at the time if everyone had access to the time and resources necessary to do such work?)

Some of what made this idea unique at the time was probably Boyle’s considerable wealth which allowed him to pay for these expensive experiments over and over again.

Boyle was credited for a gas law that we will discuss later in this text. The work of discovering it was largely done by others, but Boyle’s prestige and wealth resulted in him getting credit for its discovery. On one hand we can see how it might be unfair for Boyle to have this natural law named after him when the work was mostly done by others. On the other hand, it can be thought of as an example of how modern science works: a community of experimentalist arriving at a result through collaboration and sometimes competition.

Through the work of these and other scholars of the time, empiricism began to emerge as the most powerful form of scientific thought. The general process of experimentation involves observing nature, manipulating nature in some way, observing nature after this manipulation, comparing these results and then trying to explain them. A more formalized experimental process will be introduced in the next subsection.

Section Summary

- In a controlled experiment, one aspect of nature is changed and then observations are made.
- Replication of experiments was one of the initial aspects of the scientific method.
- In an experiment nature is observed, some change is made to it, additional observations are made, and then the results are compared.

This page titled 1.2.2: Experiments is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.2.3: A Scientific Method

Learning Objectives

- Explain the steps of the scientific method.
- Correctly apply the following terms to descriptions of a scientific process: observations, hypothesis, experiment, independent variable, dependent variable, experimental group, control group, repetition.

As we shall see soon enough, there are many ways of doing science in the modern world. For much of science there is a very formal way of approaching questions about the natural world referred to as the scientific method. This is the method often taught in public schools. It is convenient at that level in that it provides a finished product at the end and follows a procedure which provides the sort of checklist that makes the lives of students and teachers at least somewhat predictable. It is also a good way of introducing students to the type of thinking and observing of the natural world which is necessary in science.

The Four Steps of the Scientific Method

The four steps of the scientific method are listed below. To help illustrate the scientific method, an example that an entomologist (a biologist who specializes in insects) might use is given in italics below each step.

Step 1: Observations & Questions

Observe something in the natural world and ask a question about how it works. The part of the natural world that is observed and investigated is usually the area that the scientist specializes in. An entomologist for example, would ask questions about how insects function.

"The life cycle of a fruit fly is about 30 days (at 29 degrees Celsius). How do changes in temperature affect the life cycle of a fruit fly?"

Step 2: Hypothesis

Make a hypothesis (an educated guess) which attempts to answer the question. A useful hypothesis is a **testable** statement.

"Decreasing the temperature of a fruit fly's environment will increase the time it takes the fruit fly to complete its life cycle."

Step 3: Experiment

Design and carry out an experiment that is capable of testing the hypothesis. In other words, the experiment must be designed so that it will produce results that either clearly support or clearly falsify (disprove) the hypothesis. It helps to use "If-Then" predictions based on your hypothesis.

"Place 100 fruit flies at 18 degrees Celsius for one generation. Also place 100 fruit flies at 29 degrees Celsius for one generation. If the hypothesis is correct, then the fruit flies that develop at 18 degrees Celsius will complete their life cycle after those fruit flies that are placed at 29 degrees Celsius."

Step 4: Analyze Results and State Conclusions

Reject the hypothesis if the results are not consistent with the hypothesis or accept the hypothesis as possibly true if the results are consistent with the hypothesis. Notice that the hypothesis is not "proven to be true" even if the results do support it. This is because there may be explanations other than the hypothesis for the experimental result.

For example, if the fruit flies placed at 18 degrees Celsius do develop slower, it may be that their food is not as soft making it more difficult for the fruit flies to eat at the lower temperature, causing them to eat less food and thus grow slower.

If the experimental results do not support the hypothesis, the hypothesis may be modified and additional experiments may be done to test the new or revised hypothesis.

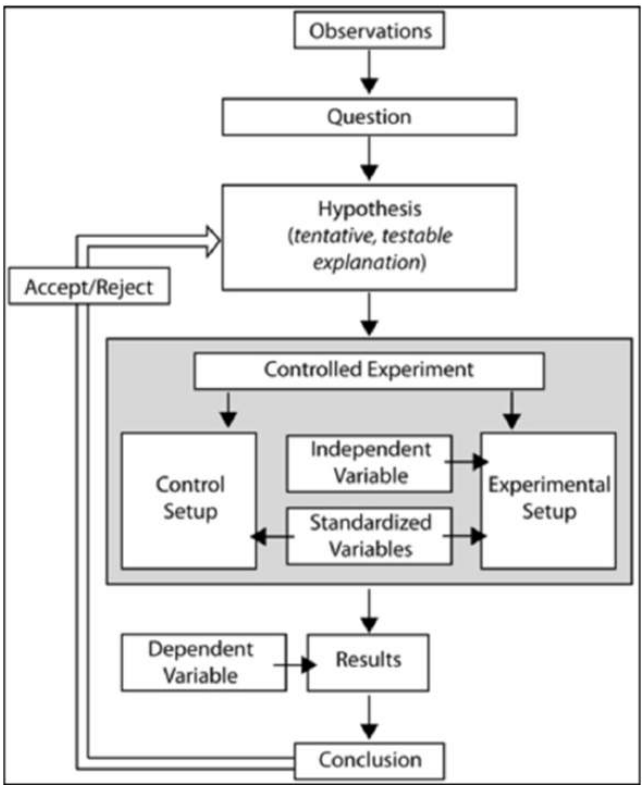


Figure 1.2.3.1: A schematic of the scientific method. Credit: Katherine Harris (Hartnell College) (CC BY-NC-SA)

Characteristics of a Well-Designed Experiment

The most challenging part of the scientific method is usually the third step, designing and carrying out an experiment to test the hypothesis. A well-designed experiment should include all of the following characteristics:

1. An independent variable. The independent variable is the part of the experiment that the scientist changes or manipulates to see what effect occurs.
"The temperature is the independent variable, since that is what the experiment changes to see its effect."

2. A dependent variable. The dependent variable is the part of the experiment that changes because of the change in the independent variable. In other words, the dependent variable is the effect while the independent variable is the cause.

"The length of the fruit flies' life cycle is the dependent variable, since the time of development is expected to change because of the temperature."

3. An experimental group. The experimental group is the group of subjects where the independent variable is set to an unusual or test level.

"The fruit flies placed at 18 degrees Celsius are the experimental group, since the effect of lower temperature on the life cycle of fruit flies is what is being tested."

4. A control group. A control group is the group of subjects in the experiment that the experimental group is compared to. For the control group, the independent variable is set to a normal or usual level (which may be zero, if that is considered a normal level).

"The fruit flies kept at 29 degrees are the control group, since this is the optimal temperature for fruit fly development."

Some experiments include things called positive controls and negative controls. These are slightly different than control groups. Positive and negative controls serve to show that the experiment is working correctly. A **positive control** is a part of the experiment that is deliberately designed to give a positive result. It shows that the experiment is capable of producing a positive result when it is supposed to. A **negative control** is a part of the experiment that is designed to give a negative result. It shows that the experiment is capable of producing a negative result when it is supposed to.

5. The experiment should contain repetition. This means that there should be more than one subject in the experimental group and the control group. Why? In general, the more repetition, the less likely that your results are due to random chance.

"The experimental group and the control group each contained 100 fruit flies."

6. The experiment should be well-defined. One aspect of "well-defined" is that the procedure (the steps) must be written down and clearly described. The true test of a well-written experimental procedure is that another scientist could duplicate it exactly using just the written directions. Another aspect of "well-defined" is that everything in the experiment, such as the materials, chemicals, equipment, environmental conditions, and the subjects (the organisms involved in the experiment), should be described as exactly as possible. All of the factors that are kept equal in the experiment and control groups are called **standardized variables**. Another aspect of "well-defined" is that all parts of the experiment should be quantified. This means they should be measured by numbers.

a) All fruit flies in this experiment are the same species (Drosophila melanogaster). All fruit flies were placed in plastic vials with standard cornmeal media and a cotton stopper. All fruit flies received artificial sunlight for 16 hours per day.

b) One group of 100 fruit flies (the experimental group) were placed in an incubator set to 18 degrees Celsius.

c) One group of 100 fruit flies (the control group) were placed in an incubator set to 29 degrees Celsius.

d) In order to accurately determine the length of the life cycle of the fruit flies, 100 adult flies were allowed to lay eggs for 2 days and then removed from vials. Vials were monitored daily to determine the stage of the life cycle of the offspring of the original adult fruit flies. The end of the life cycle was recorded as the time when half of the fruit flies had undergone eclosion (adults emerge from the pupa case).

Section Summary

- In the scientific method, the research makes observations, formulates a testable hypothesis, performs an experiment, and then analyzes the results of the experiment.
- A valid hypothesis is testable, and can be used to rule out a possible explanation for a phenomenon.
- In a controlled experiment, an independent variable is manipulated while a dependent variable is observed. The manipulation is done on an experimental group and then compared to a control group.
- A good experiment should contain repetition and the steps should be clearly defined.

Contributions and Attributions

This page was remixed by Jamie MacArthur (Madera Community College) from Scientific Method Tutorial by Dr. Katherine Harris

This page titled 1.2.3: A Scientific Method is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.3: Processes of Science

In the previous section we introduce a scientific method. This method is widely used in grade school education as a way of introducing children to the process of science, but it is not the only way that science functions. In this section we will introduce some terms that could be thought of as the outcomes of science, as well as a process of science that includes many possible approaches.

This page titled 1.3: Processes of Science is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.3.1: Laws and Theories in Science

Learning Objectives

- Identify examples of laws and theories in science.

We have already discussed some of the methods of science, but what is the end result of science? For a single scientist, the end result would be a collection of all of the work they have done over their career, which may have variable results. For the scientific community as a whole, the end result is often going to be laws, theories, and principles. Although there are many scientific laws which bear the name of a single researcher, it is never really a single person who discovers anything, but the scientific community as a whole. (In some fields this is recognized by including multiple contributors in the naming of a phenomena. For example: the Soave-Redlich-Kwong equation of state is used in thermodynamics.) This can be no better epitomized than by considering the words of Isaac Newton, for whom three of the most important laws of physics have been named: "if I have seen further....it is by standing on the shoulders of giants."

Scientists search for answers to questions and solutions to problems by using a procedure called the scientific method, as discussed previously. This procedure consists of making observations, formulating hypotheses, and designing experiments, which in turn lead to additional observations, hypotheses, and experiments in repeated cycles (Figure 1.3.1.1). Eventually, after the work of many scientists, those cycles might result in laws or theories.

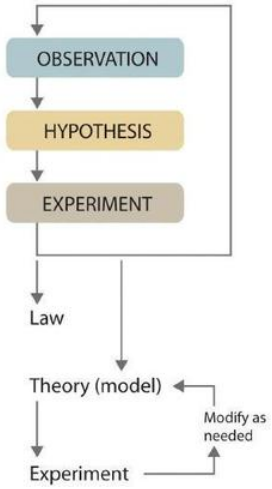


Figure 1.3.1.1: The Scientific Method. As depicted in this flowchart, the scientific method consists of making observations, formulating hypotheses, and designing experiments.

Scientific Laws

After many experimental data have been collected and analyzed, the scientific community may begin to think that the results are sufficiently reproducible (i.e., dependable) to merit being summarized in a **law**, a verbal or mathematical description of a phenomenon that allows for general predictions. A law simply says what happens; it does not address the question of why. Notice that this is very different from our use of the word law in our every day lives. We might drive a certain speed because it is the law. In this sense we can consider the law to be prescriptive: we do something because the law tells us to. Natural laws are not prescriptive but descriptive. An apple does not fall from a tree because it is written on a sign somewhere. Instead, someone writes down the law of gravity because they observed that the movement of apples and other objects followed a pattern. Laws are unlikely to change greatly over time unless later laws are found which apply to contexts which were not considered at the time.

Definition: Scientific Law

A verbal or mathematical description of a phenomenon that allows for general predictions.

One example of a law, the Law of Definite Proportions, was discovered by the French scientist Joseph Proust (1754–1826). This law states that a chemical substance always contains the same proportions of elements by mass. Thus sodium chloride (table salt) always contains the same proportion by mass of sodium to chlorine, in this case 39.34% sodium and 60.66% chlorine by mass, and sucrose (table sugar) is always 42.11% carbon, 6.48% hydrogen, and 51.41% oxygen by mass. The law does not say why the elements exist in the same proportion every time, simply that they do.

Scientific Theories

Whereas a law states only what happens, a **theory** explains why nature behaves as it does. How is a theory different than a hypothesis? When one makes a hypothesis, there is generally a reason to believe something about what will happen, but it is not necessarily generalizable beyond the particular situation which is being investigated. To be accepted as a scientific theory, there needs to be a preponderance of evidence to support a particular explanation for observed phenomena. Theories are not considered to be more or less valid than laws by those in the scientific community, they are simply different types of information. This is in contrast to how we use these terms in our regular lives. We often use the word theory to represent something that is at best a hypothesis, and more likely a conjecture. For example "My theory is that at the end of this mystery novel, Moriarty will be the culprit." This use of the word theory is merely conjecture about a fictional character, not a combination of explanations coming from multiple experiments performed by several research groups. So when we use the term "it was just a theory" we should be careful not to apply that to scientific theories. Theories don't "graduate" into a law when they have sufficient evidence. They are an explanation as opposed to a description. In a few cases, technological advancements have allowed us to see things that we couldn't before. For example, the atomic theory was considered a well established fact long before the first atoms were observed with electron microscopes. However, we still refer to it as the atomic theory because it has explanatory power for a few scientific laws, such as the law of definite proportions that we just mentioned.

Definition: Scientific Theory

When a preponderance of evidence supports a particular explanation for observed phenomena.

The theory developed to explain the extinction of the dinosaurs, for example, is that Earth occasionally encounters small- to medium-sized asteroids, and these encounters may have unfortunate implications for the continued existence of most species. The hypothesis that this might be the case was proposed by Luis and Walter Alvarez based on their discovery of a rare element in a geological site in Italy in the 1980s. Over the course of a few decades, this hypothesis became recognized as a theory as additional data emerged to support it. This theory is consistent with the bulk of evidence amassed to date by multiple researchers over many different disciplines. Figure **1.3.1.2** summarizes the application of the scientific method in this case.

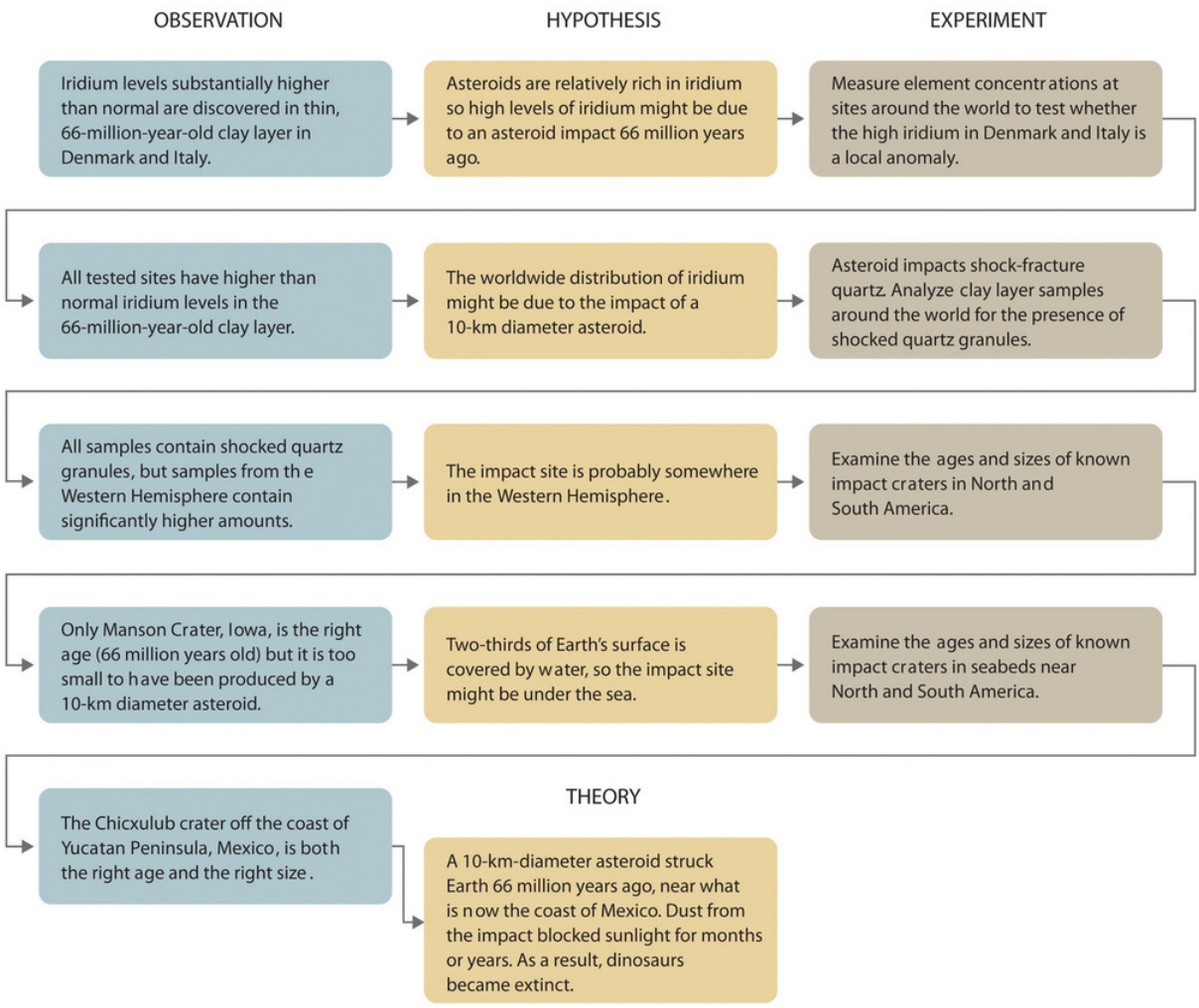


Figure **1.3.1.2**: A Summary of How the Scientific Method Was Used in Developing the Asteroid Impact Theory to Explain the Disappearance of the Dinosaurs from Earth

Since scientists can enter the cycle shown in Figure **1.3.1.1** at any point, the actual application of the scientific method to different topics can take many different forms, as we shall see in the next subsection.

It is important to remember that scientists have a tendency to formulate hypotheses in familiar terms simply because it is difficult to propose something that has never been encountered or imagined before. As a result, scientists sometimes discount or overlook unexpected findings that disagree with the basic assumptions behind the hypothesis or theory being tested. Fortunately, truly important findings are immediately subject to independent verification by scientists in other laboratories, which science is a self-correcting discipline. When the Alvarezes originally suggested that an extraterrestrial impact caused the extinction of the dinosaurs, the response was almost universal skepticism and scorn. In only 20 years, however, the persuasive nature of the evidence overcame the skepticism of many scientists, and their initial hypothesis has now evolved into a theory that has revolutionized paleontology and geology.

Section Summary

- Scientific laws are verbal or mathematical descriptions of a phenomenon that allow for general predictions.
- A scientific theory exists when a preponderance of evidence supports a particular explanation for observed phenomena.

Contributors

Curated from resources found in this text by Libretexts.

This page titled 1.3.1: Laws and Theories in Science is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.1: The Scientific Method** is licensed CC BY-NC-SA 4.0.

1.3.2: The Scientific Process

Learning Objectives

- Explain different steps within the process of science.

The scientific method alone is not enough to make real progress in accumulation of scientific knowledge, but using it as the hub of a cyclic process has led to the massive rate of scientific and technological advancement we have seen over the last century. Science can be thought of as a continuous process guided by the scientific method, as discussed earlier in this text. Modern science is done according to a complex process of checks and balances, such as replication and peer-review. This complexity emerged to help ensure the integrity of scientific results, but the process remains rooted in the basic scientific method. Notice in **Figure 1.3.2.1**: the large amount of work done within the boxes. The results of that work must be combined with the work of many other scientists to result in either a law or theory.

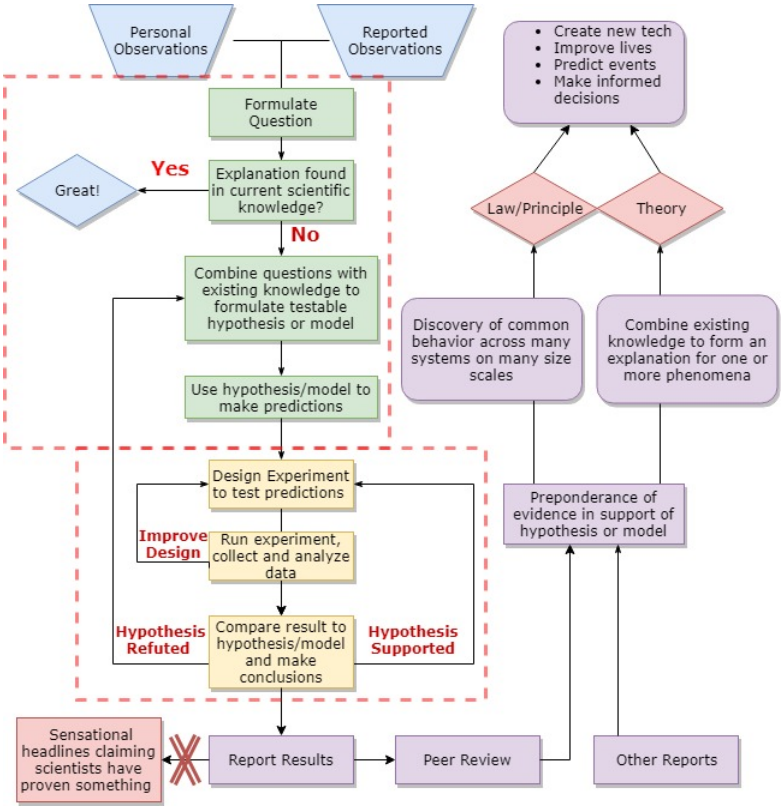


Figure 1.3.2.1: The complex modern scientific process built around the basic scientific method (within dashed lines).

In some fields of research, no single scientist can even keep up with the work inside of the dashed lines. The process of science involves many people working together to better understand the world. Some scientists are very good at making observations (often involving complex or expensive instrumentation). Others are great at finding models (usually mathematical, and sometimes requiring expensive computer programs) that explain the data that have already been collected. In some cases it is not possible to design an experiment that will test everything you would like to. For example, we cannot create another Earth populated with dinosaurs in order to test the impact theory we discussed in the previous subsection. And oftentimes the most interesting results are not at all related to the hypothesis but hints at some aspect of reality that nobody would have imagined otherwise. Some scientists have even won Nobel Prizes for discovering things they were not even looking for! Two examples of this are the Big Bang and buckyballs.

The previous diagram illustrates the complex scientific process, but also highlights the basic scientific method on which the whole process is built. After observation, the basic scientific method follows the green and yellow boxes within the dotted line in the

diagram below. Generally speaking, the green boxes comprise theoretical science and the yellow boxes comprise experimental science. Because of the degrees of specialization in each of these areas, the need to collaborate with other scientists to complete the process is quite important.

The uncertainty associated with all measurements means that science cannot *prove* anything, despite what the media often claims. Instead, the scientific process produces reviewed and reproduced conclusions that account for uncertainty. Scientific conclusions provide evidence for or against hypotheses.

The following video includes some additional ways of thinking about the process of science:



Section Summary

- The scientific method guides the progress of much of the progress of science.
- There is often specialization among researchers which focuses their work on a limited number of the steps within the process.
- Not all scientific questions can be answered with experiments, and not all experiments lead to the expected results.
- Science does not prove something is true, but can provide evidence in support of or opposed to a given hypothesis.

Contributions and Attributions

- Remixed and edited by Jamie MacArthur (Madera Community College)
- Adapted from The Scientific Process by Lawrence Davis (Umpqua Community College)

This page titled 1.3.2: The Scientific Process is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Lawrence Davis (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

This page titled 1.3.2: The Scientific Process is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: The Scientific Process** by Lawrence Davis is licensed CC BY-NC-SA 4.0. Original source: https://openoregon.pressbooks.pub/bpsupmat/.

1.4: The Limits of Science

We have discussed much of what science can do, but there are questions in the world that science cannot answer. And there are methods of doing science that should not be used, though science itself does not hold the answer to that. We will explore these ideas in the following subsections.

This page titled 1.4: The Limits of Science is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.4.1: Other Ways of Knowing

Learning Objectives

- Decide on the most applicable way of knowing in order to answer important questions.

We have talked a lot about the process of science in this text so far, and you might be inclined to think that science has all the answers to every question ever posed. It turns out that it does not. We have many choices to make in our lives that science is not equipped to answer. What is the purpose of our life? Should we ask someone out on a date? Does God exist? Do you like the new song that just came out? Science cannot answer any of these questions. Science is not the only way of knowing.

Other ways of knowing include things like personal choices, art, religion, and philosophy. None of these are necessarily less than science or greater than science. They are ways of knowing that address different aspects of life, so it makes no sense to compare them to each other in any hierarchical way. (Though of course you can compare them to each other in terms of your enjoyment of the subject or how it might help you obtain your goals in life.)

In short we can summarize the point of these various ways of knowing. Science is the way of knowing to use when trying to understand the natural world. Art is the way of knowing to use when trying to understand if something is beautiful. Religion is the way of knowing that some people use when trying to understand the purpose in life. Personal choices can be influenced by any of these ways of knowing and are ultimately about our inner sense of who we are. These different ways of knowing are not necessarily contradictory, and as we will see it is important that we have all of them.

Unfortunately, there are times when it appears that some of these ways of knowing are in conflict with one another. There are many ways this might occur, but one trend that has continued to arise is the conflict between science and religion. We can see this throughout history, and it often occurs when science is able to answer questions that had previously seemed mysterious. Examples include the laws that govern the movements of the planets and stars, the evolution of life, and the nature of static electricity. But do not worry, there are still plenty of mysteries left in the world despite all of the questions that have been answered with science.

Section Summary

- Science is the way of knowing to use when trying to understand the natural world.
- Art is the way of knowing to use when trying to understand if something is beautiful.
- Religion is the way of knowing that some people use when trying to understand the purpose in life.

This page titled 1.4.1: Other Ways of Knowing is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.4.2: What Should Science be Allowed to do?

Learning Objectives

- Discuss how research involving human subjects is regulated.
- Summarize the processes of informed consent and debriefing.
- Explain how research involving animal subjects is regulated.

When scientists seek to understand the natural world, they do not always know what they are going to find. And they certainly do not know what will be done with their discoveries. For example, Lise Meitner discovered nuclear fission but was a staunch pacifist who wanted to have nothing to do with a bomb. Unfortunately within a few decades, her discovery led to nations developing nuclear weapons capable of destroying all human life on Earth. Scientists are ultimately not responsible for what happens in the world based on the knowledge that they discover. However if the citizens of the world are not capable of making moral choices with the new knowledge that is generated, this new knowledge could be more dangerous to us than it is helpful to us. This is where the other ways of knowing become so important: they provide us with a basic respect for human dignity and safety.

Unfortunately, not all scientists have had the moral convictions of Lise Meitner. There are too many incidents of scientists throughout history performing experiments that have harmed and continue to harm people, especially people from marginalized groups. Scientists are no more or less likely to make moral decisions than anyone else is, but a poor choice by a scientist can have lasting effects. This is why having other ways of knowing is important to create the types of regulations that prevent scientists from harming people with their work. Modern researchers must demonstrate that the research they perform is ethically sound. This section presents how ethical considerations affect the design and implementation of research conducted today.

Research Involving Human Participants

Any experiment involving the participation of human subjects is governed by extensive, strict guidelines designed to ensure that the experiment does not result in harm. Any research institution that receives federal support for research involving human participants must have access to an **institutional review board (IRB)**. The IRB is a committee of individuals often made up of members of the institution's administration, scientists, and community members. The purpose of the IRB is to review proposals for research that involves human participants. The IRB reviews these proposals with the principles mentioned above in mind, and generally, approval from the IRB is required in order for the experiment to proceed.

An institution's IRB requires several components in any experiment it approves. For one, each participant must sign an informed consent form before they can participate in the experiment. An **informed consent** form provides a written description of what participants can expect during the experiment, including potential risks and implications of the research. It also lets participants know that their involvement is completely voluntary and can be discontinued without penalty at any time. Furthermore, the informed consent guarantees that any data collected in the experiment will remain completely confidential. In cases where research participants are under the age of 18, the parents or legal guardians are required to sign the informed consent form.

Link to Learning

View this example of a consent form to learn more.

While the informed consent form should be as honest as possible in describing exactly what participants will be doing, sometimes deception is necessary to prevent participants' knowledge of the exact research question from affecting the results of the study. **Deception** involves purposely misleading experiment participants in order to maintain the integrity of the experiment, but not to the point where the deception could be considered harmful. For example, if we are interested in how our opinion of someone is affected by their attire, we might use deception in describing the experiment to prevent that knowledge from affecting participants' responses. In cases where deception is involved, participants must receive a full **debriefing** upon conclusion of the study—complete, honest information about the purpose of the experiment, how the data collected will be used, the reasons why deception was necessary, and information about how to obtain additional information about the study.

DIG DEEPER: Ethics and the Tuskegee Syphilis Study

Unfortunately, the ethical guidelines that exist for research today were not always applied in the past. In 1932, poor, rural, black, male sharecroppers from Tuskegee, Alabama, were recruited to participate in an experiment conducted by the U.S. Public Health Service, with the aim of studying syphilis in black men (**Figure 1.4.2.1**). In exchange for free medical care, meals, and burial insurance, 600 men agreed to participate in the study. A little more than half of the men tested positive for syphilis, and they served as the experimental group (given that the researchers could not randomly assign participants to groups, this represents a quasi-experiment). The remaining syphilis-free individuals served as the control group. However, those individuals that tested positive for syphilis were never informed that they had the disease.

While there was no treatment for syphilis when the study began, by 1947 penicillin was recognized as an effective treatment for the disease. Despite this, no penicillin was administered to the participants in this study, and the participants were not allowed to seek treatment at any other facilities if they continued in the study. Over the course of 40 years, many of the participants unknowingly spread syphilis to their wives (and subsequently their children born from their wives) and eventually died because they never received treatment for the disease. This study was discontinued in 1972 when the experiment was discovered by the national press (Tuskegee University, n.d.). The resulting outrage over the experiment led directly to the National Research Act of 1974 and the strict ethical guidelines for research on humans described in this chapter.

Why is this study unethical? How were the men who participated and their families harmed as a function of this research?



Figure **1.4.2.1**: A participant in the Tuskegee Syphilis Study receives an injection.

Link to Learning

Visit this website about the Tuskegee Syphilis Study to learn more.

Research Involving Animal Subjects

Many psychologists conduct research involving animal subjects. Often, these researchers use rodents or birds as the subjects of their experiments—the APA estimates that 90% of all **animal research** in psychology uses these species (American Psychological Association, n.d.). Because many basic processes in animals are sufficiently similar to those in humans, these animals are acceptable substitutes for research that would be considered unethical in human participants.



Figure **1.4.2.2**: Rats, like the one shown here, often serve as the subjects of animal research.

This does not mean that animal researchers are immune to ethical concerns. Indeed, the humane and ethical treatment of animal research subjects is a critical aspect of this type of research. Researchers must design their experiments to minimize any pain or distress experienced by animals serving as research subjects.

Whereas IRBs review research proposals that involve human participants, animal experimental proposals are reviewed by an **Institutional Animal Care and Use Committee (IACUC)**. An IACUC consists of institutional administrators, scientists, veterinarians, and community members. This committee is charged with ensuring that all experimental proposals require the humane treatment of animal research subjects. It also conducts semi-annual inspections of all animal facilities to ensure that the research protocols are being followed. No animal research project can proceed without the committee's approval.

Section Summary

- Regulatory processes have been created to prevent scientific research from having adverse effects on people.
- Any research institution that receives federal support for research involving human participants must have access to an institutional review board.
- Animal experimental proposals are reviewed by an Institutional Animal Care and Use Committee.

Contributions and Attributions

This page was remixed and edited by Jamie MacArthur (Madera Community College) from the OpenStax psychology text found here.

This page titled 1.4.2: What Should Science be Allowed to do? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Rose M. Spielman, William J. Jenkins, Marilyn D. Lovett, et al. via source content that was edited to the style and standards of the LibreTexts platform.

- **2.4: Ethics** by Rose M. Spielman, William J. Jenkins, Marilyn D. Lovett, et al. is licensed CC BY 4.0. Original source: https://openstax.org/details/books/psychology-2e.

1.4.3: Pseudoscience

Learning Objectives

- Explain characteristics of pseudoscience which distinguish it from science.
- Identify examples of pseudoscience that have these characteristics.

Pseudoscience, basically "fake-science," consists of scientific claims which are made to appear factual when they are actually false. Many people question whether pseudoscience should even contain the word "science" as pseudoscience is not really science; as it disregards the scientific method altogether. Also known as alternative or fringe-science, pseudoscience relies on invalid arguments called sophisms, a word Webster dictionary defines as "an argument apparently correct in form but actually invalid; especially : such an argument used to deceive." Pseudoscience usually lacks supporting evidence and does not abide by the scientific method. That is, pseudo-theories fail to use carefully cultivated and controlled experiments to test a hypothesis. A scientific hypothesis must include observable, empirical and testable data, and must allow other experts to test the hypothesis. Pseudoscience does not accomplish these goals. Several examples of pseudoscience include phrenology, astrology, flat earth, vaccine denial, and intelligent design.

Distinguishing Pseudoscience

In order to distinguish a pseudoscience, one must look at the definition of science, and the aspects that make science what it is. Science is a process based on observations, conjectures, and assessments to provide better understanding of the natural phenomena of the world. Science generally always follows a formal system of inquiry which consists of observations, explanations, experiments, and lastly, hypothesis and predictions. Scientific theories are always challenged by experts and revised to fit new theories. Pseudoscience, however, is mostly based on beliefs and it greatly opposes contradictions. Their hypotheses are never revised to fit new data or information. Scientist continually disprove ideas to achieve a better understanding of the physical world, whereas pseudo-scientists focus on proving theories to make their claims seem plausible.

Pseudoscience beliefs often tend to be greatly exaggerated and very vague. Complicated technical language is often used to sound impressive but it is usually meaningless. For example, a phrase like "energy vibrations" is used to sound remarkable but a phrase like this is insignificant and does not really explain anything. Furthermore, pseudoscience often consists of outrageous, yet unprovable claims. Thus, pseudoscientists tend to focus on confirming their ideas, rather than finding evidence that refutes them. The following dialogue contains the thought-processes behind pseudoscience.

1. My friend and I think unicorns exist
2. Science has no evidence about unicorns.
3. Science therefore cannot prove if unicorns do or do not exist.
4. One day my friend, a very trustworthy person, said she saw a unicorn in the field by her house. There is no other evidence, other than the fact that my friend saw it.
5. Unicorns exist and any scientist who tries to deny the existence of unicorns is a fun-sucking, hostile human being.

The dialogue above features many key characteristics of pseudoscience. The speaker makes their point valid though the two facts alone that their friend had a personal experience and that science has no proof to prove the theory wrong. Finally, the speaker insults anyone who would challenge the theory. In science, challenges to a theory are accepted as everyone has the same common goal of improving the understanding of the natural word. Below is a table that lays out the key characteristics of Science and Pseudo-Science

SCIENCE	PSEUDOSCIENCE
Science never proves anything.	Pseudoscience aims to prove an idea.
Self-correcting methodology which involves critical thinking.	Starts with a conclusion and gives easy answers to complex problems.
An on-going process to develop a better understanding of the physical world by testing all possible hypotheses.	Often driven by social, political or commercial goals.
Involves a continual expansion of knowledge due to intense research.	A field has not evolved a lot since the beginning. If any research is done, it is done to justify the claims, rather that expand them.
Scientists constantly attempt to refute other scientists' works.	An attempt to disprove the beliefs is considered hostile and unacceptable.
When results or observations are not consistent with a scientific understanding, intense research follows.	Results or observations that are not consistent with current beliefs are ignored.
Remains questionable at any time. There are two types of theories: those that have been proven wrong by experimentation and data, and every other theory. Thus, no theory can be proven correct; every theory is also subject to being refuted.	Beliefs of the field can not usually be tested empirically so will likely not ever be proven wrong; Thus, pseudoscientists believe that they are right just because no one can prove them wrong.
Concepts are based on previous understandings or knowledge.	Pseudoscientists are often not in touch with main-stream science and are often driven by the egos of the "scientists". Furthermore, famous names and testimonials are often used for support rather than scientific evidence.
Findings must be stated in unambiguous, clear language.	Pseudoscience often uses very vague, yet seemingly technical terms.

Phrenology

Phrenology, also known as craniology, was a "science" popular during the early 1800s that was centered around the idea that the brain was an organ of the mind. During this time, most people believed that the brain was divided into distinct sections that all controlled different parts of a person's personality or intelligence. The basis of phrenology revolves around the concept that the brain mirrors a muscle and those parts of the brain which are "exercised" the most, will be proportionally larger than those parts of the brain that aren't often used. Thus, the scientists pictured the brain as a bumpy surface, with the make-up of the surface differing for every person depending on their personality and intelligence. By the mid 19th century, automated phrenology machines existed, which were basically a set of spring loaded probes that were placed on the head to measure the topography of one's skull. The machine then gave an automated reading about a person's characteristics based on this.

Let's consider some of the key characteristics of pseudoscience from our chart, and see how they apply to phrenology.

- Pseudoscientists are often not in touch with mainstream science: Scientific research since the 1800s has shown how though the brain is indeed divided into sections, each section does not determine a characteristic or personality trait, but instead controls a specific function such as memory or motor skills. Likewise, it has been concluded that the brain conforms to the shape of the skull, rather than the skull conforming to the shape of the brain (meaning the bumps of a persons skull have nothing to do with the shape of the brain). Back in the 1800s, little knowledge existed about the realities of brain structure and function, so the concept was not as reflexive of pseudoscience as it is today. However, some people still believe in the basic tenets of phrenology. Phrenology today exists as a classic form of pseudoscience as it goes against the common understanding about how the brain functions.
- Often driven by social, political or commercial goals: Indeed, the main goal of phrenology was a political and social one: to prove the dominance of the white race over other races. "Scientists" measured the brains of both races and concluded that the brains of white people were larger than that of people of African descent. Therefore, they concluded, they were smarter and superior. It was later revealed that the scientists were biased while conducting the experiment and that they were previously

aware of what race each brain belonged to. The experiment was repeated and this time the scientists were not aware of the race and they concluded that the brains were of equal size. The second experiment better conforms to the scientific method, as in this case the scientists objectively measured the brains, while in the first case the bias of the scientists led to their conclusions. Thus, this situation demonstrates a two-fold level of defective science because not only was the idea of measuring the brains to determine personality and intelligence not correct altogether, but the methods in which the scientists were doing this was also flawed. Phrenology was also commercially driven, since phrenology parlors where very widespread and many devices were on the market to be used to measure.

- Pseudo-Scientists are often driven by the egos of the "scientists": In the book *Phrenology and the origins of Victorian Scientific Naturalism* by John Van Whye, Van Whye quotes about the main discoverer of Phrenology Franz Joseph Gall, that " the peculiar incentive behind Gall's fascination with explaining individuals' differences may have lain in his hubris" (Van Whye 18). Of the 12 children in his family, Gall was the sharpest and brightest and naturally interested in distinguishing factors between children. Even as a young school boy, Gall noticed that the other children who were just as good at memorization as he was all had protruding eyes, which lead him to the idea of the basis of phrenology, that the characteristics of one's head indicates their intelligence.
- Beliefs of the field can not usually be tested empirically so will likely not ever be proven wrong: At the time that phrenology was popular, we did not have the ability to analyze what happens inside of people's brains, as we do now. Brain scans can now tell us which portions of the brain become more active related to certain thoughts or activities. We also now know through DNA testing that there is no biological basis for racial differences in health outcomes, the disparities we see between racial groups must come from something else: systemic racism.

The Vaccine-Autism Fraud

You may have heard that certain vaccines put the health of young children at risk. This persistent idea is not supported by scientific evidence or accepted by the vast majority of experts in the field. It stems largely from an elaborate medical research fraud that was reported in a 1998 article published in the respected British medical journal, *The Lancet.* The main author of the article was a British physician named Andrew Wakefield. In the article, Wakefield and his colleagues described case histories of 12 children, most of whom were reported to have developed autism soon after the administration of the MMR (measles, mumps, rubella) vaccine.

Several subsequent peer-reviewed studies failed to show any association between the MMR vaccine and autism. It also later emerged that Wakefield had received research funding from a group of people who were suing vaccine manufacturers. In 2004, ten of Wakefield’s 12 coauthors formally retracted the conclusions in their paper. In 2010, editors of *The Lancet* retracted the entire paper. That same year, Wakefield was charged with deliberate falsification of research and barred from practicing medicine in the United Kingdom. Unfortunately, by then the damage had already been done. Parents afraid that their children would develop autism had refrained from having them vaccinated. British MMR vaccination rates fell from nearly 100 percent to 80 percent in the years following the study. The consensus of medical experts today is that Wakefield’s fraud put hundreds of thousands of children at risk because of the lower vaccination rates and also diverted research efforts and funding away from finding the true cause of autism.

Because sufficient numbers of children were not being vaccinated, it was possible to collect a large amount of data on links to vaccinations and autism. (This sort of study would be unethical if a scientist were to put these children into a test group and control group in order to determine any effects from the vaccines. But it was the parents of the children, not the researchers, who put these children in danger by refusing to have them vaccinated.) Data from over a half million children were analyzed, and there was no link found between vaccination and autism, nor in the timing of an autism diagnosis. Follow up studies with increasingly larger sample sizes in the decades following this have continued to show the same result. Unfortunately, the pseudoscientific concern about the link between vaccinations and autism had become prevalent and has likely caused many deaths over the past few decades.

When Popular Beliefs Become Pseudoscience

Not all forms of pseudoscience come from supposed scientists, or science adjacent people (as we discussed with phrenology and the vaccine-autism fraud), although these are often the most dangerous forms of pseudoscience. Their adjacency to scientific practice can give them an air of legitimacy amongst those not trained in understanding scientific practice. Some forms of

pseudoscience are ancient ideas that still have practitioners even after science has determined that the basis for these ideas are not consistent with the laws of nature. Some of these were addressed earlier in this text, and include things such as astrology, intelligent design, and the belief that the world is flat. Unlike phrenology or vaccine aversion it might be possible to include these beliefs into a religious worldview that doesn't harm anyone. But it is important that they not be labeled as science, and as such should not be discussed as valid scientific ideas within a science classroom. (In the public education system it is important that we don't discuss religious beliefs inappropriately, due to the power dynamic with our students.)

An example of such a belief is astrology. Astrology is the study of the movements and relative positions of celestial objects as a means for divining information about human affairs and terrestrial events. Many ancient cultures attached importance to astronomical events, and some developed elaborate systems for predicting terrestrial events from celestial observations. Throughout most of its history in the West, astrology was considered a scholarly tradition and was common in academic circles. With the advent of modern Western science, astrology was called into question. It was challenged on both theoretical and experimental grounds, and it was eventually shown to have no scientific validity or explanatory power.

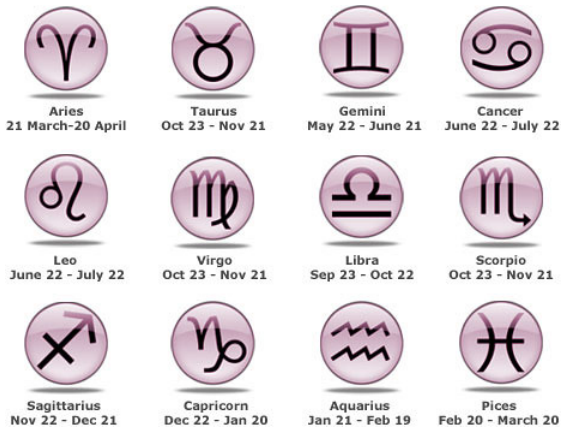


Figure 1.4.3.1: Zodiac signs.

Today, astrology is considered a pseudoscience, yet it continues to have many devotees. Most people know their astrological sign, and many people are familiar with the personality traits supposedly associated with their sign. Astrological readings and horoscopes are readily available online and in print media, and a lot of people read them, even if only occasionally. About a third of all adult Americans actually believe that astrology is scientific. Studies suggest that the persistent popularity of pseudosciences such as astrology reflects a high level of scientific illiteracy. It seems that many Americans do not have an accurate understanding of scientific principles and methodology. They are not convinced by scientific arguments against their beliefs.

Section Summary

- In contrast to science, pseudoscience aims to prove a goal, starts with a conclusion, and ignores results that are inconsistent with its beliefs.
- Pseudoscience is often driven by social, political, or commercial goals which can be dangerous to members of society.
- Some forms of pseudoscience are based on understandings of the world held by many people before scientific research showed they are not valid.

Contributors, Attributions, and References

1. This page was curated and edited by Jamie MacArthur (Madera Community College) using resources created by the following authors:
2. Stephen Lower (Simon Fraser University) whose page addressing pseudoscience can be found here.
3. Tara Jo Holmberg (Northwestern Connecticut Community College) whose page addressing pseudoscience can be found here.

This page titled 1.4.3: Pseudoscience is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **Limitations of the Scientific Method** by Stephen Lower is licensed CC BY-NC-SA 4.0. Original source: http://www.chem1.com/acad/webtext/virtualtextbook.html.

1.5: Organization of Science

Now that we have learned a little bit about what science is (and is not) we can look at a few categories of how science functions. We will explore these categories in each of the following subsections.

This page titled 1.5: Organization of Science is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

1.5.1: Pure and Applied Science

Learning Objectives

- Determine if a specific type of scientific activity is pure research or applied research.

"What is happening in the field of science today? What can I do with a science degree?"—these are all good questions, and should be asked by students interested in science. Research in science is interesting and challenging. But there are different directions a student can take as they explore research opportunities.

Pure and Applied Research

The study of modern physical science can be split into two types of research: pure and applied. Scientists who study **pure science** do research primarily to advance humankind's understanding of science. Pure science is concerned with a greater understanding of the theories behind how nature works. Pure scientists tend to be less concerned with direct applications of the research that they are doing. That is not to say that pure science can never lead to a real-world application but, rather, that a potential application is not the primary motivation for doing the research in the first place. **Applied science** is science that is directed toward a specific practical goal or application. The video below further describes applied research.



Pure Research Examples

The early history of science contains many examples of pure research. Studies on the elements (especially after Mendeleev's periodic table was published) were primarily "pure" research types of experiments. Does this element exist? What are its properties? The scientists did not have any practical application in mind, but were curious about the world around them. Below are some examples of questions that where pure research would be used:

- How was the universe formed?
- Is there life on Mars?
- What are protons, neutrons, and electrons composed of?
- What are the properties of boron?

Applied Research Examples

There is a great deal of "applied" research taking place today. In general, no new science principles are discovered, but existing knowledge is used to develop a new product. Research on laundry detergents will probably not give us any new concepts about soap, but will help us to develop materials that get our clothes cleaner, use less water, and create lower amounts of pollution. Petroleum companies use applied research to find better ways to power vehicles, better lubricants to cut down on engine wear, and better ways to lower air pollution. These companies will use information that is readily available to come up with new products.



Figure **1.5.1.1**: U.S. Air Force Technical Sergeant Shanice Spearman at Aerospace Fuels Laboratory NCO prepares fuel samples for flashpoint testing. (Credit: Airman 1st Class Greg Erwin; Source: https://www.kadena.af.mil/News/Photos/igphoto/2001869405/(opens in new window); License: Public Domain)

"In-Between" Examples

The line between pure science and applied science is not always distinct. What may start out as simply asking a question may result is some very useful information. If scientists are studying the biochemistry of a microorganism that causes a disease, they may soon find information that would suggest a way to make a chemical that would inactivate the microorganism. The compound could be used to learn more about the biochemistry, but could also be used to cure the disease.

Hemoglobin is a protein in red blood cells that transports oxygen in the bloodstream. Scientists studied hemoglobin simply to learn how it worked. Out of this research came an understanding of how the protein changes shape when oxygen attaches to it. This information was then applied to help patients with sickle cell anemia, a disorder caused by an abnormal hemoglobin structure that makes hemoglobin molecules clump up when oxygen leaves the protein. Basic knowledge of protein structure led to an improved understanding of a wide-spread disease and opened the door for development of treatments.

Section Summary

- Pure research focuses on understanding basic properties and processes.
- Applied research focuses on the use of information to create useful materials.
- Sometimes there is no clear line between pure and applied research.

Contributors

Curated from resources found in Introductory Chemistry (CK-12)

This page titled 1.5.1: Pure and Applied Science is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **1.5: Pure and Applied Chemistry** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

1.5.2: Fields of Science

Learning Objectives

- Know the subject matter field for physics, chemistry, earth science, and astronomy.

Science has become so vast that it now includes many separate fields of study, each of which contain their own sub-disciplines. This course is focused on physical sciences, which are natural sciences focused on non-living things. We will provide a brief summary of some of these disciplines. This text will primarily focus on the first two of these: physics and chemistry. The other two, earth science and astronomy, depend on an understanding of chemistry and physics.

Physics

Science consists of the theories and laws that are the general truths of nature as well as the body of knowledge they encompass. Scientists are continually trying to expand this body of knowledge and to perfect the expression of the laws that describe it. **Physics** is concerned with describing the interactions of energy, matter, space, and time, and it is especially interested in what fundamental mechanisms underlie every phenomenon. The concern for describing the basic phenomena in nature essentially defines the *realm of physics*.

Physics aims to describe the function of everything around us, from the movement of tiny charged particles to the motion of people, cars, and spaceships. In fact, almost everything around you can be described quite accurately by the laws of physics. Consider a smart phone as in **Figure 1.5.2.1**: Physics describes how electricity interacts with the various circuits inside the device. This knowledge helps engineers select the appropriate materials and circuit layout when building the smart phone. Next, consider a Global Positioning System (GPS). Physics describes the relationship between the speed of an object, the distance over which it travels, and the time it takes to travel that distance. When you use a GPS device in a vehicle, it utilizes these physics equations to determine the travel time from one location to another.



Figure 1.5.2.1: The Apple “iPhone” is a common smart phone with a GPS function. Physics describes the way that electricity flows through the circuits of this device. Engineers use their knowledge of physics to construct an iPhone with features that consumers will enjoy. One specific feature of an iPhone is the GPS function. GPS uses physics equations to determine the driving time between two locations on a map. (credit: @gletham GIS, Social, Mobile Tech Images)

Chemistry

Chemistry is the study of the composition of matter and the changes that matter undergoes. Matter is anything that has mass and takes up space. Virtually everything around us is matter, including both living and nonliving things. Chemistry affects nearly everything we see and every action we take. Chemistry explains why milk that is left in the refrigerator for too long turns sour. Chemistry explains why certain pollutants called chlorofluorocarbons have done lasting damage to the ozone layer of our planet. Chemistry explains why the leaves of deciduous trees turn from green in the summer to various shades of red and yellow in the autumn **Figure 1.5.2.2**.



Figure 1.5.2.2 Chemical reactions in the leaves of deciduous trees cause them to change color from green to red, orange, or yellow before dropping to the ground. (Credit: User:Jongleur100/Wikimedia Commons; Source: https://commons.wikimedia.org/wiki/File:Country_lane.jpg; License: Public Domain)

Chemistry touches every area of our lives. The medicines we take, the food we eat, the clothes we wear—all of these materials and more are, in some way or another, products of chemistry.

Chemists look at the world in two ways, often simultaneously. The two worlds of the chemist are the **macroscopic** world and the **microscopic** world. Macroscopic refers to substances and objects that can be seen, touched, and measured directly. Microscopic refers to the small particles that make up all matter. Chemists must observe matter and do experiments macroscopically; then make generalizations and propose explanations that are microscopic in nature. For example, anyone can observe the physical change in appearance that occurs as an iron object, such as a tractor, is left out in the elements and gradually turns to rust. However, a chemist looking at the rusting tractor considers the individual atoms that make up the iron, and how they are changing as a result of exposure to oxygen in the air, and water from rain. Throughout the study of chemistry, there is often a switch back and forth between the macroscopic and microscopic worlds.

Earth Science

Earth is the mighty planet upon which we all live. Only recently have humans begun to understand the complexity of this planet. In fact, it was only a few hundred years ago that we discovered that Earth was just a tiny part of an enormous galaxy, which in turn is a small part of an even greater universe. Earth Science deals with any and all aspects of the Earth. Our Earth has molten lava, icy mountain peaks, steep canyons and towering waterfalls. Earth scientists study the atmosphere high above us as well as the planet's core far beneath us. Earth scientists study parts of the Earth as big as continents and as small as the tiniest atom. In all its wonder, Earth scientists seek to understand the beautiful sphere on which we thrive (**Figure 1.5.2.3**).

Because the Earth is so large and science is so complex, Earth scientists specialize in studying just a small aspect of our Earth. Since all of the branches are connected together, specialists work together to answer complicated questions. Important branches of Earth Science include geology, oceanography, climatology, and meteorology.

Astronomy

Astronomers have proven that our Earth and solar system are not the only set of planets in the universe. As of June 2015, over a thousand planets outside our solar system had been discovered. Although no one can be sure how many there are, astronomers estimate that there are billions of other planets. In addition, the universe contains black holes, other galaxies, asteroids, comets, and nebula. As big as Earth seems to us, the entire universe is vastly greater. Our Earth is an infinitesimally small part of our universe.

Astronomers use resources on the Earth to study physical things beyond the Earth. They use a variety of instruments like optical telescopes and radio telescopes to see things far beyond what the human eye can see. Spacecraft travel great distances in space to send us information on faraway places, while telescopes in orbit observe astronomical bodies from the darkness of space (**Figure 1.5.2.4**).

Astronomers ask a wide variety of questions. Astronomers could study how an object or energy outside of Earth could affect us. An impact from an asteroid could have terrible effects for life on Earth. Strong bursts of energy from the sun, called solar flares, can knock out a power grid or disturb radio, television or cell phone communications. But astronomers ask bigger questions too. How was the universe created? Are there other planets on which we might live? Are there resources that we could use? Is there other life

out there? Astronomy also relies on Earth Science, when scientists compare what we know about life on Earth to the chances of finding life beyond this planet.

Section Summary

- Physics is concerned with describing the interactions of energy, matter, space, and time, and it is especially interested in what fundamental mechanisms underlie every phenomenon.
- Chemistry is the study of the composition of matter and the changes that matter undergoes.
- Earth Science deals with any and all aspects of the Earth.
- Astronomy is the study of the universe beyond the Earth.

Contributions and Attributions

This page was curated by Jamie MacArthur (Madera Community College) using resources from the following locations.

- CK-12: Defining Chemistry
- OpenStax: Physics- An Introduction
- Libretexts: Earth Science and Its Branches

This page titled 1.5.2: Fields of Science is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

CHAPTER OVERVIEW

2: Matter and Motion

- 2.1: Introduction to Motion
- 2.2: Modern Ideas about Motion
- 2.3: Direction
- 2.4: Vectors, Scalars, and Coordinate Systems
- 2.5: Speed and Velocity
- 2.6: Acceleration
- 2.7: Determining how to solve a motion equation problem
- 2.8: Motion Graphs
  - 2.8.1: Motion graphs of position, velocity, and speed versus time
  - 2.8.2: Motion graphs of acceleration, velocity, and position versus time.
  - 2.8.3: Predicting Motion graphs
- 2.9: Types of Acceleration
  - 2.9.1: Falling Objects
  - 2.9.2: Centripetal Acceleration
  - 2.9.3: Projectile Motion
- 2.10: Measuring the World
- 2.11: Development of Force Concept
- 2.12: Galileo and the Beginnings of Modern Science

This page titled 2: Matter and Motion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

2.1: Introduction to Motion



Figure **2.1.1**: 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)

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?* 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.

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 we will study only the *motion* of a football, for example, without worrying about what forces cause or change its motion. Towards the end of the chapter we will also begin to introduce ideas about force to explain that motion. But this concept will be further expanded in later chapters.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.1: Introduction to Motion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.1: Introduction to One-Dimensional Kinematics** by OpenStax is licensed CC BY 4.0.

2.2: Modern Ideas about Motion

Modern ideas about motion originated during the Scientific Revolution. Prior to this time, logical reasoning alone was held to be the way to understand the natural world. During the Scientific Revolution, the need for measurements of the world around us became an important of understanding the world we live in.

The predominant ideas about motion were built on the thoughts of the Greek philosopher Aristotle. Aristotle held the belief that heavy things fall faster than lighter things. There were other ideas at the time and proposed since that time, but this was the dominant view of motion for around 2000 years.

There were many scientific endeavors even before the Scientific Revolution across the world that gradually built up the evidence for our modern understanding of motion. Some of the names of people who did this work before and during the Scientific Revolution are Abu'l-Barakāt al-Baghdādī, Al-Biruni, Albert of Saxony, Bhāskarāchārya, Brahmagupta, Jean Buridan, Jan Comets de Groot, Johannes Kepler, and Leonardo da Vinci.

There was no single scientist who brought about our modern understanding, however the story that gained the most notoriety was that involving Galileo Galilei and the Leaning Tower of Pisa*. If two objects made of the same material but of differing masses were dropped from the Tower, they would arrive at the bottom at the same time. Galileo was not the first person to make this assertation, but along with his other contributions to the understanding of motion, it was ultimately convincing.

In the rest of this chapter we will examine a few more ideas about motion. What is motion and how do we keep track of the changing of motion? What is mass, and what can cause a mass to move?

*It is unclear if Galileo ever performed this experiment or whether it was a thought experiment. However, many such experiments have been performed from a variety of towers since that time to confirm that this is correct.

This page titled 2.2: Modern Ideas about Motion is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

2.3: Direction

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.



Figure **2.3.1**: 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. (See Figure **2.3.2**.) 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. (See Figure **2.3.3**.)

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 Δ always means "change in" whatever quantity follows it; thus, Δx means *change in position*. Always solve for displacement by subtracting initial position x_0 from final position x_f. In order to solve such problems, you may need to define a frame of reference for both the initial and final positions of the object that moved. Further discussion of how to assign frames of reference will be discussed later in this chapter. For now, see if you can solve the problems below based on the frame of reference provided.

Note that the International System of Units (SI) unit for displacement is the meter (m), 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.

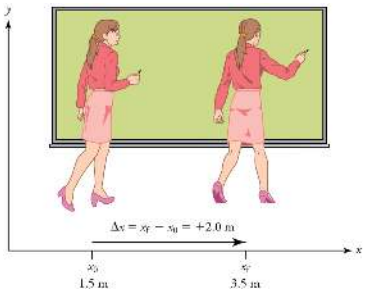


Figure **2.3.2**: A professor paces left and right while lecturing. Her position relative to the left side of the room is given by x. The **+2.0 m** displacement of the professor relative to the left side of the room is represented by an arrow pointing to the right.

Note that displacement has a direction as well as a magnitude. The professor's displacement is 2.0 m to the right, and the airline passenger's displacement is 4.0 m toward the rear. 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 $x_0 = 1.5$ m and her final position is $x_f = 3.5$ m. Thus her displacement is

$$\Delta x = x_f - x_0 = 3.5 \text{ m} - 1.5 \text{ m} = +2.0 \text{ m}.$$

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 $x_0 = 6.0$ m and his final position is $x_f = 2.0$ m, so his displacement is

$$\Delta x = x_f - x_0 = 2.0 \text{ m} - 6.0 \text{ m} = -4.0 \text{ m}.$$

His displacement is negative because his motion is toward the rear of the plane, or in the negative x direction in our coordinate system.

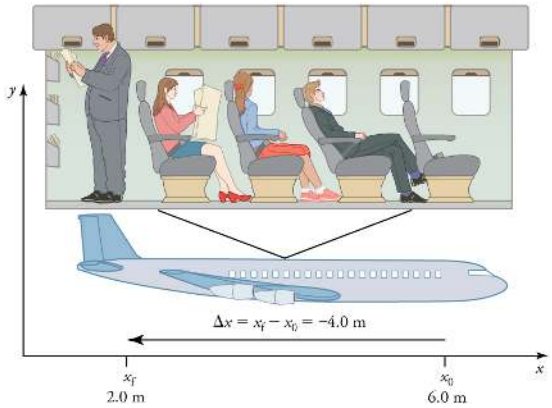


Figure **2.3.3**: A passenger moves from his seat to the back of the plane. His location relative to the airplane is given by x. The **−4.0-m** displacement of the passenger relative to the plane is represented by an arrow toward the rear of the plane. Notice that the arrow representing his displacement is twice as long as the arrow representing the displacement of the professor (he moves twice as far) in Figure **2.3.2**.

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* can be greater than the magnitude of the displacement (by magnitude, we mean 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 2.3.1

A cyclist rides 3 km west and then turns around and rides 2 km east. (NOTE: treat the eastward direction as positive for the sake of this problem.) (a) What is her displacement? (b) What distance does she ride? (c) What is the magnitude of her displacement?

Answer

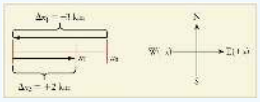


Figure **2.3.4**

(a) The rider's displacement is $\Delta x = x_f - x_0 = -1$ **km**. (The displacement is negative because we take east to be positive and west to be negative.)

(b) The distance traveled is **3 km** + **2 km** = **5 km**.

(c) The magnitude of the displacement is **1 km**.

Section 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 x_f 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

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

distance traveled

the total length of the path traveled between two positions

This page titled 2.3: Direction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.2: Displacement** by OpenStax is licensed CC BY 4.0.

2.4: Vectors, Scalars, and Coordinate Systems

Learning Objectives

- Define and distinguish between scalar and vector quantities.
- Assign a coordinate system for a scenario involving one-dimensional motion.



Figure **2.4.1**: 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)

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 kilometers per hour (km/h) east and a force of 500 newtons straight down.

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 **2.4.1**, 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.

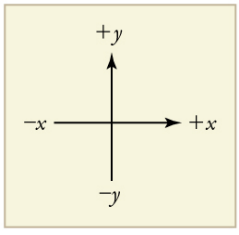


Figure **2.4.2**: It is usually convenient to consider motion upward or to the right as positive (+) and motion downward or to the left as negative (−).

? Exercise 2.4.1

A person's speed can stay the same as they round 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).

Section 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

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

vector

a quantity that is described by both magnitude and direction

This page titled 2.4: Vectors, Scalars, and Coordinate Systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.3: Vectors, Scalars, and Coordinate Systems** by OpenStax is licensed CC BY 4.0.

2.5: Speed and Velocity

Learning Objectives

- 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.
- Interpret graphs of position vs. time and velocity vs. time.



Figure **2.5.1**: The motion of these racing snails can be described by their speeds and their velocities. (credit: tobitasflickr, Flickr)

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.

Time

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.)

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*,

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0},$$

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.

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

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{-4 \text{ m}}{5 \text{ s}} = -0.8 \text{ m/s}.$$

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.

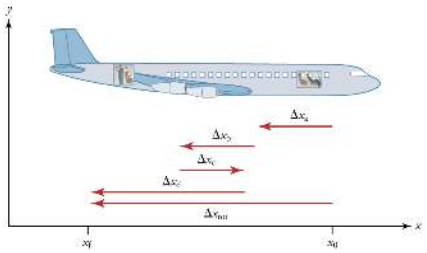


Figure **2.5.2**: A more detailed record of an airplane passenger heading toward the back of the plane, showing smaller segments of his trip.

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.

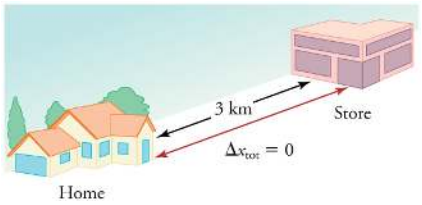


Figure **2.5.3**: During a 30-minute round trip to the store, the total distance traveled is 6 km. The average speed is 12 km/h. The displacement for the round trip is zero, since there was no net change in position. Thus the average velocity is zero.

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 **2.5.4**. (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.) We will discuss motion graphs in more detail later in this chapter. For now, you may use this as another way of visualizing how position, time, speed, and velocity are all changing together.

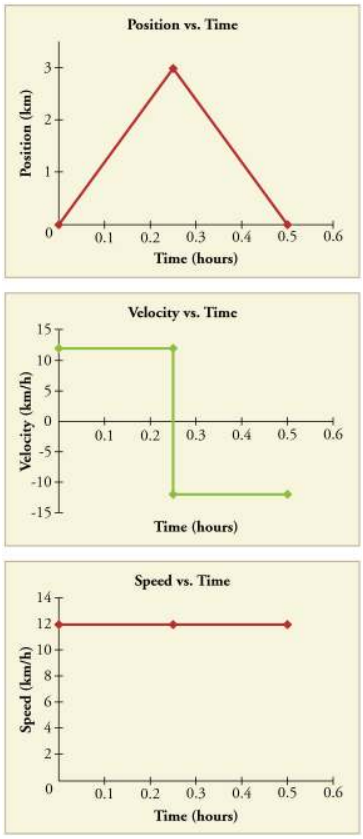


Figure **2.5.4**: Position vs. time, velocity vs. time, and speed vs. time on a trip. Note that the velocity for the return trip is negative.

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 2.5.1

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.

$$\frac{\text{distance}}{\text{time}} = \frac{80 \text{ miles}}{105 \text{ minutes}} \tag{2.5.1}$$

$$\frac{80 \text{ miles}}{105 \text{ minutes}} \times \frac{5280 \text{ feet}}{1 \text{mile}} \times \frac{1 \text{ meter}}{3.28 \text{ feet}} \times \frac{1 \text{ minute}}{60 \text{ seconds}} = 20 \text{ m/s} \tag{2.5.2}$$

Section 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.
- 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

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

elapsed time

the difference between the ending time and beginning time

This page titled 2.5: Speed and Velocity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

2.6: Acceleration

Learning Objectives

- Define and distinguish between velocity and acceleration, and between instantaneous and average acceleration.
- Calculate acceleration given initial time, initial velocity, final time, and final velocity.



Figure 2.6.1: A plane slows down as it comes in for landing in St. Maarten. It is accelerating in a direction opposite to its velocity. (credit: Steve Conry, Flickr)

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.

Definition: AVERAGE ACCELERATION

Average acceleration is *the rate at which velocity changes*,

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0},$$

where $\bar{a}$ is average acceleration, v is velocity, and t is time. (The bar over the a means *average* acceleration.)

Because acceleration is velocity in m/s divided by time in s, the SI units for acceleration are $\mathrm{m/s^2}$, 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*. If acceleration is in a direction opposite to the direction of motion, the object slows down.



Figure 2.6.2: A subway train in Sao Paulo, Brazil, slows down as it comes into a station. It is accelerating in a direction opposite to its direction of motion. (credit: Yusuke Kawasaki, Flickr)

Example 2.6.1: 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 2.6.3

Strategy

First we draw a sketch and assign a coordinate system to the problem. 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. (NOTE: you could have assigned west as positive as well, it is simply important to be aware of the direction as part of the answer because the average acceleration is a vector quantity.)

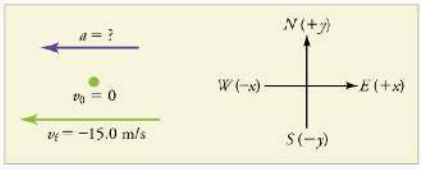


Figure 2.6.4

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. $v_0 = 0$, $v_f = -15.0\ \text{m/s}$ (the negative sign indicates direction toward the west), $\Delta t = 1.80\ \text{s}$.

2. Find the change in velocity. Since the horse is going from zero to $-15.0\ \text{m/s}$, its change in velocity equals its final velocity: $\Delta v = v_f = -15.0\ \text{m/s}$.

3. Plug in the known values (Δv and Δt) and solve for the unknown $\bar{a}$.

$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-15.0\ \text{m/s}}{1.80\ \text{s}} = -8.33\ \text{m/s}^2$.

Discussion

The negative sign for acceleration indicates that acceleration is toward the west. An acceleration of $8.33\ \text{m/s}^2$ 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 $8.33\ \text{m/s}^2$. 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—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 2.6.5 shows graphs of instantaneous acceleration versus time for two very different motions. In Figure 2.6.5 (a), 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 $1.8\ \text{m/s}^2$). In Figure 2.6.5 (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 $+3.0\ \mathrm{m/s^2}$ and $-2.0\ \mathrm{m/s^2}$, respectively.

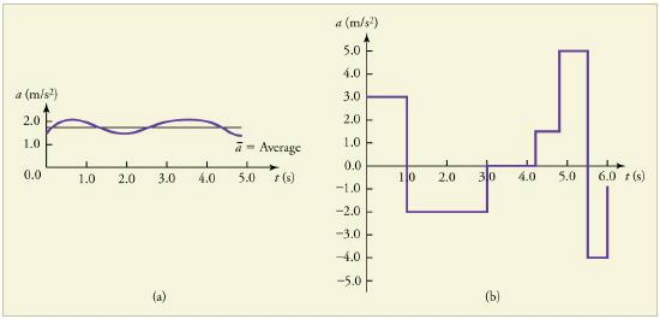


Figure 2.6.5: Graphs of instantaneous acceleration versus time for two different one-dimensional motions. (a) Here acceleration varies only slightly and is always in the same direction, since it is positive. The average over the interval is nearly the same as the acceleration at any given time. (b) Here 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.

? Exercise 2.6.1

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 slowing down: its acceleration is opposite in direction to its velocity.

Section 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 $\mathrm{m/s^2}$.
- 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 a is the acceleration at a specific instant in time.
- When an acceleration is in a direction opposite to that of the velocity of an object, the object slows down.

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

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.6: Acceleration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

2.7: Determining how to solve a motion equation problem

Learning Objectives

- Calculate any of the following from a relevant word problem about the motion of an object: distance, displacement, average velocity, or average acceleration.



Figure **2.7.1**: Kinematic equations can help us describe and predict the motion of moving objects such as these kayaks racing in Newbury, England. (credit: Barry Skeates, Flickr)

We have considered several terms to describe the motion of an object. In the next two sections we will look at relationships between three of those terms: displacement, velocity, and acceleration. Each of these terms is a vector quantity related to change in position. For displacement, this is a complete description of the term. For velocity and acceleration we are also considering time. Velocity is the change in displacement over a period of time. Acceleration is the change in velocity over a period of time. Additional relationships that involve multiple values can be derived. For example, there is a specific equation that relates acceleration and displacement. These equations are covered in more detail in a textbook for a full year course sequence about physics. For the purpose of this textbook, we will confine our interests to those equations already developed.

The rest of this section will consist of example problems pulled from content covered in the earlier sections of this chapter. As you consider these example problems, think about which of the movement concepts are most applicable to the problem before devising a solution method.

The next several examples consider the motion of the subway train shown in Figure **2.7.6**. In (a) the shuttle moves to the right, and in (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.

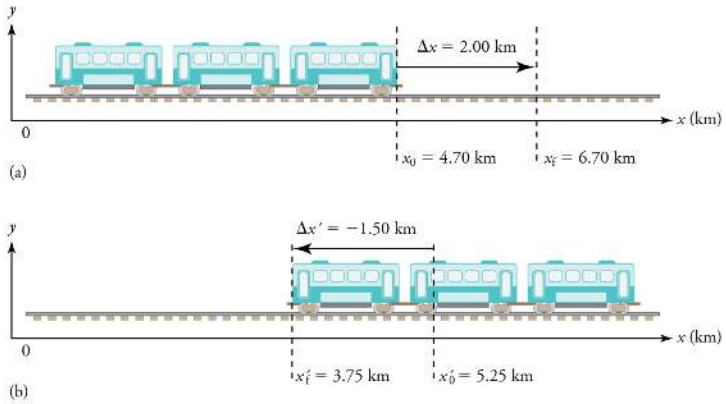


Figure **2.7.6**: One-dimensional motion of a subway train considered in Example **2.7.2**, Example **2.7.3**, Example **2.7.4**, Example **2.7.5**, Example **2.7.6**, and Example **2.7.7**. Here we have chosen the x-axis so that + means to the right and − means to the left for displacements, velocities, 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 $\Delta x'$ is −1.5 km. (Note the apostrophe symbol (′) read aloud as "prime" in this context, 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 2.7.2: 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 2.7.6?

Strategy

A drawing with a coordinate system is already provided, so we do not 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 $x_f = 6.70\ \text{km}$ and $x_0 = 4.70\ \text{km}$ for part (a), and $x_f' = 3.75\ \text{km}$ and $x_0' = 5.25\ \text{km}$ for part (b).

2. Solve for displacement in part (a).

$$\Delta x = x_f - x_0 = 6.70\ \text{km} - 4.70\ \text{km} = +2.00\ \text{km}$$

3. Solve for displacement in part (b).

$$\Delta x' = x_f' - x_0' = 3.75\ \text{km} - 5.25\ \text{km} = -1.50\ \text{km}$$

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 2.7.3: 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 2.7.6?

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 2.7.2. Distance traveled is the total length of the path traveled between the two positions. In the case of the subway train shown in Figure 2.7.6, 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 $-1.5\ \text{km}$. 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 2.7.4: Calculating Acceleration: A Subway Train Speeding Up

Suppose the train in Figure 2.7.6 (a) 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:

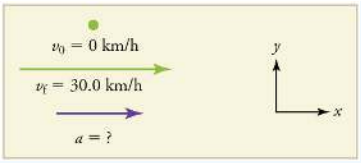


Figure **2.7.7**

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), $v_f = 30.0\ \text{km/h}$, and $\Delta t = 20.0\ \text{s}$.

2. Calculate Δv. Since the train starts from rest, its change in velocity is $\Delta v = +30.0\ \text{km/h}$, where the plus sign means velocity to the right.

3. Plug in known values and solve for the unknown, $\bar{a}$.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+30.0\ \text{km/h}}{20.0\ \text{s}}$$

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.

$$\bar{a} = \left(\frac{+30\ \text{km/h}}{20.0\ \text{s}}\right)\left(\frac{10^3\ \text{m}}{1\ \text{km}}\right)\left(\frac{1\ \text{h}}{3600\ \text{s}}\right) = 0.417\ \text{m/s}^2$$

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 2.7.5: Calculate Acceleration: A Subway Train Slowing Down

Now suppose that at the end of its trip, the train in Figure **2.7.6** (a) slows to a stop from a speed of 30.0 km/h in 8.00 s. What is its average acceleration while stopping?

Strategy

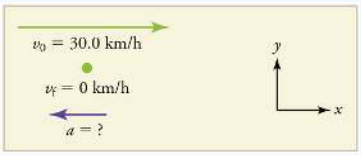


Figure **2.7.8**

In this case, the train is slowing down 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. $v_0 = 30.0\ \text{km/h}$, $v_f = 0\ \text{km/h}$ (the train is stopped, so its velocity is 0), and $\Delta t = 8.00\ \text{s}$.

2. Solve for the change in velocity, Δv.

$$\Delta v = v_f - v_0 = 0 - 30.0\ \text{km/h} = -30.0\ \text{km/h}$$

3. Plug in the knowns, Δv and Δt, and solve for $\bar{a}$.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-30.0\ \text{km/h}}{8.00\ \text{s}}$$

4. Convert the units to meters and seconds.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \left(\frac{-30.0 \text{ km/h}}{8.00 \text{ s}}\right)\left(\frac{10^3 \text{ m}}{1 \text{ km}}\right)\left(\frac{1 \text{ h}}{3600 \text{ s}}\right) = -1.04 \text{ m/s}^2.$$

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.

The graphs of position, velocity, and acceleration vs. time for the trains in Example **2.7.4** and Example **2.7.5** are displayed in Figure **2.7.9**. (We have taken the velocity to remain constant from 20 to 40 s, after which the train slows down.)

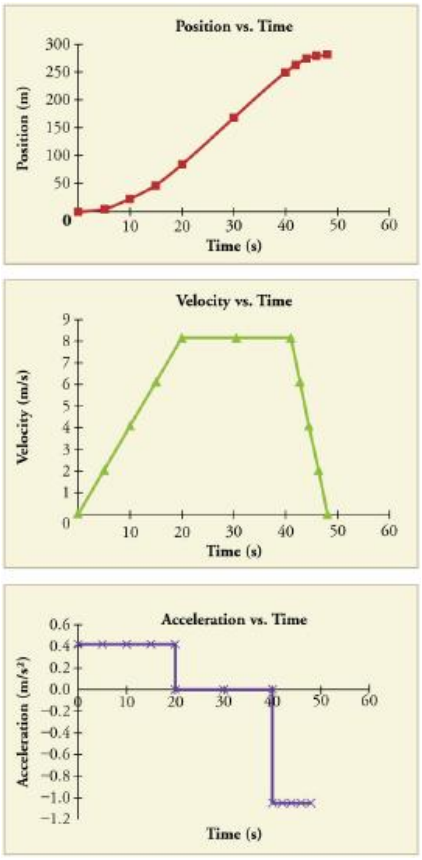


Figure **2.7.9**: (a) Position of the train over time. Notice that the train's position changes slowly at the beginning of the journey, then more and more quickly as it picks up speed. Its position then changes more slowly as it slows down at the end of the journey. In the middle of the journey, while 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 slows down 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 **2.7.6**: Calculating Average Velocity: The Subway Train

What is the average velocity of the train in part b of Example **2.7.2**, and shown again below, if it takes 5.00 min to make its trip?

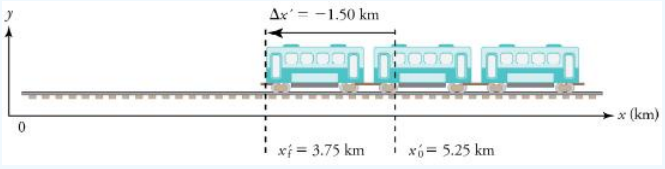


Figure **2.7.10**

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. $x'_f = 3.75\ \text{km},\ x'_0 = 5.25\ \text{km},\ \Delta t = 5.00\ \text{min}$.

2. Determine displacement, $\Delta x'$. We found $\Delta x'$ to be $-1.5\ \text{km}$ in Example 2.7.2.

3. Solve for average velocity.

$$\bar{v} = \frac{\Delta x'}{\Delta t} = \frac{-1.50\ \text{km}}{5.00\ \text{min}}$$

4. Convert units.

$$\bar{v} = \frac{\Delta x'}{\Delta t} = \left(\frac{-1.50\ \text{km}}{5.00\ \text{min}}\right)\left(\frac{60\ \text{min}}{1\ \text{h}}\right) = -18.0\ \text{km/h}$$

Discussion

The negative velocity indicates motion to the left.

Example 2.7.7: Calculating Deceleration: The Subway Train

Finally, suppose the train in Figure 2.7.10 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:

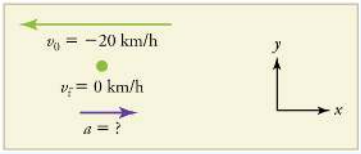


Figure 2.7.11

As before, we must find the change in velocity and the change in time to calculate average acceleration.

Solution

1. Identify the knowns. $v_0 = -20\ \text{km/h},\ v_f = 0\ \text{km/h},\ \Delta t = 10.0\ \text{s}$.

2. Calculate Δv. The change in velocity here is actually positive, since

$$\Delta v = v_f - v_0 = 0 - (-20\ \text{km/h}) = +20\ \text{km/h}$$

3. Solve for $\bar{a}$.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+20.0\ \text{km/h}}{10.0\ \text{s}}$$

4. Convert units.

$$\bar{a} = \left(\frac{+20.0\ \text{km/h}}{10.0\ \text{s}}\right)\left(\frac{10^3\ \text{m}}{1\ \text{km}}\right)\left(\frac{1\ \text{h}}{3600\ \text{s}}\right) = +0.556\ \text{m/s}^2$$

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).

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 **2.7.7**, 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 **2.7.10** is sped up by an acceleration to the left. In that case, both $\boldsymbol{v}$ and $\boldsymbol{a}$ are negative. The plus and minus signs give the directions of the accelerations. If acceleration has the same sign as the change in velocity, the object is speeding up. If acceleration has the opposite sign of the change in velocity, the object is slowing down.

Section Summary

- Consider which values are provided in a motion problem and how they are related to one another.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.7: Determining how to solve a motion equation problem is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.6: Motion Equations for Constant Acceleration in One Dimension** by OpenStax is licensed CC BY 4.0.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

2.8: Motion Graphs

Learning Objectives

- Interpret motion graphs and the relationship between different types of motion graphs.
- Predict motion graphs.

In the previous sections of this text we looked at several example problems in which motion graphs were used to model the solution to those problems. In this section, we will revisit some of those problems, and focus specifically on the trends observed within the motion graphs created for those problems. Finally, we will also hint at some forthcoming material related to a common type of motion as we consider some motion graphs we haven't seen yet. These ideas are explored in three subsections which follow.

This page titled 2.8: Motion Graphs is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

2.8.1: Motion graphs of position, velocity, and speed versus time

Learning Objectives

- Interpret motion graphs of velocity, and position versus time as well as the relationship between these graphs.

Earlier in this chapter, we looked an at example of comparing position, speed, and velocity by way of motions graphs. Let us revisit that example here.

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.

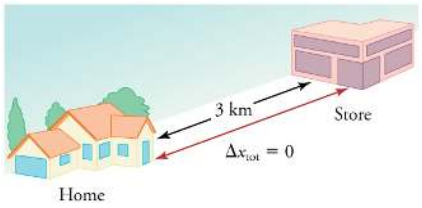


Figure **2.8.1.1**: During a 30-minute round trip to the store, the total distance traveled is 6 km. The average speed is 12 km/h. The displacement for the round trip is zero, since there was no net change in position. Thus, the average velocity is zero.

Let us now revisit the motion graphs and consider in more detail what is plotted on those graphs. In each case, the x-axis indicates time. Although there are other ways of graphically showing trends in motion, we will exclusively be plotting time on the x-axis throughout this text. The difference is then the value plotted on the y-axis. What is of interest to us in this section of the chapter is understanding how each type of motion graph is related.

The first graph shows the changing position with time: we call this the position versus time graph. Note that at the start of the position versus time graph this is indicated by an upward sloping line, indicating that the movement is initially at a constant speed in the same direction. Because the direction is the same, it also means that the velocity is also constant. If we then move to the second graph of velocity versus time we see that throughout this section there is a flat line with a positive value showing the velocity. This is exactly what we should expect based on the position versus time graph. The slope of the position versus time graph is the value of the y-axis on the velocity versus time graph. Because the slope is constant on the position versus time graph, the value on the y-axis is a constant on the velocity versus time graph. Later in this section we will investigate how a non-constant slope on the position versus time graph would translate to the velocity versus time graph.

If we move towards the end of the position versus time graph we see that the line is now sloping downward instead of upwards. The corresponding velocity versus time graph should indicate a negative value in this range because the slope of the line in the position versus time graph is negative. The value in the velocity versus time graph is also constant because the slope is constant.

The final graph shows speed versus time. For this graph, we are not considering direction, so whether the line is sloping upwards or downwards in the position versus time graph, the speed will be a positive value. Notice that the speed versus time graph is constant all the way across. There are two reasons for this. The position versus time graph has a constant slope whether it is positive or negative, so this will result in a straight line on the speed versus time graph just as it does on the velocity versus time graph. Because the magnitude of that slope remains the same whether the velocity is positive or negative in this example, the magnitude of the speed remains the same across the entire graph.

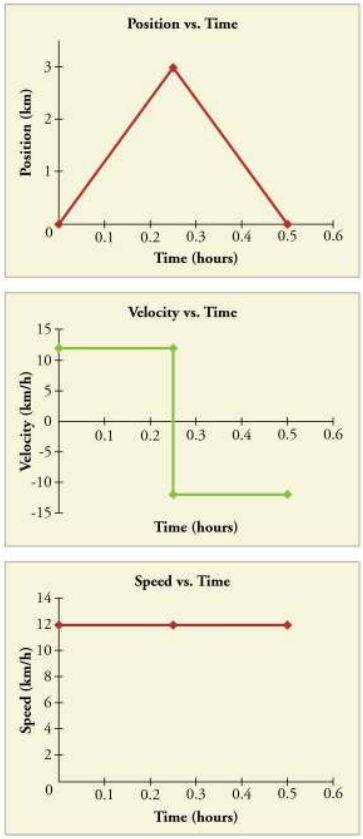


Figure **2.8.1.2**: Position vs. time, velocity vs. time, and speed vs. time on a trip. Note that the velocity for the return trip is negative.

We can consider other possibilities for position versus velocity graphs. What will happen to the velocity versus time graph if the position stops changing? What will happen to the velocity versus time graph if the position starts changing at a different rate? We will consider some of these possibilities as we continue in this chapter.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.8.1: Motion graphs of position, velocity, and speed versus time is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

2.8.2: Motion graphs of acceleration, velocity, and position versus time.

Learning Objectives

- Interpret motion graphs of acceleration, velocity, and position versus time as well as the relationship between these graphs.

Earlier in this chapter we examined some example problems about a moving train. We are now going to further examine the motion graphs that acompanied those example problems. The relevant motion graphs are repeated here in Figure **2.8.2.1**.

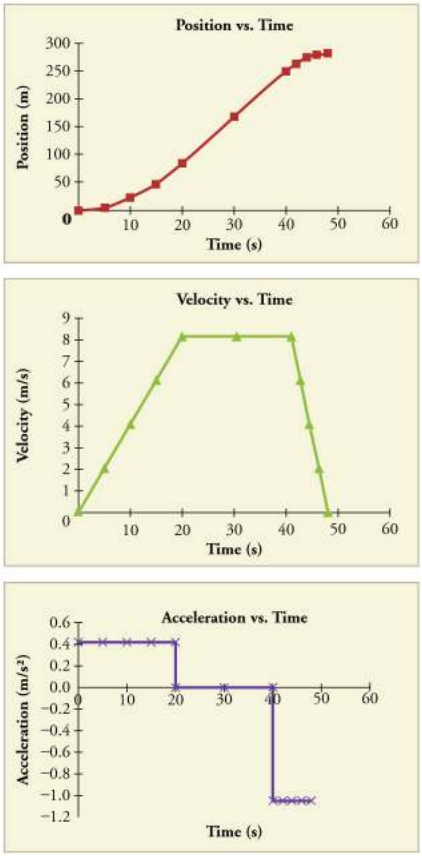


Figure **2.8.2.1**: (a) Position of the train over time. Notice that the train's position changes slowly at the beginning of the journey, then more and more quickly as it picks up speed. Its position then changes more slowly as it slows down at the end of the journey. In the middle of the journey, while 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 slows down 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.

In the position versus time graph, you will notice that the position increases throughout the region graphed, but not at a constant rate. At the start of the graph the slope of the curve is increasing as time increases. Eventually this slope appears to be constant for awhile, And then towards the end of the graph, the slope seems to decrease. It can sometimes be a challenge to interpret a graph of changing slope so here are a few things to consider in that regard. If we look at the first two marks on the graph, it appears as if the position has changed barely at all. We can imagine if the curve continued with this slope that it would take a very long time to even go as far as 50 m. If we look at the second and third points, the position has increased quite a bit more over a similar period of time. This indicates that the slope of the line is increasing.

In the velocity versus time graph we can see how the trends in the position versus time graph are indicated. The portion at the start of the velocity versus time graph is increasing, which corresponds to the portion of the position versus time graph where the slope is changing. This makes sense because the velocity is the slope of the position versus time graph. If the slope is not constant, then the velocity should not be constant either. However, it does appear to be constantly increasing. In the middle section of the position

versus time graph where the slope is constant we see that we have a constant velocity on the velocity versus time graph. This is consistent with what we saw on motion graphs earlier in this chapter. Finally, we see at the end of the velocity versus time graph a downward sloping straight line. This is consistent with what we had seen earlier for the position versus time graph with a decrease in the slope of the graph towards the end. Because the slope is decreasing, the velocity is also decreasing. Note that the velocity is not negative here, simply becoming smaller. (What would you expect to see if the velocity were negative?)

The final graph is acceleration versus time. In this graph we see only flat lines, but different values for those flat lines. At the start we see a positive flat line for the value of acceleration. This corresponds to the section of the velocity versus time graph where the velocity is increasing. In the middle section where the velocity is constant, we see a flat line at exactly zero acceleration. At the end of the graph where the velocity is decreasing, we see a negative flat line. If we consider that acceleration is a measure of the change in velocity with time, all of these results should be consistent. When the velocity is increasing, the acceleration is positive. When the velocity is constant, the acceleration is zero. When the velocity is decreasing, the acceleration is negative.

You might be wondering what would happen to all of these graphs if the acceleration were changing incrementally as we had seen for velocity and position in previous examples. That is possible, but beyond the scope of this class. We will only consider examples of constant acceleration or stepping between constant accelerations. Sometimes that acceleration will be positive. Sometimes it will be negative. And sometimes it will be zero. We will next consider how both position and velocity might change over a longer period of time when they have a constant acceleration

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.8.2: Motion graphs of acceleration, velocity, and position versus time. is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

2.8.3: Predicting Motion graphs

Learning Objectives

- Predict the velocity versus time graph that would result from a position versus time graph.
- Predict the acceleration versus time graph that would result from a velocity versus time graph.

We will consider one final example of interpreting motion graphs. As you look at the position versus time graph, imagine what you will see on the velocity versus time graph. Once you have made the prediction, read the discussion and look at the actual velocity versus time graph.

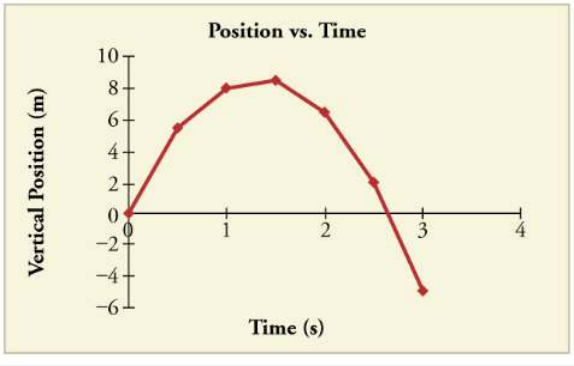


Figure **2.8.3.1**: Vertical position versus time graph for an object that starts by moving positive in the vertical direction before gradually slowing and then moving negative in the vertical direction while gradually speeding up.

There is no point on the position versus time graph at which the slope appears to be constant. If we look at the change in the slope as we move across the graph it seems to decrease until we get to a point where the position starts to go down instead of going up. This at first might seem similar to Figure 2.8.1.2 where the position increased and then abruptly began to decrease. However this is different in that the change in position with time is not a straight line leading up to the point where it starts to decrease. In fact, if there were more points on this graph we could see that it was a slow gradual curve with no abrupt points along the way. This is similar to what we saw in Figure 2.8.2.1. If we imagine that same trend being allowed to continue, the position would not just stop increasing, but eventually begin to decrease. This is indicative of a negative velocity. Regardless of whether the initial change in position is positive or negative, the fact that each successive interval is showing a smaller (or more negative) change in position indicates that the velocity is becoming more negative with time. Because we see a curved line for the position versus time graph, we might expect to see a straight line in the velocity versus time graph. (NOTE: We can say this is true because we are only considering constant acceleration in this text.)

The velocity versus time graph is indeed negatively sloped. Notice that the velocity values are positive at times where the position values are increasing, and becomes negative where the position values are decreasing. What does that tell us that we might expect to see in the acceleration versus time graph? Make a prediction and then read the discussion and look at the actual acceleration versus time graph.

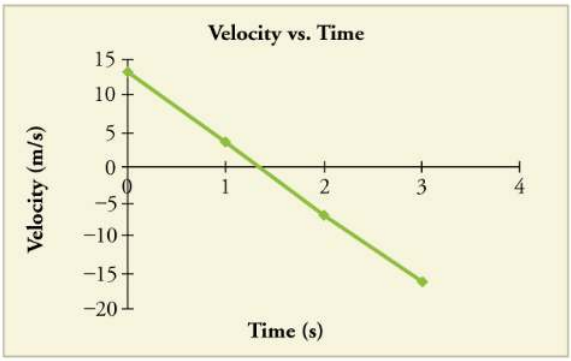


Figure 2.8.3.2: Velocity versus time graph for an object that has a velocity decreasing at a constant slope.

When we have a straight line with a negative slope for velocity versus time, that would indicate two things about acceleration: 1) the acceleration is constant because the velocity versus time graph is a straight line, and 2) the acceleration is negative because that line has a negative slope. Graphically we indicate a constant by drawing a flat line where the acceleration does not change as the time changes. Because we are only considering motion graphs with constant acceleration for this class there are essentially three options for the acceleration versus time graph: 1) a flat line at a positive number for an object which is accelerating in the positive direction, 2) a flat line at a zero value for an object which is not accelerating, and 3) a flatline at a negative number for an object which is accelerating in the negative direction (as we see in this example.)

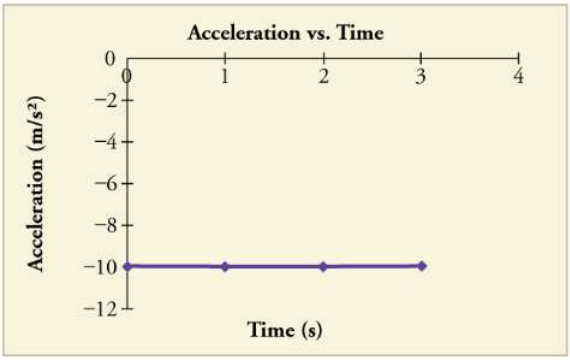


Figure 2.8.3.3: acceleration versus time graph for an object that has a constant negative acceleration.

You may continue to practice interpreting motion graphs in your study groups. Make a position versus time graph and then ask your study partners to make velocity versus time graphs and acceleration versus time graphs to accompany your graph.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.8.3: Predicting Motion graphs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.7: Falling Objects** by OpenStax is licensed CC BY 4.0.

2.9: Types of Acceleration

There are many types of acceleration that we experience in our lives. A few of them are common or interesting enough to be worth focusing on in a bit more detail. We will see these in the following subsections. There will be a few example problems in these subsections. They are simply applications of the same ideas of acceleration that we have already covered previously in this chapter. In later chapters we will learn more about why acceleration takes place.

This page titled 2.9: Types of Acceleration is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

2.9.1: Falling Objects

Learning Objectives

- Describe the effects of gravity on objects in motion.
- Describe the motion of objects that are in free fall.

Falling objects form an interesting class of motion problems. For example, we can estimate the depth of a vertical mine shaft by dropping a rock into it and listening for the rock to hit the bottom. By applying the kinematics developed so far to falling objects, we can examine some interesting situations and learn much about gravity in the process.

Gravity

The most remarkable and unexpected fact about falling objects is that, if air resistance and friction are negligible, then in a given location all objects fall toward the center of Earth with the *same constant acceleration, independent of their mass*. This experimentally determined fact is unexpected, because we are so accustomed to the effects of air resistance and friction that we expect light objects to fall slower than heavy ones.

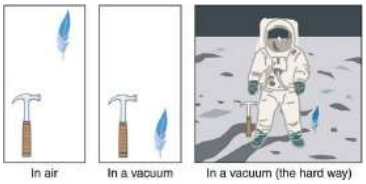


Figure **2.9.1.1**: A hammer and a feather will fall with the same constant acceleration if air resistance is considered negligible. This is a general characteristic of gravity not unique to Earth, as astronaut David R. Scott demonstrated on the Moon in 1971, where the acceleration due to gravity is only $1.67\ \mathrm{m/s^2}$.

In the real world, air resistance can cause a lighter object to fall slower than a heavier object of the same size. A tennis ball will reach the ground after a hard baseball dropped at the same time. (It might be difficult to observe the difference if the height is not large.) Air resistance opposes the motion of an object through the air, while friction between objects—such as between clothes and a laundry chute or between a stone and a pool into which it is dropped—also opposes motion between them. For the ideal situations of these first few chapters, an object *falling without air resistance or friction* is defined to be in **free-fall**.

The force of gravity causes objects to fall toward the center of Earth. The acceleration of free-falling objects is therefore called the **acceleration due to gravity**. The acceleration due to gravity is *constant*, which means we can apply the kinematics equations to any falling object where air resistance and friction are negligible. This opens a broad class of interesting situations to us. The acceleration due to gravity is so important that its magnitude is given its own symbol, g. It is constant at any given location on Earth and has the average value

$$g = 9.80\ \mathrm{m/s^2}.$$

Although g varies from $9.78\ \mathrm{m/s^2}$ to $9.83\ \mathrm{m/s^2}$, depending on latitude, altitude, underlying geological formations, and local topography, the average value of $9.80\ \mathrm{m/s^2}$ will be used in this text unless otherwise specified. The direction of the acceleration due to gravity is *downward (towards the center of Earth)*. In fact, its direction *defines* what we call vertical. Note that whether the acceleration a in the kinematic equations has the value $+g$ or $-g$ depends on how we define our coordinate system. If we define the upward direction as positive, then $a = -g = -9.80\ \mathrm{m/s^2}$, and if we define the downward direction as positive, then $a = g = 9.80\ \mathrm{m/s^2}$. In certain examples and homework exercises, slightly inaccurate but more convenient value of $g \simeq 10\ \mathrm{m/s^2}$ will be used with appropriate specifications.

Section Summary

- An object in free-fall experiences constant acceleration if air resistance is negligible.
- On Earth, all free-falling objects have an acceleration due to gravity g, which averages

$$g = 9.80\ \mathrm{m/s^2}.$$

- Whether the acceleration a should be taken as $+g$ or $-g$ is determined by your choice of coordinate system. If you choose the upward direction as positive, $a = -g = -9.80\ \mathrm{m/s^2}$ is negative. In the opposite case, $a = +g = 9.80\ \mathrm{m/s^2}$ is positive.
- For objects in free-fall, up is normally taken as positive for displacement, velocity, and acceleration.

Glossary

free-fall

the state of movement that results from gravitational force only

acceleration due to gravity

acceleration an object undergoes while in free-fall

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.9.1: Falling Objects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.7: Falling Objects** by OpenStax is licensed CC BY 4.0.

2.9.2: Centripetal Acceleration

Learning Objectives

- Explain what centripetal acceleration is.
- Use the formula for centripetal acceleration in simple situations.

We defined acceleration as a change in velocity, either in its magnitude or in its direction, or both. When an object moves along a circular path, the direction of its velocity changes constantly, so there is always an associated acceleration, even if the speed of the object is constant. You experience this acceleration yourself when you turn a corner in your car. What you notice is a sideways acceleration because you and the car are changing direction. The sharper the curve and the greater your speed, the more noticeable this acceleration will become. In this section we briefly examine the direction and magnitude of that acceleration.

Figure **2.9.2.1** shows an object moving in a circular path at constant speed, called **uniform circular motion**. The direction of the instantaneous velocity is shown at two points along the path. Acceleration is in the direction of the change in velocity, which points directly toward the center of rotation (the center of the circular path). This pointing is shown with the vector diagram in the figure. We call the acceleration of an object moving in uniform circular motion (resulting from a net external force) the **centripetal acceleration**(a_c); centripetal means "toward the center" or "center seeking."

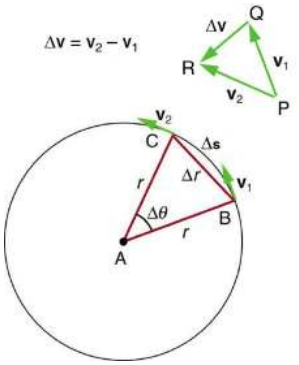


Figure **2.9.2.1**: The directions of the velocity of an object at two different points are shown, and the change in velocity $\Delta \mathbf{v}$ is seen to point directly toward the center of curvature. (See small inset.) Because $\mathbf{a_c = \Delta v/\Delta t}$, the acceleration is also toward the center; a_c is called centripetal acceleration. (For small time differences, $\Delta \theta$ is very small, and the arc length Δs is approximately equal to the chord length Δr.)

Section Summary

- Centripetal acceleration a_c is the acceleration experienced while in uniform circular motion. It always points toward the center of rotation. It is perpendicular to the linear velocity v.
- The unit of centripetal acceleration is $\mathrm{m/s^2}$.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax

This page titled 2.9.2: Centripetal Acceleration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.9: Centripetal Acceleration** by OpenStax is licensed CC BY 4.0.

2.9.3: Projectile Motion

Learning Objectives

- Identify and explain the properties of a projectile, such as acceleration due to gravity, range, and 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 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 know here is that *motions along perpendicular axes are independent* and thus can be considered separately. In particular, it is often convenient to consider vertical and horizontal motions separately, or independently. 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. And we will assume all forces except gravity are negligible (that is, we ignore forces due to air resistance, wind, friction, etc.). The vertical component of acceleration is, then, $a_y = -g = -9.80\ \mathrm{m/s^2}$, where we take upward direction as the positive y direction. And because there is no force in the horizontal direction, $a_x = 0$. Figure 2.9.3.1 illustrates this approach to analyzing projectile motion.

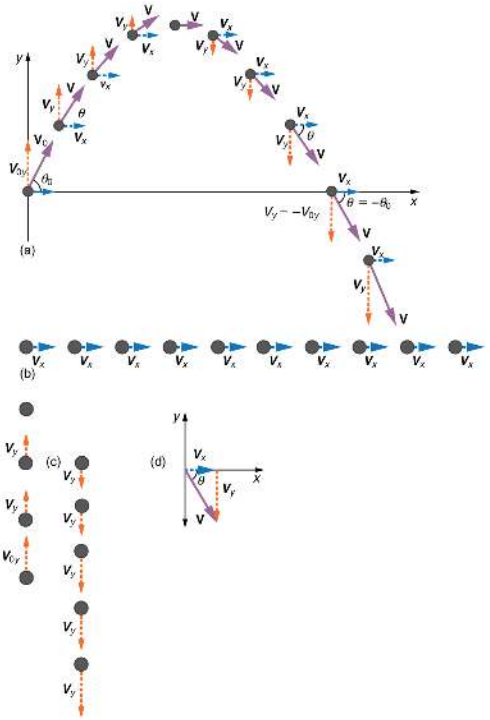


Figure 2.9.3.1: (a) We analyze two-dimensional projectile motion by breaking it into two independent one-dimensional motions along the vertical and horizontal axes. (b) The horizontal motion is simple, because $a_x = 0$ and v_x is thus constant. (c) The velocity in the vertical direction begins to decrease 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. This is exactly same as in the case of falling objects (or objects thrown directly upward). (d) The x - and y -motions can be recombined to give the total velocity at any given point on the trajectory.

One of the conceptual aspects of projectile motion we can discuss without a detailed analysis is the range. On level ground, we define **range** to be the horizontal distance R traveled by a projectile. The range of projectiles describes everyday phenomena such as how far a football can be thrown or kicked. Also, 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.

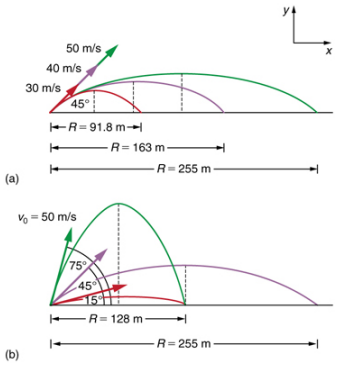


Figure **2.9.3.2**: Trajectories of projectiles on level ground. (a) The greater the initial speed v_0, the greater the range for a given initial angle. (b) The effect of initial angle θ_0 on the range of a projectile with a given initial speed. Note that the range is the same for **15°** and **75°**, although the maximum heights of those paths are different.

How does the initial velocity of a projectile affect its range? It is intuitive to guess that, the greater the initial speed v_0, the greater the range, as shown in Figure **2.9.3.2** (a). The initial angle θ_0 also has a dramatic effect on the range, as illustrated in Figure **2.9.3.2** (b). For a fixed initial speed, such as might be produced by a cannon, the maximum range is obtained with $\theta_0 = 45°$. Interestingly, for every initial angle except **45°**, there are two angles that give the same range. 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

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 **2.9.3.3**.) If the initial speed is great enough, the projectile goes into orbit. 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.

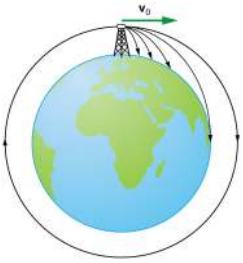


Figure **2.9.3.3**: Projectile to satellite. In each case shown here, a projectile is launched from a very high tower to avoid air resistance. With increasing initial speed, the range increases and becomes longer than it would be on level ground because the Earth curves away underneath its path. With a large enough initial speed, orbit is achieved.

Summary

- Projectile motion is the motion of an object through the air that is subject only to the acceleration of gravity.
- The most important fact regarding projectile motion is that motions along vertical direction and the horizontal direction are independent.
- The maximum horizontal distance traveled by a projectile is called the **range**.

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.

Projectile

an object subject only to the acceleration of gravity.

Range

the maximum horizontal distance that a projectile travels.

Trajectory

the path of a projectile through space.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.9.3: Projectile Motion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.8: Projectile Motion** by OpenStax is licensed CC BY 4.0.

2.10: Measuring the World

Learning Objectives

- Define mass, volume, and matter.
- Understand the relationship between mass, volume, and matter.
- Know the difference between base units and derived units.

The focus of this chapter is on movement, but sometimes knowledge of what is moving is important to completely understand the nature of the movement. Most movement that we will be studying in this course happens by objects we refer to as matter. There are several ways in which the movement of matter and the nature of that matter are related. We will investigate one of those in this chapter (force) and introduce others in later chapters. But to understand these relationships, we first must understand matter. Matter can be defined as anything that has a mass and takes up space.

Mass

The property of a body to remain at rest or to remain in motion with constant velocity is called **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 mass concept will be explored in more detail when we consider Newton's Laws in the next chapter.

Volume

We have largely been ignoring the space taken up by objects so far. We have been considering the movement of objects as if they were particles. This simply means that we are not considering the size or shape of the object when we think about how it moves. There are obviously situations where this model will not work at all, but it is a good starting point to understanding motion. When we think about occupying space, we are thinking about something called volume.

We have already considered three measurements: distance, time, and mass. We have also discussed velocity, speed, and acceleration, which are all related to both distance and time. Volume is similar to these other measurements in that it is related to multiple other measurements. We can say that it is made up of other measurements. In this case, it is multiple measures of distance. Volume is a measure of the three dimensions which make up an object. These might be labeled length, width, and height in some cases. But objects have a volume whether it can be easily measured in this way or not.

Derived Units

Volume, velocity, and acceleration are all examples of what are called derived units. Derived units are things that we measure as they are related to other measurable properties, usually by the product or ratio of some other property. Derived units can be composed of other derived units. For example, acceleration is a derived unit of velocity and time. However, some properties of matter (and energy as we will later see) are not dependent upon other properties. The measure of these properties are called base units. So far we have encountered the base units of mass, length, and time. We will learn additional base units as we move through this text. We have also learned the derived units of acceleration, velocity, and volume. We will learn many, many more derived units as we move through this course. As we learn about new measurements of the world, consider what base units those measures are derived from and how they relate to other measures we have previously discussed.

Glossary

Derived Units

something measured about an object related to other measurable properties of that object.

Inertia

the property of an object to stay at rest or to remain in motion with a constant velocity.

Mass

a measure of the inertia of an object.

Matter

Any object that has mass and occupies space.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

Volume

A measure of the space occupied by the three length dimensions which make up an object.

This page titled 2.10: Measuring the World is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- **1.8: Projectile Motion** by OpenStax is licensed CC BY 4.0.

2.11: Development of Force Concept

Learning Objectives

- Understand the definition of force.

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

A more quantitative definition of force can be based on some standard force, just as distance is measured in units relative to a standard distance. One possibility is to stretch a spring a certain fixed distance, as illustrated in Figure **2.11.1**, and use the force it exerts to pull itself back to its relaxed shape—called a *restoring force*—as a standard. The magnitude of all other forces can be stated as multiples of this standard unit of force. Many other possibilities exist for standard forces. Some alternative definitions of force will be given later in this text.

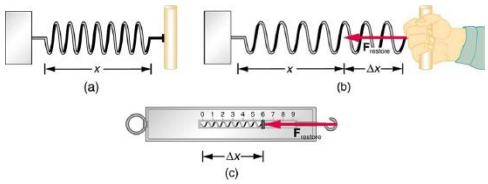


Figure **2.11.1**: The force exerted by a stretched spring can be used as a standard unit of force. (a) This spring has a length x when undistorted. (b) When stretched a distance Δx, the spring exerts a restoring force, $\mathbf{F}_{\text{restore}}$, which is reproducible. (c) A spring scale is one device that uses a spring to measure force. The force $\mathbf{F}_{\text{restore}}$ is exerted on whatever is attached to the hook. Here $\mathbf{F}_{\text{restore}}$ 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?

Section 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.

Glossary

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

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

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 2.11: Development of Force Concept is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.2: Development of Force Concept** by OpenStax is licensed CC BY 4.0.

2.12: Galileo and the Beginnings of Modern Science

Galileo Galilei was both a controversial figure and one of the first scientists. Do these two things necessarily go together? That depends on who you ask. The world might not have been ready for science yet, but Galileo wasn't going to let that stop him from sharing his thoughts about his observations.

Galileo was a meticulous observer, both in his work in understanding the motion of terrestrial objects and in his work discovering the movement of celestial objects. In terms of the impact of his work on the process of science as we know it today, it turns out that he had the greatest effect where he was the least controversial. Galileo is most famous in most circles for his contentious relationship with the Catholic church over the movement of the planets around the Sun (and the movement of moons around planets, some of which he was the first to ever observe). However, his overall contribution to the field of astronomy was not as influential as his contemporaries Tycho Brahe and Johannes Kepler. And though his instrumentation allowed him to see things that nobody else had ever seen before, the idea that the planets went around the Sun had been proposed by Copernicus a hundred years before *(and even by some early Babylonian astronomers whose names are now lost to history).*

Galileo's greatest contributions to science lay in his work on understanding motion. Galileo was able to show that the increase in velocity of a falling object was proportional to the time it was falling (hence defining acceleration, and in all cases he observed constant acceleration). He was able to measure this through the use of a very long device of variable inclination that he rolled steel balls along. To keep track of time in the days before stopwatches, he devised an elaborate system of weighing the water released from a jug.

In both his work on astronomy and in the development of the field of kinematics, Galileo's innovations were in the design and creation of novel instruments and in the meticulous observations he was able to make with those instruments. In astronomy he created telescopes. In kinematics he created the aforementioned inclined plane. Also important was his ability to observe these experiments with high precision. His ability to observe changes in velocity of an object rolling along an inclined plane was truly remarkable given the technology available to him at the time. Physics students of today often have trouble measuring as precisely as Galileo did back then, even with the technology that these students have available to them now! In addition to the new designs related to astronomy and kinematics, Galileo created more accurate thermometers, and compasses.

In addition to his groundbreaking work in instrumentation, observations, and experimentations, he needed to make sense of the data he was obtaining. For this, Galileo needed to have a comprehensive understanding of the mathematical techniques of the day. He did not really create new ways of thinking about math, but he needed to use the existing ways in order to formulate his idea. Galileo had a deep respect for the field of mathematics and was one of the first to claim that natural phenomena could be fully explained by mathematical principles.

There were several factors which lead to Galileo's eventual trial and sentencing by the Inquisition. He had been instructed by the Inquisition to no longer write about heliocentrism in 1616. The Inquisition's arguments against Galileo's beliefs largely relied on the ideas of Kepler, which are still understood to be true today. (Kepler proposed an elliptical instead of circular orbit, which could be thought of as the Earth and Sun revolving around each other.) Galileo didn't discuss heliocentrism again for decades, but did have a bitter dispute with a Jesuit about a comet in the meantime. Eventually Pope Urban VIII was elected, and based on their friendship Galileo thought he could now write about heliocentrism. However, the resulting work led to a new trial in 1633 which resulted in house arrest for the remainder of his life and a ban on any of his writings.

Galileo spent the last ten years of his life under this house arrest. It was during this time that he wrote about his work on kinematics which he had performed decades earlier: Two New Sciences. If he had not been so confined, he might not have taken the time to write down his greatest contributions to the field of science. On the other hand, everything he wrote at this point in his life was still being banned and he had to publish it in a non-Catholic country.

But perhaps Galileo's greatest contribution is not just an aspect of science, but science itself. Albert Einstein called Galileo the father of modern science. Other prominent scientists have echoed this sentiment or said very similar things. (Although it should be noted that Galileo was not the only one focused on this line of thought. Other prominent researchers of the 17th century include Francis Bacon, René Descartes, and Robert Boyle.) Prior to the work of Galileo, scientific thought in Europe was not distinct from religious and philosophical thought (as can be seen from its historical name: natural philosophy). It was perhaps mostly a curiosity within the fields of religion and philosophy that was beholden to the principles which governed these fields. The concept that it was

possible to discover natural laws through the observations of nature was itself a novel concept and so it is not surprising that it would lead to some controversy. In our modern world, philosophy, religion, and science are their own separate fields but there are still those who sometimes conflate them. This is especially true when science discovers (or sometimes rediscovers) aspects of reality that other fields of understanding make false assumptions about as part of constructing their worldview. As we discussed in chapter 1, there are multiple ways of knowing, and science is only one of them. But science is the way of knowing best equipped to provide clarity on how natural processes work based on observations of those processes.



This page titled 2.12: Galileo and the Beginnings of Modern Science is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

CHAPTER OVERVIEW

3: Types of Motion and Their Causes

3.1: Introduction to Dynamics and Momentum.
3.2: Newton’s First Law of Motion- Inertia
3.3: Newton's Second Law of Motion- Force and Acceleration
3.4: Newton’s Third Law of Motion- Symmetry in Forces
3.5: Examples of Force
3.5.1: Normal Force
3.5.2: Tension
3.5.3: Spring Force- Hooke’s Law
3.5.4: Friction
3.5.5: Centripetal Force
3.6: Summary of Newton's Laws
3.7: Newton’s Universal Law of Gravitation
3.8: Linear Momentum and Force
3.9: Conservation of Momentum
3.10: Rotation
3.11: Angular Momentum and Its Conservation

This page titled 3: Types of Motion and Their Causes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

3.1: Introduction to Dynamics and Momentum.

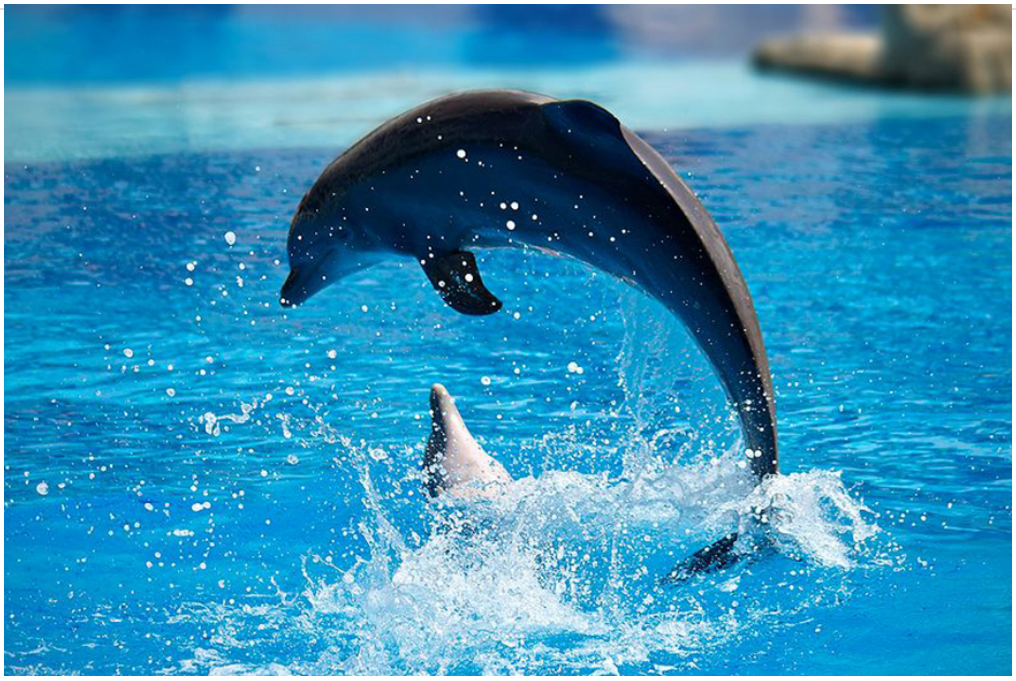
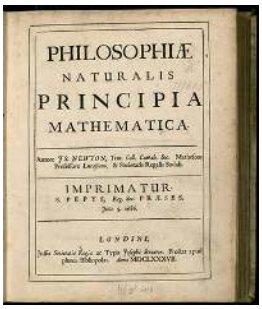


Figure **3.1.1**: Newton's laws of motion describe the motion of the dolphin's path. (credit: Jin Jang)

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

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


PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA

IMPRIMATUR.

LONDINI

Figure **3.1.2**: Issac Newton's monumental work, *Philosophiae Naturalis Principia Mathematica*, was published in 1687. It proposed scientific laws that are still used today to describe the motion of objects. (credit: Service commun de la documentation de l'Université de Strasbourg)

Galileo was instrumental in establishing *observation* as the absolute determinant of truth, rather than "logical" argument. Galileo's use of the telescope was his most notable achievement in demonstrating the importance of observation. He discovered moons orbiting Jupiter and made other observations that were inconsistent with certain ancient ideas and religious dogma. For this reason, and because of the manner in which he dealt with those in authority, Galileo was tried by the Inquisition and punished. He spent the

final years of his life under a form of house arrest. Because others before Galileo had also made discoveries by *observing* the nature of the universe, and because repeated observations verified those of Galileo, his work could not be suppressed or denied. After his death, his work was verified by others, and his ideas were eventually accepted by the church and scientific communities.

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



Figure **3.1.1**: Each rugby player has great momentum, which will affect the outcome of their collisions with each other and the ground. (credit: ozzzie, Flickr)

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

Momentum is important because it is conserved. Only a few physical quantities are conserved in nature, and studying them yields fundamental insight into how nature works, as we shall see in our study of momentum. The idea of momentum was around prior to the work of Newton, but could not be completely formalized until Newton defined mass. Competing ideas of movement still abounded at this time and, as we shall see in a later chapter, sometimes they were both correct.



MAKING CONNECTIONS: PAST AND PRESENT PHILOSOPHY

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

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 3.1: Introduction to Dynamics and Momentum. is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.1: Introduction to Dynamics- Newton's Laws of Motion** by OpenStax is licensed CC BY 4.0.
- **4.1: Introduction to Linear Momentum and Collisions** by OpenStax is licensed CC BY 4.0.

3.2: 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

As was introduced in the previous chapter, 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. 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. People have been aware of "amounts of stuff", and weights and were able to measure weight since before recorded history. Scientists had even begun to consider the idea of weight in theories of motion prior to Newton's time. But Newton was the first to consider mass as something separate from (but proportional to) weight. This difference will be explored later in the chapter.

We now know that the quantity or amount of matter in an object is determined by the numbers of atoms and molecules of various types it contains. During the time of Newton, the existence of atoms was not known, so mass was simply thought of as the quantity of matter in an object. 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 **3.2.1**

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.

Section 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.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

This page titled 3.2: Newton's First Law of Motion- Inertia is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.3: Newton's First Law of Motion- Inertia** by OpenStax is licensed CC BY 4.0.

3.3: Newton's Second Law of Motion- Force and Acceleration

Learning Objectives

- 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 **3.3.1**(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 **3.3.1**(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.

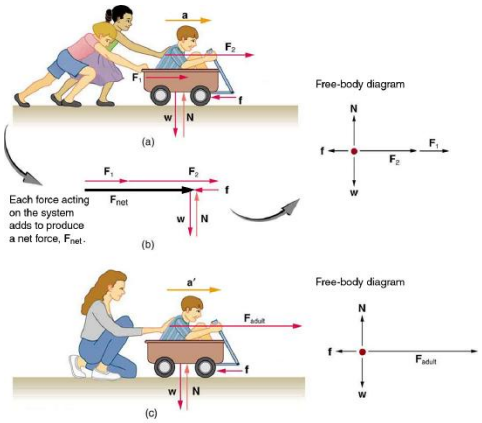


Figure **3.3.1**: Different forces exerted on the same mass produce different accelerations. (a) Two children push a wagon with a child in it. Arrows representing all external forces are shown. The system of interest is the wagon and its rider. The weight $\boldsymbol{w}$ of the system and the support of the ground $\boldsymbol{N}$ are also shown for completeness and are assumed to cancel. The vector $\boldsymbol{f}$ 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, $\boldsymbol{F_{\text{net}}}$. 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 ($|\boldsymbol{a'}| > |\boldsymbol{a}|$) when an adult pushes the child.

Figure **3.3.1** 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. 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.

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 **3.3.1**. 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 say that friction is a force that opposes the motion past each other of objects that are touching. Figure **3.3.1**(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

$$a \propto F_{\text{net}}$$

where the symbol $\propto$ means "proportional to," and F_{net} is the **net external force**, the vector sum of all external forces. 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, so that the internal forces can be ignored (again, as we will see later, they exactly cancel out, allowing for a tremendous simplification).

Now, hopefully it also seems reasonable that acceleration should be inversely proportional to the mass of the system. That is, if the system has more "stuff", then for the same net external force applied, the acceleration is less. And indeed, as illustrated in Figure **3.3.2**, 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

$$a \propto \frac{1}{m}$$

where m is the mass of the system. In fact, this is how we will *define* mass. That is, for an object B to have double the mass of an object A, the same net external force applied on object B produces half the acceleration it does on object A.

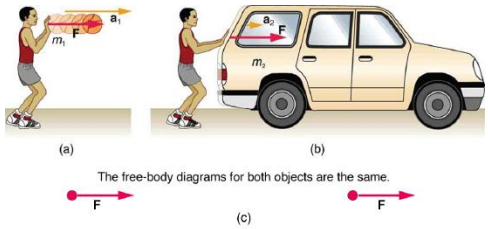


Figure **3.3.2**: The same force exerted on systems of different masses produces different accelerations. (a) A basketball player pushes on a basketball to make a pass. (The effect of gravity on the ball is ignored.) (b) The same player exerts an identical force on a stalled SUV and produces a far smaller 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.

As it turns out, 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

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

This is often written in the more familiar form

$$F_{\text{net}} = ma,$$

with the vector notation indicating that the net external force is in the same direction as acceleration. When only the magnitude of force and acceleration are considered, this equation is simply (note the lack of vector notations)

$$F_{\text{net}} = ma.$$

Although these 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, acceleration is *caused* by a net external force, not vice versa, as the second equation might mistakenly imply.

Units of Force

$F_{\text{net}} = ma$ is used to define the units of force in terms of the three basic units for mass, length, and time. This means force is a devired unit, similar to velocity and volume, as discussed in the previous chapter. The SI unit of force is called the **newton** (abbreviated N) and 1 N is the force needed to accelerate a system of mass 1 kg at the rate of $1\ \text{m/s}^2$. Putting these together,

$$1\ \text{N} = 1\ \text{kg} \cdot \text{m/s}^2.$$

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

As we learned in the previous chapter, gravity is a very common form of acceleration we experience in the world. 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, so w is directed downward. 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 g. Using Galileo's result and Newton's second law, we can derive an equation for the magnitude of 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 $F_{\text{net}} = ma$.

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

Definition: WEIGHT

This is the equation for *weight*—the gravitational force on a mass m:

$$w = mg.$$

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

$$w = mg = (1.0\ \text{kg})\left(9.80\ \text{m/s}^2\right) = 9.8\ \text{N}.$$

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. On the Moon, for example, the acceleration due to gravity is only $1.67\ \text{m/s}^2$. 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, 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). This interchangeable use is why you might have heard an expression such as, "1 kilogram is 2.2 pounds." The *correct* statement is "1 kilogram mass weighs 2.2 pounds on Earth."

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 $1.67\ \mathrm{m/s^2}$ (which is much less than the acceleration due to gravity on Earth, $9.80\ \mathrm{m/s^2}$). 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 smaller. 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 in newtons 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 3.3.1: 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 **3.3.3**: 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 $F_{\mathrm{net}} = ma$.

Solution

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

$$a = \frac{51\ \mathrm{N}}{24\ \mathrm{kg}}$$

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

$$a = \frac{51\ \mathrm{kg \cdot m/s^2}}{24\ \mathrm{kg}} = 2.1\ \mathrm{m/s^2}.$$

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 forward must be greater than the friction opposing the motion (directed backward) to result in a forward net force, and the vertical forces must cancel if there is to be no acceleration in the vertical direction. The acceleration found is small enough to be reasonable for a person pushing a mower. When the person reaches the top speed, the acceleration will be zero (no change of velocity) and the force exerted by the person pushing the mower will be equal to the friction opposing the motion.

Example 3.3.2: 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 3.3.4. The sled's initial acceleration is $49\ \mathrm{m/s^2}$ the mass of the system is 2100 kg, and the force of friction opposing the motion is known to be 650 N.

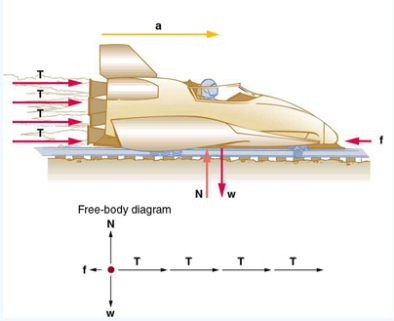


Figure 3.3.4: A sled experiences a rocket thrust that accelerates it to the right. Each rocket creates an identical thrust T. As in other situations where there is only horizontal acceleration, the vertical forces cancel. The ground exerts an upward force N on the system which balances out the weight force, w. The system here is the sled, its rockets, and rider, so none of the internal forces *between* these objects are considered. f represents the friction force (figure not drawn to 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

$$F_{\text{net}} = ma,$$

where F_{net} is the net force along the horizontal direction. We can see from Figure 3.3.4 that the engine thrusts add together, while friction opposes the thrust. In equation form, the net external force is

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

Substituting this into Newton's second law gives

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

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

$$4T = ma + f.$$

Substituting known values yields

$$4T = ma + f = (2100\ \text{kg})\left(49\ \text{m/s}^2\right) + 650\ \text{N}.$$

So the total thrust is

$$4T = 1.04 \times 10^5\ \text{N},$$

and the individual thrusts are

$$T = \frac{1.04 \times 10^5\ \text{N}}{4} = 2.6 \times 10^4\ \text{N}.$$

Discussion

The numbers are quite large, so to put the result in perspective, note that acceleration of $49\ \text{m/s}^2$ is approximately 5 times the gravitational acceleration on Earth. It is a large acceleration, which requires a large net external force to produce.

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 Missing superscript or subscript argument's. (Recall that g, the acceleration due to gravity, is $9.80\ \text{m/s}^2$. When we say that an acceleration is Missing superscript or subscript argument's, it is $45 \times 9.80\ \text{m/s}^2$, which is approximately $440\ \text{m/s}^2$.) 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 gives 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.

Section Summary

- Acceleration, a, is defined as a rate of change in velocity, resulting from a change in the magnitude and/or the direction of the velocity.
- 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 $a = \frac{F_{\text{net}}}{m}$, often written in the more familiar form: $F_{\text{net}} = ma$.
- The weight w of an object is defined as the force of gravity acting on an object of mass m. Given acceleration due to gravity g, the magnitude of weight is:

$$w = mg$$

Glossary

acceleration
the rate at which an object's velocity changes over a period of time

free-fall
a situation in which the only force acting on an object is the force due to gravity

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

net external force
the vector sum of all external forces acting on an object or system; causes a mass to accelerate

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

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.

system

the object or the group of objects under consideration

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

weight

the force due to gravity; $w = mg$ for objects on Earth

This page titled 3.3: Newton's Second Law of Motion- Force and Acceleration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.4: Newton's Second Law of Motion- Force and Acceleration** by OpenStax is licensed CC BY 4.0.

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

Learning Objectives

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

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 by the second body that is equal in magnitude and opposite in direction to the force that it itself 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 **3.4.1**. 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 bodies*. In this case, there are two bodies to consider: the swimmer and the wall. If we define the swimmer as the system of interest, as in the figure, then $\boldsymbol{F}_{\text{wall on feet}}$ is an external force on this system and affects its motion. The swimmer moves in the direction of $\boldsymbol{F}_{\text{wall on feet}}$. In contrast, the force $\boldsymbol{F}_{\text{feet on wall}}$ acts on the wall and not on our system of interest. Thus $\boldsymbol{F}_{\text{feet on wall}}$ does not affect the motion of the system and does not counteract $\boldsymbol{F}_{\text{wall on feet}}$. Note that the swimmer pushes the wall in the direction opposite to that in which she wishes to move. The reaction force by the wall on her pushes her in the desired direction.

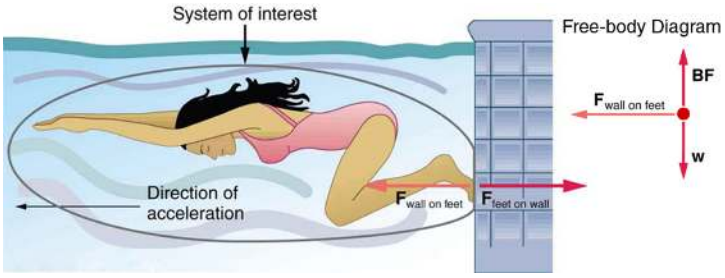


Figure **3.4.1**: When the swimmer exerts a force $\boldsymbol{F}_{\text{feet on wall}}$ on the wall, she accelerates in the direction opposite to that of her push. This means the net external force on her is in the direction opposite to $\boldsymbol{F}_{\text{feet on wall}}$. This opposition occurs because, in accordance with Newton’s third law of motion, the wall exerts a force $\boldsymbol{F}_{\text{wall on feet}}$ 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 $\boldsymbol{F}_{\text{feet on wall}}$ does not act on this system (the swimmer) and, thus, does not cancel $\boldsymbol{F}_{\text{wall on feet}}$. Thus the free-body diagram shows only $\boldsymbol{F}_{\text{wall on feet}}$, $\boldsymbol{w}$, the gravitational force, and **BF**, the buoyant force of the water supporting the swimmer’s weight. The vertical forces $\boldsymbol{w}$ and **BF** cancel since there is no vertical motion.

Other examples of Newton’s third law are easy to find. As a professor paces in front of a whiteboard, they exerts a force backward on the floor. The floor exerts a reaction force forward on the professor that causes them 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. The rockets propel themselves by pushing their own exhaust gases, and that is how they work even in a vacuum. Helicopters

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. Professional cage fighters experience reaction forces when they punch, sometimes breaking their hand by hitting an opponent's body.

Example 3.4.1: 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 3.4.2. Their 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.

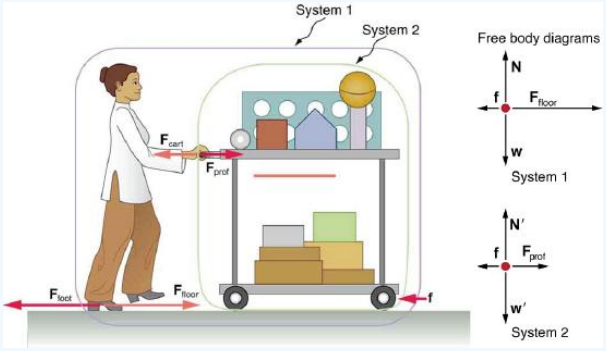


Figure 3.4.2: A professor pushes a cart of demonstration equipment. The lengths of the arrows are proportional to the magnitudes of the forces (except for f, since it is too small to draw to scale). Different questions are asked in each example; thus, the system of interest must be defined differently for each. System 1 is appropriate for this example, since it asks for the acceleration of the entire group of bodies. Only $\boldsymbol{F}_{\text{floor}}$ and $\boldsymbol{f}$ 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 appropriate for Example 3.4.2 so that $\boldsymbol{F}_{\text{prof}}$ will be an external force and enter into Newton's second law.

Strategy

Since they accelerate as a unit, we define the system to be the professor, cart, and equipment. This is System 1 in Figure 3.4.2. The professor pushes backward with a force $\boldsymbol{F}_{\text{foot}}$ of 150 N. According to Newton's third law, the floor exerts a forward reaction force $\boldsymbol{F}_{\text{floor}}$ of 150 N on System 1. Because all motion is horizontal, we can assume there is no net force in the vertical direction, and treat this as a one-dimensional problem in the horizontal direction. As noted, $\boldsymbol{f}$ opposes the motion and is thus in the opposite direction of $\boldsymbol{F}_{\text{floor}}$. Note that we do not include the forces $\boldsymbol{F}_{\text{prof}}$ or $\boldsymbol{F}_{\text{cart}}$ because these are internal forces, and we do not include $\boldsymbol{F}_{\text{foot}}$ 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

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

The net external force on System 1 is deduced from Figure 3.4.2 and the discussion above to be

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

The mass of System 1 is

$$m = (65.0 + 12.0 + 7.0)\text{kg} = 84\ \text{kg}.$$

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

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

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 them. 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 **3.4.2**: Force on the Cart—Choosing a New System

Calculate the force the professor exerts on the cart in Figure **3.4.2**, 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 **3.4.2**), then the net external force on System 2 is the force the professor exerts on the cart minus friction. The force they exerts on the cart, $\mathbf{F}_{\text{prof}}$, is an external force acting on System 2. $\mathbf{F}_{\text{prof}}$ 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 $\mathbf{F}_{\text{prof}}$. Starting with

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

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

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

we solve for $\mathbf{F}_{\text{prof}}$, the desired quantity:

$$F_{\text{prof}} = F_{\text{net}} + f.$$

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

$$F_{\text{net}} = ma,$$

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

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

Now we can find the desired force:

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

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).

COMMON MISCONCEPTION: REACTION VS. RESULT

The phrases "action" and "reaction" in reference to the pair of forces described by Newton's third law often leads to unfortunate misunderstandings regarding the nature of the reaction force. Because of the common English usage of the word "reaction" (root word: react)—for example, in film, "reaction shot" means a shot of actors' faces as they react to an event that just happened—many students intuitively guess that there is a cause-and-effect relationship between "action force" and "reaction force."

There is no such relationship. Newton's third law describes *symmetry* in forces. Action force and reaction force are symmetric—they occur simultaneously; they are two sides of one and the same interaction. Action force does not cause reaction force, no more than the head of a coin causes the tail of the coin (they merely occur together, always). Here is a quick check to see if you understood the pairing of action force and reaction force correctly: try swapping the labels. That is, if you originally called A "action force" and B "reaction force", try calling B "action force" and A "reaction force." Does this re-labeling make some sense? If not, chances are, you have mis-identified the pair of action force and reaction force. (Here, a little weirdness is O.K., but if it feels like complete nonsense to swap the labels—like reversal of a cause-and-effect relationship—that tells you you used "reaction" in the common English usage, not how the word is used in Newton's third law.)

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 by the second body that is equal in magnitude and opposite in direction to the force that it itself exerts.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

thrust

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

---

This page titled 3.4: Newton's Third Law of Motion- Symmetry in Forces is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.5: Newton's Third Law of Motion- Symmetry in Forces** by OpenStax is licensed CC BY 4.0.

3.5: Examples of Force

Learning Objectives

- Recognize different types of forces.
- Apply Newton's laws of motion to analyze forces.

Forces are given many names, such as push, pull, thrust, lift, weight, friction, and tension. Traditionally, forces have been grouped into several categories and given names relating to their source, how they are transmitted, or their effects. The most important of these categories are discussed in this section, together with some interesting applications in each subsection. The goal as you read through these subsections is to understand the differences between the types of forces which are being described. There are also example problems in many of these subsections. It is best to view these example problems as opportunities to further understand the concepts which had been previously developed in this chapter as opposed to a new set of rules to determine how force works in each case. In each case we are just looking at applications of the three laws of motion which we have already developed at the start of this chapter.

This page titled 3.5: Examples of Force is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.6: Normal Force and Tension** by OpenStax is licensed CC BY 4.0.

3.5.1: Normal Force

Learning Objectives

- Define normal force.
- Explain the relationship between normal force and apparent weight.

Weight (also called force of gravity) is a pervasive force that acts at all times and must be counteracted to keep an object from falling. You definitely notice that you must support the weight of a heavy object by pushing up on it when you hold it stationary, as illustrated in Figure **3.5.1.1**(a). But how do inanimate objects like a table support the weight of a mass placed on them, such as shown in Figure **3.5.1.1**(b)? When the bag of dog food is placed on the table, the table actually sags slightly under the load. This would be noticeable if the load were placed on a card table, but even rigid objects deform when a force is applied to them. Unless the object is deformed beyond its limit, it will exert a restoring force much like a deformed spring (or trampoline or diving board). The greater the deformation, the greater the restoring force. So when the load is placed on the table, the table sags until the restoring force becomes as large as the weight of the load. At this point the net external force on the load is zero. That is the situation when the load is stationary on the table. The table sags quickly, and the sag is slight so we do not notice it. But it is similar to the sagging of a trampoline when you climb onto it.

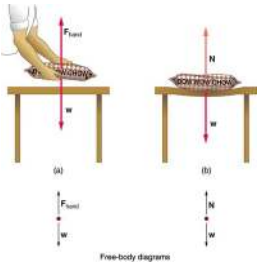


Figure **3.5.1.1**: (a) The person holding the bag of dog food must supply an upward force $\boldsymbol{F}_{\text{hand}}$ equal in magnitude and opposite in direction to the weight of the food $\boldsymbol{w}$. (b) The card table sags when the dog food is placed on it, much like a stiff trampoline. Elastic restoring forces in the table grow as it sags until they supply a force **N** equal in magnitude and opposite in direction to the weight of the load.

We must conclude that whatever supports a load, be it animate or not, must supply an upward force equal to the weight of the load, as we assumed in a few of the previous examples. The force supporting a load is perpendicular to the surface of contact between the load and its support, and this force is called a **normal force**, often indicated with symbol **N** (Please do not confuse this with the abbreviation for newton, the unit of force, N). The word *normal* means perpendicular to a surface. The normal force is not always equal to the object's weight, if there are other forces acting on the object, or if the object is accelerating, so that the net force is not zero.

Normal force also goes by another name, **apparent weight**. This is because all the forces you feel, especially your sensation of weight, are directly attributed to contact forces, such as the normal force. When NASA astronauts are weightlessly floating in the International Space Station, their actual weight (gravitational force on them by the Earth) is not zero, but their apparent weight (sensation of weight due to contact forces) is zero.

Section Summary

- When objects rest on a surface, the surface applies a force to the object that supports the weight of the object. This supporting force acts perpendicular to and away from the surface. It is called a normal force, and it also goes by the name "apparent weight."

Glossary

normal force

the force that a surface applies to an object to support the weight of the object; acts perpendicular to the surface on which the object rests

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

apparent weight

sensation of weight due to contact forces; usually the same as normal force

---

This page titled 3.5.1: Normal Force is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.6: Normal Force and Tension** by OpenStax is licensed CC BY 4.0.

3.5.2: Tension

Learning Objectives

- Describe properties of tension force.

A **tension** is a force along the length of a medium, especially a force carried by a flexible medium, such as a rope or cable. The word “tension” comes from a Latin word meaning “to stretch.” Not coincidentally, the flexible cords that carry muscle forces to other parts of the body are called *tendons*. Any flexible connector, such as a string, rope, chain, wire, or cable, can exert pulls only parallel to its length; thus, a force carried by a flexible connector is a tension with direction parallel to the connector. It is important to understand that tension is a pull in a connector. In contrast, consider the phrase: “You can’t push a rope.” The tension force pulls outward along the two ends of a rope.

Consider a person holding a mass on a rope as shown in Figure **3.5.2.2**.

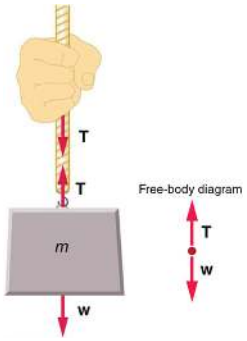


Figure **3.5.2.2**: When a perfectly flexible connector (one requiring no force to bend it) such as this rope transmits a force $\boldsymbol{T}$, that force must be parallel to the length of the rope, as shown. The pull such a flexible connector exerts is a tension. Note that the rope pulls with equal force but in opposite directions on the hand and the supported mass (neglecting the weight of the rope). This is an example of Newton’s third law. The rope is the medium that carries the equal and opposite forces between the two objects. The tension anywhere in the rope between the hand and the mass is equal. Once you have determined the tension in one location, you have determined the tension at all locations along the rope.

Tension in the rope must equal the weight of the supported mass, as we can prove using Newton’s second law. If the 5.00-kg mass in the figure is stationary, then its acceleration is zero, and thus $\boldsymbol{F_{\text{net}} = 0}$. The only external forces acting on the mass are its weight $\boldsymbol{w}$ and the tension $\boldsymbol{T}$ supplied by the rope. Thus,

$$\boldsymbol{F_{\text{net}} = T - w = 0,}$$

where $\boldsymbol{T}$ and $\boldsymbol{w}$ are the magnitudes of the tension and weight and their signs indicate direction, with up being positive here. Thus, just as you would expect, the tension equals the weight of the supported mass:

$$\boldsymbol{T = w = mg.}$$

For a 5.00-kg mass, then (neglecting the mass of the rope) we see that

$$\boldsymbol{T = mg = (5.00\ \text{kg})\left(9.80\ \text{m/s}^2\right) = 49.0\ \text{N}.}$$

If we cut the rope and insert a spring, the spring would extend a length corresponding to a force of 49.0 N, providing a direct observation and measure of the tension force in the rope.

Flexible connectors are often used to transmit forces around corners, such as in a hospital traction system, a finger joint, or a bicycle brake cable. If there is no friction, the tension is transmitted undiminished. Only its direction changes, and it is always parallel to the flexible connector. This is illustrated in Figure **3.5.2.3** (a) and (b).

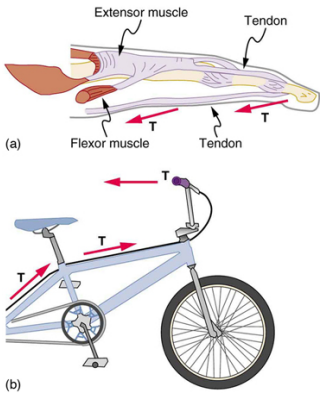


Figure **3.5.2.3**: (a) Tendons in the finger carry force ***T*** from the muscles to other parts of the finger, usually changing the force's direction, but not its magnitude (the tendons are relatively friction free). (b) The brake cable on a bicycle carries the tension ***T*** from the handlebars to the brake mechanism. Again, the direction but not the magnitude of ***T*** is changed.



Figure **3.5.2.4**: Unless an infinite tension is exerted, any flexible connector—such as the chain at the bottom of the picture—will sag under its own weight, giving a characteristic curve when the weight is evenly distributed along the length, known as a catenary curve. Suspension bridges—such as the Golden Gate Bridge shown in this image—are essentially very heavy flexible connectors. The weight of the bridge is evenly distributed along the length of flexible connectors, usually cables, which take on the characteristic shape. (credit: Leaflet, Wikimedia Commons)

Normal force and tension are examples of forces that are determined not by a specific formula but by enforcing a constraint on motion state of a body. In the example of the bag on a table above, the normal force is equal to weight, because that is the value needed to ensure that the acceleration of the bag is zero. What happens if the body experiences a non-zero acceleration? We will consider example below of a person standing on a scale (which measures his apparent weight) while riding in an elevator.

Example **3.5.2.1**: What Does the Bathroom Scale Read in an Elevator?

Figure **3.5.2.5** shows a 75.0-kg man (weight of about 165 lb.) standing on a bathroom scale in an elevator. Calculate the scale reading: (a) if the elevator accelerates upward at a rate of $\mathbf{1.20\ m/s^2}$, and (b) if the elevator moves upward at a constant speed of 1 m/s.

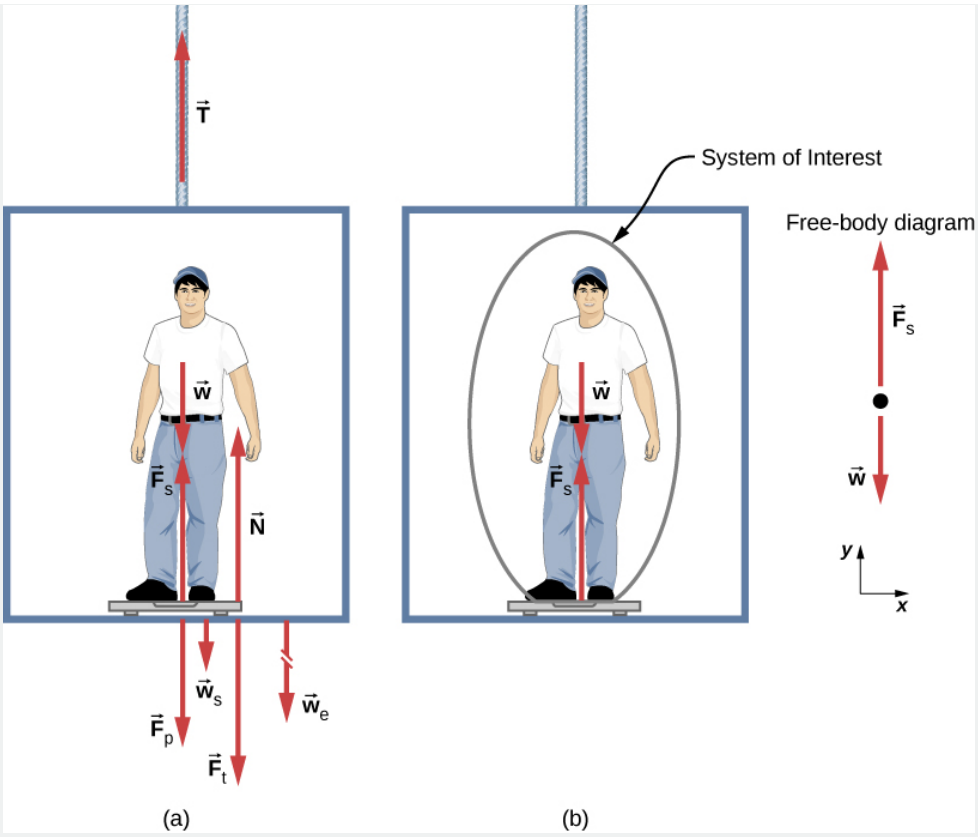


Figure 3.5.2.5: (a) The various forces acting when a person stands on a bathroom scale in an elevator. The arrows are approximately correct for when the elevator is accelerating upward—broken arrows represent forces too large to be drawn to scale. $\vec{\mathbf{T}}$ is the tension in the supporting cable, $\vec{\mathbf{w}}$ is the weight of the person, $\vec{\mathbf{w}}_s$ is the weight of the scale, $\vec{\mathbf{w}}_e$ is the weight of the elevator, $\vec{\mathbf{F}}_s$ is the force of the scale on the person, $\vec{\mathbf{F}}_p$ is the force of the person on the scale, $\vec{\mathbf{F}}_t$ is the force of the scale on the floor of the elevator, and $\vec{\mathbf{N}}$ is the force of the floor upward on the scale. (b) The free-body diagram shows only the external forces acting *on* the designated system of interest—the person—and is the diagram we use for the solution of the problem.

Strategy

If the scale at rest is accurate, its reading equals $\vec{\mathbf{F}}_p$, the magnitude of the force the person exerts downward on it. Figure 3.5.2.5(a) shows the numerous forces acting on the elevator, scale, and person. It makes this one-dimensional problem look much more formidable than if the person is chosen to be the system of interest and a free-body diagram is drawn, as in Figure 3.5.2.5(b). Analysis of the free-body diagram using Newton's laws can produce answers to both Figure 3.5.2.5(a) and (b) of this example, as well as some other questions that might arise. The only forces acting on the person are his weight $\vec{\mathbf{w}}$ and the upward force of the scale $\vec{\mathbf{F}}_s$. According to Newton's third law, $\vec{\mathbf{F}}_p$ and $\vec{\mathbf{F}}_s$ are equal in magnitude and opposite in direction, so that we need to find F_s in order to find what the scale reads. We can do this, as usual, by applying Newton's second law,

$$\vec{\mathbf{F}}_{\text{net}} = m\vec{\mathbf{a}}.$$

From the free-body diagram, we see that the net force $\vec{\mathbf{F}}_{\text{net}}$ is sum of the scale force $\vec{\mathbf{F}}_s$ and the weight $\vec{\mathbf{w}}$. Since the scale force and the weight are pointing in opposite directions, in terms of the magnitudes of the vectors,

$$F_{\text{net}} = F_s - w = ma.$$

Solving for F_s gives us an equation with only one unknown:

$$F_s = ma + w,$$

or, because $w = mg$, simply

$$F_s = ma + mg.$$

No assumptions were made about the acceleration, so this solution should be valid for a variety of accelerations in addition to those in this situation. (*Note:* We are considering the case when the elevator is accelerating upward. If the elevator is accelerating downward, Newton's second law becomes $F_s - w = -ma$.)

Solution

a. We have $a = 1.20\ \text{m/s}^2$, so that

$$F_s = (75.0\ \text{kg})\,(9.80\ \text{m/s}^2) + (75.0\ \text{kg})\,(1.20\ \text{m/s}^2)$$

yielding

$$F_s = 825\ \text{N}.$$

b. Now, what happens when the elevator reaches a constant upward velocity? Will the scale still read more than his weight? For any constant velocity—up, down, or stationary—acceleration is zero because $a = \frac{\Delta v}{\Delta t}$ and $\Delta v = 0$. Thus,

$$F_s = ma + mg = 0 + mg$$

or

$$F_s = (75.0\ \text{kg})\,(9.80\ \text{m/s}^2),$$

which gives

$$F_s = 735\ \text{N}.$$

Significance

The scale reading in Figure 3.5.2.5(a) is about 185 lb. What would the scale have read if he were stationary? Since his acceleration would be zero, the force of the scale would be equal to his weight:

$$\begin{aligned} F_{\text{net}} &= ma = 0 = F_s - w \\ F_s &= w = mg \\ F_s &= (75.0\ \text{kg})\,(9.80\ \text{m/s}^2) = 735\ \text{N}. \end{aligned}$$

Thus, the scale reading in the elevator is greater than his 735-N (165-lb.) weight. This means that the scale is pushing up on the person with a force greater than his weight, as it must in order to accelerate him upward. Clearly, the greater the acceleration of the elevator, the greater the scale reading, consistent with what you feel in rapidly accelerating versus slowly accelerating elevators. In Figure 3.5.2.5(b), the scale reading is 735 N, which equals the person's weight. This is the case whenever the elevator has a constant velocity—moving up, moving down, or stationary.

Section Summary

- The pulling force that acts along a stretched flexible connector, such as a rope or cable, is called tension, T. When a rope supports the weight of an object that is at rest, the tension in the rope is equal to the weight of the object.

Glossary

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

tension

the pulling force that acts along a medium, especially a stretched flexible connector, such as a rope or cable; when a rope supports the weight of an object, the force on the object due to the rope is called a tension force

This page titled 3.5.2: Tension is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.6: Normal Force and Tension** by OpenStax is licensed CC BY 4.0.

3.5.3: Spring Force- Hooke's Law

Learning Objectives

- Describe the restoring force and displacement.
- Explain oscillatory motion under a spring force.

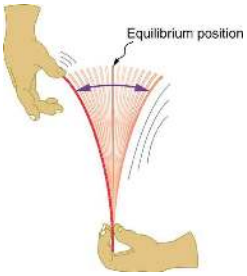


Figure **3.5.3.1**: When displaced from its vertical equilibrium position, this plastic ruler oscillates back and forth because of the restoring force opposing displacement. When the ruler is on the left, there is a force to the right, and vice versa.

Newton's first law implies that an object oscillating back and forth is experiencing forces. Without force, the object would move in a straight line at a constant speed rather than oscillate. Consider, for example, plucking a plastic ruler to the left as shown in Figure **3.5.3.1**. The deformation of the ruler creates a force in the opposite direction, known as a **restoring force**. Once released, the restoring force causes the ruler to move back toward its stable equilibrium position, where the net force on it is zero. However, by the time the ruler gets there, it gains momentum and continues to move to the right, producing the opposite deformation. It is then forced to the left, back through equilibrium, and the process is repeated until dissipative forces dampen the motion. These forces remove mechanical energy from the system, gradually reducing the motion until the ruler comes to rest.

The simplest oscillations occur when the restoring force is directly proportional to displacement. This is called Hooke's law force, or spring force:

$$F = -kx.$$

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

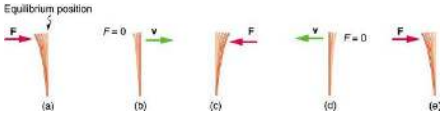


Figure **3.5.3.2**: (a) The plastic ruler has been released, and the restoring force is returning the ruler to its equilibrium position. (b) The net force is zero at the equilibrium position, but the ruler has momentum and continues to move to the right. (c) The restoring force is in the opposite direction. It stops the ruler and moves it back toward equilibrium again. (d) Now the ruler has momentum to the left. (e) In the absence of damping (caused by frictional forces), the ruler reaches its original position. From there, the motion will repeat itself.

The **spring constant** k is related to the rigidity (or stiffness) of a system—the larger the spring constant, the greater the restoring force, and the stiffer the system. The units of k are newtons per meter (N/m). Figure **3.5.3.3** shows a graph of the absolute value of the restoring force versus the displacement for a system that can be described by Hooke's law—a simple spring in this case. The slope of the graph equals the spring constant k in newtons per meter.

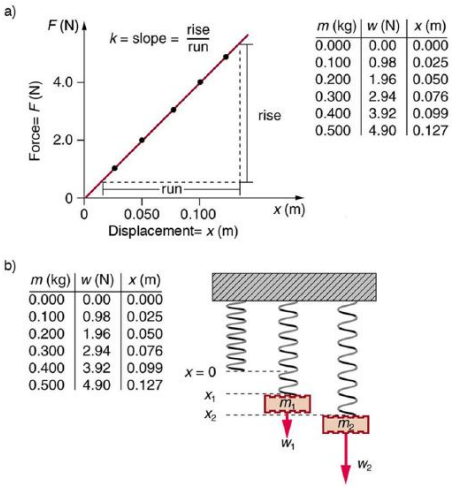


Figure **3.5.3.3**: (a) A graph of absolute value of the restoring force versus displacement is displayed. The fact that the graph is a straight line means that the system obeys Hooke's law. The slope of the graph is the spring constant k. (b) The data in the graph were generated by measuring the displacement of a spring from equilibrium while supporting various weights. The restoring force equals the weight supported, if the mass is stationary.

Example **3.5.3.1**: How Stiff Are Car Springs?



Figure **3.5.3.4**: The mass of a car increases due to the introduction of a passenger. This affects the displacement of the car on its suspension system. (credit: exfordy on Flickr)

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

Strategy

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

Solution

1. $$k = -\frac{F}{x}.$$

   Substitute known values and solve k:

$$k = -\frac{784\ \text{N}}{-1.20 \times 10^{-2}\ \text{m}}$$
$$= 6.53 \times 10^{4}\ \text{N/m}.$$

Discussion

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

Section Summary

- An oscillation is a back and forth motion of an object between two points of deformation.
- An oscillation may create a wave, which is a disturbance that propagates from where it was created.
- The simplest type of oscillations and waves are related to systems that can be described by Hooke's law:

$$F = -kx,$$

where F is the restoring force, x is the displacement from equilibrium or deformation, and k is the spring constant of the system.

Glossary

deformation

displacement from equilibrium

spring constant

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

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

restoring force

force acting in opposition to the force caused by a deformation

---

This page titled 3.5.3: Spring Force- Hooke's Law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.7: Spring Force- Hooke's Law** by OpenStax is licensed CC BY 4.0.

3.5.4: Friction

Learning Objectives

- Discuss the general characteristics of friction.
- Describe the various types of friction.
- Explain effect of friction on motion.

Friction is a force that is around us all the time that opposes relative motion between systems in contact but also allows us to move (which you have discovered if you have ever tried to walk on ice). While a common force, the behavior of friction is actually very complicated and is still not completely understood. We have to rely heavily on observations for whatever understandings we can gain. However, we can still deal with its more elementary general characteristics and understand the circumstances in which it behaves.

Definition: FRICTION

Friction is a force that opposes relative motion between systems in contact.

One of the simpler characteristics of friction is that it is parallel to the contact surface between systems and always in a direction that opposes motion or attempted motion of the systems relative to each other. If two systems are in contact and moving relative to one another, then the friction between them is called **kinetic friction**. For example, friction slows a hockey puck sliding on ice. But when objects are stationary, **static friction** can act between them; the static friction is usually greater than the kinetic friction between the objects.

Definition: KINETIC FRICTION

If two systems are in contact and moving relative to one another, then the friction between them is called kinetic friction.

Imagine, for example, trying to slide a heavy crate across a concrete floor—you may push harder and harder on the crate and not move it at all. This means that the static friction responds to what you do—it increases to be equal to and in the opposite direction of your push. But if you finally push hard enough, the crate seems to slip suddenly and starts to move. Once in motion it is easier to keep it in motion than it was to get it started, indicating that the kinetic friction force is less than the static friction force. If you add mass to the crate, say by placing a box on top of it, you need to push even harder to get it started and also to keep it moving. Furthermore, if you oiled the concrete you would find it to be easier to get the crate started and keep it going (as you might expect).

Figure **3.5.4.1** is a crude pictorial representation of how friction occurs at the interface between two objects. Close-up inspection of these surfaces shows them to be rough. So when you push to get an object moving (in this case, a crate), you must raise the object until it can skip along with just the tips of the surface hitting, break off the points, or do both. A considerable force can be resisted by friction with no apparent motion. The harder the surfaces are pushed together (such as if another box is placed on the crate), the more force is needed to move them. Part of the friction is due to adhesive forces between the surface molecules of the two objects, which explain the dependence of friction on the nature of the substances. Adhesion varies with substances in contact and is a complicated aspect of surface physics. Once an object is moving, there are fewer points of contact (fewer molecules adhering), so less force is required to keep the object moving. At small but nonzero speeds, friction is nearly independent of speed.

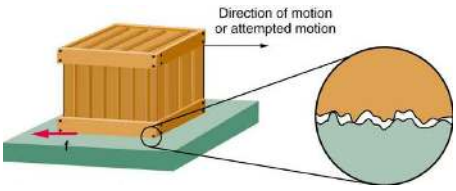


Figure **3.5.4.1**: Frictional forces, such as f, always oppose motion or attempted motion between objects in contact. Friction arises in part because of the roughness of the surfaces in contact, as seen in the expanded view. In order for the object to move, it must rise to where the peaks can skip along the bottom surface. Thus, a force is required just to set the object in motion. Some of the peaks will be broken off, also requiring a force to maintain motion. Much of the friction is actually due to attractive forces between molecules making up the two objects, so that even perfectly smooth surfaces are not friction-free. Such adhesive forces also depend on the substances the surfaces are made of, explaining, for example, why rubber-soled shoes slip less than those with leather soles.

The magnitude of the frictional force has two forms: one for static situations (static friction), the other for when there is motion (kinetic friction).

When there is no motion between the objects, the **magnitude of static friction f_s** is

$$f_s \leq \mu_s N,$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force (the force perpendicular to the surface).

MAGNITUDE OF STATIC FRICTION

Magnitude of static friction f_s is

$$f_s \leq \mu_s N,$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force.

The symbol ≤ means *less than or equal to*, implying that static friction can have a minimum and a maximum value of $\mu_s N$. Static friction is a responsive force that increases to be equal and opposite to whatever force is exerted, up to its maximum limit. Once the applied force exceeds $f_{s(max)}$, the object will move. Thus

$$f_{s(max)} = \mu_s N.$$

Once an object is moving, the **magnitude of kinetic friction** f_k is given by

$$f_k = \mu_k N,$$

where μ_k is the coefficient of kinetic friction. A system in which $f_k = \mu_k N$ is described as a system in which *friction behaves simply*.

MAGNITUDE OF KINETIC FRICTION

The magnitude of kinetic friction f_k is given by

$$f_k = \mu_k N,$$

where μ_k is the coefficient of kinetic friction.

As seen in Table 3.5.4.1, the coefficients of kinetic friction are less than their static counterparts. That values of μ in Table 3.5.4.1 are stated to only one or, at most, two digits is an indication of the approximate description of friction given by the above two equations.

Table 3.5.4.1: Coefficients of Static and Kinetic Friction

System	Static friction μ_s	Kinetic friction μ_k
Rubber on dry concrete	1.0	0.7
Rubber on wet concrete	0.7	0.5
Wood on wood	0.5	0.3
Waxed wood on wet snow	0.14	0.1
Metal on wood	0.5	0.3
Steel on steel (dry)	0.6	0.3
Steel on steel (oiled)	0.05	0.03
Teflon on steel	0.04	0.04
Bone lubricated by synovial fluid	0.016	0.015
Shoes on wood	0.9	0.7
Shoes on ice	0.1	0.05
Ice on ice	0.1	0.03
Steel on ice	0.4	0.02

The equations given earlier include the dependence of friction on materials and the normal force. The direction of friction is always opposite that of motion, parallel to the surface between objects, and perpendicular to the normal force. For example, if the crate you try to push (with a force parallel to the floor) has a mass of 100 kg, then the normal force would be equal to its weight, $W = mg = (100\ \text{kg})\ (9.80\ \text{m/s}^2) = 980\ \text{N}$, perpendicular to the floor. If the coefficient of static friction is 0.45, you would have to exert a force parallel to the floor greater than $f_{s(\text{max})} = \mu_s N = (0.45)(980\ \text{N}) = 440\ \text{N}$ to move the crate. Once there is motion, friction is less and the coefficient of kinetic friction might be 0.30, so that a force of only 290 N ($f_k = \mu_k N = (0.30)(980\ \text{N}) = 290\ \text{N}$) would keep it moving at a constant speed. If the floor is lubricated, both coefficients are considerably less than they would be without lubrication. Coefficient of friction is a unitless quantity with a magnitude usually between 0 and 1.0. The coefficient of the friction depends on the two surfaces that are in contact.

TAKE-HOME EXPERIMENT

Find a small plastic object (such as a food container) and slide it on a kitchen table by giving it a gentle tap. Now spray water on the table, simulating a light shower of rain. What happens now when you give the object the same-sized tap? Now add a few drops of (vegetable or olive) oil on the surface of the water and give the same tap. What happens now? This latter situation is particularly important for drivers to note, especially after a light rain shower. Why?

Many people have experienced the slipperiness of walking on ice. However, many parts of the body, especially the joints, have much smaller coefficients of friction—often three or four times less than ice. A joint is formed by the ends of two bones, which are connected by thick tissues. The knee joint is formed by the lower leg bone (the tibia) and the thighbone (the femur). The hip is a ball (at the end of the femur) and socket (part of the pelvis) joint. The ends of the bones in the joint are covered by cartilage, which provides a smooth, almost glassy surface. The joints also produce a fluid (synovial fluid) that reduces friction and wear. A damaged or arthritic joint can be replaced by an artificial joint (Figure 3.5.4.2). These replacements can be made of metals (stainless steel or titanium) or plastic (polyethylene), also with very small coefficients of friction.



Figure **3.5.4.2**: Artificial knee replacement is a procedure that has been performed for more than 20 years. In this figure, we see the post-op X-rays of the right knee joint replacement. (credit: Mike Baird, Flickr)

Other natural lubricants include saliva produced in our mouths to aid in the swallowing process, and the slippery mucus found between organs in the body, allowing them to move freely past each other during heartbeats, during breathing, and when a person moves. Artificial lubricants are also common in hospitals and doctor's clinics. For example, when ultrasonic imaging is carried out, the gel that couples the transducer to the skin also serves to to lubricate the surface between the transducer and the skin—thereby reducing the coefficient of friction between the two surfaces. This allows the transducer to mover freely over the skin.

We have discussed that when an object rests on a horizontal surface, there is a normal force supporting it equal in magnitude to its weight. Furthermore, simple friction is always proportional to the normal force.

MAKING CONNECTIONS: SUBMICROSCOPIC EXPLANATIONS OF FRICTION

The simpler aspects of friction dealt with so far are its macroscopic (large-scale) characteristics. Great strides have been made in the atomic-scale explanation of friction during the past several decades. Researchers are finding that the atomic nature of friction seems to have several fundamental characteristics. These characteristics not only explain some of the simpler aspects of friction—they also hold the potential for the development of nearly friction-free environments that could save hundreds of billions of dollars in energy which is currently being lost in overcoming the forces of friction.

Figure **3.5.4.3** illustrates one macroscopic characteristic of friction that is explained by microscopic (small-scale) research. We have noted that friction is proportional to the normal force, but not to the area in contact, a somewhat counterintuitive notion. When two rough surfaces are in contact, the actual contact area is a tiny fraction of the total area since only high spots touch. When a greater normal force is exerted, the actual contact area increases, and it is found that the friction is proportional to this area.

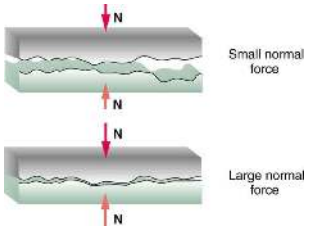


Figure **3.5.4.3**: Two rough surfaces in contact have a much smaller area of actual contact than their total area. When there is a greater normal force as a result of a greater applied force, the area of actual contact increases as does friction.

But the atomic-scale view promises to explain far more than the simpler features of friction. The mechanism for how heat is generated is now being determined. In other words, why do surfaces get warmer when rubbed? Essentially, atoms are linked with one another to form lattices. When surfaces rub, the surface atoms adhere and cause atomic lattices to vibrate—essentially creating sound waves that penetrate the material. The sound waves diminish with distance and their energy is converted into heat. Chemical reactions that are related to frictional wear can also occur between atoms and molecules on the surfaces. Figure **3.5.4.4** shows how the tip of a probe drawn across another material is deformed by atomic-scale friction. The force needed to drag the tip can be measured and is found to be related to shear stress, which will be discussed later in this chapter. The variation in shear stress is

remarkable (more than a factor of 10^{12}) and difficult to predict theoretically, but shear stress is yielding a fundamental understanding of a large-scale phenomenon known since ancient times—friction.

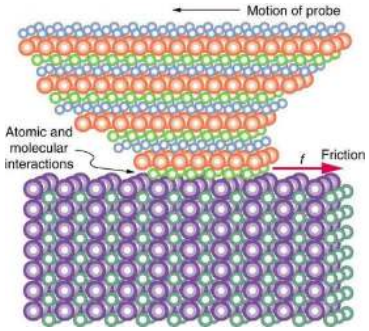


Figure 3.5.4.4: The tip of a probe is deformed sideways by frictional force as the probe is dragged across a surface. Measurements of how the force varies for different materials are yielding fundamental insights into the atomic nature of friction.

Section Summary

- Friction is a contact force between systems that opposes the motion or attempted motion between them. Simple friction is proportional to the normal force N pushing the systems together. (A normal force is always perpendicular to the contact surface between systems.) Friction depends on both of the materials involved. The magnitude of static friction f_s between systems stationary relative to one another is given by

$$f_s \leq \mu_s N,$$

  where μ_s is the coefficient of static friction, which depends on both of the materials.
- The kinetic friction force f_k between systems moving relative to one another is given by

$$f_k = \mu_k N,$$

  where μ_k is the coefficient of kinetic friction, which also depends on both materials.

Glossary

friction
a force that opposes relative motion or attempts at motion between systems in contact

kinetic friction
a force that opposes the motion of two systems that are in contact and moving relative to one another

static friction
a force that opposes the motion of two systems that are in contact and are not moving relative to one another

magnitude of static friction
$f_s \leq \mu_s N$, where μ_s is the coefficient of static friction and N is the magnitude of the normal force

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

magnitude of kinetic friction
$f_k = \mu_k N$, where μ_k is the coefficient of kinetic friction

This page titled 3.5.4: Friction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.8: Friction** by OpenStax is licensed CC BY 4.0.

3.5.5: Centripetal Force

Learning Objectives

- Explain the role of centripetal force in a uniform circular motion.
- Calculate centripetal force and acceleration for simple situations.

Any force or combination of forces can cause a centripetal or radial acceleration. Just a few examples are the tension in the rope on a tether ball, the force of Earth's gravity on the Moon, friction between roller skates and a rink floor, a banked roadway's force on a car, and forces on the tube of a spinning centrifuge.

Any net force causing uniform circular motion is called a **centripetal force**. The direction of a centripetal force is toward the center of curvature, the same as the direction of centripetal acceleration. According to Newton's second law of motion, net force is mass times acceleration: $F_{\text{net}} = ma$. For uniform circular motion, the acceleration is the centripetal acceleration— $a = a_c$ discussed in the previous chapter. Thus, the magnitude of centripetal force F_c is

$$F_c = ma_c.$$

If we analyze the mathematics of this in detail, we would see that for a given mass and velocity, a large centripetal force causes a small radius of curvature—that is, a tight curve.

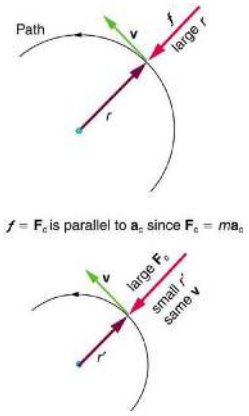


Figure 3.5.5.1: The frictional force supplies the centripetal force and is numerically equal to it. Centripetal force is perpendicular to velocity and causes uniform circular motion. The larger the F_c, the smaller the radius of curvature r and the sharper the curve. The second curve has the same v, but a larger F_c produces a smaller r'.

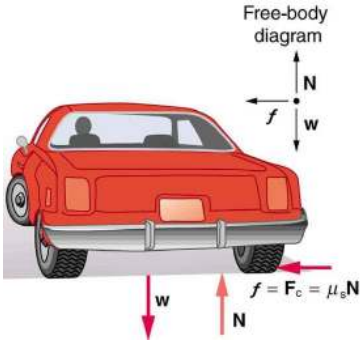


Figure 3.5.5.2: This car on level ground is moving away and turning to the left. The centripetal force causing the car to turn in a circular path is due to friction between the tires and the road. A minimum coefficient of friction is needed, or the car will move in a larger-radius curve and leave the roadway.

In the case of **banked curves**, where the slope of the road helps you negotiate the curve, some or all of the necessary centripetal force is provided by the normal force. See Figure 3.5.5.3. The greater the angle θ, the faster you can take the curve. Race tracks for bikes as well as cars, for example, often have steeply banked curves. In an "ideally banked curve," the angle θ is such that you can negotiate the curve at a certain speed without the aid of friction between the tires and the road. Conceptually, for **ideal banking**, the

net external force equals the horizontal centripetal force in the absence of friction. The components of the normal force N in the horizontal and vertical directions must equal the centripetal force and the weight of the car, respectively.

Figure 3.5.5.3 shows a free body diagram for a car on a frictionless banked curve. If the angle θ is ideal for the speed and radius, then the net external force will equal the necessary centripetal force. The only two external forces acting on the car are its weight w and the normal force of the road N. (A frictionless surface can only exert a force perpendicular to the surface—that is, a normal force.) These two forces must add to give a net external force that is horizontal toward the center of curvature and has magnitude mv^2/r. We omit detailed calculations, which require trigonometry.

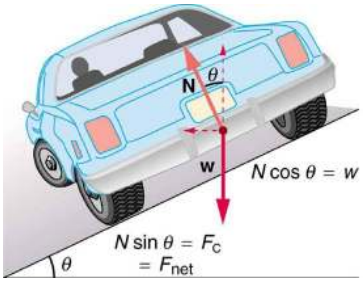


Figure 3.5.5.3: The car on this banked curve is moving away and turning to the left.

Section Summary

- Centripetal force $\mathbf{F_c}$ is any force causing uniform circular motion. It is a "center-seeking" force that always points toward the center of rotation. It is perpendicular to linear velocity v and has magnitude

$$F_c = ma_c,$$

which can also be expressed as

$$F_c = m\frac{v^2}{r}.$$

Glossary

centripetal force
any net force causing uniform circular motion

ideal banking
the sloping of a curve in a road, where the angle of the slope allows the vehicle to negotiate the curve at a certain speed without the aid of friction between the tires and the road; the net external force on the vehicle equals the horizontal centripetal force in the absence of friction

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

banked curve
the curve in a road that is sloping in a manner that helps a vehicle negotiate the curve

This page titled 3.5.5: Centripetal Force is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.10: Centripetal Force** by OpenStax is licensed CC BY 4.0.

3.6: Summary of Newton's Laws

Understanding Newton's laws of motion has been very important for our society in that this understanding has led to numerous technological advances. Think of all of the human constructed moving objects in this world. Each of these were designed and created by scientists and engineers with a deep understanding of Newton's Laws: airplanes, boats, cars, cannons, elevators, rockets, satellites, space shuttles, trains, trucks, and many more. Some of these vehicles we even trust with our lives! And it is not just the moving objects that require an understanding of Newton's laws of motion, Newton's laws are also important in order to understand how to build stationary objects that will stand the test of time. Bridges, buildings, fences, roads, swimming pools, towers and many more constructions can only be created to be reliable with a proper understanding of Newton's laws. Newton's laws are fundamental for creating the world in which we live. We will take a minute now to review these laws. Every example we have seen so far in this chapter can be solved with a proper understanding of Newton's laws of motion.

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.

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)*. Please review earlier portions of this chapter for more on Newton's first law of motion.

As it turns out, 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

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

This is often written in the more familiar form

$$F_{\text{net}} = ma,$$

with the vector notation indicating that the net external force is in the same direction as acceleration. When only the magnitude of force and acceleration are considered, this equation is simply (note the lack of vector notations)

$$F_{\text{net}} = ma.$$

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. This was also discussed previously in this chapter.

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 by the second body that is equal in magnitude and opposite in direction to the force that it itself 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. Again, this is review from material previously covered.

This page titled 3.6: Summary of Newton's Laws is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- 2.3: Newton's First Law of Motion- Inertia by OpenStax is licensed CC BY 4.0.

- **2.3: Newton's First Law of Motion- Inertia** by OpenStax is licensed CC BY 4.0.
- **2.4: Newton's Second Law of Motion- Force and Acceleration** by OpenStax is licensed CC BY 4.0.
- **2.5: Newton's Third Law of Motion- Symmetry in Forces** by OpenStax is licensed CC BY 4.0.

3.7: Newton's Universal Law of Gravitation

Learning Objectives

- Explain Earth's gravitational force.
- Describe the gravitational effect of the Moon on Earth.
- Explain sensation of weightlessness in space.

What do aching feet, a falling apple, and the orbit of the Moon have in common? Each is caused by the gravitational force. Our feet are strained by supporting our weight—the force of Earth's gravity on us. An apple falls from a tree because of the same force acting a few meters above Earth's surface. And the Moon orbits Earth because gravity is able to supply the necessary centripetal force at a distance of hundreds of millions of meters. In fact, the same force causes planets to orbit the Sun, stars to orbit the center of the galaxy, and galaxies to cluster together. Gravity is another example of underlying simplicity in nature. It is the weakest of the four basic forces found in nature, and in some ways the least understood. It is a force that acts at a distance, without physical contact, and is expressed by a formula that is valid everywhere in the universe, for masses and distances that vary from the tiny to the immense.

Sir Isaac Newton was the first scientist to precisely define the gravitational force, and to show that it could explain both falling bodies and astronomical motions. See Figure 3.7.1. But Newton was not the first to suspect that the same force caused both our weight and the motion of planets. His forerunner Galileo Galilei had contended that falling bodies and planetary motions had the same cause. Some of Newton's contemporaries, such as Robert Hooke, Christopher Wren, and Edmund Halley, had also made some progress toward understanding gravitation. But Newton was the first to propose an exact mathematical form and to use that form to show that the motion of heavenly bodies should be conic sections—circles, ellipses, parabolas, and hyperbolas. This theoretical prediction was a major triumph—it had been known for some time that moons, planets, and comets follow such paths, but no one had been able to propose a mechanism that caused them to follow these paths and not others.



Figure 3.7.1: According to early accounts, Newton was inspired to make the connection between falling bodies and astronomical motions when he saw an apple fall from a tree and realized that if the gravitational force could extend above the ground to a tree, it might also reach the Moon. The inspiration of Newton's apple is a part of worldwide folklore and may even be based in fact. Great importance is attached to it because Newton's universal law of gravitation and his laws of motion answered very old questions about nature and gave tremendous support to the notion of underlying simplicity and unity in nature. Scientists still expect underlying simplicity to emerge from their ongoing inquiries into nature.

The gravitational force is relatively simple. It is always attractive, and it depends only on the masses involved and the distance between them. Stated in modern language, **Newton's universal law of gravitation** states that every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

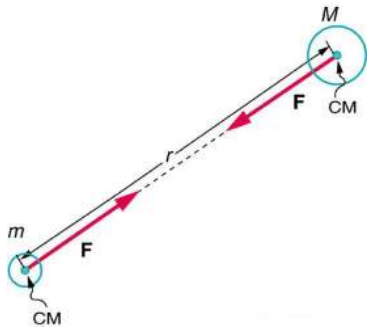


Figure 3.7.2: Gravitational attraction is along a line joining the centers of mass of these two bodies. The magnitude of the force is the same on each, consistent with Newton's third law.

MISCONCEPTION ALERT

The magnitude of the force on each object (one has larger mass than the other) is the same, consistent with Newton's third law.

The bodies we are dealing with tend to be large. To simplify the situation we assume that the body acts as if its entire mass is concentrated at one specific point called the **center of mass** (CM). For two bodies having masses m and M with a distance r between their centers of mass, the equation for Newton's universal law of gravitation is

$$F = G\frac{mM}{r^2},$$

where F is the magnitude of the gravitational force and G is a proportionality factor called the **gravitational constant**. G is a universal gravitational constant—that is, it is thought to be the same everywhere in the universe. It has been measured experimentally to be

$$G = 6.673 \times 10^{-11}\frac{\text{N}\cdot\text{m}^2}{\text{kg}^2}$$

in SI units. Note that the units of G are such that a force in newtons is obtained from $F = G\frac{mM}{r^2}$, when considering masses in kilograms and distance in meters. For example, two 1.000 kg masses separated by 1.000 m will experience a gravitational attraction of 6.673×10^{-11} N. This is an extraordinarily small force. The small magnitude of the gravitational force is consistent with everyday experience. We are unaware that even large objects like mountains exert gravitational forces on us. In fact, our body weight is the force of attraction of the *entire Earth* on us with a mass of 5.98×10^{24} kg.

Recall that the acceleration due to gravity g is about $9.80\ \text{m/s}^2$ on Earth. We can now determine why this is so. The weight of an object mg is the gravitational force between it and Earth. Substituting mg for F in Newton's universal law of gravitation gives

$$mg = G\frac{mM}{r^2},$$

where m is the mass of the object, M is the mass of Earth, and r is the distance to the center of Earth (the distance between the centers of mass of the object and Earth). See Figure 3.7.3. The mass m of the object cancels, leaving an equation for g:

$$g = G\frac{M}{r^2}.$$

Substituting known values for Earth's mass and radius (to three significant figures),

$$g = \left(6.67 \times 10^{-11}\frac{\text{N}\cdot\text{m}^2}{\text{kg}^2}\right) \times \frac{5.98 \times 10^{24}\ \text{kg}}{\left(6.38\times10^{6}\ \text{m}\right)^2},$$

and we obtain a value for the acceleration of a falling body:

$$g = 9.80\ \text{m/s}^2.$$

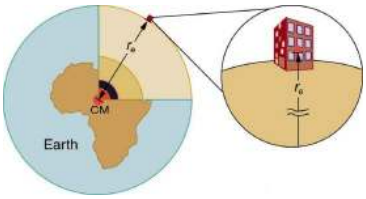


Figure **3.7.3**: The distance between the centers of mass of Earth and an object on its surface is very nearly the same as the radius of Earth, because Earth is so much larger than the object.

This is the expected value *and is independent of the body's mass*. Newton's law of gravitation takes Galileo's observation that all masses fall with the same acceleration a step further, explaining the observation in terms of a force that causes objects to fall—in fact, in terms of a universally existing force of attraction between masses.

TAKE-HOME EXPERIMENT

Take a marble, a ball, and a spoon and drop them from the same height. Do they hit the floor at the same time? If you drop a piece of paper as well, does it behave like the other objects? Explain your observations.

Example **3.7.1**: Earth's Gravitational Force on a Mass

(a) Determine the weight of a 5.00 kg rock when on Earth's surface.

(b) Determine the weight of a 5.00 kg rock when 3620 km above the surface of Earth.

Strategy for (a)

Use acceleration due to gravity near Earth's surface and Newton's second law.

Solution for (a)

$$F = mg = 5.00\ \text{kg} \times 9.80\ \text{m/s}^2 = 49.0\ \text{N}$$

Strategy for (b)

Use Newton's universal law of gravitation. Remember that distance is from the center of Earth. In this case the distance is 6380 km + 3620 km = 10,000 km = 1.00×10^7 m.

Solution for (b)

$$F = G\frac{mM}{r^2} = \left(6.673 \times 10^{-11}\frac{\text{N}\cdot\text{m}^2}{\text{kg}^2}\right)\frac{(5.00\ \text{kg})(5.98 \times 10^{24}\ \text{kg})}{(1.00\times10^7\ \text{m})^2} = 20.0\ \text{N}$$

Tides

Ocean tides are one very observable result of the Moon's gravity acting on Earth. Figure **3.7.4** is a simplified drawing of the Moon's position relative to the tides. Because water easily flows on Earth's surface, a high tide is created on the side of Earth nearest to the Moon, where the Moon's gravitational pull is strongest. Why is there also a high tide on the opposite side of Earth? The answer is that Earth is pulled toward the Moon more than the water on the far side, because Earth is closer to the Moon. So the water on the side of Earth closest to the Moon is pulled away from Earth, and Earth is pulled away from water on the far side. As Earth rotates, the tidal bulge (an effect of the tidal forces between an orbiting natural satellite and the primary planet that it orbits) keeps its orientation with the Moon. Thus, there are two tides per day (the actual tidal period is about 12 hours and 25.2 minutes), because the Moon moves in its orbit each day as well).



Figure **3.7.4**: The Moon causes ocean tides by attracting the water on the near side more than Earth, and by attracting Earth more than the water on the far side. The distances and sizes are not to scale. For this simplified representation of the Earth-Moon system, there are two high and two low tides per day at any location, because Earth rotates under the tidal bulge.

The Sun also affects tides, although it has about half the effect of the Moon. However, the largest tides, called spring tides, occur when Earth, the Moon, and the Sun are aligned. The smallest tides, called neap tides, occur when the Sun is at a 90° angle to the Earth-Moon alignment.

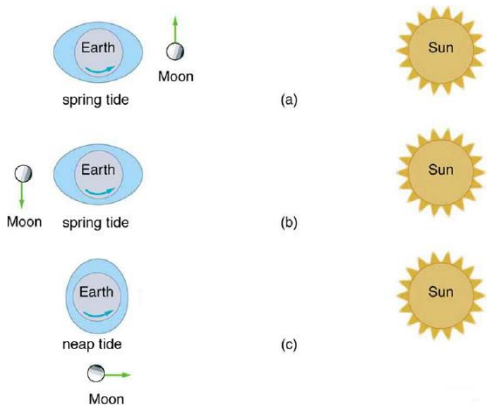


Figure 3.7.5: (a, b) Spring tides: The highest tides occur when Earth, the Moon, and the Sun are aligned. (c) Neap tide: The lowest tides occur when the Sun lies at 90° to the Earth-Moon alignment. Note that this figure is not drawn to scale.

Tides are not unique to Earth but occur in many astronomical systems. The most extreme tides occur where the gravitational force is the strongest and varies most rapidly, such as near black holes (see Figure 3.7.6). A few likely candidates for black holes have been observed in our galaxy. These have masses greater than the Sun but have diameters only a few kilometers across. The tidal forces near them are so great that they can actually tear matter from a companion star.

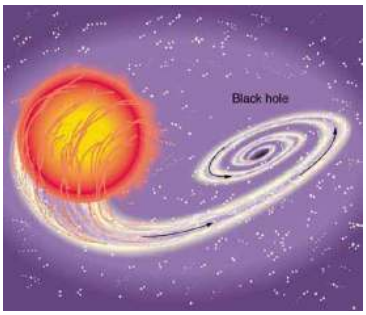


Figure 3.7.6: A black hole is an object with such strong gravity that not even light can escape it. This black hole was created by the supernova of one star in a two-star system. The tidal forces created by the black hole are so great that it tears matter from the companion star. This matter is compressed and heated as it is sucked into the black hole, creating light and X-rays observable from Earth.

”Weightlessness” and Microgravity

You may have seen images of astronauts in International Space Station, floating in their environment. What is the effect of “weightlessness” upon an astronaut who is in orbit for months? Or what about the effect of weightlessness upon plant growth? Weightlessness does not mean that an astronaut is not being acted upon by the gravitational force. There is no “zero gravity” in an astronaut’s orbit. The term just means that the astronaut is in free-fall, accelerating with the acceleration due to gravity. If an elevator cable breaks, the passengers inside will be in free fall and will experience weightlessness. You can experience short periods of weightlessness in some rides in amusement parks, and the commercial “zero-G” airplane rides simulate this experience of weightlessness about 20 seconds at a time.



Figure **3.7.7**: Astronauts experiencing weightlessness on board the International Space Station. (credit: NASA)

Microgravity refers to an environment in which the acceleration of a body due to non-gravitational forces is small compared with that produced by Earth on its surface. Many interesting biology and physics topics have been studied over the past three decades in the presence of microgravity. Of immediate concern is the effect on astronauts of extended times in outer space, such as at the International Space Station. Researchers have observed that muscles will atrophy (waste away) in this environment. There is also a corresponding loss of bone mass. Research continues on cardiovascular adaptation to space flight. On Earth, blood pressure is usually higher in the feet than in the head, because the higher column of blood exerts a downward force on it, due to gravity. When standing, 70% of your blood is below the level of the heart, while in a horizontal position, just the opposite occurs. What difference does the absence of this pressure differential have upon the heart?

Some findings in human physiology in space can be clinically important to the management of diseases back on Earth. On a somewhat negative note, spaceflight is known to affect the human immune system, possibly making the crew members more vulnerable to infectious diseases. Experiments flown in space also have shown that some bacteria grow faster in microgravity than they do on Earth. However, on a positive note, studies indicate that microbial antibiotic production can increase by a factor of two in space-grown cultures. One hopes to be able to understand these mechanisms so that similar successes can be achieved on the ground. In another area of physics space research, inorganic crystals and protein crystals have been grown in outer space that have much higher quality than any grown on Earth, so crystallography studies on their structure can yield much better results.

Plants have evolved with the stimulus of gravity and with gravity sensors. Roots grow downward and shoots grow upward. Plants might be able to provide a life support system for long duration space missions by regenerating the atmosphere, purifying water, and producing food. Some studies have indicated that plant growth and development are not affected by gravity, but there is still uncertainty about structural changes in plants grown in a microgravity environment.

Section Summary

- Newton's universal law of gravitation: Every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. In equation form, this is

$$F = G\frac{mM}{r^2},$$

  where F is the magnitude of the gravitational force. G is the gravitational constant, given by $G = 6.673 \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$.
- Newton's law of gravitation applies universally.

Glossary

gravitational constant, G

a proportionality factor used in the equation for Newton's universal law of gravitation; it is a universal constant—that is, it is thought to be the same everywhere in the universe

center of mass

the point where the entire mass of an object can be thought to be concentrated

microgravity

an environment in which the acceleration of a body due to non-gravitational forces is small compared with that produced by Earth on its surface

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

Newton’s universal law of gravitation

every particle in the universe attracts every other particle with a force along a line joining them; the force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them

---

This page titled 3.7: Newton’s Universal Law of Gravitation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **2.9: Newton’s Universal Law of Gravitation** by OpenStax is licensed CC BY 4.0.

3.8: Linear Momentum and Force

Learning Objectives

- 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.

Linear Momentum

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. In symbols, linear momentum is expressed as

$$\boldsymbol{p} = m\boldsymbol{v}$$

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 $\mathbf{kg \cdot m/s}$.

Definition: LINEAR MOMENTUM

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

$$\boldsymbol{p} = m\boldsymbol{v}.$$

Example 3.8.1: Calculating Momentum: A Football Player and a Football

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

Strategy

No information is given regarding direction, and so we can calculate only the magnitude of the momentum, p. (As usual, a symbol that is in italics is a magnitude, whereas one that is italicized and is boldfaced or 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 the equation, 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.

$$p_{\text{player}} = (110\ \text{kg})(8.00\ \text{m/s}) = 880\ \text{kg} \cdot \text{m/s}$$

Solution for (b)

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

$$p_{\text{ball}} = (0.410\ \text{kg})(25.0\ \text{m/s}) = 10.3\ \text{kg} \cdot \text{m/s}$$

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

$$\frac{p_{\text{player}}}{p_{\text{ball}}} = \frac{880}{10.3} = 85.9.$$

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 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. Using symbols, this law is

$$\boldsymbol{F}_{\text{net}} = \frac{\Delta \boldsymbol{p}}{\Delta t},$$

where $\boldsymbol{F}_{\text{net}}$ is the net external force, $\Delta \boldsymbol{p}$ is the change in momentum, and Δt is the change in time.

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: $\boldsymbol{F}_{\text{net}} = \Delta \boldsymbol{p}/\Delta t$.

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 $\boldsymbol{F}_{\text{net}} = m\boldsymbol{a}$ as a special case. We can derive this form as follows. First, note that the change in momentum $\Delta \boldsymbol{p}$ is given by

$$\Delta \boldsymbol{p} = \Delta(m\boldsymbol{v}).$$

If the mass of the system is constant, then

$$\Delta(m\boldsymbol{v}) = m\Delta \boldsymbol{v}.$$

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

$$\boldsymbol{F}_{\text{net}} = \frac{\Delta \boldsymbol{p}}{\Delta t} = m\frac{\Delta \boldsymbol{v}}{\Delta t}$$

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

$$\boldsymbol{F}_{\text{net}} = m\boldsymbol{a}$$

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 3.8.2: 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 $\boldsymbol{F}_{\text{net}} = \Delta \boldsymbol{p}/\Delta t$.

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

$$\Delta p = m\Delta v = m\left(v_{\text{f}} - v_{\text{i}}\right).$$

In this example, the velocity just after impact and the change in time are given; thus, once Δp is calculated, $F_{\text{net}} = \frac{\Delta p}{\Delta t}$ 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.

$$\begin{aligned}\Delta p &= m\,(v_{\text{f}} - v_{\text{i}}) \\ &= (0.057\ \text{kg})(58\ \text{m/s} - 0\ \text{m/s}) \\ &= 3.306\ \text{kg} \cdot \text{m/s} \approx 3.3\ \text{kg} \cdot \text{m/s}\end{aligned}$$

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

$$\begin{aligned}F_{\text{net}} &= \frac{\Delta p}{\Delta t} = \frac{3.306\ \text{kg} \cdot \text{m/s}}{5.0 \times 10^{-3}\ \text{s}} \\ &= 661\ \text{N} \approx 660\ \text{N},\end{aligned}$$

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 $F_{\text{net}} = ma$, but one additional step would be required compared with the strategy used in this example.

Section Summary

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

$$\boldsymbol{p} = m\boldsymbol{v},$$

  where mm is the mass of the system and $\boldsymbol{v}$ is its velocity.
- The SI unit for momentum is $\text{kg} \cdot \text{m/s}$.
- 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

$$F_{\text{net}} = \frac{\Delta \boldsymbol{p}}{\Delta t},$$

  where $\boldsymbol{F}_{\text{net}}$ is the net external force, $\Delta \boldsymbol{p}$ is the change in momentum, and Δt is the change time.

Glossary

linear momentum

the product of mass and velocity

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

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 3.8: Linear Momentum and Force is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **4.2: Linear Momentum and Force** by OpenStax is licensed CC BY 4.0.

3.9: Conservation of Momentum

Learning Objectives

- Describe the principle of conservation of momentum.
- Explain conservation of momentum with examples.

In the previous sections in this chapter, changes in momentum for an object were considered in terms of impulse and force. But for certain conditions, the total momentum is conserved (stays constant). What are these conditions?

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

Consider what happens if the masses of two colliding objects are more similar (i.e, both objects are about the same mass) than the masses of a football player and Earth—for example, one car bumping into another, as shown in Figure 3.9.1. Both cars are coasting in the same direction when the lead car (labeled m_2) is bumped by the trailing car (labeled m_1). The only unbalanced force on each car is the force of the collision. (Assume that the effects due to friction are negligible.) Car 1 slows down as a result of the collision, losing some momentum, while car 2 speeds up and gains some momentum. We shall now show that the total momentum of the two-car system remains constant.

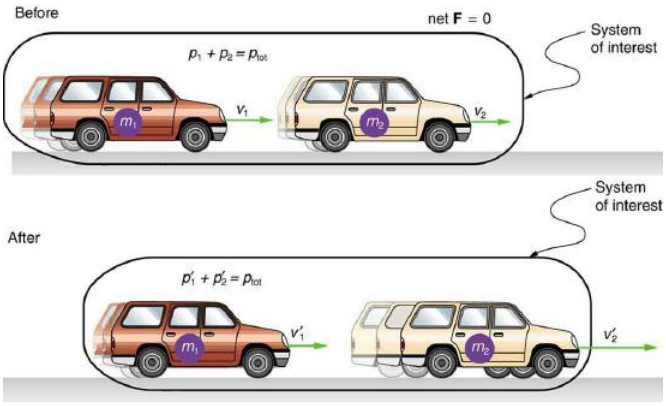


Figure 3.9.1: A car of mass m_1 moving with a velocity of v_1 bumps into another car of mass m_2 and velocity v_2 that it is following. As a result, the first car slows down to a velocity of v_1' and the second speeds up to a velocity of v_2'. The momentum of each car is changed, but the total momentum p_{tot} of the two cars is the same before and after the collision (if you assume friction is negligible).

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

$$\Delta p_1 = F_1 \Delta t,$$

where F_1 is the force on car 1 due to car 2, and Δt is the time the force acts (the duration of the collision). Similarly, the change in momentum of car 2 is

$$\Delta p_2 = F_2 \Delta t,$$

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

$$\Delta p_2 = -F_1 \Delta t = -\Delta p_1.$$

Thus, the changes in momentum cancel out, and

$$\Delta p_1 + \Delta p_2 = 0.$$

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

$$p_1 + p_2 = \text{ constant } .$$

Because of Newton's third law, when two objects in a system interact with each other, it does not change the total momentum of the system. In the absence of an external force (that is, force due to an object outside of the system), the total momentum of the system is constant, or conserved. We call this the **conservation of momentum principle**.

Note: Conservation Laws

The conservation of momentum is the first example we have seen in this text of a conservation law. It turns out that such laws are very important in many fields of science, and we will see more of them as we continue into later chapters. The reason why conservation laws are so useful is because they allow us to make important assumptions in order to solve problems that arise. For example, conservation of momentum is quite useful in describing collisions. Conservation of momentum is crucial to our understanding of atomic and subatomic particles because much of what we know about these particles comes from collision experiments.

We have noted that the motion in perpendicular directions—x, y, and —are independent. Momentum conservation follows this paradigm, and components of total momentum along each direction are conserved separately. For example, during projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero and momentum is unchanged. But along the vertical direction, the net vertical force is not zero and the momentum of the projectile is not conserved. (See Figure **3.9.2**.) However, if the momentum of the projectile-Earth system is considered in the vertical direction, we find that the total momentum is conserved.

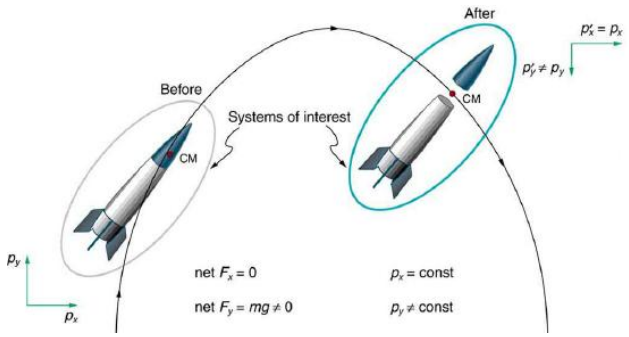


Figure **3.9.2**: The horizontal component of a projectile's momentum is conserved if air resistance is negligible, even in this case where a space probe separates. The forces causing the separation are internal to the system, so that the net external horizontal force Fx–net is still zero. The vertical component of the momentum is not conserved, because the net vertical force Fy–net is not zero. In the vertical direction, the space probe-Earth system needs to be considered and we find that the total momentum is conserved. The center of mass of the space probe takes the same path it would if the separation did not occur.

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

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION—DROP OF TENNIS BALL AND A BASKETBALL

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

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION—TWO TENNIS BALLS IN A BALLISTIC TRAJECTORY

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

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

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

Section Summary

- The conservation of momentum principle says that the total momentum is conserved, or constant, in the absence of a net external force, as a direct consequence of Newton's third law.
- During projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero.

Glossary

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

conservation of momentum principle

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

---

This page titled 3.9: Conservation of Momentum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **4.4: Conservation of Momentum** by OpenStax is licensed CC BY 4.0.

3.10: Rotation

Learning Objectives

- Describe uniform circular motion.
- Know the rotational analogs for translational measures of motion, force, and momentum.

So far we have considered in detail the movement of particles in straight lines: changes in their direction, velocity, acceleration, and momentum, as well as the forces acting on those objects to cause them to move (or in some cases prevent them from moving). This is often referred to as translational motion. But we also know that objects can rotate and move in circular paths. Can the laws of motion address this sort of movement as well? It turns out that they can, but to do so requires a bit more effort. We are not going to develop all of the relationships here, but we will discuss a few of the definitions and principles involved. With these in place we will be able to then discuss some of the implications of this type of movement.

Previously, we looked at uniform circular motion, where an object moves in a circular path at a constant speed (but changing directions, so it undergoes centripetal acceleration). We can also describe this in terms of a rotation at a constant rate. For example, an object completing a circle every second can be said to rotate at a constant rate of 360 degrees per second about the center of the circle (since there are 360 degrees in one complete circle). This type of motion is referred to as angular motion. As we discuss angular motion, we will consider things like angular velocity, angular acceleration, and angular momentum. We will also introduce new analogs for displacement and mass in the concepts of change of angle and moment of inertia.

For the formulas and equations you will see in this section, we will introduce a mathematically convenient unit of angle called **radians**. An angular measure in radian is defined by the arc length covered by the angle divided by the radius of the circle that the arc length is part of. It turns out that if we do this, we can then define the angular velocity as the speed of the object rotating divided by the radius about which it rotates. This will then affect our definitions of angular acceleration and angular momentum (which are summarized in a table later in this section).

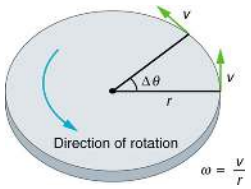


Figure **3.10.1**: This figure shows uniform circular motion and some of its defined quantities.

The other thing we must consider for the rotation of objects is that it often becomes more challenging to consider them to be mere particles. As a reminder, a particle is when we consider objects as a point mass where we can ignore the size and shape of the object itself. This is an assumption we make in order to simplify the calculations involved. For many rotating objects this is assumption is no longer valid and we must instead consider how the shape of the object affects the way in which it will rotate. In order to do this we must define a way of keeping track of the inertia of a moving object more complicated than a point mass. The inertia of this object is instead understood to be the summation of all the point masses that make it up modified by the relationship of those objects to the radius around which they are rotating. The result of this calculation is referred to as the moment of inertia, and is often given the symbol *I*. This value, *I*, will be the angular motion analog to the *m* in translational motion. Calculating the moment of inertia is rather complicated, but you can learn more about it in other texts if you are interested.

At this point we can define the remaining rotational measures in terms of analogs of translational dynamics. There are many lower case Greek letters which are often used for these measures. θ (theta) is used to indicate a measure of the angle. ω (omega) is used to indicate a measure of the angular velocity. α is used to indicate a measure of the angular acceleration. Force effected to cause turning is referred to as torque, and τ (tau) is used to indicate a measure of this torque. These and the remaining rotational analaogs of translational movement are summarized in the following table. For the purpose of this text we are mainly concerned with realizing that each aspect of dynamics we have considered for translation has an analog for rotation.

Rotational and Translational Quantities

Rotational	Translational	Relationship
θ	x	$\theta = \frac{x}{r}$
ω	v	$\omega = \frac{v}{r}$
α	a	$\alpha = \frac{a_t}{r}$
τ	F	$\tau = mr^2\alpha$
I	m	depends on shape
L	p	L=Iω

MAKING CONNECTIONS: TAKE-HOME EXPERIMENT

Sit down with your feet on the ground on a chair that rotates. Lift one of your legs such that it is unbent (straightened out). Using the other leg, begin to rotate yourself by pushing on the ground. Stop using your leg to push the ground but allow the chair to rotate. From the origin where you began, sketch the angle, angular velocity, and angular acceleration of your leg as a function of time in the form of three separate graphs. Estimate the magnitudes of these quantities.

Section Summary

- Uniform circular motion is the motion with a constant angular velocity $\omega = \frac{\Delta\theta}{\Delta t}$.
- In non-uniform circular motion, the velocity changes with time and the rate of change of angular velocity (i.e. angular acceleration) is $\alpha = \frac{\Delta\omega}{\Delta t}$.
- Linear or tangential acceleration refers to changes in the magnitude of velocity but not its direction, given as $a_t = \frac{\Delta v}{\Delta t}$.
- Linear or tangential motion variables x, v, and a_t are proportional to the rotational quantities θ, ω, and α. They are also proportional to the radius r of the rotational motion.

Glossary

angular acceleration

the rate of change of angular velocity with time

angular velocity

the rate of change of angular position with time

change in angular velocity

the difference between final and initial values of angular velocity

radian

a unit of angular measure defined by the arc length covered by the angle divided by the radius of the circle that the arc length is part of; one full circle is 2π radians.

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

tangential acceleration

the acceleration in a direction tangent to the circle at the point of interest in circular motion

This page titled 3.10: Rotation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **6.2: Angular Acceleration** by OpenStax is licensed CC BY 4.0.

3.11: Angular Momentum and Its Conservation

Learning Objectives

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

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

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

$$L = I\omega.$$

This equation is an analog to the definition of linear momentum as $p = mv$. Units for linear momentum are $\mathrm{kg \cdot m/s}$ while units for angular momentum are $\mathrm{kg \cdot m^2/s}$. As we would expect, an object that has a large moment of inertia I, such as Earth, has a very large angular momentum. An object that has a large angular velocity ω, such as a centrifuge, also has a rather large angular momentum.

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

$$\text{net } \tau = \frac{\Delta L}{\Delta t}.$$

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

Conservation of Angular Momentum

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

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

$$\text{net } \tau = 0$$

implying that

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

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

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

or

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

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

MAKING CONNECTIONS: CONSERVATION LAWS

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

Before looking at some examples of the conservation of angular momentum, let us consider what it means that angular momentum is conserved. For any rotating object experiencing no net torque, we can express the conservation of angular momentum with the following equation:

$$L = L'.$$

Here, the primed quantity might represent the angular momentum after some change occurs to the system. (Primed quantity here means the same symbol but with a ' symbol after it. This is a common convention in physics to show how measures of the same variable relate after some change has occurred.)

We must ask ourselves what change could possibly be occurring in this system if there is not net torque. We also saw changes in systems where linear momentum is conserved. For example, a part might fall off of a rocket, decreasing the mass of the remaining rocket. The linear momentum of the system composed of both the rocket and the piece that fell from it will remain the same even as those two objects are separated. Just as linear momentum is composed of mass and velocity, angular momentum is composed of angular velocity and moment of inertia. We can express the equation as such,

$$I\omega = I'\omega',$$

Looking at this equation, we can see how changes might be possible to angular velocity if we make changes to the moment of inertia. Unlike mass, moment of inertia is not conserved, so it is possible to effect great changes in the angular velocity simply by moving the parts of the system relative to the axis of rotation. It turns out that moment of inertia becomes smaller as the mass of a spinning object moves closer to the axis of rotation. By changing the location of the masses that make up a spinning object, we can effect great changes in the angular velocity without expending much force to do so. We will discuss some examples of this in phenomena that you may be aware of.

An example of conservation of angular momentum is seen in Figure **3.11.2**, in which an ice skater is executing a spin. The net torque on her is very close to zero, because there is relatively little friction between her skates and the ice and because the friction is exerted very close to the pivot point. (Both $\boldsymbol{F}$ and $\boldsymbol{R}$ are small, and so τ is negligibly small.) Consequently, she can spin for quite some time. She can do something else, too: she can increase her rate of spin by pulling her arms and legs in. Why does pulling her arms and legs in increase her rate of spin? The answer is that her angular momentum is constant, so that changing the moment of inertia can result in a corresponding change to angular momentum, as shown in the previous equation. When the skater pulls her arms in I', becomes smaller as more of her mass gets closer to the axis of rotation. Because I' is smaller, the angular velocity ω' must increase to keep the angular momentum constant. The change can be dramatic, as the following example shows.

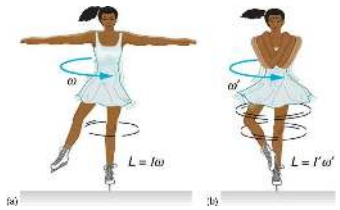


Figure **3.11.2**: (a) An ice skater is spinning on the tip of her skate with her arms extended. Her angular momentum is conserved because the net torque on her is negligibly small. In the next image, her rate of spin increases greatly when she pulls in her arms, decreasing her moment of inertia. The work she does to pull in her arms results in an increase in rotational kinetic energy.

There are several other examples of objects that increase their rate of spin because something reduced their moment of inertia. Tornadoes are one example. Storm systems that create tornadoes are slowly rotating. When the radius of rotation narrows, even in a local region, angular velocity increases, sometimes to the furious level of a tornado. Earth is another example. Our planet was born from a huge cloud of gas and dust, the rotation of which came from turbulence in an even larger cloud. Gravitational forces caused the cloud to contract, and the rotation rate increased as a result. (See Figure **3.11.3**.)

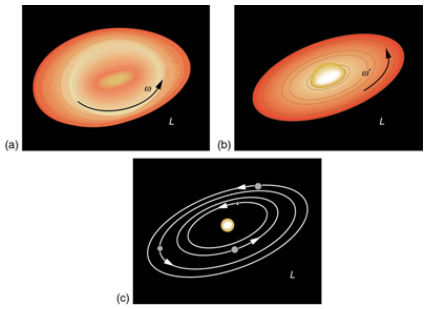


Figure 3.11.3: The Solar System coalesced from a cloud of gas and dust that was originally rotating. The orbital motions and spins of the planets are in the same direction as the original spin and conserve the angular momentum of the parent cloud.

In case of human motion, one would not expect angular momentum to be conserved when a body interacts with the environment as its foot pushes off the ground. Astronauts floating in space aboard the International Space Station have no angular momentum relative to the inside of the ship if they are motionless. Their bodies will continue to have this zero value no matter how they twist about as long as they do not give themselves a push off the side of the vessel.

Exercise 3.11.1

Is angular momentum completely analogous to linear momentum? What, if any, are their differences?

Answer

Yes, angular and linear momentums are completely analogous. While they are exact analogs they have different units and are not directly inter-convertible like forms of energy are.

Section Summary

- Every rotational phenomenon has a direct translational analog , likewise angular momentum $\boldsymbol{L}$ can be defined as $\boldsymbol{L = I\omega}$.
- This equation is an analog to the definition of linear momentum as $\boldsymbol{p = mv}$. The relationship between torque and angular momentum is $\textbf{net}\ \tau = \frac{\Delta L}{\Delta t}$.
- Angular momentum, like energy and linear momentum, is conserved. This universally applicable law is another sign of underlying unity in physical laws. Angular momentum is conserved when net external torque is zero, just as linear momentum is conserved when the net external force is zero.

Glossary

Contributors

Curated from resources found in Introduction to Physics published by OpenStax.

rotational kinetic energy

the kinetic energy due to the rotation of an object. This is part of its total kinetic energy

---

This page titled 3.11: Angular Momentum and Its Conservation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **6.5: Angular Momentum and Its Conservation** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

4: Fluid Mechanics and Waves

- 4.1: Introduction to Fluids and Waves
- 4.2: What Is a Fluid?
- 4.3: Density
- 4.4: Pressure
  - 4.4.1: Pressure Due to the Weight of Fluid
  - 4.4.2: Boyle's Law
- 4.5: Fluid Principles
  - 4.5.1: Archimedes' Principle
  - 4.5.2: Pascal's Principle
  - 4.5.3: Bernoulli's Principle
- 4.6: Wave Mechanics
  - 4.6.1: Period and Frequency in Oscillations
  - 4.6.2: Simple Harmonic Motion- A Special Periodic Motion
  - 4.6.3: Waves
- 4.7: Properties of Waves
  - 4.7.1: Transverse and Longitudinal Waves
  - 4.7.2: Forced Oscillations and Resonance
  - 4.7.3: Wave Interference- Standing Waves and Beats
  - 4.7.4: When Waves Sources Move- Doppler effects and Bow Wakes
- 4.8: Sound
  - 4.8.1: Speed of Sound, Frequency, and Wavelength

This page titled 4: Fluid Mechanics and Waves is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

4.1: Introduction to Fluids and Waves



Figure **4.1.1**: The fluid essential to all life has a beauty of its own. It also helps support the weight of this swimmer. (credit: Terren, Wikimedia Commons)

Much of what we value in life is fluid: a breath of fresh winter air; the hot blue flame in our gas cooker; the water we drink, swim in, and bathe in; the blood in our veins. What exactly is a fluid? Can we understand fluids with the laws already presented, or will new laws emerge from their study? The physical characteristics of fluids and some of the laws that govern their behavior are one of the topics of this chapter.

Also by their very definition, fluids flow. Examples come easily—a column of smoke rises from a camp fire, water streams from a fire hose, blood courses through your veins. Why does rising smoke curl and twist? How does a nozzle increase the speed of water emerging from a hose? How does the body regulate blood flow? The physics of fluids in motion—fluid dynamics—allows us to answer these and many other questions.

We will also introduce waves in this chapter. When you hear the word wave, you might think of water, which is a fluid. Waves can travel through fluids, but they can also travel through other media, as we will learn.



Figure **4.1.2**: There are at least four types of waves in this picture—only the water waves are evident. There are also sound waves, light waves, and waves on the guitar strings. (credit: John Norton)

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

Some oscillations create **waves**. A guitar creates sound waves. You can make water waves in a swimming pool by slapping the water with your hand. You can no doubt think of other types of waves. Some, such as water waves, are visible. Some, such as sound waves, are not. But *every wave is a disturbance that moves from its source and carries energy (which we will learn more*

about in a later chapter). Other examples of waves include earthquakes and visible light. Even subatomic particles, such as electrons, can behave like waves.

By studying oscillatory motion and waves, we shall find that a small number of underlying principles describe all of them and that wave phenomena are more common than you have ever imagined.

Glossary

oscillate

moving back and forth regularly between two points

wave

a disturbance that moves from its source and carries energy

This page titled 4.1: Introduction to Fluids and Waves is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **7.1: Introduction to Fluids** by OpenStax is licensed CC BY 4.0.
- **5.1: Introduction to Oscillatory Motion and Waves** by OpenStax is licensed CC BY 4.0.

4.2: What Is a Fluid?

Learning Objectives

- State the common phases of matter.
- Explain the physical characteristics of solids, liquids, and gases.
- Describe the arrangement of atoms in solids, liquids, and gases.

Matter most commonly exists as a solid, liquid, or gas; these states are known as the three common *phases of matter* and will be explored in more detail in a later chapter. Solids have a definite shape and a specific volume, liquids have a definite volume but their shape changes depending on the container in which they are held, and gases have neither a definite shape nor a specific volume as their molecules move to fill the container in which they are held. (See Figure **4.2.1**.) Liquids and gases are considered to be fluids because they yield to shearing forces, whereas solids resist them. Note that the extent to which fluids yield to shearing forces (and hence flow easily and quickly) depends on a quantity called the viscosity. We can understand the phases of matter and what constitutes a fluid by considering the forces between atoms that make up matter in the three phases.

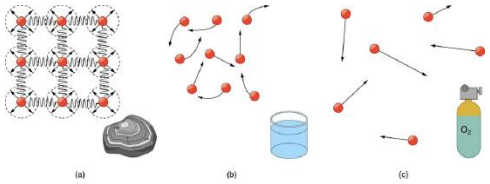


Figure **4.2.1**: (a) Atoms in a solid always have the same neighbors, held near home by forces represented here by springs. These atoms are essentially in contact with one another. A rock is an example of a solid. This rock retains its shape because of the forces holding its atoms together. (b) Atoms in a liquid are also in close contact but can slide over one another. Forces between them strongly resist attempts to push them closer together and also hold them in close contact. Water is an example of a liquid. Water can flow, but it also remains in an open container because of the forces between its atoms. (c) Atoms in a gas are separated by distances that are considerably larger than the size of the atoms themselves, and they move about freely. A gas must be held in a closed container to prevent it from moving out freely.

Atoms in *solids* are in close contact, with forces between them that allow the atoms to vibrate but not to change positions with neighboring atoms. (These forces can be thought of as springs that can be stretched or compressed, but not easily broken.) Thus a solid *resists* all types of stress. A solid cannot be easily deformed because the atoms that make up the solid are not able to move about freely. Solids also resist compression, because their atoms form part of a lattice structure in which the atoms are a relatively fixed distance apart. Under compression, the atoms would be forced into one another. Most of the examples we have studied so far have involved solid objects which deform very little when stressed.

CONNECTIONS: SUBMICROSCOPIC EXPLANATION OF SOLIDS AND LIQUIDS

Atomic and molecular characteristics explain and underlie the macroscopic characteristics of solids and fluids. This present section is devoted entirely to the submicroscopic explanation of solids and liquids.

In contrast, *liquids* deform easily when stressed and do not spring back to their original shape once the force is removed because the atoms are free to slide about and change neighbors—that is, they *flow* (so they are a type of fluid), with the molecules held together by their mutual attraction. When a liquid is placed in a container with no lid on, it remains in the container (providing the container has no holes below the surface of the liquid!). Because the atoms are closely packed, liquids, like solids, resist compression.

Atoms in *gases* are separated by distances that are large compared with the size of the atoms. The forces between gas atoms are therefore very weak, except when the atoms collide with one another. Gases thus not only flow (and are therefore considered to be fluids) but they are relatively easy to compress because there is much space and little force between atoms. When placed in an open container gases, unlike liquids, will escape. The major distinction is that gases are easily compressed, whereas liquids are not. We shall generally refer to both gases and liquids simply as **fluids**, and make a distinction between them only when they behave differently.

Section Summary

- A fluid is a state of matter that yields to sideways or shearing forces. Liquids and gases are both fluids. Fluid statics is the physics of stationary fluids.

Glossary

fluids

liquids and gases; a fluid is a state of matter that yields to shearing forces

---

This page titled 4.2: What Is a Fluid? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **7.2: What Is a Fluid?** by OpenStax is licensed CC BY 4.0.

4.3: Density

Learning Objectives

- Define density.
- Calculate the mass of an object from its density and volume.
- Compare and contrast the densities of various substances.

Which weighs more, a ton of feathers or a ton of bricks? This old riddle plays with the distinction between mass and density. A ton is a ton, of course; but bricks have much greater density than feathers, and so we are tempted to think of them as heavier. (See Figure 4.3.1.)

Density, as you will see, is an important characteristic of substances. It is crucial, for example, in determining whether an object sinks or floats in a fluid. Density is the mass per unit volume of a substance or object. In equation form, density is defined as

$$\rho = \frac{m}{V},$$

where the Greek letter ρ (rho) is the symbol for density, m is the mass, and V is the volume occupied by the substance.

Definition: DENSITY

Density is mass per unit volume.

$$\rho = \frac{m}{V},$$

where ρ is the symbol for density, m is the mass, and V is the volume occupied by the substance.

In the riddle regarding the feathers and bricks, the masses are the same, but the volume occupied by the feathers is much greater, since their density is much lower. The SI unit of density is $\mathbf{kg/m^3}$, representative values are given in Table 4.3.1. The metric system was originally devised so that water would have a density of $\mathbf{1\ g/cm^3}$, equivalent to $\mathbf{10^3\ kg/m^3}$. Thus the basic mass unit, the kilogram, was first devised to be the mass of 1000 mL of water, which has a volume of 1000 cm^3.

Table 4.3.1: Densities of Various Substances

Substance	$\rho\,(10^3\ \text{kg/m}^3 \text{ or } \text{g/mL})$	Substance	$\rho\,(10^3\ \text{kg/m}^3 \text{ or } \text{g/mL})$	Substance	$\rho\,(10^3\ \text{kg/m}^3 \text{ or } \text{g/mL})$
Solids		**Liquids**		**Gases**	
Aluminum	2.7	Water (4ºC)	1.000	Air	1.29×10^{-3}
Brass	8.44	Blood	1.05	Carbon dioxide	1.98×10^{-3}
Copper (average)	8.8	Sea water	1.025	Carbon monoxide	1.25×10^{-3}
Gold	19.32	Mercury	13.6	Hydrogen	0.090×10^{-3}
Iron or steel	7.8	Ethyl alcohol	0.79	Helium	0.18×10^{-3}
Lead	11.3	Petrol	0.68	Methane	0.72×10^{-3}
Polystyrene	0.10	Glycerin	1.26	Nitrogen	1.25×10^{-3}
Tungsten	19.30	Olive oil	0.92	Nitrous oxide	1.98×10^{-3}
Uranium	18.70			Oxygen	1.43×10^{-3}
Concrete	2.30–3.0			Steam (100°C)	0.60×10^{-3}
Cork	0.24				
Glass, common (average)	2.6				
Granite	2.7				
Earth’s crust	3.3				
Wood	0.3–0.9				
Ice (0°C)	0.917				
Bone	1.7–2.0				



Figure 4.3.1: A ton of feathers and a ton of bricks have the same mass, but the feathers make a much bigger pile because they have a much lower density.

As you can see by examining Table 4.3.1, the density of an object may help identify its composition. The density of gold, for example, is about 2.5 times the density of iron, which is about 2.5 times the density of aluminum. Density also reveals something about the phase of the matter and its substructure. Notice that the densities of liquids and solids are roughly comparable, consistent with the fact that their atoms are in close contact. The densities of gases are much less than those of liquids and solids, because the atoms in gases are separated by large amounts of empty space.

TAKE-HOME EXPERIMENT SUGAR AND SALT

A pile of sugar and a pile of salt look pretty similar, but which weighs more? If the volumes of both piles are the same, any difference in mass is due to their different densities (including the air space between crystals). Which do you think has the greater density? What values did you find? What method did you use to determine these values?

Example 4.3.1: Calculating the Mass of a Reservoir From Its Volume

A reservoir has a surface area of $50.0\ \text{km}^2$ and an average depth of 40.0 m. What mass of water is held behind the dam? (See Figure 4.3.2 for a view of a large reservoir—the Three Gorges Dam site on the Yangtze River in central China.)

Strategy

We can calculate the volume V of the reservoir from its dimensions, and find the density of water ρ in Table 4.3.1. Then the mass m can be found from the definition of density

$$\rho = \frac{m}{V}.$$

Solution

Solving equation $\rho = m/V$ for m gives $m = \rho V$.

The volume V of the reservoir is its surface area A times its average depth h:

$$\begin{aligned} V = Ah &= (50.0\ \text{km}^2)\,(40.0\ \text{m}) \\ &= \left[(50.0\ \text{km}^2)\left(\frac{10^3\ \text{m}}{1\ \text{km}}\right)^2\right](40.0\ \text{m}) = 2.00 \times 10^9\ \text{m}^3 \end{aligned}$$

The density of water ρ from Table 4.3.1 is $1.000 \times 10^3\ \text{kg/m}^3$. Substituting V and ρ into the expression for mass gives

$$\begin{aligned} m &= \left(1.00 \times 10^3\ \text{kg/m}^3\right)\left(2.00 \times 10^9\ \text{m}^3\right) \\ &= 2.00 \times 10^{12}\ \text{kg}. \end{aligned}$$

Discussion

A large reservoir contains a very large mass of water. In this example, the weight of the water in the reservoir is $mg = 1.96 \times 10^{13}\ \text{N}$, where g is the acceleration due to the Earth's gravity (about $9.80\ \text{m/s}^2$). It is reasonable to ask whether the dam must supply a force equal to this tremendous weight. The answer is no. As we shall see in the following sections, the force the dam must supply can be much smaller than the weight of the water it holds back.



Figure 4.3.2: Three Gorges Dam in central China. When completed in 2008, this became the world's largest hydroelectric plant, generating power equivalent to that generated by 22 average-sized nuclear power plants. The concrete dam is 181 m high and 2.3 km across. The reservoir made by this dam is 660 km long. Over 1 million people were displaced by the creation of the reservoir. (credit: Le Grand Portage)

Section Summary

- Density is the mass per unit volume of a substance or object. In equation form, density is defined as

$$\rho = \frac{m}{V}.$$

- The SI unit of density is kg/m^3.

Glossary

density
the mass per unit volume of a substance or object

This page titled 4.3: Density is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- 7.3: Density by OpenStax is licensed CC BY 4.0.

https://chem.libretexts.org/@go/page/472540

4.4: Pressure

Learning Objectives

- Define pressure.
- Explain the relationship between pressure and force.
- Calculate force given pressure and area.

You have no doubt heard the word **pressure** being used in relation to blood (high or low blood pressure) and in relation to the weather (high- and low-pressure weather systems). These are only two of many examples of pressures in fluids. Pressure P is defined as

$$P = \frac{F}{A}$$

where F is a force applied to an area A that is perpendicular to the force.

Definition: PRESSURE

Pressure is defined as the force divided by the area perpendicular to the force over which the force is applied, or

$$P = \frac{F}{A}.$$

A given force can have a significantly different effect depending on the area over which the force is exerted, as shown in Figure 4.4.1. Note that in this example, the pressure is not from a fluid. Pressure is a concept that can be applied to all of the physical states. For the sake of this course, however, we will largely focus on examples of pressure specifically for fluids.

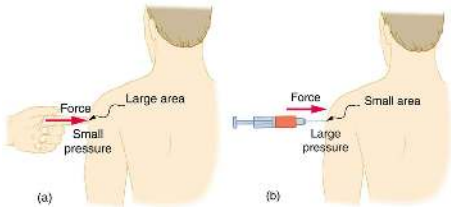


Figure 4.4.1: (a) While the person being poked with the finger might be irritated, the force has little lasting effect. (b) In contrast, the same force applied to an area the size of the sharp end of a needle is great enough to break the skin.

The SI unit for pressure is the pascal, where

$$1\ \text{Pa} = 1\ \text{N/m}^2.$$

In addition to the pascal, there are many other units for pressure that are in common use. In meteorology, atmospheric pressure is often described in units of millibar (mb), where

$$100\text{mb} = 1 \times 10^5\ \text{Pa}.$$

Pounds per square inch (lb/in^2 or psi) is still sometimes used as a measure of tire pressure, and millimeters of mercury (mm Hg) is still often used in the measurement of blood pressure. Pressure is defined for all states of matter but is particularly important when discussing fluids.

Example 4.4.1: Calculating Force Exerted by the Air: What Force Does a Pressure Exert?

An astronaut is working outside the International Space Station where the atmospheric pressure is essentially zero. The pressure gauge on her air tank reads 6.90×10^6 Pa. What force does the air inside the tank exert on the flat end of the cylindrical tank, a disk 0.150 m in diameter?

Strategy

We can find the force exerted from the definition of pressure given in $P = \frac{F}{A}$, provided we can find the area A acted upon.

Solution

By rearranging the definition of pressure to solve for force, we see that

$$F = PA.$$

Here, the pressure P is given, as is the area of the end of the cylinder A, given by $A = \pi r^2$. Thus,

$$\begin{aligned} F &= \left(6.90 \times 10^6 \text{ N/m}^2\right)(3.14)(0.0750 \text{ m})^2 \\ &= 1.22 \times 10^5 \text{ N}. \end{aligned}$$

Discussion

Wow! No wonder the tank must be strong. Since we found $F = PA$, we see that the force exerted by a pressure is directly proportional to the area acted upon as well as the pressure itself.

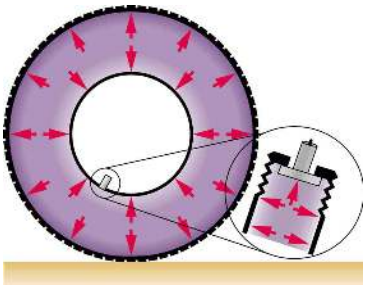


Figure **4.4.2**: Pressure inside this tire exerts forces perpendicular to all surfaces it contacts. The arrows give representative directions and magnitudes of the forces exerted at various points. Note that static fluids do not exert shearing forces.

The force exerted on the end of the tank is perpendicular to its inside surface. This direction is because the force is exerted by a static or stationary fluid. We have already seen that fluids cannot *withstand* shearing (sideways) forces; they cannot *exert* shearing forces, either. Fluid pressure has no direction, being a scalar quantity. The forces due to pressure have well-defined directions: they are always exerted perpendicular to any surface. (See the tire in Figure **4.4.2**, for example.) Finally, note that pressure is exerted on all surfaces. Swimmers, as well as the tire, feel pressure on all sides. (See Figure **4.4.3**.)

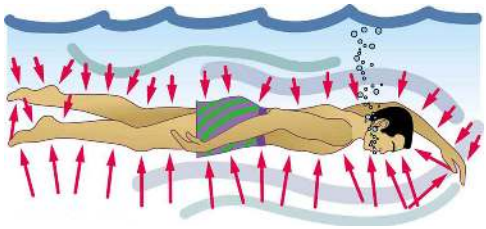


Figure **4.4.3**: Pressure is exerted on all sides of this swimmer, since the water would flow into the space he occupies if he were not there. The arrows represent the directions and magnitudes of the forces exerted at various points on the swimmer. Note that the forces are larger underneath, due to greater depth, giving a net upward or buoyant force that is balanced by the weight of the swimmer.

In the following subsections we will consider specific situations where there are certain other variables which affect the pressure of fluids in particular.

Section Summary

- Pressure is the force per unit perpendicular area over which the force is applied. In equation form, pressure is defined as

$$P = \frac{F}{A}.$$

- The SI unit of pressure is pascal and

$$1 \text{ Pa} = 1 \text{ N/m}^2.$$

pressure

the force per unit area perpendicular to the force, over which the force acts

---

This page titled 4.4: Pressure is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **7.4: Pressure** by OpenStax is licensed CC BY 4.0.

4.4.1: Pressure Due to the Weight of Fluid

Learning Objectives

- Define pressure in terms of weight.
- Explain the variation of pressure with depth in a fluid.
- Calculate density given pressure and altitude.

If your ears have ever popped on a plane flight or ached during a deep dive in a swimming pool, you have experienced the effect of depth on pressure in a fluid. At the Earth's surface, the air pressure exerted on you is a result of the weight of air above you. This pressure is reduced as you climb up in altitude and the weight of air above you decreases. Under water, the pressure exerted on you increases with increasing depth. In this case, the pressure being exerted upon you is a result of both the weight of water above you *and* that of the atmosphere above you. You may notice an air pressure change on an elevator ride that transports you many stories, but you need only dive a meter or so below the surface of a pool to feel a pressure increase. The difference is that water is much denser than air, about 775 times as dense.

Consider the container in Figure 4.4.1.1. Its bottom supports the weight of the fluid in it. Let us calculate the pressure exerted on the bottom by the weight of the fluid. That **pressure** is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):

$$P = \frac{mg}{A}.$$

We can find the mass of the fluid from its volume and density:

$$m = \rho V.$$

The volume of the fluid V is related to the dimensions of the container. It is

$$V = Ah,$$

where A is the cross-sectional area and h is the depth. Combining the last two equations gives

$$m = \rho Ah.$$

If we enter this into the expression for pressure, we obtain

$$P = \frac{(\rho Ah)g}{A}.$$

The area cancels, and rearranging the variables yields

$$P = \rho gh.$$

This value is the **pressure due to the weight of a fluid**. The equation has general validity beyond the special conditions under which it is derived here. Even if the container were not there, the surrounding fluid would still exert this pressure, keeping the fluid static. Thus the equation $P = \rho gh$ represents the pressure due to the weight of any fluid of *average density* ρ at any depth h below its surface. For liquids, which are nearly incompressible, this equation holds to great depths.

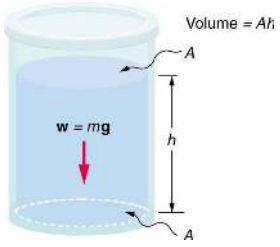


Figure 4.4.1.1: The bottom of this container supports the entire weight of the fluid in it. The vertical sides cannot exert an upward force on the fluid (since it cannot withstand a shearing force), and so the bottom must support it all.

Example 4.4.1.1: Calculating the Average Pressure and Force Exerted: What Force Must a Dam Withstand?

In Example 4.3.1, we calculated the mass of water in a large reservoir. We will now consider the pressure and force acting on the dam retaining water. (See Figure 4.4.1.2.) The dam is 500 m wide, and the water is 80.0 m deep at the dam. (a) What is the average pressure on the dam due to the water? (b) Calculate the force exerted against the dam and compare it with the weight of water in the dam (previously found to be 1.96×1013N).

Strategy for (a)

The average pressure $\overline{P}$ due to the weight of the water is the pressure at the average depth $\bar{h}$ of 40.0 m, since pressure increases linearly with depth.

Solution for (a)

The average pressure due to the weight of a fluid is

$$\overline{P} = \rho g \bar{h}.$$

Entering the density of water from Table 8.2.1 and taking $\bar{h}$ to be the average depth of 40.0 m, we obtain

$$\begin{aligned} \overline{P} &= \left(10^3 \frac{\text{kg}}{\text{m}^3}\right)\left(9.80 \frac{\text{m}}{\text{s}^2}\right)(40.0\ \text{m}) \\ &= 3.92 \times 10^5 \frac{\text{N}}{\text{m}^2} = 392\text{kPa}. \end{aligned}$$

Strategy for (b)

The force exerted on the dam by the water is the average pressure times the area of contact:

$$F = \overline{P}A$$

Solution for (b)

We have already found the value for $\overline{P}$. The area of the dam is $A = 80.0\ \text{m} \times 500\ \text{m} = 4.00 \times 10^4\ \text{m}^2$, so that

$$\begin{aligned} F &= \left(3.92 \times 10^5\ \text{N/m}^2\right)\left(4.00 \times 10^4\ \text{m}^2\right) \\ &= 1.57 \times 10^{10}\ \text{N}. \end{aligned}$$

Discussion

Although this force seems large, it is small compared with the 1.96×10^{13} N weight of the water in the reservoir—in fact, it is only 0.0800% of the weight. Note that the pressure found in part (a) is completely independent of the width and length of the lake—it depends only on its average depth at the dam. Thus the force depends only on the water's average depth and the dimensions of the dam, *not* on the horizontal extent of the reservoir. In the diagram, the thickness of the dam increases with depth to balance the increasing force due to the increasing pressure.

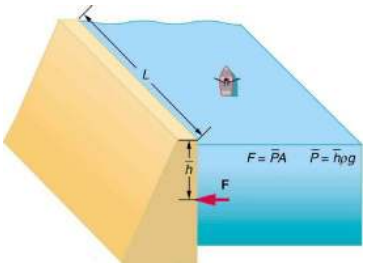


Figure 4.4.1.2: The dam must withstand the force exerted against it by the water it retains. This force is small compared with the weight of the water behind the dam.

Atmospheric pressure is another example of pressure due to the weight of a fluid, in this case due to the weight of *air* above a given height. The atmospheric pressure at the Earth's surface varies a little due to various weather patterns. However, the average pressure at sea level is given by the *standard atmospheric pressure* P_{atm}, measured to be

$$1\ \text{atmosphere (atm)} = P_{\text{atm}} = 1.01 \times 10^5\ \text{N/m}^2 = 101\ \text{kPa}$$

This relationship means that, on average, at sea level, a column of air above $1.00\ \text{m}^2$ of the Earth's surface has a weight of $1.01 \times 10^5\ \text{N}$, equivalent to 1 atm. (See Figure 4.4.1.3.) As one increases elevation from sea level, the total size of that column of air will decrease. This is why you might experience your ears popping in an airplane, or even when driving through steep hills in the mountains.

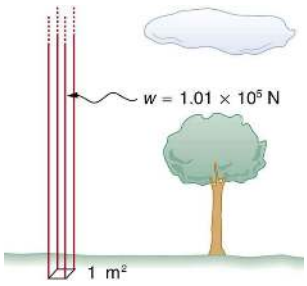


Figure 4.4.1.3: Atmospheric pressure at sea level averages $1.01 \times 10^5\ \text{Pa}$ (equivalent to 1 atm), since the column of air over this $1\ \text{m}^2$, extending to the top of the atmosphere, weighs $1.01 \times 10^5\ \text{N}$.

Example 4.4.1.2: Calculating Depth Below the Surface of Water: What Depth of Water Creates the Same Pressure as the Entire Atmosphere?

Calculate the depth below the surface of water at which the pressure due to the weight of the water equals 1.00 atm.

Strategy

We begin by solving the equation $P = \rho g h$ for depth h:

$$h = \frac{P}{\rho g}.$$

Then we take P to be 1.00 atm and ρ to be the density of the water that creates the pressure.

Solution

Entering the known values into the expression for h gives

$$h = \frac{1.01 \times 10^5\ \text{N/m}^2}{(1.00 \times 10^3\ \text{kg/m}^3)\,(9.80\ \text{m/s}^2)} = 10.3\ \text{m}.$$

Discussion

Just 10.3 m of water creates the same pressure as 120 km of air (height of uppermost layers of atmosphere). Since water is nearly incompressible, we can neglect any change in its density over this depth.

What do you suppose is the *total* pressure at a depth of 10.3 m in a swimming pool? Does the atmospheric pressure on the water's surface affect the pressure below? The answer is yes. This seems only logical, since both the water's weight and the atmosphere's weight must be supported. So the *total* pressure at a depth of 10.3 m is 2 atm—half from the water above and half from the air above.

Section Summary

- Pressure is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):

$$P = \frac{mg}{A}.$$

- Pressure due to the weight of a liquid is given by

$$P = \rho g h,$$

where P is the pressure, ρ is the density of the liquid, g is the acceleration due to gravity, and h is the height of the liquid.

Glossary

pressure due to the weight of fluid

pressure at a depth below a fluid surface due to its weight; given by $P = \rho gh$

This page titled 4.4.1: Pressure Due to the Weight of Fluid is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **7.5: Pressure Due to the Weight of Fluid** by OpenStax is licensed CC BY 4.0.

4.4.2: Boyle's Law

Each day, hundreds of weather balloons are launched. Made of a synthetic rubber and carrying a box of instruments, each helium-filled balloon rises up into the sky. As a balloon gains altitude, the atmospheric pressure becomes less and the balloon expands. At some point the balloon bursts due to the expansion; the instruments drop (aided by a parachute) to be retrieved and studied for information about the weather.

Boyle's Law

Robert Boyle (1627-1691), an English chemist, is widely considered to be one of the founders of the modern experimental science of chemistry. He reported that doubling the pressure of an enclosed sample of gas, while keeping its temperature constant, caused the volume of the gas to be reduced by half. **Boyle's law** states that the volume of a given mass of gas varies inversely with the pressure when the temperature is kept constant. An inverse relationship is described in this way: as one variable increases in value, the other variable decreases.

> **Note: Applicability of Boyle's Law**
>
> Although we are considering all fluids in this chapter, Boyle's Law only applies to gases. The molecules in a liquid are much too close together to be affected by pressure in this way.

Physically, what is happening? The gas molecules are moving and are a certain distance apart from one another. An increase in pressure pushes the molecules closer together, reducing the volume. If the pressure is decreased, the gases are free to move about in a larger volume.



Figure **4.4.2.1**: Robert Boyle. (CC BY-NC; CK-12)

Mathematically, Boyle's law can be expressed by the equation:

$$P \times V = k$$

The k is a constant for a given sample of gas and depends only on the mass of the gas and the temperature. The table below shows pressure and volume data for a set amount of gas at a constant temperature. The third column represents the value of the constant (k) for this data and is always equal to the pressure multiplied by the volume. As one of the variables changes, the other changes in such a way that the product of $P \times V$ always remains the same. In this particular case, that constant is $500\ \text{atm} \cdot \text{mL}$, as can be seen in Table **4.4.2.1**.

Table **4.4.2.1**: Pressure-Volume Data

Pressure (atm)	Volume (mL)	$P \times V = k$ (atm · mL)
0.5	1000	500
0.625	800	500
1.0	500	500
2.0	250	500
5.0	100	500
8.0	62.5	500
10.0	50	500

A graph of the data in the table further illustrates the inverse relationship nature of Boyle's Law (see figure below). Volume is plotted on the x-axis, with the corresponding pressure on the y-axis.

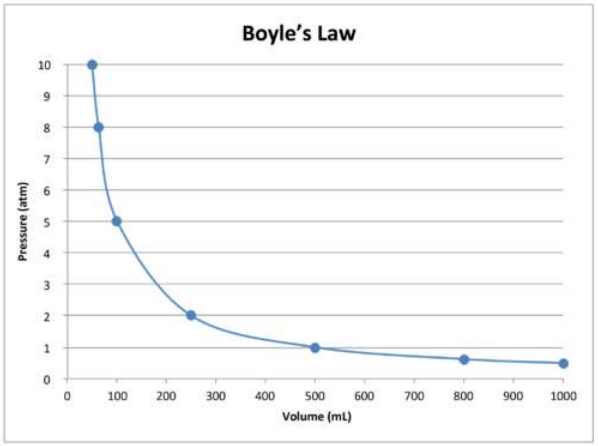


Figure 4.4.2.2: The pressure of a gas decreases as the volume increases, making Boyle's law an inverse relationship. (CC BY-NC; CK-12)

Boyle's Law can be used to compare changing conditions for a gas. We use P_1 and V_1 to stand for the initial pressure and initial volume of a gas. After a change has been made, P_2 and V_2 stand for the final pressure and volume. The mathematical relationship of Boyle's Law becomes:

$$P_1 \times V_1 = P_2 \times V_2$$

This equation can be used to calculate any one of the four quantities if the other three are known.

Example 4.4.2.1

A sample of oxygen gas has a volume of $425\ \text{mL}$ when the pressure is equal to $387\ \text{kPa}$. The gas is allowed to expand into a $1.75\ \text{L}$ container. Calculate the new pressure of the gas.

Solution

Step 1: List the known quantities and plan the problem.

Known

- $P_1 = 387\ \text{kPa}$
- $V_1 = 425\ \text{mL}$
- $V_2 = 1.75\ \text{L} = 1750\ \text{mL}$

Unknown

Use Boyle's Law to solve for the unknown pressure (P_2). It is important that the two volumes (V_1 and V_2) are expressed in the same units, so V_2 has been converted to mL.

Step 2: Solve.

First, rearrange the equation algebraically to solve for P_2.

$$P_2 = \frac{P_1 \times V_1}{V_2}$$

Now substitute the known quantities into the equation and solve.

$$P_2 = \frac{387\ \text{kPa} \times 425\ \text{mL}}{1750\ \text{mL}} = 94.0\ \text{kPa}$$

Step 3: Think about your result.

The volume has increased to slightly over 4 times its original value and so the pressure is decreased by about one fourth. The pressure is in **kPa** and the value has three significant figures. Note that any pressure or volume units can be used as long as they are consistent throughout the problem.

Section Summary

- Pressure and volume are inversely related.

This page titled 4.4.2: Boyle's Law is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **14.3: Boyle's Law** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

4.5: Fluid Principles

There are a few principles about the behavior of fluids that we will discuss in this section. We will start with the oldest and work towards the more recent. Notice as we do so that there is a level of increasing complexity as more and more properties of the fluid are considered as science progresses. Also, that sometimes relationships are discovered which are subsets of other relationships that have not yet been discovered. For example, you will see that Archimede's Principle can be viewed as an application of Newton's Laws of Motion. But Archimede's Principle precedes the formulate of Newton's Laws by nearly 2000 years! Likewise, you should be able to see how Pascal's Principle is an application of the Bernoulli Principle. There are other, more complex relationships beyond the scope of this class which incorporate more and more variables into how we understand fluid mechanics: The Bernoulli equation, which we hint at, the Euler equations, and the Navier-Stokes equations for example.

This page titled 4.5: Fluid Principles is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

4.5.1: Archimedes' Principle

Learning Objectives

- Define buoyant force.
- State Archimedes' principle.
- Understand why objects float or sink.
- Understand the relationship between density and Archimedes' principle.

When you rise from lounging in a warm bath, your arms feel strangely heavy. This is because you no longer have the buoyant support of the water. Where does this buoyant force come from? Why is it that some things float and others do not? Do objects that sink get any support at all from the fluid? Is your body buoyed by the atmosphere, or are only helium balloons affected? (See Figure **4.5.1.1**.)

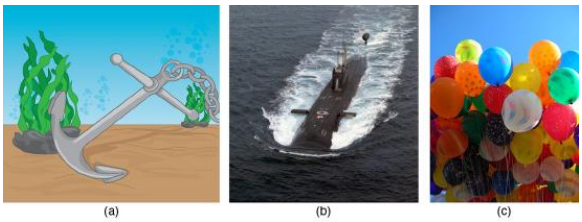


Figure **4.5.1.1**: (a) Even objects that sink, like this anchor, are partly supported by water when submerged. (b) Submarines have adjustable density (ballast tanks) so that they may float or sink as desired. (credit: Allied Navy) (c) Helium-filled balloons tug upward on their strings, demonstrating air's buoyant effect. (credit: Crystl)

Answers to all these questions, and many others, are based on the fact that pressure increases with depth in a fluid. This means that the upward force on the bottom of an object in a fluid is greater than the downward force on the top of the object. There is a net upward, or **buoyant force** on any object in any fluid. (See Figure **4.5.1.2**.) If the buoyant force is greater than the object's weight, the object will rise to the surface and float. If the buoyant force is less than the object's weight, the object will sink. If the buoyant force equals the object's weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.

Definition: BUOYANT FORCE

The buoyant force is the net upward force on any object in any fluid.

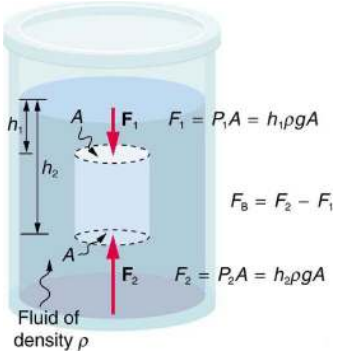


Figure **4.5.1.2**: Pressure due to the weight of a fluid increases with depth since $P = \rho gh$. This pressure and associated upward force on the bottom of the cylinder are greater than the downward force on the top of the cylinder. Their difference is the buoyant force $\mathbf{F_B}$. (Horizontal forces cancel.)

Just how great is this buoyant force? To answer this question, think about what happens when a submerged object is removed from a fluid, as in Figure **4.5.1.3**.

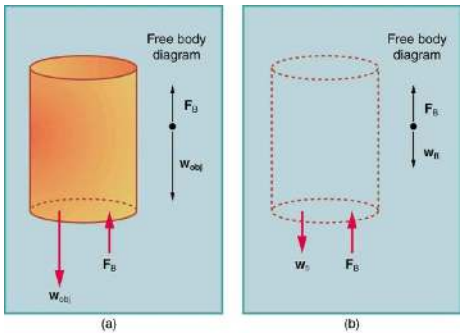


Figure **4.5.1.3**: (a) An object submerged in a fluid experiences a buoyant force $\mathbf{F_B}$. If $\mathbf{F_B}$ is greater than the weight of the object, the object will rise. If $\mathbf{F_B}$ is less than the weight of the object, the object will sink. (b) If the object is removed, it is replaced by fluid having weight w_{fl}. Since this weight is supported by surrounding fluid, the buoyant force must equal the weight of the fluid displaced. That is, $F_B = w_{fl}$,a statement of Archimedes' principle.

The space it occupied is filled by fluid having a weight w_{fl}. This weight is supported by the surrounding fluid, and so the buoyant force must equal w_{fl}, the weight of the fluid displaced by the object. It is a tribute to the genius of the Greek mathematician and inventor Archimedes (ca. 287–212 B.C.E.) that he stated this principle long before concepts of force were well established. Stated in words, **Archimedes' principle** is as follows: The buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

$$F_B = w_{fl},$$

where $\mathbf{F_B}$ is the buoyant force and w_{fl} is the weight of the fluid displaced by the object. Archimedes' principle is valid in general, for any object in any fluid, whether partially or totally submerged.

Definition: ARCHIMEDES' PRINCIPLE

According to this principle the buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

$$F_B = w_{fl},$$

where $\mathbf{F_B}$ is the buoyant force and w_{fl} is the weight of the fluid displaced by the object.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION

The density of aluminum foil is 2.7 times the density of water. Take a piece of foil, roll it up into a ball and drop it into water. Does it sink? Why or why not? Can you make it sink?

Floating and Sinking

Drop a lump of clay in water. It will sink. Mold the lump of clay into the shape of a boat and it will float. Because of its shape, the boat displaces more water than the lump and experiences a greater buoyant force. The same is true of steel ships.

Example **4.5.1.1**: Calculating buoyant force: dependency on shape

(a) Calculate the buoyant force on 10,000 metric tons ($\mathbf{1.00 \times 10^7\ kg}$) of solid steel completely submerged in water, and compare this with the steel's weight. (b) What is the maximum buoyant force that water could exert on this same steel if it were shaped into a boat that could displace $\mathbf{1.00 \times 10^5\ m^3}$ of water?

Strategy for (a)

To find the buoyant force, we must find the weight of water displaced. We can do this by using the densities of water and steel given earlier (see "Density"). We note that, since the steel is completely submerged, its volume and the water's volume are the same. Once we know the volume of water, we can find its mass and weight.

Solution for (a)

First, we use the definition of density $\rho = \frac{m}{V}$ to find the steel's volume, and then we substitute values for mass and density. This gives

$$V_{st} = \frac{m_{st}}{\rho_{st}} = \frac{1.00 \times 10^7 \text{ kg}}{7.8 \times 10^3 \text{ kg/m}^3} = 1.28 \times 10^3 \text{ m}^3.$$

Because the steel is completely submerged, this is also the volume of water displaced, V_w. We can now find the mass of water displaced from the relationship between its volume and density, both of which are known. This gives

$$\begin{aligned} m_w &= \rho_w V_w = \left(1.000 \times 10^3 \text{ kg/m}^3\right)\left(1.28 \times 10^3 \text{ m}^3\right) \\ &= 1.28 \times 10^6 \text{ kg} \end{aligned}.$$

By Archimedes' principle, the weight of water displaced is $m_w g$, so the buoyant force is

$$\begin{aligned} F_B = w_w = m_w g &= \left(1.28 \times 10^6 \text{ kg}\right)\left(9.80 \text{ m/s}^2\right) \\ &= 1.3 \times 10^7 \text{ N} \end{aligned}.$$

The steel's weight is $m_w g = 9.80 \times 10^7$ N, which is much greater than the buoyant force, so the steel will remain submerged. Note that the buoyant force is rounded to two digits because the density of steel is given to only two digits.

Strategy for (b)

Here we are given the maximum volume of water the steel boat can displace. The buoyant force is the weight of this volume of water.

Solution for (b)

The mass of water displaced is found from its relationship to density and volume, both of which are known. That is,

$$\begin{aligned} m_w = \rho_w V_w &= \left(1.000 \times 10^3 \text{ kg/m}^3\right)\left(1.00 \times 10^5 \text{ m}^3\right) \\ &= 1.00 \times 10^8 \text{ kg} \end{aligned}.$$

The maximum buoyant force is the weight of this much water, or

$$\begin{aligned} F_B = w_w = m_w g &= \left(1.00 \times 10^8 \text{ kg}\right)\left(9.80 \text{ m/s}^2\right) \\ &= 9.80 \times 10^8 \text{ N} \end{aligned}.$$

Discussion

The maximum buoyant force is ten times the weight of the steel, meaning the ship can carry a load nine times its own weight without sinking.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION

A piece of household aluminum foil is 0.016 mm thick. Use a piece of foil that measures 10 cm by 15 cm. (a) What is the mass of this amount of foil? (b) If the foil is folded to give it four sides, and paper clips or washers are added to this "boat," what shape of the boat would allow it to hold the most "cargo" when placed in water? Test your prediction.

Density and Archimedes' Principle

Density plays a crucial role in Archimedes' principle. The average density of an object is what ultimately determines whether it floats. If its average density is less than that of the surrounding fluid, it will float. This is because the fluid, having a higher density, contains more mass and hence more weight in the same volume. The buoyant force, which equals the weight of the fluid displaced, is thus greater than the weight of the object. Likewise, an object denser than the fluid will sink.

The extent to which a floating object is submerged depends on how the object's density is related to that of the fluid. In Figure 4.5.1.4, for example, the unloaded ship has a lower density and less of it is submerged compared with the same ship loaded. We can derive a quantitative expression for the fraction submerged by considering density. The fraction submerged is the ratio of the volume submerged to the volume of the object, or

$$\text{fraction submerged} = \frac{V_{sub}}{V_{obj}} = \frac{V_{fl}}{V_{obj}}.$$

The volume submerged equals the volume of fluid displaced, which we call V_{fl}. Now we can obtain the relationship between the densities by substituting $\rho = \frac{m}{V}$ into the expression. This gives

$$\frac{V_{\text{fl}}}{V_{\text{obj}}} = \frac{m_{\text{fl}}/\rho_{\text{fl}}}{m_{\text{obj}}/\bar{\rho}_{\text{obj}}},$$

where $\bar{\rho}_{\text{obj}}$ is the average density of the object and ρ_{fl} is the density of the fluid. Since the object floats, its mass and that of the displaced fluid are equal, and so they cancel from the equation, leaving

$$\text{fraction submerged} = \frac{\bar{\rho}_{\text{obj}}}{\rho_{\text{fl}}}.$$

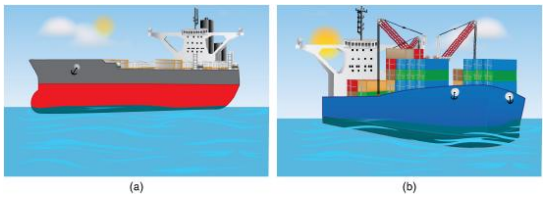


Figure 4.5.1.4: An unloaded ship (a) floats higher in the water than a loaded ship (b).

We use this last relationship to measure densities. This is done by measuring the fraction of a floating object that is submerged—for example, with a hydrometer. It is useful to define the ratio of the density of an object to a fluid (usually water) as **specific gravity**:

$$\text{specific gravity} = \frac{\bar{\rho}}{\rho_{\text{w}}},$$

where $\bar{\rho}$ is the average density of the object or substance and ρ_{w} is the density of water at 4.00°C. Specific gravity is dimensionless, independent of whatever units are used for ρ. If an object floats, its specific gravity is less than one. If it sinks, its specific gravity is greater than one. Moreover, the fraction of a floating object that is submerged equals its specific gravity. If an object's specific gravity is exactly 1, then it will remain suspended in the fluid, neither sinking nor floating. Scuba divers try to obtain this state so that they can hover in the water. We measure the specific gravity of fluids, such as battery acid, radiator fluid, and urine, as an indicator of their condition. One device for measuring specific gravity is shown in Figure 4.5.1.5.

Definition: SPECIFIC GRAVITY

Specific gravity is the ratio of the density of an object to a fluid (usually water).

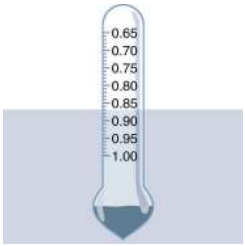


Figure 4.5.1.5: This hydrometer is floating in a fluid of specific gravity 0.87. The glass hydrometer is filled with air and weighted with lead at the bottom. It floats highest in the densest fluids and has been calibrated and labeled so that specific gravity can be read from it directly.

Example 4.5.1.2: Calculating Average Density: Floating Woman

Suppose a 60.0-kg woman floats in freshwater with **97.0%** of her volume submerged when her lungs are full of air. What is her average density?

Strategy

We can find the woman's density by solving the equation

$$\text{fraction submerged} = \frac{\bar{\rho}_{\text{obj}}}{\rho_{\text{fl}}}$$

for the density of the object. This yields

$$\bar{\rho}_{\text{obj}} = \bar{\rho}_{\text{person}} = (\text{ fraction submerged }) \cdot \rho_{\text{fl}}.$$

We know both the fraction submerged and the density of water, and so we can calculate the woman's density.

Solution

Entering the known values into the expression for her density, we obtain

$$\bar{\rho}_{\text{person}} = 0.970 \cdot \left(10^3 \frac{\text{kg}}{\text{m}^3}\right) = 970 \frac{\text{kg}}{\text{m}^3}.$$

Discussion

Her density is less than the fluid density. We expect this because she floats. Body density is one indicator of a person's percent body fat, of interest in medical diagnostics and athletic training. (See Figure 4.5.1.6.)



Figure 4.5.1.6: Subject in a "fat tank," where he is weighed while completely submerged as part of a body density determination. The subject must completely empty his lungs and hold a metal weight in order to sink. Corrections are made for the residual air in his lungs (measured separately) and the metal weight. His corrected submerged weight, his weight in air, and pinch tests of strategic fatty areas are used to calculate his percent body fat.

There are many obvious examples of lower-density objects or substances floating in higher-density fluids—oil on water, a bit of cork in wine, an iceberg, and hot wax in a "lava lamp," to name a few. This same principle also applies to samples of gas that are lighter than the surrounding atmosphere and float through the air—a hot-air balloon, helium ballons, and hydrogen balloons. Less obvious examples include lava rising in a volcano and mountain ranges floating on the higher-density crust and mantle beneath them. Even seemingly solid Earth has fluid characteristics.

More Density Measurements

One of the most common techniques for determining density is shown in Figure 4.5.1.7.

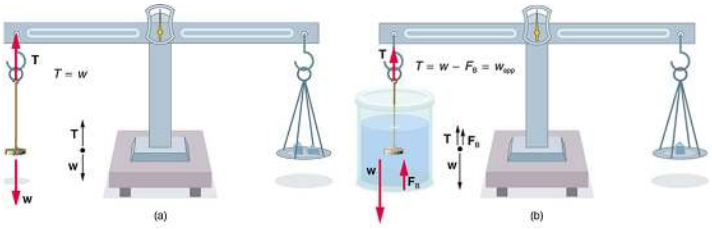


Figure 4.5.1.7: (a) A coin is weighed in air. (b) The apparent weight of the coin is determined while it is completely submerged in a fluid of known density. These two measurements are used to calculate the density of the coin.

An object, here a coin, is weighed in air and then weighed again while submerged in a liquid. The density of the coin, an indication of its authenticity, can be calculated if the fluid density is known. This same technique can also be used to determine the density of the fluid if the density of the coin is known. All of these calculations are based on Archimedes' principle.

Archimedes' principle states that the buoyant force on the object equals the weight of the fluid displaced. This, in turn, means that the object *appears* to weigh less when submerged; we call this measurement the object's *apparent weight*. The object suffers an *apparent weight loss* equal to the weight of the fluid displaced. Alternatively, on balances that measure mass, the object suffers an *apparent mass loss* equal to the mass of fluid displaced. That is

$$\text{apparent weight loss} = \text{weight of fluid displaced}$$

or

$$\text{apparent mass loss} = \text{mass of fluid displaced.}$$

The next example illustrates the use of this technique.

Example 4.5.1.3: Calculating Density: Is the Coin Authentic?

The mass of an ancient Greek coin is determined in air to be 8.630 g. When the coin is submerged in water as shown in Figure 4.5.1.7, its apparent mass is 7.800 g. Calculate its density, given that water has a density of $1.000\ \text{g/cm}^3$ and that effects caused by the wire suspending the coin are negligible.

Strategy

To calculate the coin's density, we need its mass (which is given) and its volume. The volume of the coin equals the volume of water displaced. The volume of water displaced V_w can be found by solving the equation for density $\rho = \frac{m}{V}$ for V.

Solution

The volume of water is $V_w = \frac{m_w}{\rho_w}$ where m_w is the mass of water displaced. As noted, the mass of the water displaced equals the apparent mass loss, which is $m_w = 8.630\ \text{g} - 7.800\ \text{g} = 0.830\ \text{g}$. Thus the volume of water is $V_w = \frac{0.830\ \text{g}}{1.000\ \text{g/cm}^3} = 0.830\ \text{cm}^3$. This is also the volume of the coin, since it is completely submerged. We can now find the density of the coin using the definition of density:

$$\rho_c = \frac{m_c}{V_c} = \frac{8.630\ \text{g}}{0.830\ \text{cm}^3} = 10.4\ \text{g/cm}^3.$$

Discussion

You see that this density is very close to that of pure silver, appropriate for this type of ancient coin. Most modern counterfeits are not pure silver.

This brings us back to Archimedes' principle and how it came into being. As the story goes, the king of Syracuse gave Archimedes the task of determining whether the royal crown maker was supplying a crown of pure gold. The purity of gold is difficult to determine by color (it can be diluted with other metals and still look as yellow as pure gold), and other analytical techniques had not yet been conceived. Even ancient peoples, however, realized that the density of gold was greater than that of any other then-known substance. Archimedes purportedly agonized over his task and had his inspiration one day while at the public baths, pondering the support the water gave his body. He came up with his now-famous principle, saw how to apply it to determine density, and ran naked down the streets of Syracuse crying "Eureka!" (Greek for "I have found it"). Similar behavior can be observed in contemporary physicists from time to time!

Section Summary

- Buoyant force is the net upward force on any object in any fluid. If the buoyant force is greater than the object's weight, the object will rise to the surface and float. If the buoyant force is less than the object's weight, the object will sink. If the buoyant force equals the object's weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.
- Archimedes' principle states that the buoyant force on an object equals the weight of the fluid it displaces.
- Specific gravity is the ratio of the density of an object to a fluid (usually water).

Glossary

Archimedes' principle

the buoyant force on an object equals the weight of the fluid it displaces

buoyant force

the net upward force on any object in any fluid

specific gravity

the ratio of the density of an object to a fluid (usually water)

This page titled 4.5.1: Archimedes' Principle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **7.6: Archimedes' Principle** by OpenStax is licensed CC BY 4.0.

4.5.2: Pascal's Principle

Learning Objectives

- Apply Pascal's Principle.

Most automobile shops are able to lift cars up using a hydraulic jack. Many manufacturing processes are dependent upon the hydraulic press. Both of these types of operation are based on pressure relationships within an incompressible fluid at rest. (By at rest, we mean that the fluid does not have a velocity.)

Blaise Pascal was a 17th century French mathematician, scientist, and theologian. He discovered the relationship that these devices are based upon, and which we now call Pascal's Principle.

Definition: PASCAL'S PRINCIPLE

A change in pressure at any point in an enclosed incompressible fluid at rest is transmitted equally and undiminished to all points within that fluid.

If we apply Pascal's principle to a U-tube filled with a fluid, we can see how it is a useful concept. Each of the openings of the U-tube will experience the same pressure. If we add a force to one end of the U-tube, the pressure exerted by that force is dependent upon the area of the opening. If both openings have the same area, then an equal force will be exerted by the fluid at the other opening. If, however, the size of the openings are different, the area will be different, and therefore the force will be different. What remains the same is the pressure.

This principle can be very useful in multiplying the effect of a force. If the area of one side of the opening is half that of the other side, then a force exerted on the smaller area will be doubled on the other side. Similarly, if the difference in areas is a factor of 10, then the difference in forces will also be a factor of ten. This is the principle upon which the hydraulic press and the hydraulic lift are based. In a later chapter we will discuss simple machines and see that the effect of Pascal's principle is very similar to the effect of a simple machine. In many cases, Pascal's principle seems to be more easily applied in our modern technology than the simple machines we will discuss later.

This page titled 4.5.2: Pascal's Principle is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

4.5.3: Bernoulli's Principle

Learning Objectives

- Calculate with Bernoulli's principle.
- List some applications of Bernoulli's principle.

Although Pascal's principle is very useful in many applications, there are many more situations where it doesn't apply because the fluids are moving. There are a number of common examples of pressure dropping in rapidly-moving fluids. Shower curtains have a disagreeable habit of bulging into the shower stall when the shower is on. The high-velocity stream of water and air creates a region of lower pressure inside the shower, and standard atmospheric pressure on the other side. The pressure difference results in a net force inward pushing the curtain in. You may also have noticed that when passing a truck on the highway, your car tends to veer toward it. The reason is the same—the high velocity of the air between the car and the truck creates a region of lower pressure, and the vehicles are pushed together by greater pressure on the outside. (See Figure 4.5.3.1.) This effect was observed as far back as the mid-1800s, when it was found that trains passing in opposite directions tipped precariously toward one another.

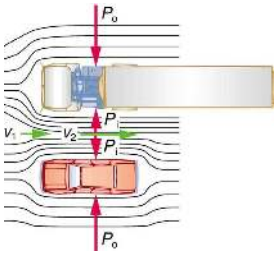


Figure 4.5.3.1: An overhead view of a car passing a truck on a highway. Air passing between the vehicles flows in a narrower channel and must increase its speed (v_2 is greater than v_1), causing the pressure between them to drop (P_i is less than P_o). Greater pressure on the outside pushes the car and truck together.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION WITH A SHEET OF PAPER

Hold the short edge of a sheet of paper parallel to your mouth with one hand on each side of your mouth. The page should slant downward over your hands. Blow over the top of the page. Describe what happens and explain the reason for this behavior.

The relationship between pressure and velocity in fluids is described quantitatively by **Bernoulli's equation**, named after its discoverer, the Swiss scientist Daniel Bernoulli (1700–1782). Bernoulli's equation applies for any incompressible, frictionless fluid. The entire equation is beyond the scope of this course and requires an understanding of concepts introduced in chapter 5. However, for most of the applications we will want to consider, we can use a simplified version of this equation that assumes no change in elevation for the moving fluid. This is called **Bernoulli's principle.**

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2.$$

As we have just discussed, pressure drops as speed increases in a moving fluid. We can see this from Bernoulli's principle. For example, if v_2 is greater than v_1 in the equation, then P_2 must be less than P_1 for the equality to hold.

Example 4.5.3.1: Calculating Pressure: Pressure Drops as a Fluid Speeds Up

In Example 8.6.2, we found that the speed of water in a hose increased from 1.96 m/s to 25.5 m/s going from the hose to the nozzle. Calculate the pressure in the hose, given that the absolute pressure in the nozzle is $1.01 \times 10^5\ \mathrm{N/m^2}$ (atmospheric, as it must be) and assuming level, frictionless flow.

Strategy

Level flow means constant depth, so Bernoulli's principle applies. We use the subscript 1 for values in the hose and 2 for those in the nozzle. We are thus asked to find P_1.

Solution

Solving Bernoulli's principle for P_1 yields

$$P_1 = P_2 + \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho\left(v_2^2 - v_1^2\right).$$

Substituting known values,

$$\begin{aligned} P_1 =& 1.01 \times 10^5 \text{ N/m}^2 \\ &+ \frac{1}{2}\left(10^3 \text{ kg/m}^3\right)\left[(25.5 \text{ m/s})^2 - (1.96 \text{ m/s})^2\right] \\ =& 4.24 \times 10^5 \text{ N/m}^2. \end{aligned}$$

Discussion

This absolute pressure in the hose is greater than in the nozzle, as expected since v is greater in the nozzle. The pressure P_2 in the nozzle must be atmospheric since it emerges into the atmosphere without other changes in conditions.

Applications of Bernoulli's Principle

There are a number of devices and situations in which fluid flows at a constant height and, thus, can be analyzed with Bernoulli's principle.

Entrainment

People have long put the Bernoulli principle to work by using reduced pressure in high-velocity fluids to move things about. With a higher pressure on the outside, the high-velocity fluid forces other fluids into the stream. This process is called *entrainment*. Entrainment devices have been in use since ancient times, particularly as pumps to raise water small heights, as in draining swamps, fields, or other low-lying areas. Some other devices that use the concept of entrainment are shown in Figure 4.5.3.2.

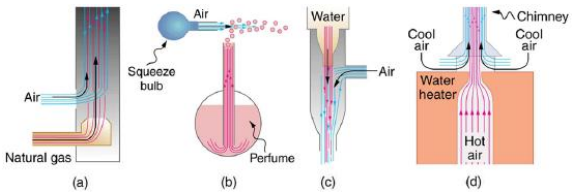


Figure 4.5.3.2: Examples of entrainment devices that use increased fluid speed to create low pressures, which then entrain one fluid into another. (a) A Bunsen burner uses an adjustable gas nozzle, entraining air for proper combustion. (b) An atomizer uses a squeeze bulb to create a jet of air that entrains drops of perfume. Paint sprayers and carburetors use very similar techniques to move their respective liquids. (c) A common aspirator uses a high-speed stream of water to create a region of lower pressure. Aspirators may be used as suction pumps in dental and surgical situations or for draining a flooded basement or producing a reduced pressure in a vessel. (d) The chimney of a water heater is designed to entrain air into the pipe leading through the ceiling.

Wings and Sails

The airplane wing is a beautiful example of Bernoulli's principle in action. Figure 4.5.3.3(a) shows the characteristic shape of a wing. The wing is tilted upward at a small angle and the upper surface is longer, causing air to flow faster over it. The pressure on top of the wing is therefore reduced, creating a net upward force or lift. (Wings can also gain lift by pushing air downward, utilizing the conservation of momentum principle. The deflected air molecules result in an upward force on the wing — Newton's third law.) Sails also have the characteristic shape of a wing. (See Figure 4.5.3.3(b).) The pressure on the front side of the sail, P_{front}, is lower than the pressure on the back of the sail, P_{back}. This results in a forward force and even allows you to sail into the wind.

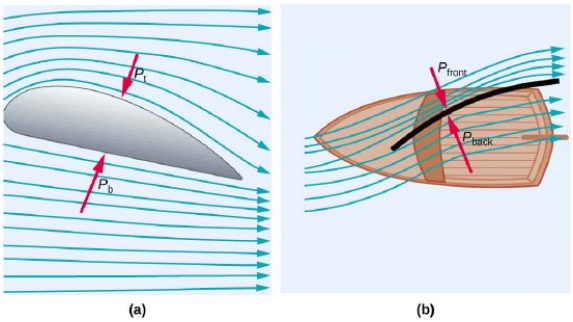


Figure **4.5.3.3**: (a) The Bernoulli principle helps explain lift generated by a wing. (b) Sails use the same technique to generate part of their thrust.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION WITH TWO STRIPS OF PAPER

For a good illustration of Bernoulli's principle, make two strips of paper, each about 15 cm long and 4 cm wide. Hold the small end of one strip up to your lips and let it drape over your finger. Blow across the paper. What happens? Now hold two strips of paper up to your lips, separated by your fingers. Blow between the strips. What happens?

Velocity measurement

Figure **4.5.3.4** shows two devices that measure fluid velocity based on Bernoulli's principle. The manometer in Figure **4.5.3.4**(a) is connected to two tubes that are small enough not to appreciably disturb the flow. The tube facing the oncoming fluid creates a dead spot having zero velocity ($v_1 = 0$) in front of it, while fluid passing the other tube has velocity v_2. This means that Bernoulli's principle as stated in $P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2$ becomes

$$P_1 = P_2 + \frac{1}{2}\rho v_2^2.$$

Thus, pressure P_2 over the second opening is reduced by $\frac{1}{2}\rho v_2^2$, and so the fluid in the manometer rises by h on the side connected to the second opening, where

$$h \propto \frac{1}{2}\rho v_2^2.$$

(Recall that the symbol $\propto$ means "proportional to.") Solving for v_2, we see that

$$v_2 \propto \sqrt{h}.$$

Figure **4.5.3.4**(b) shows a version of this device that is in common use for measuring various fluid velocities; such devices are frequently used as air speed indicators in aircraft.

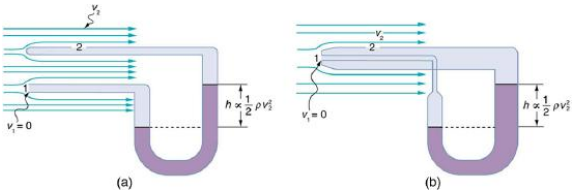


Figure **4.5.3.4**: Measurement of fluid speed based on Bernoulli's principle. (a) A manometer is connected to two tubes that are close together and small enough not to disturb the flow. Tube 1 is open at the end facing the flow. A dead spot having zero speed is created there. Tube 2 has an opening on the side, and so the fluid has a speed v across the opening; thus, pressure there drops. The difference in pressure at the manometer is $\frac{1}{2}\rho v_2^2$, and so h is proportional to $\frac{1}{2}\rho v_2^2$. (b) This type of velocity measuring device is a Prandtl tube, also known as a pitot tube.

Section Summary

- Bernoulli's principle is Bernoulli's equation applied to situations in which depth is constant. The terms involving depth (or height h) subtract out, yielding

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2.$$

- Bernoulli’s principle has many applications, including entrainment, wings and sails, and velocity measurement.

Glossary

Bernoulli’s principle

Bernoulli’s equation applied at constant depth: $P_1 + 1/2pv_1^2 = P_2 + 1/2pv_2^2$

This page titled 4.5.3: Bernoulli’s Principle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **7.8: Bernoulli’s Equation** by OpenStax is licensed CC BY 4.0.

4.6: Wave Mechanics



Figure **4.6.1**: Waves in the ocean behave similarly to all other types of waves. (credit: Steve Jurveston, Flickr)

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

In the following subsections we will discuss some of the underlying principles for oscillation, which are the basis for understanding the mechanics of how waves move. In later chapters we will learn more about the driving forces for a variety of types of waves.

This page titled 4.6: Wave Mechanics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.5: Waves** by OpenStax is licensed CC BY 4.0.

4.6.1: Period and Frequency in Oscillations

Learning Objectives

- Understand the relationship between the frequency and the period of oscillations.
- Determine the frequency of oscillations.



Figure **4.6.1.1**: The strings on this guitar vibrate at regular time intervals. (credit: JAR)

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

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

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

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

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

Example 4.6.1.1: Determine the Frequency of Two Oscillations: Medical Ultrasound and the Period of Middle C

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

Strategy

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

Solution a

1. Substitute 0.400 μs for T in $f = \frac{1}{T}$:

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

Solve to find

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

Discussion a

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

Solution b

1. Identify the known values:

   The time for one complete oscillation is the period T:

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

2. Solve for T:

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

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

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

Discussion

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

Exercise 4.6.1.1

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

Answer

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

Section Summary

- Periodic motion is a repetitious oscillation.
- The time for one oscillation is the period T.
- The number of oscillations per unit time is the frequency f.
- These quantities are related by

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

Glossary

period
time it takes to complete one oscillation

periodic motion
motion that repeats itself at regular time intervals

frequency
number of events per unit of time

This page titled 4.6.1: Period and Frequency in Oscillations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.2: Period and Frequency in Oscillations** by OpenStax is licensed CC BY 4.0.

4.6.2: Simple Harmonic Motion- A Special Periodic Motion

Learning Objectives

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

The oscillations of a system in which the net force can be described by Hooke's law are of special importance, because they are very common. They are also the simplest oscillatory systems. **Simple Harmonic Motion** (SHM) is the name given to oscillatory motion for a system where the net force can be described by Hooke's law, and such a system is called a **simple harmonic oscillator**. If the net force can be described by Hooke's law and there is no *damping* (by friction or other non-conservative forces), then a simple harmonic oscillator will oscillate with equal displacement on either side of the equilibrium position, as shown for an object on a spring in Figure **4.6.2.1**. The maximum displacement from equilibrium is called the **amplitude** X. The units for amplitude and displacement are the same, but depend on the type of oscillation. For the object on the spring, the units of amplitude and displacement are meters; whereas for sound oscillations, they have units of pressure (and other types of oscillations have yet other units). Because amplitude is the maximum displacement, it is related to the energy in the oscillation.

TAKE-HOME EXPERIMENT: SHM AND THE MARBLE

Find a bowl or basin that is shaped like a hemisphere on the inside. Place a marble inside the bowl and tilt the bowl periodically so the marble rolls from the bottom of the bowl to equally high points on the sides of the bowl. Get a feel for the force required to maintain this periodic motion. What is the restoring force and what role does the force you apply play in the simple harmonic motion (SHM) of the marble?

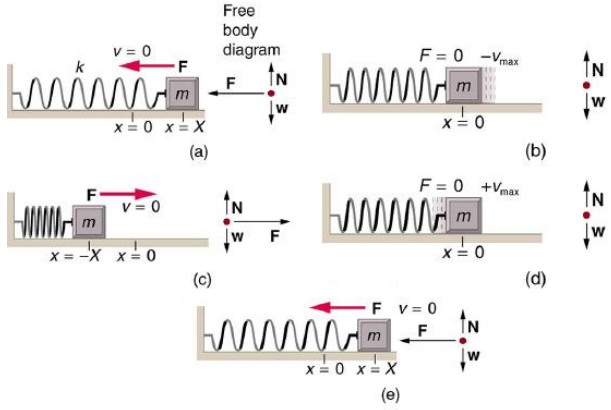


Figure **4.6.2.1**: An object attached to a spring sliding on a frictionless surface is an uncomplicated simple harmonic oscillator. When displaced from equilibrium, the object performs simple harmonic motion that has an amplitude X and a period T. The object's maximum speed occurs as it passes through equilibrium. The stiffer the spring is, the smaller the period T. The greater the mass of the object is, the greater the period T.

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

The Link between Simple Harmonic Motion and Waves

If a time-exposure photograph of the bouncing car were taken as it drove by, the headlight would make a wavelike streak, as shown in Figure **4.6.2.2**. Similarly, Figure **4.6.2.3** shows an object bouncing on a spring as it leaves a wavelike trace of its position on a moving strip of paper. Both waves are sine functions. All simple harmonic motion is intimately related to sine and cosine waves.

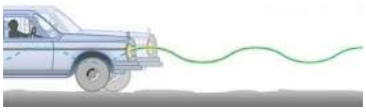


Figure **4.6.2.2**: The bouncing car makes a wavelike motion. If the restoring force in the suspension system can be described only by Hooke's law, then the wave is a sine function. (The wave is the trace produced by the headlight as the car moves to the right.)

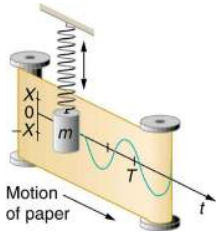


Figure **4.6.2.3**: The vertical position of an object bouncing on a spring is recorded on a strip of moving paper, leaving a sine wave.

The displacement $x(t)$, the velocity $v(t)$, and the acceleration $a(t)$ of simple harmonic motion sketches out characteristic shapes (called sinusoidal functions) as a function of time. You just need to scale them horizontally and vertically, so that they have correct period and amplitude, respectively.

Figure **4.6.2.4** shows the simple harmonic motion of an object on a spring and presents graphs of $x(t)$,$v(t)$, and $a(t)$ versus time.

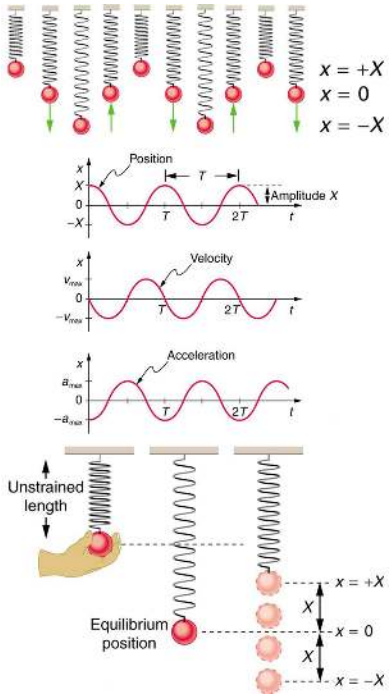


Figure **4.6.2.4**: Graphs of $x(t)$,$v(t)$, and $a(t)$ versus t for the motion of an object on a spring. The net force on the object can be described by Hooke's law, and so the object undergoes simple harmonic motion. Note that the initial position has the vertical displacement at its maximum value X; v is initially zero and then negative as the object moves down; and the initial acceleration is negative, back toward the equilibrium position and becomes zero at that point.

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

Exercise 4.6.2.1

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

Answer

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

Exercise 4.6.2.2

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

Answer

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

Section Summary

- Simple harmonic motion is oscillatory motion for a system that can be described only by Hooke's law. Such a system is also called a simple harmonic oscillator.
- Displacement, velocity, and acceleration in simple harmonic motion as a function of time are represented by sinusoidal functions in time, that is functions that are similar to $\sin(t)$ and $\cos(t)$.

Glossary

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

simple harmonic motion
the oscillatory motion in a system where the net force can be described by Hooke's law

simple harmonic oscillator
a device that implements Hooke's law, such as a mass that is attached to a spring, with the other end of the spring being connected to a rigid support such as a wall

This page titled 4.6.2: Simple Harmonic Motion- A Special Periodic Motion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.3: Simple Harmonic Motion- A Special Periodic Motion** by OpenStax is licensed CC BY 4.0.

4.6.3: Waves

Learning Objectives

- State the characteristics of a wave.
- Calculate the velocity of wave propagation, given frequency and wavelength.



Figure **4.6.3.1**: Waves in the ocean behave similarly to all other types of waves. (credit: Steve Jurveston, Flickr)

A wave is a disturbance that propagates, or moves from the place it was created. The simplest waves repeat themselves for several cycles and are associated with simple harmonic motion. Let us start by considering the simplified water wave in Figure **4.6.3.2**. The wave is an up and down disturbance of the water surface. It causes a sea gull to move up and down in simple harmonic motion as the wave crests and troughs (peaks and valleys) pass under the bird. The time for one complete up and down motion is the wave's period T. The wave's frequency is $f = 1/T$, as usual. The wave itself moves to the right in the figure. This movement of the wave is actually the disturbance moving to the right, not the water itself (or the bird would move to the right). We define **wave velocity** v_w to be the speed at which the disturbance moves. Wave velocity is sometimes also called the *propagation velocity or propagation speed*, because the disturbance propagates from one location to another.

MISCONCEPTION ALERT

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

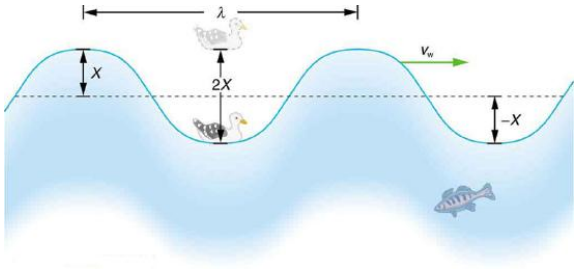


Figure **4.6.3.2**: An idealized ocean wave passes under a sea gull that bobs up and down in simple harmonic motion. The wave has a wavelength λ, which is the distance between adjacent identical parts of the wave. The up and down disturbance of the surface propagates parallel to the surface at a speed v_w.

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

$$v_w = \frac{\lambda}{T}$$

or

$$v_w = f\lambda. \tag{4.6.3.1}$$

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

One note of caution with Equation (4.6.3.1): despite the appearance, wave speed v_w is a *not* a function of frequency f and wavelength λ. When f changes (or λ changes), v_w does not change (you will see this explicitly later with sound and light waves). Wave speed is a *property of the medium* the wave travels in. Unless the medium itself changes, the wave speed remains constant. This is true of sound and light waves, as well as almost all other types of waves you will see in this textbook (water waves, waves on a string, etc.). So, what happens when frequency or wavelength does change, then? The wavelength of frequency changes to compensate. For example, if the frequency doubles for a given wave, instead of the wave speed doubling, the wavelength will decrease to half, so that the product $f\lambda$ remains constant.

TAKE-HOME EXPERIMENT: WAVES IN A BOWL

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

Calculate the Velocity of Wave Propagation: Gull in the Ocean

Calculate the wave velocity of the ocean wave in Figure **4.6.3.2** if the distance between wave crests is 10.0 m and the time for a sea gull to bob up and down is 5.00 s.

Strategy

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

Solution

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

$$v_w = \frac{10.0\ \text{m}}{5.00\ \text{s}}.$$

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

Discussion

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

Section Summary

- A wave is a disturbance that moves from the point of creation with a wave velocity v_w.
- A wave has a wavelength λ, which is the distance between adjacent identical parts of the wave.
- Wave velocity and wavelength are related to the wave's frequency and period by $v_w = \frac{\lambda}{T}$ or $v_w = f\lambda$. However, wave velocity is a property of the medium and remains constant as wavelength or frequency changes.

Glossary

wave velocity

the speed at which the disturbance moves; also called wave speed, propagation velocity, or propagation speed

wavelength

the distance between adjacent identical parts of a wave

This page titled 4.6.3: Waves is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.5: Waves** by OpenStax is licensed CC BY 4.0.

4.7: Properties of Waves

We have learned the basics of waves as a disturbance that travels through a medium, as well as the simple harmonic motion that will govern any calculations related to wave movement. In this section we will address some additional ways in which we can characterize waves. What form does the disturbance take? Are there some ways that simple harmonic motion interacts with materials in special ways? What happens when waves interact with each other? What happens when the source of the wave is itself moving relative to your frame of reference? We will explore all of these ideas in this section.

This page titled 4.7: Properties of Waves is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

4.7.1: Transverse and Longitudinal Waves

Learning Objectives

- Understand the difference between a transverse and a longitudinal wave.

A simple wave consists of a periodic disturbance that propagates from one place to another. The wave in Figure **4.7.1.3** propagates in the horizontal direction while the surface is disturbed in the vertical direction. Such a wave is called a **transverse wave** or shear wave; in such a wave, the disturbance is perpendicular to the direction of propagation. In contrast, in a **longitudinal wave** or compressional wave, the disturbance is parallel to the direction of propagation. Figure **4.7.1.4** shows an example of a longitudinal wave. The size of the disturbance is its amplitude X and is completely independent of the speed of propagation v_w.

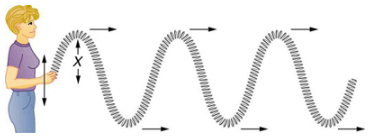


Figure **4.7.1.3**: In this example of a transverse wave, the wave propagates horizontally, and the disturbance in the cord is in the vertical direction.

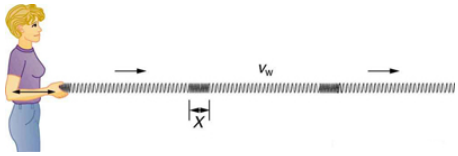


Figure **4.7.1.4**: In this example of a longitudinal wave, the wave propagates horizontally, and the disturbance in the cord is also in the horizontal direction.

Waves may be transverse, longitudinal, or *a combination of the two*. (Water waves are actually a combination of transverse and longitudinal. The simplified water wave illustrated in Figure **4.7.1.2** shows no longitudinal motion of the bird.) The waves on the strings of musical instruments are transverse—so are electromagnetic waves, such as visible light.

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

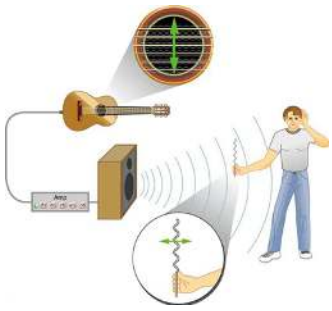


Figure **4.7.1.5**: The wave on a guitar string is transverse. The sound wave rattles a sheet of paper in a direction that shows the sound wave is longitudinal.

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

Exercise 4.7.1.1

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

Answer

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

Section Summary

- A transverse wave has a disturbance perpendicular to its direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

Glossary

longitudinal wave

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

transverse wave

a wave in which the disturbance is perpendicular to the direction of propagation

---

This page titled 4.7.1: Transverse and Longitudinal Waves is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.5: Waves** by OpenStax is licensed CC BY 4.0.

4.7.2: Forced Oscillations and Resonance

Learning Objectives

- Observe the resonance phenomena in several examples.
- Observe amplitude of a damped harmonic oscillator.
- Understand the origin of damping of resonance.



Figure **4.7.2.1**: You can cause the strings in a piano to vibrate simply by producing sound waves from your voice. (credit: Matt Billings, Flickr)

Sit in front of a piano sometime and sing a loud brief note at it with the dampers off its strings. It will sing the same note back at you—the strings, having the same frequencies as your voice, are resonating in response to the forces from the sound waves that you sent to them. Your voice and a piano's strings is a good example of the fact that objects—in this case, piano strings—can be forced to oscillate but oscillate best at their natural frequency. In this section, we shall briefly explore applying a *periodic driving force* acting on a simple harmonic oscillator. The driving force puts energy into the system at a certain frequency, not necessarily the same as the natural frequency of the system. The **natural frequency** is the frequency at which a simple harmonic oscillator oscillates, if there is no damping force, and other than an initial kick to get the motion started, there is no driving force. A **damping force** is a frictional force (friction, air resistance, and other forces that lead to decrease in mechanical energy) which converts the mechanical energy of the oscillatory motion into thermal energy.

Most of us have played with toys involving an object supported on an elastic band, something like the paddle ball suspended from a finger in Figure **4.7.2.2**. Imagine the finger in the figure is your finger. At first you hold your finger steady, and the ball bounces up and down with a small amount of damping. If you move your finger up and down slowly, the ball will follow along without bouncing much on its own. As you increase the frequency at which you move your finger up and down, the ball will respond by oscillating with increasing amplitude. When you drive the ball at its natural frequency, the ball's oscillations increase in amplitude with each oscillation for as long as you drive it. The phenomenon of driving a system with a frequency equal to its natural frequency is called **resonance**. A system being driven at its natural frequency is said to **resonate**. As the driving frequency gets progressively higher than the resonant or natural frequency, the amplitude of the oscillations becomes smaller, until the oscillations nearly disappear and your finger simply moves up and down with little effect on the ball.

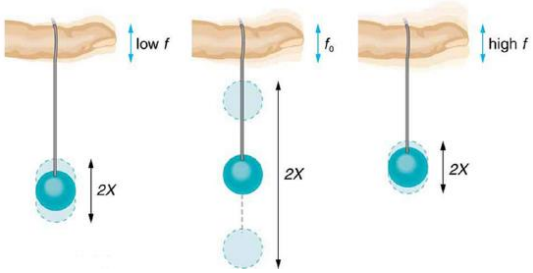


Figure **4.7.2.2**: The paddle ball on its rubber band moves in response to the finger supporting it. If the finger moves with the natural frequency f_0 of the ball on the rubber band, then a resonance is achieved, and the amplitude of the ball's oscillations increases dramatically. At higher and lower driving frequencies, energy is transferred to the ball less efficiently, and it responds with lower-amplitude oscillations.

Figure **4.7.2.3** shows a graph of the amplitude of a damped harmonic oscillator as a function of the frequency of the periodic force driving it. There are three curves on the graph, each representing a different amount of damping. All three curves peak at the point where the frequency of the driving force equals the natural frequency of the harmonic oscillator. The highest peak, or greatest response, is for the least amount of damping, because less energy is removed by the damping force.

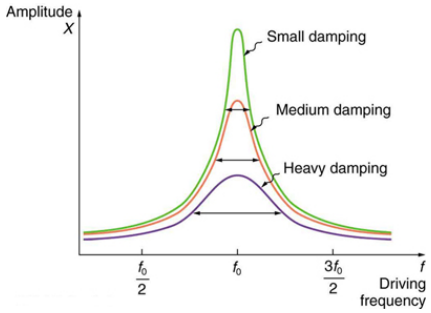


Figure **4.7.2.3**: Amplitude of a harmonic oscillator as a function of the frequency of the driving force. The curves represent the same oscillator with the same natural frequency but with different amounts of damping. Resonance occurs when the driving frequency equals the natural frequency, and the greatest response is for the least amount of damping. The narrowest response is also for the least damping.

It is interesting that the widths of the resonance curves shown in Figure **4.7.2.3** depend on damping: the less the damping, the narrower the resonance. The message is that if you want a driven oscillator to resonate at a very specific frequency, you need as little damping as possible. Little damping is the case for piano strings and many other musical instruments. Conversely, if you want small-amplitude oscillations, such as in a car's suspension system, then you want heavy damping. Heavy damping reduces the amplitude, but the tradeoff is that the system responds at more frequencies.

These features of driven harmonic oscillators apply to a huge variety of systems. When you tune a radio, for example, you are adjusting its resonant frequency so that it only oscillates to the desired station's broadcast (driving) frequency. The more selective the radio is in discriminating between stations, the smaller its damping. Magnetic resonance imaging (MRI) is a widely used medical diagnostic tool in which atomic nuclei (mostly hydrogen nuclei) are made to resonate by incoming radio waves (on the order of 100 MHz). A child on a swing is driven by a parent at the swing's natural frequency to achieve maximum amplitude. In all of these cases, the efficiency of energy transfer from the driving force into the oscillator is best at resonance. Speed bumps and gravel roads prove that even a car's suspension system is not immune to resonance. In spite of finely engineered shock absorbers, which ordinarily convert mechanical energy to thermal energy almost as fast as it comes in, speed bumps still cause a large-amplitude oscillation. On gravel roads that are corrugated, you may have noticed that if you travel at the "wrong" speed, the bumps are very noticeable whereas at other speeds you may hardly feel the bumps at all. Figure **4.7.2.4** shows a photograph of a famous example (the Tacoma Narrows Bridge) of the destructive effects of a driven harmonic oscillation. The Millennium Bridge in London was closed for a short period of time for the same reason while inspections were carried out.

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



Figure **4.7.2.4**: In 1940, the Tacoma Narrows Bridge in Washington state collapsed. Heavy cross winds drove the bridge into oscillations at its resonant frequency. Damping decreased when support cables broke loose and started to slip over the towers, allowing increasingly greater amplitudes until the structure failed (credit: PRI's *Studio 360*, via Flickr)

Exercise **4.7.2.1**

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

Answer

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

Section Summary

- A system’s natural frequency is the frequency at which the system will oscillate if not affected by driving or damping forces.
- A periodic force driving a harmonic oscillator at its natural frequency produces resonance. The system is said to resonate.
- The less damping a system has, the higher the amplitude of the forced oscillations near resonance. The more damping a system has, the broader response it has to varying driving frequencies.

Glossary

natural frequency

the frequency at which a simple harmonic oscillator oscillates if it is set in motion without a driving force

damping force

a frictional force which converts the mechanical energy of the oscillatory motion into thermal energy

resonance

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

resonate

a system being driven at its natural frequency

This page titled 4.7.2: Forced Oscillations and Resonance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.4: Forced Oscillations and Resonance** by OpenStax is licensed CC BY 4.0.

4.7.3: Wave Interference- Standing Waves and Beats

Learning Objectives

- Distinguish between constructive interference and destructive interference.
- Explain standing waves in terms of interference, nodes, and antinodes.
- Explain the beat phenomenon in terms of interference.



Figure **4.7.3.1**: These waves result from the superposition of several waves from different sources, producing a complex pattern. (credit: waterborough, Wikimedia Commons)

Most waves do not look very simple. They look more like the waves in Figure **4.7.3.1** than like the simple water wave considered in "Waves". (Simple waves may be created by a simple harmonic oscillation, and thus have a sinusoidal shape). Complex waves are more interesting, even beautiful, but they look formidable. Most waves appear complex because they result from several simple waves adding together. Luckily, the rules for adding waves are quite simple.

When two or more waves arrive at the same point, they superimpose themselves on one another. More specifically, the disturbances of waves are superimposed when they come together—a phenomenon called **superposition**. Each disturbance corresponds to a force, and forces add. If the disturbances are along the same line, then the resulting wave is a simple addition of the disturbances of the individual waves—that is, their amplitudes add. Figure **4.7.3.2** and Figure **4.7.3.3** illustrate superposition in two special cases, both of which produce simple results.

Figure **4.7.3.2** shows two identical waves that arrive at the same point exactly in phase. The crests of the two waves are precisely aligned, as are the troughs. This superposition produces pure **constructive interference**. Because the disturbances add, pure constructive interference produces a wave that has twice the amplitude of the individual waves, but has the same wavelength.

Figure **4.7.3.3** shows two identical waves that arrive exactly out of phase—that is, precisely aligned crest to trough—producing pure **destructive interference**. Because the disturbances are in the opposite direction for this superposition, the resulting amplitude is zero for pure destructive interference—the waves completely cancel.

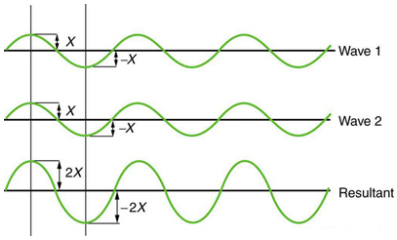


Figure **4.7.3.2**: Pure constructive interference of two identical waves produces one with twice the amplitude, but the same wavelength.

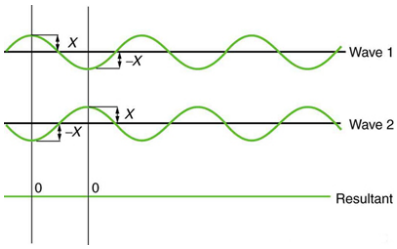


Figure **4.7.3.3**: Pure destructive interference of two identical waves produces zero amplitude, or complete cancellation.

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

An example of the superposition of two dissimilar waves is shown in Figure **4.7.3.4**. Here again, the disturbances add and subtract, producing a more complicated looking wave.

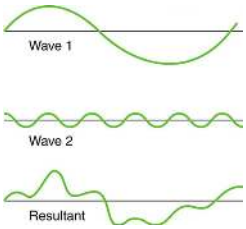


Figure **4.7.3.4**: Superposition of non-identical waves exhibits both constructive and destructive interference.

Standing Waves

Sometimes waves do not seem to move; rather, they just vibrate in place. Unmoving waves can be seen on the surface of a glass of milk in a refrigerator, for example. Vibrations from the refrigerator motor create waves on the milk that oscillate up and down but do not seem to move across the surface. These waves are formed by the superposition of two or more moving waves, such as illustrated in Figure **4.7.3.5** for two identical waves moving in opposite directions. The waves move through each other with their disturbances adding as they go by. If the two waves have the same amplitude and wavelength, then they alternate between constructive and destructive interference. The resultant looks like a wave standing in place and, thus, is called a **standing wave**. Waves on the glass of milk are one example of standing waves. There are other standing waves, such as on guitar strings and in organ pipes. With the glass of milk, the two waves that produce standing waves may come from reflections from the side of the glass.

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

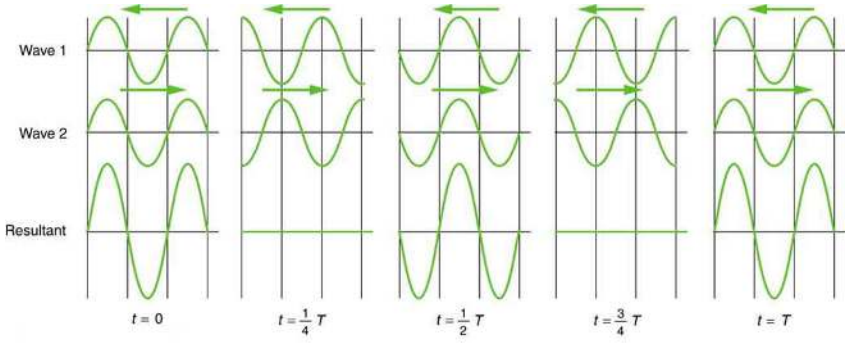


Figure **4.7.3.5**: Standing wave created by the superposition of two identical waves moving in opposite directions. The oscillations are at fixed locations in space and result from alternately constructive and destructive interference.

Standing waves are also found on the strings of musical instruments and are due to reflections of waves from the ends of the string. Figure 4.7.3.6 and Figure 4.7.3.7 show three standing waves that can be created on a string that is fixed at both ends. **Nodes** are the points where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave. The fixed ends of strings must be nodes, too, because the string cannot move there. The word **antinode** is used to denote the location of maximum amplitude in standing waves. Standing waves on strings have a frequency that is related to the propagation speed v_w of the disturbance on the string. The wavelength λ is determined by the distance between the points where the string is fixed in place.

The lowest frequency, called the **fundamental frequency**, is thus for the longest wavelength, which is seen to be $\lambda_1 = 2L$. Therefore, the fundamental frequency is $f_1 = v_w/\lambda_1 = v_w/2L$. In this case, the **overtones** or harmonics are multiples of the fundamental frequency. As seen in Figure 4.7.3.7, the first harmonic can easily be calculated since $\lambda_2 = L$. Thus, $f_2 = v_w/\lambda_2 = v_w/2L = 2f_1$. Similarly, $f_3 = 3f_1$, and so on. All of these frequencies can be changed by adjusting the tension in the string. The greater the tension, the greater v_w is and the higher the frequencies. This observation is familiar to anyone who has ever observed a string instrument being tuned. We will see in later chapters that standing waves are crucial to many resonance phenomena, such as in sounding boxes on string instruments.

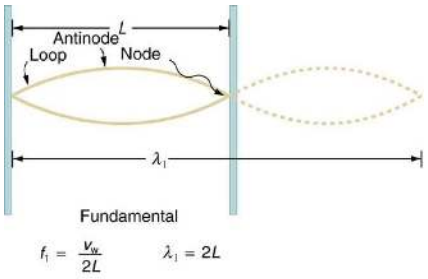


Figure 4.7.3.6: The figure shows a string oscillating at its fundamental frequency.

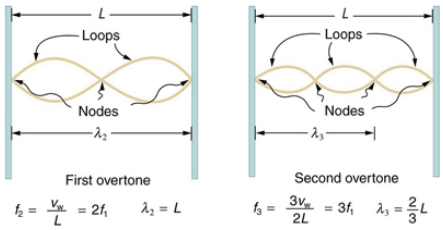


Figure 4.7.3.7: First and second harmonic frequencies are shown.

Beats

Striking two adjacent keys on a piano produces a warbling combination usually considered to be unpleasant. The superposition of two waves of similar but not identical frequencies is the culprit. Another example is often noticeable in jet aircraft, particularly the two-engine variety, while taxiing. The combined sound of the engines goes up and down in loudness. This varying loudness happens because the sound waves have similar but not identical frequencies. The discordant warbling of the piano and the fluctuating loudness of the jet engine noise are both due to alternately constructive and destructive interference as the two waves go in and out of phase. Figure 4.7.3.8 illustrates this graphically.

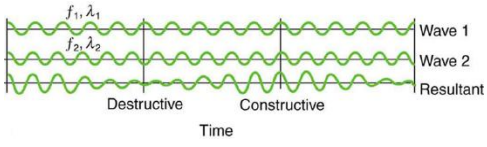


Figure 4.7.3.8: Beats are produced by the superposition of two waves of slightly different frequencies but identical amplitudes. The waves alternate in time between constructive interference and destructive interference, giving the resulting wave a time-varying amplitude.

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

$$f_B = |f_1 - f_2|$$

is the beat frequency. The resultant wave has the average frequency of the two superimposed waves, but it also fluctuates in overall amplitude at the beat frequency f_B. This result is valid for all types of waves. However, if it is a sound wave, providing the two frequencies are similar, then what we hear is an average frequency that gets louder and softer (or warbles) at the beat frequency.

MAKING CAREER CONNECTIONS

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

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

Exercise 4.7.3.1

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

Answer

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

Exercise 4.7.3.2

Define nodes and antinodes.

Answer

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

Exercise 4.7.3.3

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

Answer

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

Section Summary

- Superposition is the combination of two waves at the same location.
- Constructive interference occurs when two waves are superimposed in phase (both disturbances in the same direction).
- Destructive interference occurs when two waves are superimposed exactly out of phase (two disturbances in opposite directions).
- A standing wave is one in which two waves superimpose to produce a wave that forms a standing pattern of nodes and antinodes, locations of no displacement and of maximum displacement.

- Standing waves form at specific resonance frequencies, with the lowest frequency being the fundamental frequency, and higher frequencies being called overtones or harmonics.
- Beats occur when waves of similar frequencies f_1 and f_2 are superimposed. The resulting amplitude oscillates with a beat frequency given by

$$f_B = |f_1 - f_2| .$$

Glossary

antinode
the location of maximum amplitude in standing waves

beat frequency
the frequency of the amplitude fluctuations of a wave

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

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

fundamental frequency
the lowest frequency of a periodic waveform

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

overtones
multiples of the fundamental frequency of a sound

superposition
the phenomenon that occurs when two or more waves arrive at the same point

---

This page titled 4.7.3: Wave Interference- Standing Waves and Beats is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.6: Wave Interference- Standing Waves and Beats** by OpenStax is licensed CC BY 4.0.

4.7.4: When Waves Sources Move- Doppler effects and Bow Wakes

Learning Objectives

- Define Doppler effect, Doppler shift, and sonic boom.
- Describe the sounds produced by objects moving faster than the speed of sound.

The characteristic sound of a motorcycle buzzing by is an example of the Doppler effect. The high-pitch scream shifts dramatically to a lower-pitch roar as the motorcycle passes by a stationary observer. The closer the motorcycle brushes by, the more abrupt the shift. The faster the motorcycle moves, the greater the shift. We also hear this characteristic shift in frequency for passing race cars, airplanes, and trains. It is so familiar that it is used to imply motion and children often mimic it in play.

The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer. Although less familiar, this effect is easily noticed for a stationary source and moving observer. For example, if you ride a train past a stationary warning bell, you will hear the bell's frequency shift from high to low as you pass by. This change in frequency due to *relative* motion of source and observer is called a **Doppler shift**. The Doppler effect is named for the Austrian physicist and mathematician Christian Johann Doppler (1803–1853), who did experiments with both moving sources and moving observers. Doppler, for example, had musicians play on a moving open train car and also play standing next to the train tracks as a train passed by. Their music was observed both on and off the train, and changes in frequency were measured.

What causes the Doppler shift? Figure **4.7.4.1**, Figure **4.7.4.2**, and Figure **4.7.4.3** compare sound waves emitted by stationary and moving sources in a stationary air mass. Each disturbance spreads out spherically from the point where the sound was emitted. If the source is stationary, then all of the spheres representing the air compressions in the sound wave centered on the same point, and the stationary observers on either side see the same wavelength and frequency as emitted by the source, as in Figure **4.7.4.1**. If the source is moving, as in Figure **4.7.4.2**, then the situation is different. Each compression of the air moves out in a sphere from the point where it was emitted, but the point of emission moves. This moving emission point causes the air compressions to be closer together on one side and farther apart on the other. Thus, the wavelength is shorter in the direction the source is moving (on the right in Figure **4.7.4.2**), and longer in the opposite direction (on the left in Figure **4.7.4.2**). Finally, if the observers move, as in Figure **4.7.4.3**, the frequency at which they receive the compressions changes. The observer moving toward the source receives them at a higher frequency, and the person moving away from the source receives them at a lower frequency.



Figure **4.7.4.1**: Sounds emitted by a source spread out in spherical waves. Because the source, observers, and air are stationary, the wavelength and frequency are the same in all directions and to all observers.

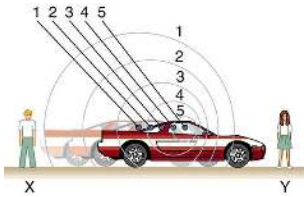


Figure **4.7.4.2**: Sounds emitted by a source moving to the right spread out from the points at which they were emitted. The wavelength is reduced and, consequently, the frequency is increased in the direction of motion, so that the observer on the right hears a higher-pitch sound. The opposite is true for the observer on the left, where the wavelength is increased and the frequency is reduced.

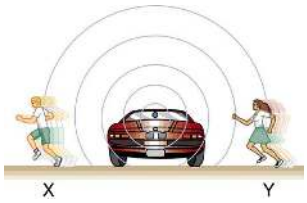


Figure 4.7.4.3: The same effect is produced when the observers move relative to the source. Motion toward the source increases frequency as the observer on the right passes through more wave crests than she would if stationary. Motion away from the source decreases frequency as the observer on the left passes through fewer wave crests than he would if stationary.

We know that wavelength and frequency are related by $v_w = f\lambda$, where v_w is the fixed speed of sound. The sound moves in a medium and has the same speed v_w in that medium whether the source is moving or not. Thus f multiplied by λ is a constant. Because the observer on the right in Figure 4.7.4.2 receives a shorter wavelength, the frequency she receives must be higher. Similarly, the observer on the left receives a longer wavelength, and hence he hears a lower frequency. The same thing happens in Figure 4.7.4.3. A higher frequency is received by the observer moving toward the source, and a lower frequency is received by an observer moving away from the source. In general, then, relative motion of source and observer toward one another increases the received frequency. Relative motion apart decreases frequency. The greater the relative speed is, the greater the effect.

THE DOPPLER EFFECT

The Doppler effect occurs not only for sound but for any wave when there is relative motion between the observer and the source. There are Doppler shifts in the frequency of sound, light, and water waves, for example. Doppler shifts can be used to determine velocity, such as when ultrasound is reflected from blood in a medical diagnostic. The recession of galaxies is determined by the shift in the frequencies of light received from them and has implied much about the origins of the universe. Modern physics has been profoundly affected by observations of Doppler shifts.

Sonic Booms to Bow Wakes

What happens to the sound produced by a moving source, such as a jet airplane, that approaches or even exceeds the speed of sound? The answer to this question applies not only to sound but to all other waves as well.

Suppose a jet airplane is coming nearly straight at you, emitting a sound of frequency f_s. The greater the plane's speed v_s, the greater the Doppler shift and the greater the value observed for f_{obs}. Now, as v_s approaches the speed of sound, f_{obs} approaches infinity. At the speed of sound, this result means that in front of the source, each successive wave is superimposed on the previous one because the source moves forward at the speed of sound. The observer gets them all at the same instant, and so the frequency is infinite. (Before airplanes exceeded the speed of sound, some people argued it would be impossible because such constructive superposition would produce pressures great enough to destroy the airplane.) If the source exceeds the speed of sound, no sound is received by the observer until the source has passed, so that the sounds from the approaching source are mixed with those from it when receding. This mixing appears messy, but something interesting happens—a sonic boom is created. (See Figure 4.7.4.4.)

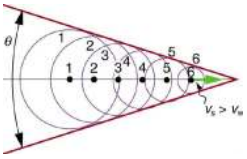


Figure 4.7.4.4: Sound waves from a source that moves faster than the speed of sound spread spherically from the point where they are emitted, but the source moves ahead of each. Constructive interference along the lines shown (actually a cone in three dimensions) creates a shock wave called a sonic boom. The faster the speed of the source, the smaller the angle θ.

There is constructive interference along the lines shown (a cone in three dimensions) from similar sound waves arriving there simultaneously. This superposition forms a disturbance called a **sonic boom**, a constructive interference of sound created by an object moving faster than sound. Inside the cone, the interference is mostly destructive, and so the sound intensity there is much less than on the shock wave. An aircraft creates two sonic booms, one from its nose and one from its tail. (See Figure 4.7.4.5.) During television coverage of space shuttle landings, two distinct booms could often be heard. These were separated by exactly the time it would take the shuttle to pass by a point. Observers on the ground often do not see the aircraft creating the sonic boom, because it has passed by before the shock wave reaches them, as seen in Figure 4.7.4.5. If the aircraft flies close by at low altitude,

pressures in the sonic boom can be destructive and break windows as well as rattle nerves. Because of how destructive sonic booms can be, supersonic flights are banned over populated areas of the United States.

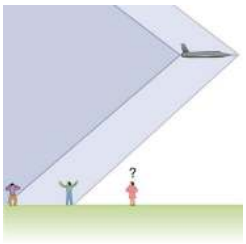


Figure 4.7.4.5: Two sonic booms, created by the nose and tail of an aircraft, are observed on the ground after the plane has passed by.

Sonic booms are one example of a broader phenomenon called bow wakes. A **bow wake**, such as the one in Figure 4.7.4.6, is created when the wave source moves faster than the wave propagation speed. Water waves spread out in circles from the point where created, and the bow wake is the familiar V-shaped wake trailing the source. A more exotic bow wake is created when a subatomic particle travels through a medium faster than the speed of light travels in that medium (the speed of light in vacuum is $c = 2.998 \times 10^8$ m/s, and this is the maximum speed anything can move, but in water, for example, speed of light slows to $0.75c$, and other particles can move faster). If the particle creates light in its passage, that light spreads on a cone with an angle indicative of the speed of the particle, as illustrated in Figure 4.7.4.7. Such a bow wake is called Cherenkov radiation and is commonly observed in particle physics.



Figure 4.7.4.6: Bow wake created by a duck. Constructive interference produces the rather structured wake, while there is relatively little wave action inside the wake, where interference is mostly destructive. (credit: Horia Varlan, Flickr)



Figure 4.7.4.7: The blue glow in this research reactor pool is Cherenkov radiation caused by subatomic particles traveling faster than the speed of light in water. (credit: U.S. Nuclear Regulatory Commission)

Doppler shifts and sonic booms are interesting sound phenomena that occur in all types of waves. They can be of considerable use. For example, the Doppler shift in ultrasound can be used to measure blood velocity, while police use the Doppler shift in radar (a microwave) to measure car velocities. In meteorology, the Doppler shift is used to track the motion of storm clouds; such “Doppler Radar” can give velocity and direction and rain or snow potential of imposing weather fronts. In astronomy, we can examine the light emitted from distant galaxies and determine their speed relative to ours. As galaxies move away from us, their light is shifted to a lower frequency, and so to a longer wavelength—the so-called red shift. Such information from galaxies far, far away has allowed us to estimate the age of the universe (from the Big Bang) as about 14 billion years.

Exercise 4.7.4.1

Why did scientist Christian Doppler observe musicians both on a moving train and also from a stationary point not on the train?

Answer

Doppler needed to compare the perception of sound when the observer is stationary and the sound source moves, as well as when the sound source and the observer are both in motion.

Exercise **4.7.4.2**

Describe a situation in your life when you might rely on the Doppler shift to help you either while driving a car or walking near traffic.

Answer

If I am driving and I hear Doppler shift in an ambulance siren, I would be able to tell when it was getting closer and also if it has passed by. This would help me to know whether I needed to pull over and let the ambulance through

Section Summary

- The Doppler shift is change in observed frequency of a wave (for example, a sound wave) due to relative motion of the source and the observer, with approaching source and/or observer increasing the observed frequency and receding source and/or observer decreasing the observed frequency.
- A sonic boom is constructive interference of sound created by an object moving faster than sound.
- A shock wave is a type of bow wake created when any wave source moves faster than the wave propagation speed.

Glossary

Doppler shift

the change in wave frequency due to relative motion of source and observer

sonic boom

a constructive interference of sound created by an object moving faster than sound

bow wake

V-shaped disturbance created when the wave source moves faster than the wave propagation speed

This page titled 4.7.4: When Waves Sources Move- Doppler effects and Bow Wakes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.9: Doppler Effect and Sonic Booms** by OpenStax is licensed CC BY 4.0.

4.8: Sound

Learning Objectives

- Define sound and hearing.
- Describe sound as a longitudinal wave.
- Distinguish between audible sound, ultrasound, and infrasound.



Figure 4.8.1: 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)

Sound can be used as a familiar illustration of waves. Because hearing is one of our most important senses, it is interesting to see how the physical properties of sound correspond to our perceptions of it. **Hearing** is the perception of sound, just as vision is the perception of visible light. But sound has important applications beyond hearing. Ultrasound, for example, is not heard but can be employed to form medical images and is also used in treatment.

The physical phenomenon of **sound** is defined to be a disturbance of matter that is transmitted from its source outward. Sound is a wave. On the atomic scale, it is a disturbance of atoms that is far more ordered than their thermal motions. In many instances, sound is a periodic wave, and the atoms undergo simple harmonic motion. In this text, we shall explore such periodic sound waves.

A vibrating string produces a sound wave as illustrated in Figure 4.8.2, Figure 4.8.3, and Figure 4.8.4. 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 4.8.4 shows a graph of gauge pressure versus distance from the vibrating string.

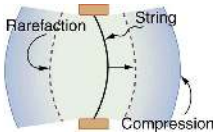


Figure 4.8.2: A vibrating string moving to the right compresses the air in front of it and expands the air behind it.

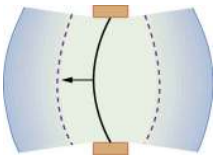


Figure 4.8.3: As the string moves to the left, it creates another compression and rarefaction as the ones on the right move away from the string.

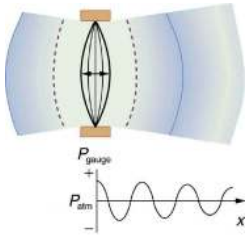


Figure **4.8.4**: After many vibrations, there are a series of compressions and rarefactions moving out from the string as a sound wave. The graph shows gauge pressure versus distance from the source. Pressures vary only slightly from atmospheric for ordinary sounds.

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 **4.8.5**, 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. 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.

When the frequency of the sound wave is in **audible range** of 20 to 20,000 Hz, the vibration of the eardrums at these frequencies can be sensed, which is what we hear. Sound waves of frequency above 20 kHz are called **ultrasound** waves. Some construction equipment produces **infrasound** waves of frequencies less than 20 Hz, which, although they can not be heard, can still affect the human body at large enough amplitudes.

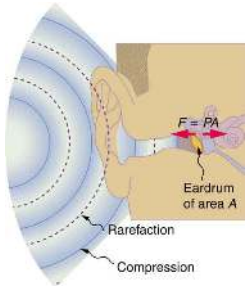


Figure **4.8.5**: Sound wave compressions and rarefactions travel up the ear canal and force the eardrum to vibrate. There is a net force on the eardrum, since the sound wave pressures differ from the atmospheric pressure found behind the eardrum. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person.

Section 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

audible range

frequencies between 20 Hz and 20,000 Hz

infrasound

sound waves below audible frequencies (less than 20 Hz)

hearing

the perception of sound

sound

a disturbance of matter that is transmitted from its source outward

ultrasound

sound waves above audible frequencies (greater than 20,000 Hz)

This page titled 4.8: Sound is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.7: Sound** by OpenStax is licensed CC BY 4.0.

4.8.1: Speed of Sound, Frequency, and Wavelength

Learning Objectives

- Define pitch.
- Describe the relationship between the speed of sound, its frequency, and its wavelength.
- Describe the effects on the speed of sound as it travels through various media.
- Describe the effects of temperature on the speed of sound.



Figure 4.8.1.1: When a firework explodes, the light energy is perceived before the sound energy. Sound travels more slowly than light does. (credit: Dominic Alves, Flickr)

Sound, like all waves, travels at a certain speed and has the properties of frequency and wavelength. You can observe direct evidence of the speed of sound while watching a fireworks display. The flash of an explosion is seen well before its sound is heard, implying both that sound travels at a finite speed and that it is much slower than light. You can also directly sense the frequency of a sound. Perception of frequency is called **pitch**. The wavelength of sound is not directly sensed, but indirect evidence is found in the correlation of the size of musical instruments with their pitch. Small instruments, such as a piccolo, typically make high-pitch sounds, while large instruments, such as a tuba, typically make low-pitch sounds. High pitch means small wavelength, and the size of a musical instrument is directly related to the wavelengths of sound it produces. So a small instrument creates short-wavelength sounds. Similar arguments hold that a large instrument creates long-wavelength sounds.

The relationship of the speed of sound, its frequency, and wavelength is the same as for all waves:

$$v_w = f\lambda,$$

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 4.8.1.2. The frequency is the same as that of the source and is the number of waves that pass a point per unit time.

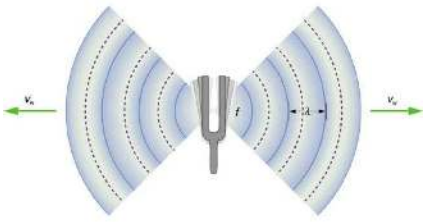


Figure 4.8.1.2: A sound wave emanates from a source vibrating at a frequency f, propagates at v_w, and has a wavelength λ.

Table 4.8.1.1 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. This observation is analogous to the fact that the frequency of a simple harmonic motion is directly proportional to the stiffness of the oscillating object. The greater the density of a medium, the slower the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is inversely proportional to the mass of the oscillating object. 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 **4.8.1.1**: Speed of Sound in Various Media

Medium	v_w(m/s)
Gases at 0ºC	
Air	331
Carbon dioxide	259
Oxygen	316
Helium	965
Hydrogen	1290
Liquids at 20ºC	
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

$$v_w = (331\ \text{m/s})\sqrt{\frac{T}{273\ \text{K}}},$$

where the temperature (denoted as T) is in units of kelvin. While not negligible, this is not a strong dependence. At 0ºC, the speed of sound is 331 m/s, whereas at 20.0ºC it is 343 m/s, less than a 4% increase. Figure **4.8.1.3** shows a use of the speed of sound by a bat to sense distances. Echoes are also used in medical imaging.



Figure 4.8.1.3: 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

$$v_w = f\lambda.$$

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 4.8.1.4 and consider the following example.

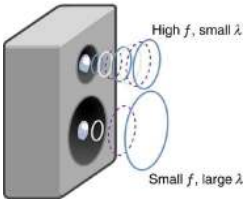


Figure 4.8.1.4: Because they travel at the same speed in a given medium, low-frequency sounds must have a greater wavelength than high-frequency sounds. Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, while the higher-frequency sounds are emitted by the small speaker, called a tweeter.

Example 4.8.1.1: 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 30.0ºC 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

$$v_w = (331\ \text{m/s})\sqrt{\frac{T}{273\ \text{K}}}.$$

2. Convert the temperature into kelvin and then enter the temperature into the equation

$$v_w = (331\ \text{m/s})\sqrt{\frac{303\ \text{K}}{273\ \text{K}}} = 348.7\ \text{m/s}.$$

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:

$$\lambda_{max} = \frac{348.7\ \text{m/s}}{20\ \text{Hz}} = 17\ \text{m}.$$

5. Enter the speed and the maximum frequency to give the minimum wavelength:

$$\lambda_{\text{min}} = \frac{348.7 \text{ m/s}}{20,000 \text{ Hz}} = 0.017 \text{ m} = 1.7 \text{ cm}.$$

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 4.8.1.1

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 4.8.1.2

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.

Section 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

$$v_w = (331 \text{ m/s})\sqrt{\frac{T}{273 \text{ K}}}.$$

v_w is the same for all frequencies and wavelengths.

Glossary

pitch

the perception of the frequency of a sound

This page titled 4.8.1: Speed of Sound, Frequency, and Wavelength is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **5.8: Speed of Sound, Frequency, and Wavelength** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

5: Work and Energy

5.1: Introduction to Work and Energy
5.2: Work- The Scientific Definition
5.3: Kinetic Energy and the Work-Energy Theorem
5.4: Potential Energy- Gravity and Springs
5.4.1: Gravitational Potential Energy
5.4.2: Spring Potential Energy
5.5: Conservative Forces, Potential Energy, and Conservation of Energy
5.6: Forms of Energy
5.7: Simple Machines
5.8: Power
5.9: Energy and Momentum

This page titled 5: Work and Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

5.1: Introduction to Work and Energy



Figure **5.1.1**: How many forms of energy can you identify in this photograph of a wind farm in Iowa? (credit: Jürgen from Sandesneben, Germany, Wikimedia Commons)

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*.

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 $E = mc^2$).

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. We will explore energy in more detail in the next chapter when we consider the other method of energy transfer, heat, and also consider why some forms of energy are unavailable to do work.

From a societal viewpoint, energy is one of the major building blocks of modern civilization. Energy resources are key limiting factors to economic growth. We will consider some of these implications at the end of both chapters 5 and 6. In this chapter we will consider how sophisticated applications of the concepts of force and work can result in simple machines which have been very effective in helping to create the world we now enjoy. At the end of chapter 6, we will consider how 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 in chapter 6.

This page titled 5.1: Introduction to Work and Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **3.1: Introduction to Work and Energy** by OpenStax is licensed CC BY 4.0.

5.2: Work- The Scientific Definition

Learning Objectives

- 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 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

$$W = F_{\parallel} d$$

,

where W is work, d is the distance the force acts, and $F_{\parallel}$ is the force parallel to the direction of motion. We can also write this more simply as

$$W = Fd$$

,

as long as one keeps in mind that the force is in the same direction as the distance.

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 would 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

$$W = Fd$$

,

where W is work, d is the distance the force acts, F is the force parallel to the direction of motion.

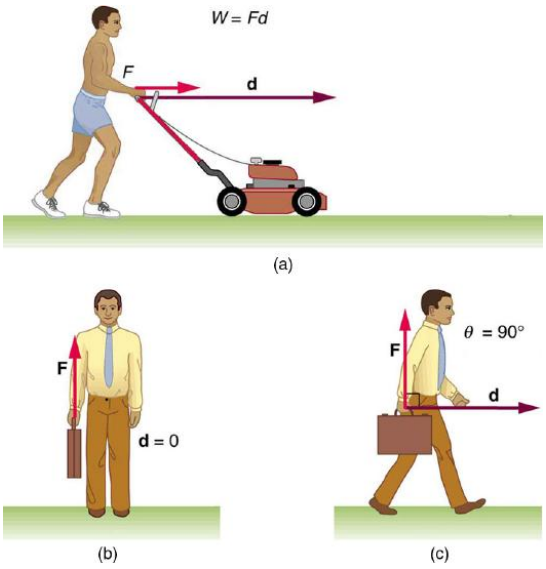


Figure 5.2.1: Examples of work. (a) Note that the force is in the direction of motion. (b) A person holding a briefcase does no work on it, because there is no motion. No energy is transferred to or from the briefcase. (c) The person moving the briefcase horizontally at a constant speed does no work on it, and transfers no energy to it. The force acting on the briefcase (against gravity) is perpendicular to the direction of motion.

To examine what the definition of work means, let us consider the other situations shown in Figure 5.2.1. The person holding the briefcase in Figure 5.2.1(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"). 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 5.2.1(c) does no work on it, because the force is perpendicular to the motion.

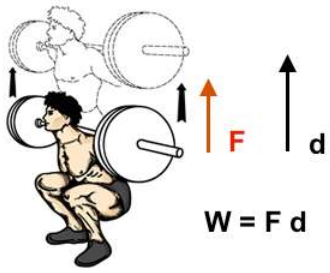


Figure 5.2.2: Work done when lifting a mass. Work is positive on the way up because force and displacement point in the same direction. On the way down work is negative due to force and displacement pointing in opposite directions (force up and displacement down).

Work can be positive or negative. In Figure 5.2.2 work done lifting the mass is positive because both force and displacement are in the same direction. Likewise, when the mass is lowered the work done is negative because the force and displacement are in *opposite* directions. We will soon see that positive work adds energy to the system and negative work removes energy from a system.

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 $1\ \text{J} = 1\ \text{N} \cdot \text{m} = 1\ \text{kg} \cdot \text{m}^2/\text{s}^2$. One joule is not a large amount of energy; it would lift a small 100-gram apple a distance of about 1 meter.

Example 5.2.1: 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 5.2.1(a) if he exerts a constant force of 75.0 N and pushes the mower 25.0 m on level ground? Compare it with this person's average daily intake of 10,000 kJ (about 2400 kcal) of food energy.

Strategy

We can solve this problem by substituting the given values into the definition of work done on a system, stated in the equation $W = Fd$. The force and displacement are given, so that only the work W is unknown. Note that force and displacement are in the same direction.

Solution

The equation for the work is

$$W = Fd.$$

Substituting the known values gives

$$W = (75.0\ \text{N})(25.0\ \text{m}) = 1875\ \text{J} = 1.88 \times 10^3\ \text{J}.$$

The ratio of the work done to the daily consumption is

$$\frac{W}{1.00 \times 10^4\ \text{kJ}} = \frac{1.88 \times 10^3\ \text{J}}{1.00 \times 10^7\ \text{J}} = 1.88 \times 10^{-4}.$$

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. We will discuss these other forms of energy in more detail later in this chapter.

Example 5.2.2: Calculating the Work Done Lifting a Mass

The 60.0 kg mass shown in Figure 5.2.2 is raised at a constant speed through a vertical distance of 0.800 m. What is the force that must be exerted? How much work is done lifting the mass?

Strategy

Constant speed tells us that the upward force must have the same magnitude as the downward force of gravity $F = mg$ (to have a zero net force). Using that information, along with the given values for mass and displacement, we can determine work using $W = Fd$.

Solution

The force is given by

$$F = mg = (60.0\ \text{kg})\left(9.80\ \text{m/s}^2\right) = 588\ \text{N}.$$

The work done is

$$W = Fd = (588\ \text{N})(0.800\ \text{m}) = 470\ \text{J}.$$

Discussion

Note that the work done lowering the mass back to its starting position would be $-470\ \text{J}$because force and displacement point in opposite directions. The total work done raising and lowering the mass is $470\ \text{J} - 470\ \text{J} = 0\ \text{J}$.

Section 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 parallel to the motion, times the magnitude d of the displacement. In symbols,

$$W = Fd.$$

- The SI unit for work and energy is the joule (J), where $1\ \text{J} = 1\ \text{N} \cdot \text{m} = 1\ \text{kg} \cdot \text{m}^2/\text{s}^2$.
- 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 5.2: Work- The Scientific Definition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **3.2: Work- The Scientific Definition** by OpenStax is licensed CC BY 4.0.

5.3: Kinetic Energy and the Work-Energy Theorem

Learning Objectives

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

Work Transfers Energy

What happens to the work done on a system? Energy is transferred into the system, but in what form? Does it remain in the system or move on? The answers depend on the situation. For example, if a lawn mower is pushed just hard enough to keep it going at a constant speed, then energy put into the mower by the person is removed continuously by friction, and eventually leaves the system in the form of heat transfer. In contrast, work done on a briefcase by a person carrying it up stairs is stored in the briefcase-Earth system and can be recovered at any time. 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 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 $\mathbf{F}_{\text{net}}$. In equation form, this is $W_{\text{net}} = F_{\text{net}}\, d$.

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 **5.3.1**.

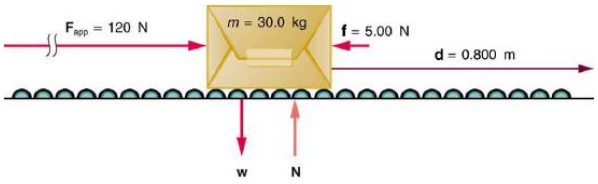


Figure **5.3.1**: A package on a roller belt is pushed horizontally through a distance **d**.

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 $\mathbf{F}_{\text{app}}$ and the horizontal friction force **f**. Thus, as expected, the net force is parallel to the displacement and the net work is given by

$$W_{\text{net}} = F_{\text{net}}\, d.$$

The effect of the net force $\mathbf{F}_{\text{net}}$ is to accelerate the package from v_0 to v. By using Newton's second law, and doing some algebra, we can reach an interesting conclusion. Substituting $F_{\text{net}} = ma$ from Newton's second law gives

$$W_{\text{net}} = mad.$$

To get a relationship between net work and the speed given to a system by the net force acting on it, we take $d = x - x_0$. Then (noting that a appears in the expression for the net work) use a motion equation for the change in speed over a distance d if the acceleration has the constant value a. The most useful motion equation for this purpose is $v^2 = v_0^2 + 2ad$ (we have not covered motion equations in great detail in this textbook, but you can find a more thorough listing of motion equations and their derivations in the text this one was curated from). 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

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

The d cancels, and we rearrange this to obtain

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

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.

THE WORK-ENERGY THEOREM

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

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$$

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. In equation form, the translational kinetic energy,

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

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. The work-energy theorem can be compactly written as,

$$W_{\text{net}} = \Delta\text{KE}$$

where ΔKE is understood to mean "change in KE."

We are aware that it takes energy to get an object, like a car or the package in Figure 5.3.1, 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 40 m/s has four times the kinetic energy it has at 20 m/s, 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 5.3.1: Calculating the Kinetic Energy of a Package

Suppose a 30.0-kg package on the roller belt conveyor system in Figure 5.3.1 is moving at 0.500 m/s. What is its kinetic energy?

Strategy

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

Solution

The kinetic energy is given by

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

Entering known values gives

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

which yields

$$\text{KE} = 3.75\ \text{kg}\cdot\text{m}^2/\text{s}^2 = 3.75\ \text{J}.$$

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 5.3.2: Determining the Work to Accelerate a Package

Suppose that you push on the 30.0-kg package in Figure 5.3.1 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 5.3.1.) As expected, the net work is the net force times distance.

Solution for (a)

The net force is the push force minus friction, or $F_{\text{net}} = 120\ \text{N} - 5.00\ \text{N} = 115\ \text{N}$. Thus, the net work is

$$W_{\text{net}} = F_{\text{net}}\, d = (115\ \text{N})(0.800\ \text{m}) = 92.0\ \text{N} \cdot \text{m} = 92.0\ \text{J}.$$

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.

$$W_{\text{app}} = F_{\text{app}}\, d = (120\ \text{N})(0.800\ \text{m}) = 96.0\ \text{J}$$

The friction force and displacement are in opposite directions. The work done by friction is therefore negative.

$$W_{\text{fr}} = -F_{\text{fr}} d = -(5.00\ \text{N})(0.800\ \text{m}) = -4.00\ \text{J}$$

So the amounts of work done by gravity, by the normal force, by the applied force, and by friction are, respectively,

$$\begin{aligned} W_{\text{gr}} &= 0, \\ W_{\text{N}} &= 0, \\ W_{\text{app}} &= 96.0\ \text{J}, \\ W_{\text{fr}} &= -4.00\ \text{J}. \end{aligned}$$

The total work done as the sum of the work done by each force is then seen to be

$$W_{\text{total}} = W_{\text{gr}} + W_{\text{N}} + W_{\text{app}} + W_{\text{fr}} = 92.0\ \text{J}.$$

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.

Section 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 $\text{KE} = \frac{1}{2}mv^2$.
- The work-energy theorem states that the net work W_{net} on a system changes its kinetic energy, $W_{\text{net}} = \Delta\text{KE} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$.

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 5.3: Kinetic Energy and the Work-Energy Theorem is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **3.3: Kinetic Energy and the Work-Energy Theorem** by OpenStax is licensed CC BY 4.0.

5.4: Potential Energy- Gravity and Springs

We've seen how energy can exist because of the movement of an object. As it turns out, energy can also exist within an object because it has the potential to have some sort of movement. There are multiple examples of this, but for now we will look at two: gravity and springs. Later in this chapter we will revisit the idea of energy and other ways in which we might categorize it.

This page titled 5.4: Potential Energy- Gravity and Springs is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

5.4.1: Gravitational Potential Energy

Learning Objectives

- Explain gravitational potential energy in terms of work done against gravity.
- Show that the gravitational potential energy of an object near surface of Earth is given by product of its mass, height, and gravitational acceleration.
- 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

Consider lifting a weight (of, say, 5 kg, which weighs about 50 N) from the floor to a table surface at a height of 1 meter from the floor. In lifting this weight, you have done work, both in the usual sense (if you do this lifting of weight over and over, you will eventually get tired) and in the scientific sense we defined in the earlier section ($W = Fd$, and you applied an upward force as you displaced the weight upward by 1 meter). One thing is different in this scenario, however. Earlier, in describing work-kinetic energy theorem, we said that net work done on an object changes the kinetic energy of the object ($\Delta \mathrm{KE} = W_{\mathrm{net}}$). But in this case, if you imagine lifting the weight very slowly up to the height of the table, you will find that the kinetic energy of the weight does not change. It starts at zero and remains at zero throughout this very slow motion. What is going on here?

The key difference is that, for this weight, there are now two forces acting on it as it is being lifted. The gravitational force pulls downward, while the force you apply is directed upward. If you are lifting the weight very slowly, then its acceleration is very close to zero, so the net force on it is zero ($F_{\mathrm{net}} = 0$), which means the *net work* done on the weight is also zero ($W_{\mathrm{net}} = F_{\mathrm{net}}\, d$; the applied upward force does positive work, while the downward gravitational force does negative work).

While this is a correct description (the negative work done by the gravity cancels out the positive work done by the applied force), this description misses something. Specifically, you can see that this weight that has been lifted by 1 meter *does* have more energy than a weight that was resting on the floor. This is how you can see it: if you let go of the weight (after removing the table), as the weight originally at 1-meter height reaches the floor, it gains a substantial kinetic energy as it continues to be pulled by gravity. Rather than saying that this kinetic energy simply appeared as the weight fell through the 1-meter height, we opt for the description that this weight already had this energy, when it was at rest at the 1-meter height, just in a form that was not kinetic energy. We call this energy associated with the height of objects the **gravitational potential energy**, PE_g. And in this example, the energy originally came from the work done by the applied force—rather than considering positive work done by the applied force to be cancelled out by the negative work done by the gravity, we say that this positive work done is *stored* as positive PE_g.

This phenomenon of potential energy turns out to be quite general and goes beyond the gravitational force. There are other forces in nature where negative work done by these forces turns into positive potential energy. We call these forces **conservative forces**, which will be the subject of a later section of this chapter. For the purpose of this section, following is the quick derivation of formula for PE_g (everything in terms of symbols now). In lifting the weight of mass m through a height h slowly, you have to apply a force equal to its weight, mg. So the work done is $W = Fd = mgh$, and since it's this work done that is stored as gravitational potential energy, the gravitational potential energy must be:

$$\mathrm{PE}_g = mgh.$$

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 5.4.1.2.) This shortcut makes it is easier to solve problems using energy (if possible) rather than explicitly using forces.

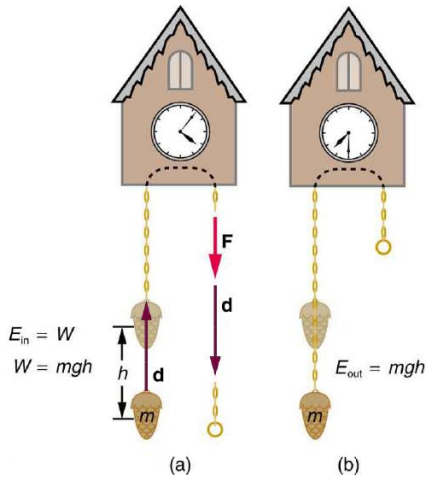


Figure 5.4.1.1: (a) The work done to lift the weight is stored in the mass-Earth system as gravitational potential energy. (b) As the weight moves downward, this gravitational potential energy is transferred to the cuckoo clock.

More precisely, we define the *change* in gravitational potential energy $\Delta \mathrm{PE_g}$ to be

$$\Delta \mathrm{PE_g} = \mathrm{PE_{gf}} - \mathrm{PE_{gi}} = mgh_f - mgh_i = mg\,(h_f - h_i) = mg\Delta h$$

Note that Δh 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

$$\begin{aligned} mg\Delta h &= (0.500\ \mathrm{kg})\left(9.80\ \mathrm{m/s^2}\right)(1.00\ \mathrm{m}) \\ &= 4.90\ \mathrm{kg \cdot m^2/s^2} = 4.90\ \mathrm{J}. \end{aligned}$$

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 $\Delta \mathrm{PE_g} = mg\Delta h$ applies for any path that has a change in height of $\Delta \mathrm{h}$, not just when the mass is lifted straight up. (See Figure 5.4.1.2.) It is much easier to calculate $mg\Delta h$ (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 $\Delta \mathrm{h}$ of a mass m is accompanied by a change in gravitational potential energy $mg\Delta h$, and we will avoid the equivalent but more difficult task of calculating work done by or against the gravitational force.

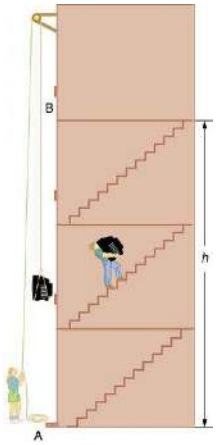


Figure 5.4.1.2: The change in gravitational potential energy ($\Delta \mathrm{PE_g}$) between points A and B is independent of the path. $\Delta \mathrm{PE_g} = mg\Delta h$ for any path between the two points. Gravity is one of a small class of forces where the work done by or against the force depends only on the starting and ending points, not on the path between them.

Example 5.4.1.1: 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 $\mathbf{PE_g}$ 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

$$W = -Fd,$$

with a minus sign because the displacement while stopping and the force from floor are in opposite directions. 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 $\mathbf{\Delta h}$:

$$\mathbf{KE} = -\Delta\mathbf{PE_g} = -mg\Delta h,$$

The distance d that the person's knees bend is much smaller than the height $\mathbf{\Delta h}$ 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:

$$W = -\mathbf{KE} = mg\Delta h.$$

Combining this equation with the expression for W gives

$$-Fd = mg\Delta h.$$

Recalling that $\mathbf{\Delta h}$ is negative because the person fell *down*, the force on the knee joints is given by

$$F = -\frac{mg\Delta h}{d} = -\frac{(60.0\ \text{kg})\,(9.80\ \text{m/s}^2)\,(-3.00\ \text{m})}{5.00 \times 10^{-3}\ \text{m}} = 3.53 \times 10^5\ \text{N}.$$

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 5.4.1.3.)



Figure 5.4.1.3: 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 5.4.1.2: Finding the Speed of a Roller Coaster from its Height

(a) What is the final speed of the roller coaster shown in Figure 5.4.1.4 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?

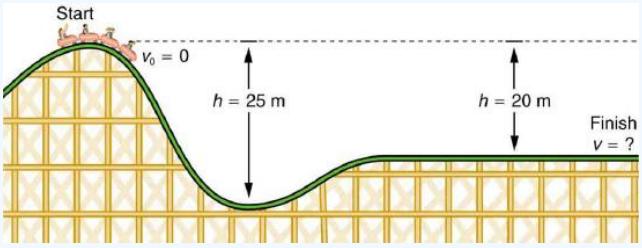


Figure 5.4.1.4: The speed of a roller coaster increases as gravity pulls it downhill and is greatest at its lowest point. Viewed in terms of energy, the roller-coaster-Earth system's gravitational potential energy is converted to kinetic energy. If work done by friction is negligible, all $\Delta\mathrm{PE_g}$ 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 $-\Delta\mathrm{PE_g} = \Delta\mathrm{KE}$. Using the equations for $\mathrm{PE_g}$ and KE, we can solve for the final speed v, which is the desired quantity.

Solution for (a)

$$-\Delta\mathrm{PE_g} = \Delta\mathrm{KE}$$

becomes

$$-(\mathrm{PE_{gf}} - \mathrm{PE_{gi}}) = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2$$

Noting that the initial speed is zero, and using the formula for gravitational potential energy, we have

$$-(mgh_f - mgh_i) = \frac{1}{2}mv_f^2$$

Taking the initial height as zero gives

$$-mgh_f = \frac{1}{2}mv_f^2$$

Solving for the final speed, we find that mass cancels and that

$$v_f = \sqrt{-2gh_f}$$

Substituting known values,

$$v_f = \sqrt{-2gh_f} = \sqrt{(-2)\,(9.80\ \mathrm{m/s^2})\,(-20.0\ \mathrm{m})} = 19.8\ \mathrm{m/s}$$

Please note that the height used above was negative. This is because we chose the initial height to be our reference level for gravitational potential energy. Our final height is 20 meters *below* that reference level.

Solution for (b)

Again

$$-\Delta\mathrm{PE_g} = \Delta\mathrm{KE}.$$

Thus,

$$-(\mathrm{PE_{gf}} - \mathrm{PE_{gi}}) = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2$$

Taking the initial height as zero leads to

$$-mgh_f = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2$$

Rearranging gives

$$\frac{1}{2}mv_f^2 = -mgh_f + \frac{1}{2}mv_i^2$$

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

$$v_f = \sqrt{-2gh_f + v_i^2}$$

Now, substituting known values gives

$$v_f = \sqrt{-2gh_f + v_i^2} = \sqrt{(-2)\,(9.80\ \text{m/s}^2)\,(-20.0\ \text{m}) + (5.00\ \text{m/s})^2} = 20.4\ \text{m/s}.$$

Discussion and Implications

First, note that mass cancels. This is quite consistent with observations 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.

Section 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 $\Delta\text{PE}_g = mg\Delta h$, with Δh being the change 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 $\Delta\text{KE} = -\Delta\text{PE}_g$.

Glossary

gravitational potential energy

energy associated with height of objects on the Earth

This page titled 5.4.1: Gravitational Potential Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **3.4: Gravitational Potential Energy** by OpenStax is licensed CC BY 4.0.

5.4.2: Spring Potential Energy

Learning Objectives

- Explain the work done in deforming a spring.
- Describe the potential energy stored in a deformed spring.

Hooke's Law, $F = -kx$, describes force exerted by a spring being deformed. Here, F is the restoring force, x is the displacement from equilibrium or **deformation**, and k is a constant related to the difficulty in deforming the system. The minus sign indicates the restoring force is in the direction opposite to the displacement.

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

$$\mathrm{PE_{el}} = \frac{1}{2}kx^2,$$

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

Connecting Ideas

Hooke's Law is the same equation that was used to develop the motion of waves. If we think about the oscillating motion of a compressed spring after the depressing force is released, this illustrates the reason for this connection.

It is possible to find the work done in deforming a system in order to find the energy stored. This work is performed by an applied force F_{app} . The applied force is exactly opposite to the restoring force (action-reaction), and so $F_{\mathrm{app}} = kx$. Figure 5.4.2.1 shows a graph of the applied force versus deformation x for a system that can be described by Hooke's law. Work done on the system is force multiplied by distance, which equals the area under the curve or $(1/2)kx^2$ (Method A in the figure). Another way to determine the work is to note that the force increases linearly from 0 to kx, so that the average force is $(1/2)kx$, the distance moved is x, and thus $W = F_{\mathrm{app}}d = [(1/2)kx](x) = (1/2)kx^2$ (Method B in the figure).

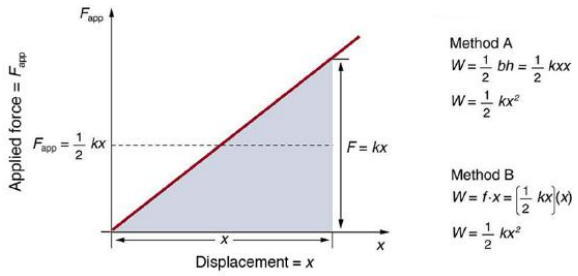


Figure 5.4.2.1: A graph of applied force versus distance for the deformation of a system that can be described by Hooke's law is displayed. The work done on the system equals the area under the graph or the area of the triangle, which is half its base multiplied by its height, or $W = (1/2)kx^2$.

Example 5.4.2.1: Calculating Stored Energy: A Tranquilizer Gun Spring

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

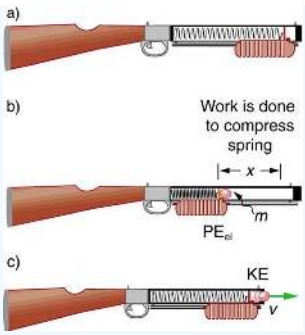


Figure 5.4.2.2: (a) In this image of the gun, the spring is uncompressed before being cocked. (b) The spring has been compressed a distance x, and the projectile is in place. (c) When released, the spring converts elastic potential energy $\mathrm{PE_{el}}$ into kinetic energy.

Strategy for a

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

Solution for a

Entering the given values for k and x yields

$$\mathrm{PE_{el}} = \frac{1}{2}kx^2 = \frac{1}{2}(50.0\ \mathrm{N/m})(0.150\ \mathrm{m})^2 = 0.563\ \mathrm{N \cdot m} = 0.563\ \mathrm{J}$$

Strategy for b

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

Solution for b

1. Identify known quantities:

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

2. Solve for v:

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

3. Convert units: 23.7m/s

Discussion

(a) and (b): This projectile speed is impressive for a tranquilizer gun (more than 80 km/h). The numbers in this problem seem reasonable. The force needed to compress the spring is small enough for an adult to manage, and the energy imparted to the dart is small enough to limit the damage it might do. Yet, the speed of the dart is great enough for it to travel an acceptable distance.

Exercise 5.4.2.1

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

Answer

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

Exercise 5.4.2.2

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

Answer

It was stored in the object as potential energy.

Section Summary

- Hooke's law describes force exerted by a spring being deformed,

$$F = -kx,$$

where F is the restoring force, x is the displacement from equilibrium or deformation, and k is the force constant of the system.

- Elastic potential energy $\mathbf{PE_{el}}$ stored in the deformation of a system that can be described by Hooke's law is given by

$$\mathbf{PE_{el}} = (1/2)kx^2.$$

Glossary

deformation

displacement from equilibrium

elastic potential energy

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

This page titled 5.4.2: Spring Potential Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **3.6: Spring Potential Energy** by OpenStax is licensed CC BY 4.0.

5.5: Conservative Forces, Potential Energy, and Conservation of Energy

Learning Objectives

- Define conservative force, potential energy, and mechanical energy.
- Apply conservation of mechanical energy to simple physical situations.
- Explain the law of the conservation of energy.

Potential Energy and Conservative Forces

Work is done by a force, and some forces, such as weight, have special characteristics. A **conservative force** is one, like the gravitational force, for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken. This happens if the force is a function of position alone and not time, velocity, or other parameters. We can define a **potential energy** (**PE**) for any conservative force, just as we did for the gravitational force. For example, when you wind up a toy, an egg timer, or an old-fashioned watch, you do work against its spring and store energy in it. (We treat these springs as ideal, in that we assume there is no friction and no production of thermal energy.) This stored energy is recoverable as work, and it is useful to think of it as potential energy contained in the spring. Indeed, the reason that the spring has this characteristic is that its force is *conservative*. That is, a conservative force results in stored or potential energy. Gravitational potential energy is one example, as is the energy stored in a spring. We will also see how conservative forces are related to the conservation of energy.

POTENTIAL ENERGY AND CONSERVATIVE FORCES

Potential energy is the energy a system has due to position, shape, or configuration. It is stored energy that is completely recoverable.

A conservative force is a force that is a function of position alone, with the result that the work done by the force depends only on the starting and ending points of a motion and not on the particular path taken.

Change in potential energy comes from the work done against a conservative force. Or flipping this around, potential energy increases as a conservative force does a negative work. That is,

$$\Delta \mathbf{PE} = -W_c,$$

where W_c is work done by a conservative force.

Conservation of Mechanical Energy

Consider an object in a system. If it gains any kinetic energy, this is a result of a net work done on the object, according to the work-kinetic energy theorem. If only conservative forces, such as the gravitational force or a spring force, do work in this system, then as kinetic energy increases with the net work done by the conservative forces, the system loses potential energy. That is, $\Delta \mathbf{KE} = W_c = -\Delta \mathbf{PE}$. In other words,

$$\Delta \mathbf{KE} + \Delta \mathbf{PE} = 0.$$

This equation means that the total kinetic and potential energy is constant for any process involving only conservative forces. That is,

$$\left.\begin{array}{l} \mathbf{KE} + \mathbf{PE} = \text{ constant} \\ \text{or} \\ \mathbf{KE_i} + \mathbf{PE_i} = \mathbf{KE_f} + \mathbf{PE_f} \end{array}\right\} \text{(conservative forces only)},$$

where i and f denote initial and final values. This equation is a form of the work-energy theorem for conservative forces; it is known as the **conservation of mechanical energy** principle. Remember that this applies to the extent that all the forces are conservative, so that friction is negligible. The total kinetic plus potential energy of a system is defined to be its **mechanical energy**, (**KE+PE**). In a system that experiences only conservative forces, there is a potential energy associated with each force, and the energy only changes form between **KE** and the various types of **PE**, with the total energy remaining constant.

Example 5.5.1: Using Conservation of Mechanical Energy to Calculate the Speed of a Toy Car

A 0.100-kg toy car is propelled by a compressed spring, as shown in Figure 5.5.1. The car follows a track that rises 0.180 m above the starting point. The compressed spring has a potential energy of 0.200 J. Assuming work done by friction to be negligible, find (a) how fast the car is going before it starts up the slope and (b) how fast it is going at the top of the slope.

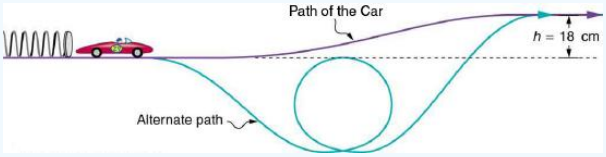


Figure 5.5.1: A toy car is pushed by a compressed spring and coasts up a slope. Assuming negligible friction, the potential energy in the spring is first completely converted to kinetic energy, and then to a combination of kinetic and gravitational potential energy as the car rises. The details of the path are unimportant because all forces are conservative—the car would have the same final speed if it took the alternate path shown.

Strategy

The spring force and the gravitational force are conservative forces, so conservation of mechanical energy can be used. Thus,

$$\mathrm{KE_i + PE_i = KE_f + PE_f}$$

or

$$\frac{1}{2}mv_\mathrm{i}^2 + mgh_\mathrm{i} + \mathrm{PE_{si}} = \frac{1}{2}mv_\mathrm{f}^2 + mgh_\mathrm{f} + \mathrm{PE_{sf}}$$

where h is the height (vertical position) and $\mathrm{PE_s}$ is the potential energy of the spring. This general statement looks complex but becomes much simpler when we start considering specific situations. First, we must identify the initial and final conditions in a problem; then, we enter them into the last equation to solve for an unknown.

Solution for (a)

This part of the problem is limited to conditions just before the car is released and just after it leaves the spring. Take the initial height to be zero, so that both h_i and h_f are zero. Furthermore, the initial speed v_i is zero and the final compression of the spring is zero, and so several terms in the conservation of mechanical energy equation are zero and it simplifies to

$$\mathrm{PE_{si}} = \frac{1}{2}mv_\mathrm{f}^2.$$

In other words, the initial potential energy in the spring is converted completely to kinetic energy in the absence of friction. Solving for the final speed and entering known values yields

$$v_\mathrm{f} = \sqrt{\frac{2\mathrm{PE_{si}}}{m}} = \sqrt{\frac{(2)(0.200\ \mathrm{J})}{0.100\ \mathrm{kg}}} = 2.00\ \mathrm{m/s}.$$

Solution for (b)

One method of finding the speed at the top of the slope is to consider conditions just before the car is released and just after it reaches the top of the slope, completely ignoring everything in between. Doing the same type of analysis to find which terms are zero, the conservation of mechanical energy becomes

$$\mathrm{PE_{si}} = \frac{1}{2}mv_\mathrm{f}^2 + mgh_\mathrm{f}.$$

This form of the equation means that the spring's initial potential energy is converted partly to gravitational potential energy and partly to kinetic energy. The final speed at the top of the slope will be less than at the bottom. Solving for v_f and substituting known values gives

$$v_\mathrm{f} = \sqrt{\frac{(2)\,(\mathrm{PE_{si}} - mgh_\mathrm{f})}{m}} = \sqrt{\frac{0.0472\ \mathrm{J}}{0.100\ \mathrm{kg}}} = 0.687\ \mathrm{m/s}$$

Discussion

Another way to solve this problem is to realize that the car's kinetic energy before it goes up the slope is converted partly to potential energy—that is, to take the final conditions in part (a) to be the initial conditions in part (b).

Note that, for conservative forces, we do not directly calculate the work they do; rather, we consider their effects through their corresponding potential energies, just as we did in Example 5.5.1. Note also that we do not consider details of the path taken—only the starting and ending points are important (as long as the path is not impossible). This assumption is usually a tremendous simplification, because the path may be complicated and forces may vary along the way.

Conservation of Total 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:

Total energy is constant in any process. It may change in form or be transferred from one system to another, but the total remains the same.

We have explored some forms of energy and some ways it can be transferred from one system to another, in the case of work done by conservative forces transforming between potential energy and kinetic energy. Together, these make up the mechanical energy (**KE**+**PE**), and the mechanical energy is not always conserved, because it can be transformed into other forms of energy through work done by non-conservative forces. We will explore other types of energy in the next section.

Section Summary

- A conservative force is one for which work depends only on the starting and ending points of a motion, not on the path taken.
- We can define potential energy (**PE**) for any conservative force, just as we defined $\mathbf{PE}_g$ for the gravitational force.
- Mechanical energy is defined to be **KE**+**PE** for a conservative force.
- When only conservative forces act on and within a system, the total mechanical energy is conserved, $\mathbf{\Delta KE + \Delta PE = 0}$.
- The law of conservation of energy states that the total energy, including the mechanical energy and other forms of 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.

Glossary

conservative force
a force that is a function of position alone, with the result that the work done by the force depends only on the starting and ending points of a motion and not on the particular path taken

potential energy
energy due to position, shape, or configuration

conservation of mechanical energy
the rule that the sum of the kinetic energies and potential energies remains constant if only conservative forces act on and within a system

mechanical energy
the sum of kinetic energy and potential energy

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

This page titled 5.5: Conservative Forces, Potential Energy, and Conservation of Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **3.5: Conservative Forces, Potential Energy, and Conservation of Energy** by OpenStax is licensed CC BY 4.0.

5.6: Forms of Energy

Learning Objectives

- Describe some of the many forms of energy.
- Categorize forms of energy as kinetic or potential.

Earlier in this chapter we introduced the concept of kinetic and potential energy as ways of understanding particular types of mechanical energy. We also introduced the law of conservation of energy and stated that it applied to more than just the mechanical energy that we had considered so far. We will now explore some of these other types of energy and how they fit within the context of kinetic and potential energy.

Some of the Many Forms of Energy

Here are some of the many forms of energy. You probably have heard of some of these before; 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 and other methods that can result in the kinetic energy of a moving car or the potential energy of climbing a mountain. 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 nuclear bombs. 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.

Transformation of Energy

As you have probably gathered from reading about these different forms 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 (a form of radiant 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. That electrical energy moves through wires to homes and businesses where it is used to perform a variety of tasks which all require 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 text.)

Another example of energy conversion occurs in a solar cell. Sunlight impinging on a solar cell (see Figure 5.6.2) 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 5.6.2: Solar energy is converted into electrical energy by solar cells, which is used to run a motor in this solar-power aircraft. (credit: NASA)

Categorizing Energy

As you read through the different forms of energy, were there aspects of it that seemed similar to the ways we have previously discussed energy in this text? There were some forms of energy that seemed to fit the pattern of storage, while others seemed to fit the pattern of movement. These are also the patterns we saw when analyzing mechanical energy in more detail earlier in the chapter. We referred to energy as either being kinetic if it was moving or potential if it was stored. We can do the same for these other forms of energy we are now introducing. We will not be able to get into many of the details on these other forms of energy just yet, but we can hopefully see these trends.

Potential Energy: the energy of position and storage

In our introduction to potential energy we looked at springs and gravity as places where energy was stored. We can very clearly see the ways in which the energy stored in these situations might be converted into mechanical energy. Now we will examine some of the other forms of energy mentioned and how some of them might fit into the category of stored energy.

One way of understanding potential energy is to think of it as the position of an object somewhere within a field of force. We had earlier established that the force of gravity varies based on the distance between objects. As an object moves to different positions within a force field the energy stored will change. Likewise, we can think of the force fields associated with **electromagnetism.** As objects move to different locations within these fields they will consequently store different amounts of energy. It turns out that this property of electromagnetism is very important to our ability in the modern world to have energy available to us on demand by the literal flip of a switch. This is a concept we will explore in greater detail later in this text.

If you read through the portions of this section on the forms of energy and transformation of energy, you will see **chemical energy** described as "being transferred," "producing," or "converted into." This language suggests that this chemical energy is stored in a way similar to the ways the potential energy in springs is stored. We will discuss chemical bonding in greater detail later in this text, but for now we can understand chemical bonds as a location between atoms where energy is stored. When chemical bonds are broken the energy stored in those bonds is released in some form. New bonds might form that store energy in a different way and by a different amount than the ways in which that energy was stored previously. We might imagine an elaborate contraption of springs that is rearranged from an initial situation where some of those springs are storing energy to a final situation where different springs are storing energy. This is actually very similar to how some chemists understand the changes that occur during a chemical reaction. The difference is that we can't see those changes going on at the molecular level the way we can see changes in springs at the macrosopic level.

You may have also noticed a mention of converting mass itself into energy. How is that possible if both mass and energy are conserved? It turns out that mass itself is a form of potential energy. This is far from intuitive and was described by Einstein in one of the most famous equations in science: $E=mc^2$. We will exploring this equation in more detail when we consider the atomic nucleus later in this text.

Kinetic Energy: the energy of movement

Just as we read through our forms of energy looking for examples of position or storage as clues for types of potential energy, we can also read through our forms of energy for examples of movement for types of kinetic energy. Although we did not include an example of it in the list above, we have previously looked at waves in this text. Waves, of course, involve movement. So any energy transferred through a wave would be a form of kinetic energy and might be referred to as **wave energy**.

When we discussed energy coming to Earth from the Sun in the form of **radiant energy**, we can clearly see that movement is involved. Radiant energy does not have to come from the Sun, in fact radiant energy comes from everything. But it is always a form of energy that is only possible when there is movement. We will learn more about radiant energy in this text when we explore the concept of heat, and even more when we discuss electromagnetic radiation. We will also learn that **nuclear energy** is a very specific type of radiant energy.

We described **thermal energy** as a form of random motion. As such, it could be classified as a form of kinetic energy. Thermal energy is the random movement of the atoms that make up a substance. As we will see later in this text, there is a connection between thermal energy and what we measure as temperature. Thermal energy is also where much of the energy ends up that is "lost" to friction when we try to do work. It is often the missing piece when conservative forces are added together and it might

appear that energy into and out of a process do not add up. But the energy is still there. In fact, that energy is still capable of doing some amount of work. We shall explore this idea in more detail later in this text.

When we discussed "electrical energy moving through wires" you might have recognized that it is also a form of kinetic energy. However, we had earlier discussed electromagnetic energy as being a form of potential energy. Is there a conflict here? Just as we can have mechanical energy exist as kinetic energy or potential energy, we can also have both kinetic and potential energy related to elecotromagnetism. The movement of an electrical charge through a wire demonstrates how **electrical energy** is a form of kinetic energy.

Section Summary

- Commonly encountered forms of energy include electric energy, chemical energy, radiant energy, nuclear energy, and thermal energy.
- Potential energy involves storage of energy, sometimes due to position.
- Kinetic energy involves movement, sometimes of matter.

Glossary

Kinetic energy
: energy due to movement

potential energy
: energy due to position, shape, or configuration

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

---

This page titled 5.6: Forms of Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

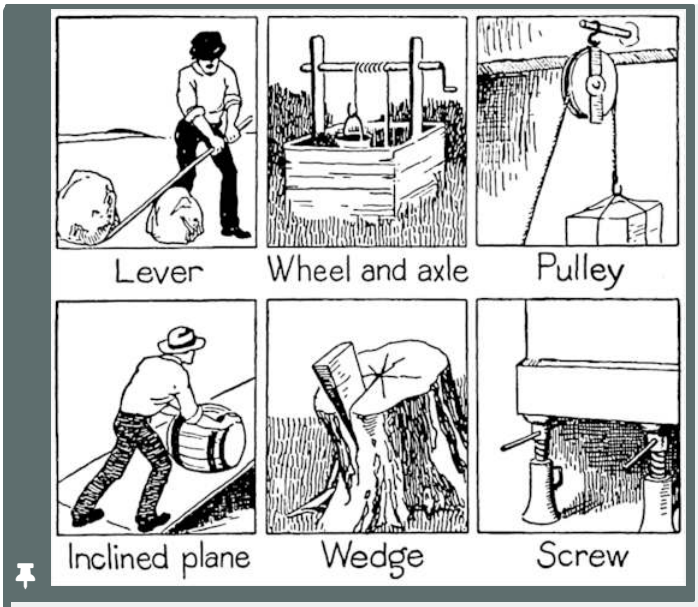
- **3.5: Conservative Forces, Potential Energy, and Conservation of Energy** by OpenStax is licensed CC BY 4.0.

5.7: Simple Machines

Learning Objectives

- Identify the six simple machines.
- Understand how a simple machine can change the application of force based on the principle of conservation of energy.

Long before the concept of energy was understood, the principles of conservation of energy were being used to develop simple machines. There is no documentation of the first machine ever used by humans or our ancestors, but it was probably a found object used in prehistoric times: perhaps a rock or stick used to pry open something to get at a source of food. What we refer to as the six classical simple machines were defined during the Renaissance period. They are the lever, the wheel and axle, the pulley, the inclined plane, the wedge, and the screw.



By Foster, Ellsworth D., ed; Hughes, James L. (James Laughlin) - The American educator; completely remodelled and rewritten from original text of the New practical reference library, with new plans and additional material (1919), Vol. 5, pg. 2290., Public Domain, [commons.wikimedia.org]

All of the simple machines function on the principle of conservation of energy, related to the work-energy theorem we discussed previously in this chapter. When considering simple machines, we typically ignore energy lost to friction even though we know that it exists. When we do that we can consider the relationship between the output of the machine and the applied force. The ratio of output to applied force is sometimes called the **mechanical advantage**. For example, a roadside service technician might use a hand or foot operated jack to push the body of a car up in order to change the tire. At no point in this process does the roadside technician apply a force with their own body capable of lifting the mass of a car in Earth's gravity. But the total energy the roadside technician puts into the machine is approximately equal to the total energy the machine expends to lift the car. The jack is an example of the lever, one of the six simple machines.

The relationship between the input and output of the machine is illustrated by the following equation:

$$W_{\text{in}} = W_{\text{out}}$$

.

Work is conserved because energy is conserved.

The energy that is conserved in simple machines relates to the force that enters and leaves the machine. As such, we can rewrite this equation to show it in terms of force and distance:

$$(Fd)_{\text{in}} = (Fd)_{\text{out}}$$

If the force that leaves the machine is greater than the force that enters the machine (as in our earlier example of the hand operated jack), then the distance over which that force operates must be proportionally smaller. For the hand operated jack, the distance the operator of the jack moves their hand in order to operate it is much greater than the distance through which the jack moves against the force of gravity on the car. We can say that the machine is multiplying the force output. It does this by increasing the distance over which the input force operates.

Example 5.7.1

Suppose a technician using a hand operated jack needs to lift one end of a car by a quarter of a meter. In order to do so, they must move a hand operated lever 200 times through a distance of 0.10 meters each time. What is the average force they are depressing the lever with over those 200 times in order to lift the car? Assume the average mass of the portion of the car being lifted is 700 kg.

Solution

This is an example of using the formulation of conservation of mechanical energy for simple machines that we introduced in this section:

$$(Fd)_{\text{in}} = (Fd)_{\text{out}}$$

We can start by listing known values and determining other values which are part of this equation from the given information.

$$(d)_{\text{out}} = 0.25\ \text{m}$$

Because the lever has to be moved 200 times, we get the total distance by multiplying this by the length of the lever:

$$(d)_{\text{in}} = 200(0.10\ \text{m}) = 20\ \text{m}$$

We can assume that lifting this car is going to involve working against the Earth's gravity, so we find that force as follows:

$$(F)_{\text{out}} = (700\ \text{kg})(10 \cdot \text{m}^2/\text{s}^2) = 7000\ \text{N}$$

We can then solve for the force being added using some algebra.

$$(F)_{\text{in}} = \frac{(Fd)_{\text{out}}}{(d)_{\text{in}}}$$

And then if we put in the numbers we have already calculated we get:

$$(F)_{\text{in}} = \frac{(7000\ \text{N})(0.25\ \text{m})}{(20\ \text{m})} = 87.5\ \text{N}$$

This is considerably less than the 7000 N the half car is exerting on the Earth due to the gravitational force.

Another way in which machines alter the force is through changing the direction of the force. Work and energy are scalar functions, but force is still a vector. The direction of the vector might be much different due to the interaction with the simple machine. Simple machines might change the direction of the force without changing its magnitude. But many times they do both.

Section Summary

- There are six simple machines, many of which have been used since before recorded history.
- Simple machines function based on the conservation of energy.
- The direction and/or magnitude of a force might be changed by a simple machine.

This page titled 5.7: Simple Machines is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

5.8: Power

Learning Objectives

- Understand the relationship among power, energy, and work done.
- 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 5.8.1.



Figure 5.8.1: 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.

Definition: POWER

Power is the rate at which work is done.

$$P = \frac{W}{t}$$

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 5.8.1: 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 5.8.2.)

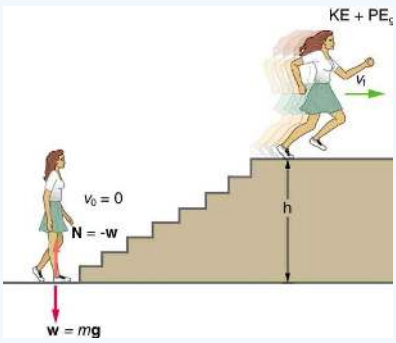


Figure 5.8.2: When this woman runs upstairs starting from rest, she converts the chemical energy originally from food into kinetic energy and gravitational potential energy. Her power output depends on how fast she does this.

Strategy and Concept

The work going into mechanical energy is **W**= **KE** + **PE**. At the bottom of the stairs, we take both **KE** and $\mathbf{PE}_g$ as initially zero; thus, $W = \mathbf{KE}_f + \mathbf{PE}_g = \frac{1}{2}mv_f^2 + mgh$, 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

$$P = \frac{W}{t} = \frac{\frac{1}{2}mv_f^2 + mgh}{t}.$$

Entering known values yields

$$\begin{aligned} P &= \frac{0.5(60.0\ \text{kg})(2.00\ \text{m/s})^2 + (60.0\ \text{kg})\,(9.80\ \text{m/s}^2)\,(3.00\ \text{m})}{3.50\ \text{s}} \\ &= \frac{120\ \text{J} + 1764\ \text{J}}{3.50\ \text{s}} \\ &= 538\ \text{W}. \end{aligned}$$

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. Do not 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 **5.8.1** for some examples.) Sunlight reaching Earth's surface carries a maximum power of about 1.3 kilowatts per square meter ($\mathbf{kW/m^2}$). 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 $\mathbf{10^6}$ **W**) 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 **5.8.3**.) Later in this text we will look at the efficiency of energy transfers in greater detail and learn why they happen the way that they do.



Figure **5.8.3**: 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)

Table **5.8.1**: Power Output or Consumption

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 $P = W/t = E/t$, where E is the energy supplied by the electricity company. So the energy consumed over a time t is

$$E = Pt.$$

Electricity bills state the energy used in units of **kilowatt-hours** ($\mathrm{kW \cdot h}$), 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 5.8.2: 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 $\mathbf{kW \cdot h}$?

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 $\mathbf{kW \cdot h}$, at the start of a problem such as this it is convenient to convert the units into $\mathbf{kW}$ and hours.

Solution

The energy consumed in $\mathbf{kW \cdot h}$ is

$$E = Pt = (0.200\ \mathbf{kW})(6.00\ \mathbf{h/d})(30.0\ \mathbf{d}) = 36.0\ \mathbf{kW \cdot h},$$

and the cost is simply given by

$$\mathbf{cost} = (36.0\ \mathbf{kW \cdot h})(\$0.120\ \mathbf{per\ kW \cdot h}) = \$4.32\ \mathbf{per\ month}.$$

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 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 later, the potential for energy to produce useful work has been “degraded” in the energy transformation.

Section 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 $1\ \mathbf{W} = 1\ \mathbf{J/s}$.
- The power of many devices such as electric motors is also often expressed in horsepower (hp), where $1\mathbf{hp} = 746\ \mathbf{W}$.

Glossary

power
the rate at which work is done

watt
(W) SI unit of power, with $1\ \mathbf{W} = 1\ \mathbf{J/s}$

horsepower
an older non-SI unit of power, with $1\ \mathbf{hp} = 746\ \mathbf{W}$

kilowatt-hour
($\mathbf{kW \cdot h}$)unit used primarily for electrical energy provided by electric utility companies

This page titled 5.8: Power is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **3.7: Power** by OpenStax is licensed CC BY 4.0.

5.9: Energy and Momentum

Learning Objectives

- Understand the differences between energy and momentum.

Energy and momentum have many similarities as well as many differences. These similarities and differences were confounding for the scientists of the 1700s and 1800s who were trying to formulate ideas about how the movements of masses followed basic repeatable principles.

Momentum and energy can both be formulated around ideas of mass and velocity. However, the formulation of energy is based on mass times velocity squared while the formulation of momentum is based on mass times velocity. Also, energy is not always based on movement while momentum is always based on movement. We have already seen how potential energy can be calculated for objects which are not moving. As you can probably imagine, there is no component of movement involved in most of the other forms of energy we have not considered in great detail yet.

The other similarity between momentum and energy is that they are both conserved. This is an important similarity, as there are very few properties in nature that are conserved. However, these two are conserved for different reasons and in different ways. Momentum is a vector quantity, while energy is a scalar quantity (an important type of scalar quantity called a state function, as we will learn later). As a vector quantity, momentum could cancel out during a collision. For example, if two cars of equal mass were moving at the same speed directly towards each other we might expect them to stop moving when they collide. In this case the momentum of the system defined as both cars is conserved because there overall momentum adds up to zero both before and after the collision. However, the energy does not cancel out. The energy of motion has been converted into potential energy as the bumpers are compressed and into thermal energy as they warm up. It does not cancel out as momentum does.

Section Summary

- Momentum is based on mass and velocity, and energy can be formulated to also be based on mass and velocity.
- Momentum and energy are both conserved.
- Momentum is a vector property, and energy is a scalar property.
- Momentum and energy are different properties, although they have sometimes been confused with each other.

This page titled 5.9: Energy and Momentum is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

CHAPTER OVERVIEW

6: Heat and Energy

6.1: Introduction to Thermodynamics
6.2: Temperature
6.2.1: Temperature and Thermometers
6.2.2: The Molecular Basis for Temperature
6.3: Heat
6.4: Heat Transfer Methods
6.5: Temperature Change and Heat Capacity
6.6: The First Law of Thermodynamics
6.7: Heat Engines and the Second Law of Thermodynamics
6.7.1: Efficiency and the Second Law of Thermodynamics
6.7.2: Carnot’s Perfect Heat Engine- The Second Law of Thermodynamics Restated
6.8: Entropy and the Second Law of Thermodynamics
6.8.1: Statistical Interpretation of Entropy and the Second Law of Thermodynamics- The Underlying Explanation
6.8.2: Entropy, Life, the Universe, and Everything
6.9: Energy Sources for Our Lives

This page titled 6: Heat and Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

6.1: Introduction to Thermodynamics



Figure **6.1.1**: The welder's gloves and helmet protect him from the electric arc that transfers enough thermal energy to melt the rod, spray sparks, and burn the retina of an unprotected eye. The thermal energy can be felt on exposed skin a few meters away, and its light can be seen for kilometers. (credit: Kevin S. O'Brien/U.S. Navy)

Heat is something familiar to each of us. We feel the warmth of the summer Sun, the chill of a clear summer night, the heat of coffee after a winter stroll, and the cooling effect of our sweat. Heat transfer is maintained by temperature differences. Manifestations of **heat transfer**—the movement of heat energy from one place or material to another—are apparent throughout the universe. Heat from beneath Earth's surface is brought to the surface in flows of incandescent lava. The Sun warms Earth's surface and is the source of much of the energy we find on it. Rising levels of atmospheric carbon dioxide threaten to trap more of the Sun's energy, perhaps fundamentally altering the ecosphere. In space, supernovas explode, briefly radiating more heat than an entire galaxy does.

What is heat? How do we define it? How is it related to temperature? What are heat's effects? How is it related to other forms of energy and to work? We will find that, in spite of the richness of the phenomena, there is a small set of underlying physical principles that unite the subjects and tie them to other fields.

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

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



Figure **6.1.2**: In a typical thermometer like this one, the alcohol, with a red dye, expands more rapidly than the glass containing it. When the thermometer's temperature increases, the liquid from the bulb is forced into the narrow tube, producing a large change in the length of the column for a small change in temperature. (credit: Chemical Engineer, Wikimedia Commons)

This page titled 6.1: Introduction to Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.1: Introduction to Thermal Physics** by OpenStax is licensed CC BY 4.0.

6.2: Temperature

Temperature might seem to be a straightforward concept, but there is quite a bit going on with it. We will discuss temperature in two subsections: how we measure it, and what is happening inside a material that gives rise to the measurement that we make. Understanding each of these will be crucial to understanding the material later in this chapter.

This page titled 6.2: Temperature is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

6.2.1: Temperature and Thermometers

Learning Objectives

- Define temperature.
- Convert temperatures between the Celsius, Fahrenheit, and Kelvin scales.
- Define thermal equilibrium.
- State the zeroth law of thermodynamics.

The concept of temperature has evolved from the common concepts of hot and cold. Human perception of what feels hot or cold is a relative one. For example, if you place one hand in hot water and the other in cold water, and then place both hands in tepid water, the tepid water will feel cool to the hand that was in hot water, and warm to the one that was in cold water. The scientific definition of temperature is less ambiguous than your senses of hot and cold. **Temperature** is operationally defined to be what we measure with a thermometer. (Many physical quantities are defined solely in terms of how they are measured. We shall see later how temperature is related to the kinetic energies of atoms and molecules, a more physical explanation.) Two accurate thermometers, one placed in hot water and the other in cold water, will show the hot water to have a higher temperature. If they are then placed in the tepid water, both will give identical readings (within measurement uncertainties). In this section, we discuss temperature, its measurement by thermometers, and its relationship to thermal equilibrium. Again, temperature is the quantity measured by a thermometer.

MISCONCEPTION ALERT: HUMAN PERCEPTION VS. REALITY

On a cold winter morning, the wood on a porch feels warmer than the metal of your bike. The wood and bicycle are in thermal equilibrium with the outside air, and are thus the same temperature. They *feel* different because of the difference in the way that they conduct heat away from your skin. The metal conducts heat away from your body faster than the wood does. This is just one example demonstrating that the human sense of hot and cold is not determined by temperature alone.

Another factor that affects our perception of temperature is humidity. Most people feel much hotter on hot, humid days than on hot, dry days. This is because on humid days, sweat does not evaporate from the skin as efficiently as it does on dry days. It is the evaporation of sweat (or water from a sprinkler or pool) that cools us off.

Any physical property that depends on temperature, and whose response to temperature is reproducible, can be used as the basis of a thermometer. Because many physical properties depend on temperature, the variety of thermometers is remarkable. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer, the old mercury thermometer, and the bimetallic strip (Figure **6.2.1.1**). Other properties used to measure temperature include electrical resistance and color, as shown in Figure **6.2.1.2**, and the emission of infrared radiation, as shown in Figure **6.2.1.3**.

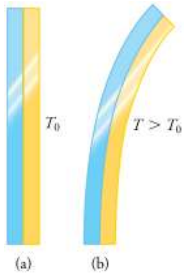


Figure **6.2.1.1**: The curvature of a bimetallic strip depends on temperature. (a) The strip is straight at the starting temperature, where its two components have the same length. (b) At a higher temperature, this strip bends to the right, because the metal on the left has expanded more than the metal on the right.



Figure **6.2.1.2**: Each of the six squares on this plastic (liquid crystal) thermometer contains a film of a different heat-sensitive liquid crystal material. Below 95ºF, all six squares are black. When the plastic thermometer is exposed to temperature that increases to 95ºF, the first liquid crystal square changes color. When the temperature increases above 96.8ºF the second liquid crystal square also changes color, and so forth. (credit: Arkrishna, Wikimedia Commons)



Figure **6.2.1.3**: Fireman Jason Ormand uses a pyrometer to check the temperature of an aircraft carrier's ventilation system. Infrared radiation (whose emission varies with temperature) from the vent is measured and a temperature readout is quickly produced. Infrared measurements are also frequently used as a measure of body temperature. These modern thermometers, placed in the ear canal, are more accurate than alcohol thermometers placed under the tongue or in the armpit. (credit: Lamel J. Hinton/U.S. Navy)

Temperature Scales

Thermometers are used to measure temperature according to well-defined scales of measurement, which use pre-defined reference points to help compare quantities. The three most common temperature scales are the Fahrenheit, Celsius, and Kelvin scales. A temperature scale can be created by identifying two easily reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used.

The **Celsius** scale (which replaced the slightly different *centigrade* scale) has the freezing point of water at 0ºC and the boiling point at 100ºC. Its unit is the **degree Celsius**(ºC). On the **Fahrenheit** scale (still the most frequently used in the United States), the freezing point of water is at 32ºF and the boiling point is at 212ºF. The unit of temperature on this scale is the **degree Fahrenheit**(ºF). Note that a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Only 100 Celsius degrees span the same range as 180 Fahrenheit degrees, thus one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale $180/100 = 9/5$.

The **Kelvin** scale is the temperature scale that is commonly used in science. It is an *absolute temperature* scale defined to have 0 K at the lowest possible temperature, called **absolute zero**. The official temperature unit on this scale is the *kelvin*, which is abbreviated K, and is not accompanied by a degree sign. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Thus, the magnitude of temperature differences is the same in units of kelvins and degrees Celsius. Unlike other temperature scales, the Kelvin scale is an absolute scale. It is used extensively in scientific work because a number of physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature. The kelvin is the SI unit used in scientific work.

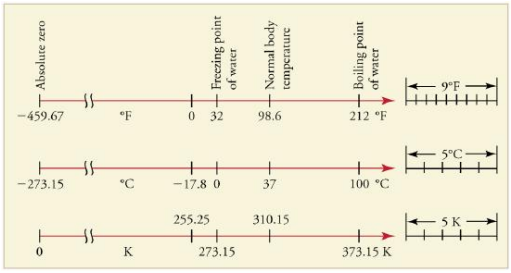


Figure **6.2.1.4**: Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree. The relative sizes of the scales are also shown.

The relationships between the three common temperature scales is shown in Figure **6.2.1.4**.

Temperature Ranges in the Universe

Figure **6.2.1.6** shows the wide range of temperatures found in the universe. Human beings have been known to survive with body temperatures within a small range, from 24ºC to 44ºC (75ºF to 111ºF). The average normal body temperature is usually given as 37.0ºC (98.6ºF), and variations in this temperature can indicate a medical condition: a fever, an infection, a tumor, or circulatory problems (see Figure **6.2.1.5**).

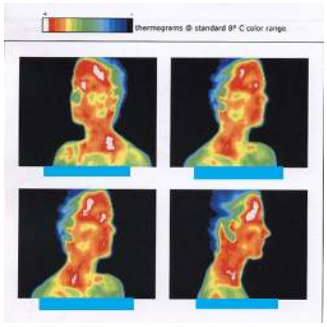


Figure **6.2.1.5**: This image of radiation from a person's body (an infrared thermograph) shows the location of temperature abnormalities in the upper body. Dark blue corresponds to cold areas and red to white corresponds to hot areas. An elevated temperature might be an indication of malignant tissue (a cancerous tumor in the breast, for example), while a depressed temperature might be due to a decline in blood flow from a clot. In this case, the abnormalities are caused by a condition called hyperhidrosis. (credit: Porcelina81, Wikimedia Commons)

The lowest temperatures ever recorded have been measured during laboratory experiments: 4.5×10^{-10} K at the Massachusetts Institute of Technology (USA), and 1.0×10^{-10} K at Helsinki University of Technology (Finland). In comparison, the coldest recorded place on Earth's surface is Vostok, Antarctica at 183 K (–89ºC), and the coldest place (outside the lab) known in the universe is the Boomerang Nebula, with a temperature of 1 K.

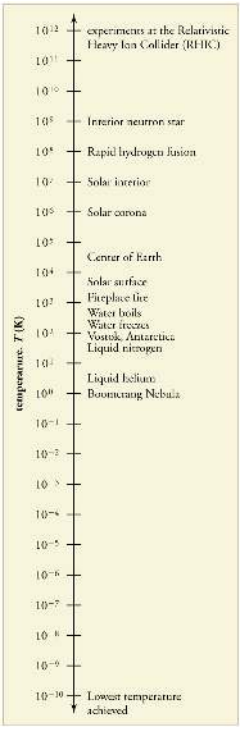


Figure **6.2.1.6**: Each increment on this logarithmic scale indicates an increase by a factor of ten, and thus illustrates the tremendous range of temperatures in nature. Note that zero on a logarithmic scale would occur off the bottom of the page at infinity.

MAKING CONNECTIONS: ABSOLUTE ZERO

What is absolute zero? Absolute zero is the temperature at which all molecular motion has ceased. The concept of absolute zero arises from the behavior of gases. Figure **6.2.1.7** shows how the pressure of gases at a constant volume decreases as temperature decreases. Various scientists have noted that the pressures of gases extrapolate to zero at the same temperature, –273.15ºC. This extrapolation implies that there is a lowest temperature. This temperature is called *absolute zero*. Today we know that most gases first liquefy and then freeze, and it is not actually possible to reach absolute zero. The numerical value of absolute zero temperature is –273.15ºC or 0 K.

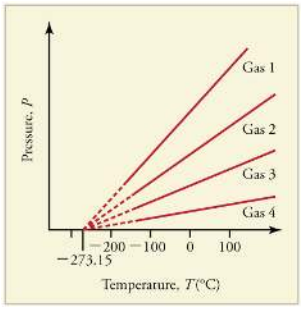


Figure **6.2.1.7**:Graph of pressure versus temperature for various gases kept at constant volume. Note that all of the graphs extrapolate to zero pressure at the same temperature.

Thermal Equilibrium and the Zeroth Law of Thermodynamics

Thermometers actually take their *own* temperature, not the temperature of the object they are measuring. This raises the question of how we can be certain that a thermometer measures the temperature of the object with which it is in contact. It is based on the fact that any two systems placed in *thermal contact* (meaning heat transfer can occur between them) will reach the same temperature. That is, heat will flow from the hotter object to the cooler one until they have exactly the same temperature. The objects are then in **thermal equilibrium**, and no further changes will occur. The systems interact and change because their temperatures differ, and the changes stop once their temperatures are the same. Thus, if enough time is allowed for this transfer of heat to run its course, the temperature a thermometer registers *does* represent the system with which it is in thermal equilibrium. Thermal equilibrium is established when two bodies are in contact with each other and can freely exchange energy.

Furthermore, experimentation has shown that if two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system C, then A is also in thermal equilibrium with C. This conclusion may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the **zeroth law of thermodynamics**.

THE ZEROTH LAW OF THERMODYNAMICS

If two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

This law was postulated in the 1930s, after the first and second laws of thermodynamics had been developed and named. It is called the *zeroth law* because it comes logically before the first and second laws. An example of this law in action is seen in babies in incubators: babies in incubators normally have very few clothes on, so to an observer they look as if they may not be warm enough. However, the temperature of the air, the cot, and the baby is the same, because they are in thermal equilibrium, which is accomplished by maintaining air temperature to keep the baby comfortable.

Exercise 6.2.1.1

Does the temperature of a body depend on its size?

Answer

No, the system can be divided into smaller parts each of which is at the same temperature. We say that the temperature is an *intensive* quantity. Intensive quantities are independent of size.

Section Summary

- Temperature is the quantity measured by a thermometer.
- Temperature is related to the average kinetic energy of atoms and molecules in a system.
- Absolute zero is the temperature at which there is no molecular motion.
- There are three main temperature scales: Celsius, Fahrenheit, and Kelvin.
- Temperatures on one scale can be converted to temperatures on another scale using the following equations:

$$T_{\circ}\mathrm{F} = \frac{9}{5}T_{\mathrm{C}}\mathrm{C} + 32$$
$$T_{\circ}\mathrm{C} = \frac{5}{9}\left(T_{\circ} - 32\right)$$
$$T_{\mathrm{K}} = T_{\circ}\mathrm{C} + 273.15$$
$$T_{\circ}\mathrm{C} = T_{\mathrm{K}} - 273.15$$

- Systems are in thermal equilibrium when they have the same temperature.
- Thermal equilibrium occurs when two bodies are in contact with each other and can freely exchange energy.
- The zeroth law of thermodynamics states that when two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

Glossary

temperature
the quantity measured by a thermometer

Celsius scale
temperature scale in which the freezing point of water is 0ºC and the boiling point of water is 100ºC

degree Celsius
unit on the Celsius temperature scale

Fahrenheit scale
temperature scale in which the freezing point of water is 32ºF and the boiling point of water is 212ºF

degree Fahrenheit
unit on the Fahrenheit temperature scale

Kelvin scale
temperature scale in which 0 K is the lowest possible temperature, representing absolute zero

absolute zero
the lowest possible temperature; the temperature at which all molecular motion ceases

thermal equilibrium
the condition in which heat no longer flows between two objects that are in contact; the two objects have the same temperature

zeroth law of thermodynamics
law that states that if two objects are in thermal equilibrium, and a third object is in thermal equilibrium with one of those objects, it is also in thermal equilibrium with the other object

---

This page titled 6.2.1: Temperature and Thermometers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.2: Temperature** by OpenStax is licensed CC BY 4.0.

6.2.2: The Molecular Basis for Temperature

Learning Objectives

- Explain how temperature relates to the motion of atoms.
- State and explain the ideal gas law using Boltzmann's constant.
- Use the ideal gas law to calculate changes in relationships between pressure, volume, temperature, or number of molecules in a gas.



Figure **6.2.2.1**: The air inside this hot air balloon flying over Putrajaya, Malaysia, is hotter than the ambient air. As a result, the balloon experiences a buoyant force pushing it upward. (credit: Kevin Poh, Flickr)

All matter is made up of atoms, and those atoms are always in motion. The motion of these atoms can take several forms: translation, rotation, and vibration. The type of motion that a particular atom has available to it within a substance depends on the type of bonding (if any) that is present between the atoms of those substances. For solids and liquids, this bonding is too complicated to address here. But for gases, we can describe this relationship in greater detail and how it relates to temperature. Monatomic gases can only move via translation: back and forth linear motion.

Because gas phase atoms and molecules have large separations, forces between them can be ignored, except when they collide with each other during collisions. The motion of atoms and molecules (at temperatures well above the boiling temperature) is fast, such that the gas occupies all of the accessible volume and the expansion of gases is rapid. In contrast, in liquids and solids, atoms and molecules are closer together and are quite sensitive to the forces between them.



Figure **6.2.2.2**: Atoms and molecules in a gas are typically widely separated, as shown. Because the forces between them are quite weak at these distances, the properties of a gas depend more on the number of atoms per unit volume and on temperature than on the type of atom.

To get some idea of how pressure, temperature, and volume of a gas are related to one another, consider what happens when you pump air into an initially deflated tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the walls limit volume expansion. If we continue to pump air into it, the pressure increases. The pressure will further increase when the car is driven and the tires move. Most manufacturers specify optimal tire pressure for cold tires. (See Figure **6.2.2.3**.)

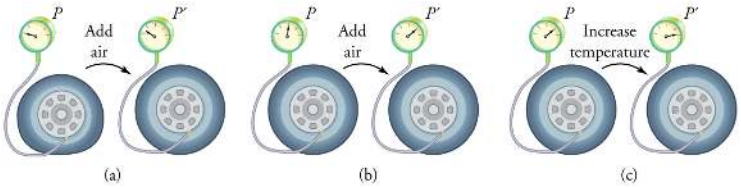


Figure **6.2.2.3**: (a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion and the pressure increases with more air. (c) Once the tire is inflated, its pressure increases with temperature.

At room temperatures, collisions between atoms and molecules can be ignored. In this case, the gas is called an ideal gas, in which case the relationship between the pressure, volume, and temperature is given by the equation of state called the ideal gas law.

IDEAL GAS LAW

The **ideal gas law** states that

$$PV = NkT,$$

where P is the absolute pressure of a gas, V is the volume it occupies, N is the number of atoms and molecules in the gas, and T is its absolute temperature. The constant k is called the **Boltzmann constant** in honor of Austrian physicist Ludwig Boltzmann (1844–1906) and has the value

$$k = 1.38 \times 10^{-23}\ \text{J/K}.$$

The ideal gas law is a combination of a few other gas laws which related binary pairs of pressure, temperature, volume, and number of particles to result in direct or inverse relationships. It is one of these relationships (pressure and temperature) which led to the prediction of absolute zero discussed previously.

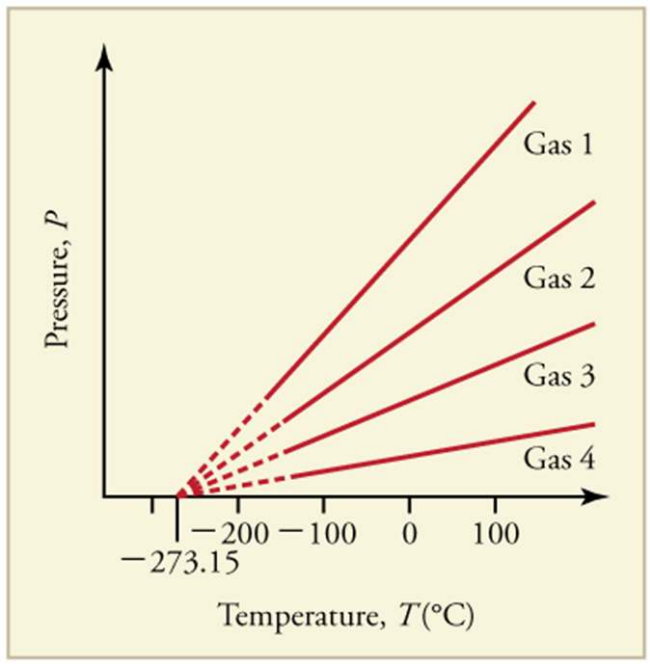


Figure **6.2.2.4**: If the relationship between Pressure and Temperature is extrapolated to zero pressure for any gas we arrive at the accepted value of absolute zero temperature.

Let us see how the ideal gas law is consistent with the behavior of filling the tire when it is pumped slowly and the temperature is constant. At first, the pressure P is essentially equal to atmospheric pressure, and the volume V increases in direct proportion to the

number of atoms and molecules N put into the tire. Once the volume of the tire is constant, the equation $PV = NkT$ predicts that the pressure should increase in proportion to *the number N of atoms and molecules*.

Example 6.2.2.1: Calculating Pressure Changes Due to Temperature Changes: Tire Pressure

Suppose your bicycle tire is fully inflated, with an absolute pressure of 7.00×10^5 Pa (a gauge pressure of just under $90.0\ \text{lb/in}^2$) at a temperature of $18.0°\text{C}$. What is the pressure after its temperature has risen to 35.0ºC? Assume that there are no appreciable leaks or changes in volume.

Strategy

The pressure in the tire is changing only because of changes in temperature. First we need to identify what we know and what we want to know, and then identify an equation to solve for the unknown.

We know the initial pressure $P_0 = 7.00 \times 10^5$ Pa, the initial temperature $T_0 = 18.0°\text{C}$, and the final temperature $T_f = 35.0°\text{C}$. We must find the final pressure P_f. How can we use the equation $PV = NkT$? At first, it may seem that not enough information is given, because the volume V and number of atoms N are not specified. What we can do is use the equation twice: $P_0V_0 = NkT_0$ and $P_fV_f = NkT_f$. If we divide P_fV_f by P_0V_0 we can come up with an equation that allows us to solve for Pf.

$$\frac{P_fV_f}{P_0V_0} = \frac{N_fkT_f}{N_0kT_0}$$

Since the volume is constant, V_f and V_0 are the same and they cancel out. The same is true for N_f and N_0, and k, which is a constant. Therefore,

$$\frac{P_f}{P_0} = \frac{T_f}{T_0}.$$

We can then rearrange this to solve for P_f:

$$P_f = P_0\frac{T_f}{T_0},$$

where the temperature must be in units of kelvins, because T_0 and T_f are absolute temperatures.

Solution

1. Convert temperatures from Celsius to Kelvin.

$$T_0 = (18.0 + 273)\text{K} = 291\ \text{K}$$
$$T_f = (35.0 + 273)\text{K} = 308\ \text{K}$$

2. Substitute the known values into the equation.

$$P_f = P_0\frac{T_f}{T_0} = 7.00 \times 10^5\ \text{Pa}\left(\frac{308\ \text{K}}{291\ \text{K}}\right) = 7.41 \times 10^5\ \text{Pa}$$

Discussion

The final temperature is about 6% greater than the original temperature, so the final pressure is about 6% greater as well. Note that *absolute* pressure and *absolute* temperature must be used in the ideal gas law.

MAKING CONNECTIONS: TAKE-HOME EXPERIMENT—REFRIGERATING A BALLOON

Inflate a balloon at room temperature. Leave the inflated balloon in the refrigerator overnight. What happens to the balloon, and why?

Section Summary

- The ideal gas law relates the pressure and volume of a gas to the number of gas molecules and the temperature of the gas.
- The ideal gas law can be written in terms of the number of molecules of gas:

$$PV = NkT,$$

where P is pressure, V is volume, T is temperature, N is number of molecules, and k is the Boltzmann constant

$$k = 1.38 \times 10^{-23}\ \text{J/K}.$$

- The ideal gas law is generally valid at temperatures well above the boiling temperature.

Glossary

ideal gas law

the physical law that relates the pressure and volume of a gas to the number of gas molecules or number of gas particles and the temperature of the gas

Boltzmann constant

k , a physical constant that relates energy to temperature; $k=1.38\times10^{-23}$ J/K

This page titled 6.2.2: The Molecular Basis for Temperature is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.3: The Ideal Gas Law** by OpenStax is licensed CC BY 4.0.

6.3: Heat

Learning Objectives

- Define heat as transfer of energy.

We have defined work as force times distance and learned that work done on an object changes its kinetic energy. We have also seen that temperature is proportional to the (average) kinetic energy of atoms and molecules. We say that a thermal system has a certain internal energy: its internal energy is higher if the temperature is higher. If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter to the colder object until equilibrium is reached and the bodies reach thermal equilibrium (i.e., they are at the same temperature). No work is done by either object, because no force acts through a distance. The transfer of energy is caused by the temperature difference, and ceases once the temperatures are equal. These observations lead to the following definition of **heat**: Heat is the spontaneous transfer of energy due to a temperature difference.

Heat is often confused with temperature. For example, we may say the heat was unbearable, when we actually mean that the temperature was high. Heat is a form of energy, whereas temperature is not. The misconception arises because we are sensitive to the flow of heat, rather than the temperature.

Owing to the fact that heat is a form of energy, it has the SI unit of *joule* (J). The *calorie* (cal) is a common unit of energy, defined as the energy needed to change the temperature of 1.00 g of water by 1.00ºC —specifically, between 14.5ºC and 15.5ºC, since there is a slight temperature dependence. Perhaps the most common unit of heat is the **kilocalorie** (kcal), which is the energy needed to change the temperature of 1.00 kg of water by 1.00ºC. Since mass is most often specified in kilograms, kilocalorie is commonly used. Food calories (given the notation Cal, and sometimes called “big calorie”) are actually kilocalories (1kilocalorie=1000 calories), a fact not easily determined from package labeling.

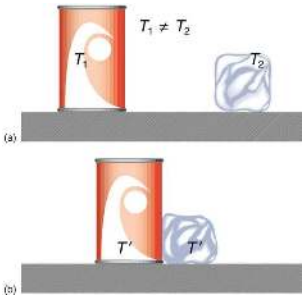


Figure **6.3.1**: In figure (a) the soft drink and the ice have different temperatures, T_1 and T_2, and are not in thermal equilibrium. In figure (b), when the soft drink and ice are allowed to interact, energy is transferred until they reach the same temperature T', achieving equilibrium. Heat transfer occurs due to the difference in temperatures. In fact, since the soft drink and ice are both in contact with the surrounding air and bench, the equilibrium temperature will be the same for both.

Mechanical Equivalent of Heat

It is also possible to change the temperature of a substance by doing work. Work can transfer energy into or out of a system. This realization helped establish the fact that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the **mechanical equivalent of heat**—*the work needed to produce the same effects as heat transfer*. In terms of the units used for these two terms, the best modern value for this equivalence is

$$1.000 \text{ kcal} = 4186 \text{ J}.$$

We consider this equation as the conversion between two different units of energy.

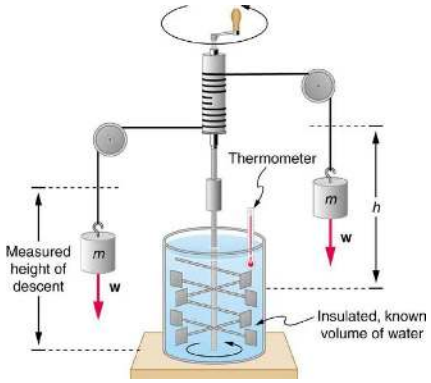


Figure **6.3.2**: Schematic depiction of Joule's experiment that established the equivalence of heat and work.

The figure above shows one of Joule's most famous experimental setups for demonstrating the mechanical equivalent of heat. It demonstrated that work and heat can produce the same effects, and helped establish the principle of conservation of energy. Gravitational potential energy (PE) (work done by the gravitational force) is converted into kinetic energy (KE), and then randomized by viscosity and turbulence into increased average kinetic energy of atoms and molecules in the system, producing a temperature increase. His contributions to the field of thermodynamics were so significant that the SI unit of energy was named after him.

Heat added or removed from a system changes its internal energy, and thus its temperature. Such a temperature increase is observed while cooking. However, adding heat does not necessarily increase the temperature. An example is melting of ice; that is, when a substance changes from one phase to another. Work done on the system or by the system can also change the internal energy of the system. Joule demonstrated that the temperature of a system can be increased by stirring. If an ice cube is rubbed against a rough surface, work is done by the frictional force. A system has a well-defined internal energy, but we cannot say that it has a certain "heat content" or "work content". We use the phrase "heat transfer" to emphasize its nature.

Exercise 6.3.1

Two samples (A and B) of the same substance are kept in a lab. Someone adds 10 kilojoules (kJ) of heat to one sample, while 10 kJ of work is done on the other sample. How can you tell to which sample the heat was added?

Answer

Heat and work both change the internal energy of the substance. However, the properties of the sample only depend on the internal energy so that it is impossible to tell whether heat was added to sample A or B.

Section Summary

- Heat and work are the two distinct methods of energy transfer.
- Heat is energy transferred solely due to a temperature difference.
- Any energy unit can be used for heat transfer, and the most common are kilocalorie (kcal) and joule (J).
- Kilocalorie is defined to be the energy needed to change the temperature of 1.00 kg of water between 14.5ºC and 15.5ºC.
- The mechanical equivalent of this heat transfer is 1.00 kcal=4186 J.

Glossary

heat

the spontaneous transfer of energy due to a temperature difference

kilocalorie

1 kilocalorie = 1000 calories

mechanical equivalent of heat

the work needed to produce the same effects as heat transfer

This page titled 6.3: Heat is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.4: Heat** by OpenStax is licensed CC BY 4.0.

6.4: Heat Transfer Methods

Learning Objectives

- Discuss the three different methods of heat transfer.

Equally as interesting as the effects of heat transfer on a system are the methods by which this occurs. Whenever there is a temperature difference, heat transfer occurs. Heat transfer may occur rapidly, such as through a cooking pan, or slowly, such as through the walls of a picnic ice chest. We can control rates of heat transfer by choosing materials (such as thick wool clothing for the winter), controlling air movement (such as the use of weather stripping around doors), or by choice of color (such as a white roof to reflect summer sunlight). So many processes involve heat transfer, so that it is hard to imagine a situation where no heat transfer occurs. Yet every process involving heat transfer takes place by only three methods:

1. **Conduction** is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale—we know there is thermal motion of the atoms and molecules at any temperature above absolute zero.) Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction.
2. **Convection** is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.
3. Heat transfer by **radiation** occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of the Earth by the Sun. A less obvious example is thermal radiation from the human body.

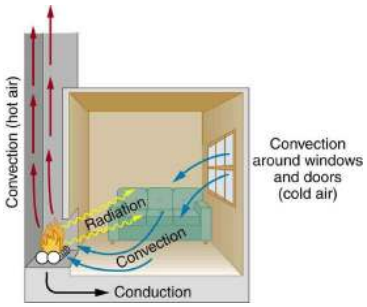


Figure **6.4.1**: In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but at a much slower rate. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

Figure **6.4.1** illustrates these three methods of heat transfer occurring in one system. While we skip detailed discussion of methods of heat transfer, the one thing all three methods-conduction, convection, and radiation-share is they are driven by temperature difference.

Exercise 6.4.1

Name an example from daily life (different from the text) for each mechanism of heat transfer.

Answer

Conduction: Heat transfers into your hands as you hold a hot cup of coffee.

Convection: Heat transfers as the barista “steams” cold milk to make hot *cocoa*.

Radiation: Reheating a cold cup of coffee in a microwave oven.

Summary

- Heat is transferred by three different methods: conduction, convection, and radiation.

Glossary

conduction

heat transfer through stationary matter by physical contact

convection

heat transfer by the macroscopic movement of fluid

radiation

heat transfer which occurs when microwaves, infrared radiation, visible light, or other electromagnetic radiation is emitted or absorbed

---

This page titled 6.4: Heat Transfer Methods is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.5: Heat Transfer Methods** by OpenStax is licensed CC BY 4.0.

6.5: Temperature Change and Heat Capacity

Learning Objectives

- Observe heat transfer and change in temperature and mass.
- Calculate final temperature after heat transfer between two objects.

One of the major effects of heat transfer is temperature change: heating increases the temperature while cooling decreases it. We assume that there is no phase change and that no work is done on or by the system. Experiments show that the transferred heat depends on three factors—the change in temperature, the mass of the system, and the substance and phase of the substance.

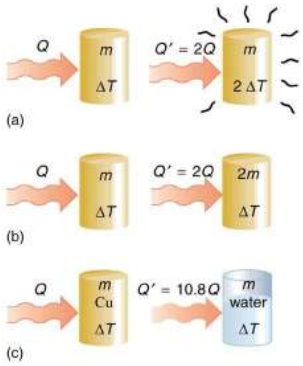


Figure 6.5.1: The heat Q transferred to cause a temperature change depends on the magnitude of the temperature change, the mass of the system, and the substance and phase involved. (a) The amount of heat transferred is directly proportional to the temperature change. To double the temperature change of a mass m, you need to add twice the heat. (b) The amount of heat transferred is also directly proportional to the mass. To cause an equivalent temperature change in a doubled mass, you need to add twice the heat. (c) The amount of heat transferred depends on the substance and its phase. If it takes an amount Q of heat to cause a temperature change ΔT in a given mass of copper, it will take 10.8 times that amount of heat to cause the equivalent temperature change in the same mass of water assuming no phase change in either substance.

The dependence on temperature change and mass are easily understood. Owing to the fact that the (average) kinetic energy of an atom or molecule is proportional to the absolute temperature, the internal energy of a system is proportional to the absolute temperature and the number of atoms or molecules. Owing to the fact that the transferred heat is equal to the change in the internal energy, the heat is proportional to the mass of the substance and the temperature change. The transferred heat also depends on the substance so that, for example, the heat necessary to raise the temperature is less for alcohol than for water. For the same substance, the transferred heat also depends on the phase (gas, liquid, or solid).

HEAT TRANSFER AND TEMPERATURE CHANGE

The quantitative relationship between heat transfer and temperature change contains all three factors:

$$Q = mc\Delta T,$$

where Q is the symbol for heat transfer, m is the mass of the substance, and ΔT is the change in temperature. The symbol c stands for **specific heat** and depends on the material and phase. The specific heat is the amount of heat necessary to change the temperature of 1.00 kg of mass by 1.00ºC. The specific heat c is a property of the substance; its SI unit is $\mathrm{J/(kg \cdot K)}$ or $\mathrm{J/(kg \cdot °C)}$. Recall that the temperature change (ΔT) is the same in units of kelvin and degrees Celsius. If heat transfer is measured in kilocalories, then *the unit of specific heat* is $\mathrm{kcal/(kg \cdot °C)}$.

Values of specific heat must generally be looked up in tables, because there is no simple way to calculate them. In general, the specific heat also depends on the temperature. Table 6.5.1 lists representative values of specific heat for various substances. Except for gases, the temperature and volume dependence of the specific heat of most substances is weak. We see from this table that the specific heat of water is five times that of glass and ten times that of iron, which means that it takes five times as much heat to raise the temperature of water the same amount as for glass and ten times as much heat to raise the temperature of water as for iron. In fact, water has one of the largest specific heats of any material, which is important for sustaining life on Earth.

Example 6.5.1: Calculating the Required Heat: Heating Water in an Aluminum Pan

A 0.500 kg aluminum pan on a stove is used to heat 0.250 liters of water from $20.0°C$ to $80.0°C$. (a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

Strategy

The pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and the pan is increased by the same amount. We use the equation for the heat transfer for the given temperature change and mass of water and aluminum. The specific heat values for water and aluminum are given in Table 6.5.1.

Solution

Because water is in thermal contact with the aluminum, the pan and the water are at the same temperature.

1. Calculate the temperature difference:

$$\Delta T = T_f - T_i = 60.0°C.$$

2. Calculate the mass of water. Because the density of water is $1000\ kg/m^3$, one liter of water has a mass of 1 kg, and the mass of 0.250 liters of water is $m_w = 0.250\ kg$.
3. Calculate the heat transferred to the water. Use the specific heat of water in Table 6.5.1:

$$Q_w = m_w c_w \Delta T = (0.250\ kg)\,(4186\ J/kg°C)\,(60.0°C) = 62.8\ kJ.$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in Table 6.5.1:

$$Q_{Al} = m_{Al} c_{Al} \Delta T = (0.500\ kg)\,(900\ J/kg°C)\,(60.0°C) = 27.0 \times 10^4\ J = 27.0\ kJ.$$

5. Compare the percentage of heat going into the pan versus that going into the water. First, find the total transferred heat:

$$Q_{Total} = Q_W + Q_{Al} = 62.8\ kJ + 27.0\ kJ = 89.8\ kJ.$$

Thus, the amount of heat going into heating the pan is

$$\frac{27.0\ kJ}{89.8\ kJ} \times 100\% = 30.1\%,$$

and the amount going into heating the water is

$$\frac{62.8\ kJ}{89.8\ kJ} \times 100\% = 69.9\%.$$

Discussion

In this example, the heat transferred to the container is a significant fraction of the total transferred heat. Although the mass of the pan is twice that of the water, the specific heat of water is over four times greater than that of aluminum. Therefore, it takes a bit more than twice the heat to achieve the given temperature change for the water as compared to the aluminum pan.



Figure 6.5.2: The smoking brakes on this truck are a visible evidence of the mechanical equivalent of heat.

Table 6.5.1: Specific Heats[1] of Various Substances

Substances	Specific heat (c)	
Solids	J/kg·ºC	kcal/kg·ºC[2]
Aluminum	900	0.215
Asbestos	800	0.19
Concrete, granite (average)	840	0.20
Copper	387	0.0924
Glass	840	0.20
Gold	129	0.0308
Human body (average at 37 °C)	3500	0.83
Ice (average, -50°C to 0°C)	2090	0.50
Iron, steel	452	0.108
Lead	128	0.0305
Silver	235	0.0562
Wood	1700	0.4
Liquids		
Benzene	1740	0.415
Ethanol	2450	0.586
Glycerin	2410	0.576
Mercury	139	0.0333
Water (15.0 °C)	4186	1.000
Gases [3]		
Air (dry)	721 (1015)	0.172 (0.242)
Ammonia	1670 (2190)	0.399 (0.523)
Carbon dioxide	638 (833)	0.152 (0.199)
Nitrogen	739 (1040)	0.177 (0.248)
Oxygen	651 (913)	0.156 (0.218)
Steam (100°C)	1520 (2020)	0.363 (0.482)

Note that Example **6.5.2** is an illustration of the mechanical equivalent of heat. Alternatively, the temperature increase could be produced by a blow torch instead of mechanically.

Example 6.5.3: Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan

Suppose you pour 0.250 kg of **20.0°C** water (about a cup) into a 0.500-kg aluminum pan off the stove with a temperature of **150°C**. Assume that the pan is placed on an insulated pad and that a negligible amount of water boils off. What is the temperature when the water and pan reach thermal equilibrium a short time later?

Strategy

The pan is placed on an insulated pad so that little heat transfer occurs with the surroundings. Originally the pan and water are not in thermal equilibrium: the pan is at a higher temperature than the water. Heat transfer then restores thermal equilibrium once the water and pan are in contact. Because heat transfer between the pan and water takes place rapidly, the mass of evaporated water is negligible and the magnitude of the heat lost by the pan is equal to the heat gained by the water. The exchange of heat stops once a thermal equilibrium between the pan and the water is achieved. The heat exchange can be written as $|Q_{\text{hot}}| = Q_{\text{cold}}$.

Solution

1. Use the equation for heat transfer $Q = mc\Delta T$ to express the heat lost by the aluminum pan in terms of the mass of the pan, the specific heat of aluminum, the initial temperature of the pan, and the final temperature:

$$Q_{\text{hot}} = m_{\text{Al}} c_{\text{Al}} (T_{\text{f}} - 150^{\circ}\text{C}).$$

2. Express the heat gained by the water in terms of the mass of the water, the specific heat of water, the initial temperature of the water and the final temperature:

$$Q_{\text{cold}} = m_{\text{W}} c_{\text{W}} (T_{\text{f}} - 20.0^{\circ}\text{C}).$$

3. Note that $Q_{\text{hot}} < 0$ and $Q_{\text{cold}} > 0$ and that they must sum to zero because the heat lost by the hot pan must be the same as the heat gained by the cold water:

$$\begin{aligned} Q_{\text{cold}} + Q_{\text{hot}} &= 0, \\ Q_{\text{cold}} &= -Q_{\text{hot}}, \\ m_{\text{W}} c_{\text{W}} (T_{\text{f}} - 20.0^{\circ}\text{C}) &= -m_{\text{Al}} c_{\text{Al}} (T_{\text{f}} - 150^{\circ}\text{C}.) \end{aligned}$$

4. This is an equation for the unknown final temperature, T_{f}.
5. Bring all terms involving T_{f} on the left hand side and all other terms on the right hand side. Solve for T_{f},

$$T_{\text{f}} = \frac{m_{\text{Al}} c_{\text{Al}} (150^{\circ}\text{C}) + m_{\text{W}} c_{\text{W}} (20.0^{\circ}\text{C})}{m_{\text{Al}} c_{\text{Al}} + m_{\text{W}} c_{\text{W}}},$$

and insert the numerical values:

$$\begin{aligned} T_{\text{f}} &= \frac{(0.500\ \text{kg})(900\ \text{J/kg}^{\circ}\text{C})(150^{\circ}\text{C}) + (0.250\ \text{kg})(4186\ \text{J/kg}^{\circ}\text{C})(20.0^{\circ}\text{C})}{(0.500\ \text{kg})(900\ \text{J/kg}^{\circ}\text{C}) + (0.250\ \text{kg})(4186\ \text{J/kg}^{\circ}\text{C})} \\ &= \frac{88430\ \text{J}}{1496.5\ \text{J/}^{\circ}\text{C}} \\ &= 59.1^{\circ}\text{C}. \end{aligned}$$

Discussion

This is a typical *calorimetry* problem—two bodies at different temperatures are brought in contact with each other and exchange heat until a common temperature is reached. Why is the final temperature so much closer to 20.0ºC than 150ºC? The reason is that water has a greater specific heat than most common substances, and thus undergoes a small temperature change for a given heat transfer. A large body of water, such as a lake, requires a large amount of heat to increase its temperature appreciably. This explains why the temperature of a lake stays relatively constant during a day even when the temperature change of the air is large. However, the water temperature does change over longer times (e.g., summer to winter).

TAKE-HOME EXPERIMENT: TEMPERATURE CHANGE OF LAND AND WATER

What heats faster, land or water?

To study differences in heat capacity:

- Place equal masses of dry sand (or soil) and water at the same temperature into two small jars. (The average density of soil or sand is about 1.6 times that of water, so you can achieve approximately equal masses by using 50% more water by volume.)
- Heat both (using an oven or a heat lamp) for the same amount of time.
- Record the final temperature of the two masses.

- Now bring both jars to the same temperature by heating for a longer period of time.
- Remove the jars from the heat source and measure their temperature every 5 minutes for about 30 minutes.

Which sample cools off the fastest? This activity replicates the phenomena responsible for land breezes and sea breezes.

Exercise 6.5.1

If 25 kJ is necessary to raise the temperature of a block from 25ºC to 30ºC, how much heat is necessary to heat the block from 45ºC to 50ºC?

Answer

The heat transfer depends only on the temperature difference. Since the temperature differences are the same in both cases, the same 25 kJ is necessary in the second case.

Summary

- The transfer of heat Q that leads to a change ΔT in the temperature of a body with mass m is $Q = mc\Delta T$, where c is the specific heat of the material. This relationship can also be considered as the definition of specific heat.

Footnotes

- 1 The values for solids and liquids are at constant volume and at 25ºC, except as noted.
- 2 These values are identical in units of cal/g·ºC.
- 3 c_v at constant volume and at 20.0ºC, except as noted, and at 1.00 atm average pressure. Values in parentheses are c_p at a constant pressure of 1.00 atm.

Glossary

specific heat

the amount of heat necessary to change the temperature of 1.00 kg of a substance by 1.00 ºC

This page titled 6.5: Temperature Change and Heat Capacity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.6: Temperature Change and Heat Capacity** by OpenStax is licensed CC BY 4.0.

6.6: The First Law of Thermodynamics

Learning Objectives

- Define the first law of thermodynamics.
- Describe how conservation of energy relates to the first law of thermodynamics.
- Identify instances of the first law of thermodynamics working in everyday situations.



Figure **6.6.1**: This boiling tea kettle represents energy in motion. The water in the kettle is turning to water vapor because heat is being transferred from the stove to the kettle. As the entire system gets hotter, work is done—from the evaporation of the water to the whistling of the kettle. (credit: Gina Hamilton)

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

$$\Delta U = Q - W.$$

Here ΔU is the *change in internal energy* U of the system. Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfer into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. We use the following sign conventions: if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system and positive W takes energy from the system. Thus $\Delta U = Q - W$. Note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this—heat transfer into them takes place so that they can do work. (See Figure **6.6.2**.) We will now examine Q, W, and ΔU further.

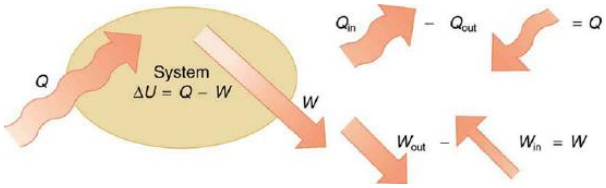


Figure **6.6.2**: The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. Q is positive for net heat transfer *into* the system. W is the total work done on and by the system. W is positive when more work is done *by* the system than on it. The change in the internal energy of the system, ΔU, is related to heat and work by the first law of thermodynamics, $\Delta U = Q - W$.

MAKING CONNECTIONS: LAW OF THERMODYNAMICS AND LAW OF CONSERVATION OF ENERGY

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

Heat Q and Work W

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

Internal Energy U

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

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

$$\Delta U = Q - W.$$

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

MAKING CONNECTIONS: MACROSCOPIC AND MICROSCOPIC

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

Section Summary

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

Glossary

first law of thermodynamics

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

internal energy

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

human metabolism

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

This page titled 6.6: The First Law of Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.8: The First Law of Thermodynamics** by OpenStax is licensed CC BY 4.0.

6.7: Heat Engines and the Second Law of Thermodynamics

Learning Objectives

- Describe the processes of a simple heat engine.
- Explain the relationship between work done by a gas and change in its volume.

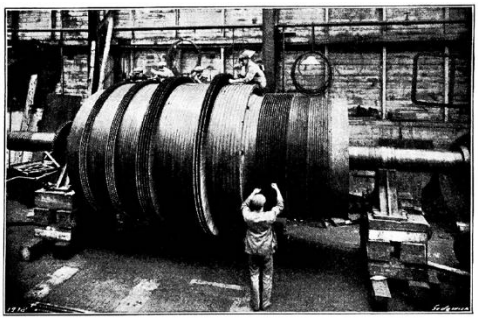


Figure **6.7.1**: Beginning with the Industrial Revolution, humans have harnessed power through the use of the first law of thermodynamics, before we even understood it completely. This photo, of a steam engine at the Turbinia Works, dates from 1911, a mere 61 years after the first explicit statement of the first law of thermodynamics by Rudolph Clausius. (credit: public domain; author unknown)

The field of thermodynamics was largely driven by the demands of the Industrial Revolution, and the interest in having mechanical devices to help do work on a very large scale. It was through this process that the Second Law of Thermodynamics was first discovered. It was not until much later that it was discovered how this law applies to all sorts of processes, even those undertaken by living creatures. In this section we will first discuss on a very broad level the types of devices which were initially used to understand thermodynamics, and then in each of the subsections we will examine the framings of the second law of thermodynamics that arose from these experiments.

The scientists of the Industrial Revolution were focused on finding ways in which to use heat transfer to do work for us. The devices they used for this investigation are called **heat engines**. Car engines and steam turbines that generate electricity are examples of heat engines. Figure **6.7.2** shows schematically how the first law of thermodynamics applies to the typical heat engine.

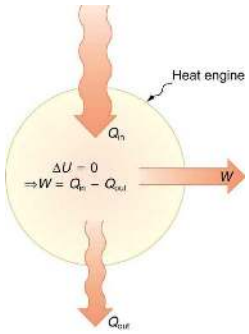


Figure **6.7.2**: Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics (and other laws of thermodynamics we will discuss later).

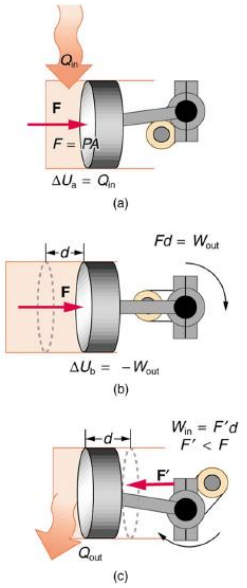


Figure **6.7.3**: (a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating that the gas's internal energy has been decreased by doing work. (c) Heat transfer to the environment further reduces pressure in the gas so that the piston can be more easily returned to its starting position.

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

Work Done by a Gas

As an example, we will examine a process by which a gas does work on a piston at constant pressure. There are other possible ways in which a gas can do work on a piston, but this is the simplest one to illustrate the point. Since the pressure is constant, the force exerted is constant and the work done is given as shown in the following derivation.

$$P\Delta V.$$

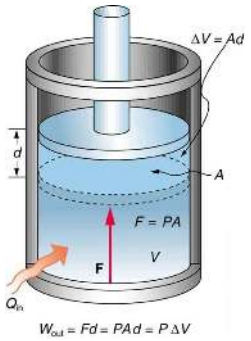


Figure **6.7.4**: An isobaric expansion of a gas requires heat transfer to keep the pressure constant. Since pressure is constant, the work done is $P\Delta V$.

Recall from mechanics that work done by a force F on an object undergoing displacement d is

$$W = Fd.$$

See the symbols as shown in Figure **6.7.4**. Now force is pressure times area ($F = PA$), and so

$$W = \text{PAd} .$$

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

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

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

(Note that the pressure involved in this work that we have called P is the pressure of the gas *inside* the tank. If we call the pressure outside the tank P_{ext}, an expanding gas would be working *against* the external pressure; the work done would therefore be $W = -P_{ext}\,\Delta V$ (isobaric process). There are some—especially chemists—who use this definition of work, and not the definition based on internal pressure, as the basis of the First Law of Thermodynamics. This definition reverses the sign conventions for work, and results in a statement of the first law that becomes $\Delta U = Q + W$. In this textbook, we will use the *physics* convention of using work done by the system on the surrounding, not the other way around.)

This is the key lesson from the above derivation: *a gas expanding under pressure does work on its surrounding, and unless additional energy is added through heat transfer, the internal energy of the gas decreases*. We will examine the experimental results that come about as a consequence of this fact later.

Thermodynamic Processes

We introduced an example process above in discussing work done by a gas. This process is an example of a **thermodynamic process**. A thermodynamic process describes a change that happens to a gas, which results in change in its pressure (P), volume (V), and/or temperature (T). There are several named thermodynamic processes: isobaric, isochoric, isothermal, and adiabatic. (It turns out the one we considered was isobaric.) These processes are given special names because they occur under some restrictions, which gives them their special properties. The details of these processes are beyond the scope of this course, but they have been well studied and lead to some very interesting results which help us to understand the Second Law of Thermodynamics.

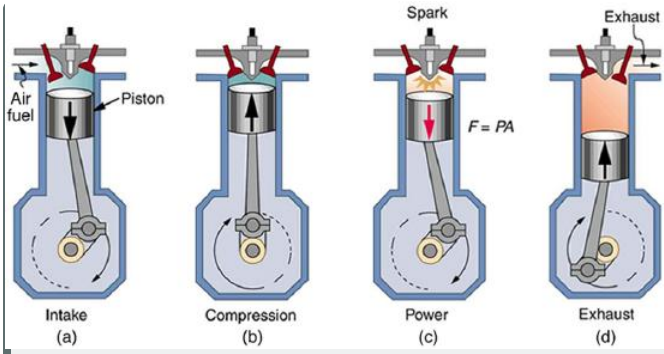


Figure 6.7.5: This thermodynamic process is called the Otto Cycle, and is an example of a cyclic process. In this cycle, a mixture of air and fuel enters the chamber and is ignited. The chemical reaction that occurs creates a large pressure which does work in part (c). However, the work done is not equal to all of the energy from the chemical reaction, as some of the energy is released as a hot exhaust gas. The internal combustion engines of gasoline powered vehicles are an example of the Otto Cycle.

Reversible Processes

Some of these thermodynamic processes are reversible in principle. A **reversible process** is one in which both the system and its environment can return to exactly the states they were in by following the reverse path. Real macroscopic processes are never exactly reversible. In the previous example, our system is a gas (like that in Figure 6.7.4), and its environment is the piston, cylinder, and the rest of the universe. If there are any energy-dissipating mechanisms, such as friction or turbulence, then heat transfer to the environment occurs for either direction of the piston. So although the gas inside the piston might be returned to its original state, the environment will not—it will have been heated. Reversibility requires the direction of heat transfer to reverse for the reverse path. Since dissipative mechanisms cannot be completely eliminated, real processes cannot be reversible.

There must be reasons that real macroscopic processes cannot be reversible. We can imagine them going in reverse. For example, heat transfer occurs spontaneously from hot to cold and never spontaneously the reverse. Yet it would not violate the first law of thermodynamics for this to happen. In fact, all spontaneous processes, such as bubbles bursting, never go in reverse. There is a

second thermodynamic law that forbids them from going in reverse. When we study this law, we will learn something about nature and also find that such a law limits the efficiency of heat engines. We will find that heat engines with the greatest possible theoretical efficiency would have to use reversible processes, and even they cannot convert all heat transfer into doing work.

Section Summary

- One of the important implications of the first law of thermodynamics is that machines can be harnessed to do work that humans previously did by hand or by external energy supplies such as running water or the heat of the Sun. A machine that uses heat transfer to do work is known as a heat engine.
- There are several simple processes, used by heat engines, that flow from the first law of thermodynamics, and they differ from one another based on how they affect pressure, volume, temperature, and heat transfer.
- If the work done is performed on the outside environment, work (W) will be a positive value. If the work done is done to the heat engine system, work (W) will be a negative value.
- Some thermodynamic processes are reversible in theory; that is, both the thermodynamic system and the environment can be returned to their initial states. However, because of loss of energy owing to the second law of thermodynamics, complete reversibility does not work in practice.

Glossary

heat engine

a device that converts heat transfer into work

thermodynamic process

describes a change to a gas, which results in changes to its pressure, volume, and/or temperature

reversible process

both the system and its environment can return to exactly the states they were in by following the reverse path.

This page titled 6.7: Heat Engines and the Second Law of Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.9: The First Law of Thermodynamics and Heat Engine Processes** by OpenStax is licensed CC BY 4.0.

6.7.1: Efficiency and the Second Law of Thermodynamics

Learning Objectives

- State the expressions of the second law of thermodynamics.
- Calculate the efficiency of a coal-fired electricity plant, using second law characteristics.



Figure **6.7.1.1**: These ice floes melt during the Arctic summer. Some of them refreeze in the winter, but the second law of thermodynamics predicts that it would be extremely unlikely for the water molecules contained in these particular floes to reform the distinctive alligator-like shape they formed when the picture was taken in the summer of 2009. (credit: Patrick Kelley, U.S. Coast Guard, U.S. Geological Survey)

The second law of thermodynamics deals with the direction taken by spontaneous processes. Many processes occur spontaneously in one direction only—that is, they are irreversible, under a given set of conditions. Here are some examples of irreversible processes seen in day-to-day life: a glass breaking, heat transferring from a hot object to a cold object, conversion of kinetic energy into thermal energy, and a puff of gas expanding from a corner of a vacuum chamber into the entire chamber. If you saw a video of glass un-breaking, or two objects in thermal contact where one gets hotter and the other gets colder, or a hot stationary object spontaneously cooling off and moving in some direction, or a gas in a chamber "regrouping" into a corner, you would say this is a video run backwards. This is because you intuitively recognize these are examples of **irreversible process**, and these processes in nature only take place in the forward direction, not reverse (See Figure **6.7.1.2**).

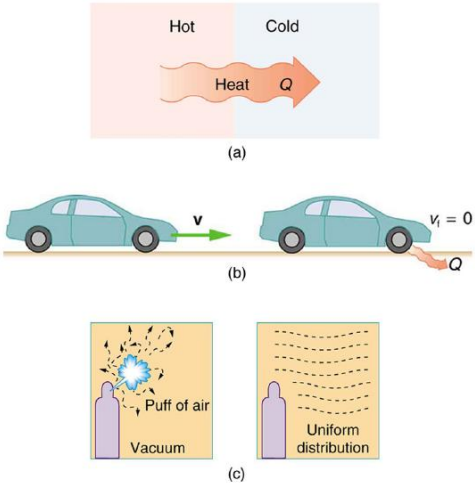


Figure **6.7.1.2**: Examples of one-way processes in nature. (a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will never return them to the corner.

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

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

THE SECOND LAW OF THERMODYNAMICS (FIRST EXPRESSION)

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

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

Heat Engines

Now let us consider a device that uses heat transfer to do work. As noted previously, such a device is called a heat engine, and one is shown schematically in Figure **6.7.1.3**(b). Gasoline and diesel engines, jet engines, and steam turbines are all heat engines that do work by using part of the heat transfer from some source. Heat transfer from the hot object (or hot reservoir) is denoted as Q_h, while heat transfer into the cold object (or cold reservoir) is Q_c, and the work done by the engine is W. The temperatures of the hot and cold reservoirs are T_h and T_c, respectively.

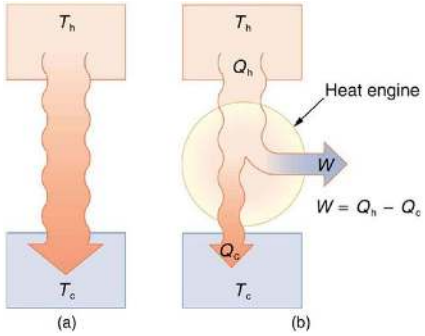


Figure **6.7.1.3**: (a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs. Q_h is the heat transfer out of the hot reservoir, W is the work output, and Q_c is the heat transfer into the cold reservoir.

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

THE SECOND LAW OF THERMODYNAMICS (SECOND EXPRESSION)

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

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

$$\Delta U = Q - W,$$

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

$$0 = Q - W,$$

so that

$$W = Q.$$

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

$$W = Q_h - Q_c \text{ (cyclical process)},$$

just as shown schematically in Figure 6.7.1.3(b). The problem is that in all processes, there is some heat transfer Q_c to the environment—and usually a very significant amount at that.

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

$$Eff = \frac{W}{Q_h}.$$

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

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

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

✔ Example 6.7.1.1: Daily Work Done by a Coal-Fired Power Station, Its Efficiency and Carbon Dioxide Emissions

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

Strategy for (a)

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

Solution for (a)

Work output is given by:

$$W = Q_h - Q_c.$$

Substituting the given values:

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

Strategy for (b)

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

Solution for (b)

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

$$Eff = \frac{1.02 \times 10^{14}\ \text{J}}{2.50 \times 10^{14}\ \text{J}} = 0.408,\ \text{or } 40.8\%$$

Strategy for (c)

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

Solution for (c)

The daily coal consumption is

$$\frac{2.50 \times 10^{14}\ \text{J}}{2.50 \times 10^{6}\ \text{J/kg}} = 1.0 \times 10^{8}\ \text{kg}.$$

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

$$1.0 \times 10^{8}\ \text{kg coal} \times \frac{44\ \text{kg}CO_2}{12\ \text{kg coal}} = 3.7 \times 10^{8}\ \text{kg}CO_2.$$

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

Discussion

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

With the information given in Example **6.7.1.1**, we can find characteristics such as the efficiency of a heat engine without any knowledge of how the heat engine operates.

Section Summary

- The two expressions of the second law of thermodynamics are: (i) Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction; and (ii) It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state. These expressions describe the direction of irreversible processes that occur in nature.
- Cyclical processes are processes that return to their original state at the end of every cycle.
- In a cyclical process, such as a heat engine, the net work done by the system equals the net heat transfer into the system, or $W = Q_h - Q_c$, where Q_h is the heat transfer from the hot object (hot reservoir), and Q_c is the heat transfer into the cold object (cold reservoir).
- Efficiency can be expressed as $Eff = \frac{W}{Q_h}$, the ratio of work output divided by the amount of energy input.

Glossary

irreversible process

a process which occurs in only one direction in nature; a process that cannot be exactly reversed

second law of thermodynamics

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

cyclical process

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

---

This page titled 6.7.1: Efficiency and the Second Law of Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.10: Introduction to the Second Law of Thermodynamics- Heat Engines and Their Efficiency** by OpenStax is licensed CC BY 4.0.

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

Learning Objectives

- Calculate maximum theoretical efficiency of any process involving the transfer of heat.
- Explain how dissipative processes affect the ideal Carnot engine.



Figure **6.7.2.1**: This novelty toy, known as the drinking bird, is an example of Carnot's engine. It contains methylene chloride (mixed with a dye) in the abdomen, which boils at a very low temperature—about 100ºF. To operate, one gets the bird's head wet. As the water evaporates, fluid moves up into the head, causing the bird to become top-heavy and dip forward back into the water. This cools down the methylene chloride in the head, and it moves back into the abdomen, causing the bird to become bottom heavy and tip up. (credit: Arabesk.nl, Wikimedia Commons)

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

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

CARNOT ENGINE

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

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

Figure **6.7.2.2** shows the PV diagram for a Carnot cycle. The cycle comprises two isothermal and two adiabatic processes. Recall that both isothermal and adiabatic processes are, in principle, reversible.

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

$$Eff = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h}.$$

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

$$Eff_C = 1 - \frac{T_c}{T_h},$$

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

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

It is also apparent that the greatest efficiencies are obtained when the ratio T_c/T_h is as small as possible. This means that efficiency is greatest for the highest possible temperature of the hot reservoir and lowest possible temperature of the cold reservoir. The actual reservoir temperatures of a heat engine are usually related to the type of heat source and the temperature of the environment into which heat transfer occurs.

Example 6.7.2.1: Maximum Theoretical Efficiency for a Nuclear Reactor

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

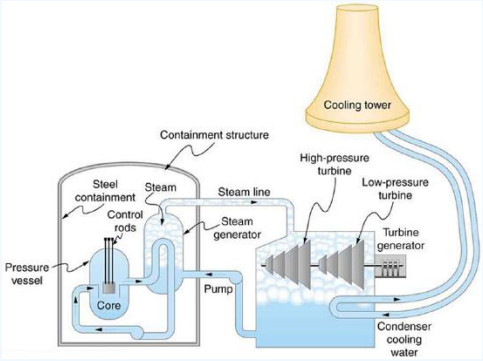


Figure 6.7.2.3: Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than operating a single generator that produces the same amount of electrical energy. The steam is condensed to liquid before being returned to the heat exchanger, to keep exit steam pressure low and aid the flow of steam through the turbines (equivalent to using a lower-temperature cold reservoir). The considerable energy associated with condensation must be dissipated into the local environment; in this example, a cooling tower is used so there is no direct heat transfer to an aquatic environment. (Note that the water going to the cooling tower does not come into contact with the steam flowing over the turbines.)

Strategy

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

Solution

The hot and cold reservoir temperatures are given as 300°C and 27.0°C, respectively. In kelvins, then, $T_h = 573\ \text{K}$ and $T_c = 300\ \text{K}$, so that the maximum efficiency is

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

Thus,

$$Eff_C = 1 - \frac{300\ \text{K}}{573\ \text{K}} = 0.476,\ \text{or}\ 47.6\%.$$

Discussion

A typical nuclear power station's actual efficiency is about 35%, a little better than 0.7 times the maximum possible value, a tribute to superior engineering. Electrical power stations fired by coal, oil, and natural gas have greater actual efficiencies (about 42%), because their boilers can reach higher temperatures and pressures. The cold reservoir temperature in any of these power stations is limited by the local environment. Figure **6.7.2.4** shows (a) the exterior of a nuclear power station and (b) the exterior of a coal-fired power station. Both have cooling towers into which water from the condenser enters the tower near the top and is sprayed downward, cooled by evaporation.

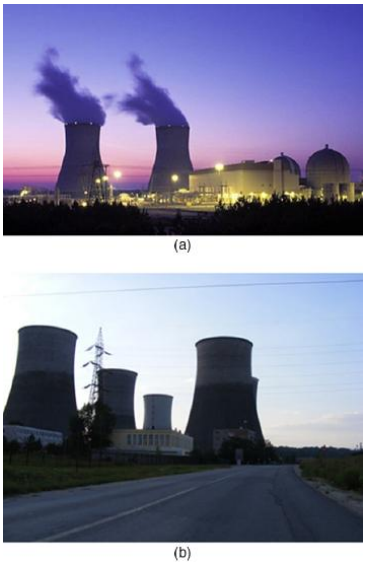


Figure **6.7.2.4**: (a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station. Both have cooling towers in which water evaporates into the environment, representing Q_c. The nuclear reactor, which supplies Q_h, is housed inside the dome-shaped containment buildings. (credit: Robert & Mihaela Vicol, publicphoto.org)

Since all real processes are irreversible, the actual efficiency of a heat engine can never be as great as that of a Carnot engine, as illustrated in Figure **6.7.2.5**(a). Even with the best heat engine possible, there are always dissipative processes in peripheral equipment, such as electrical transformers or car transmissions. These further reduce the overall efficiency by converting some of the engine's work output back into heat transfer, as shown in Figure **6.7.2.5**(b).

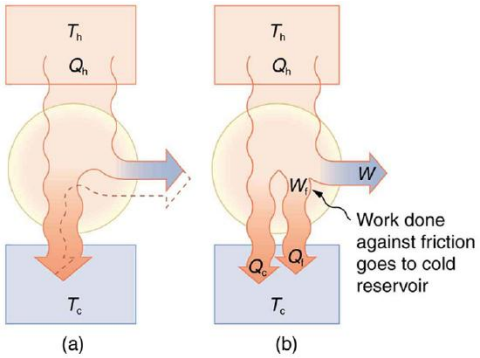


Figure **6.7.2.5**: Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some of its work output into heat transfer to the environment.

Section Summary

- The Carnot cycle is a theoretical cycle that is the most efficient cyclical process possible. Any engine using the Carnot cycle, which uses only reversible processes, is known as a Carnot engine.
- Any engine that uses the Carnot cycle enjoys the maximum theoretical efficiency.

- While Carnot engines are ideal engines, in reality, no engine achieves Carnot's theoretical maximum efficiency, since dissipative processes, such as friction, play a role. Carnot cycles without heat loss may be possible at absolute zero, but this has never been seen in nature.

Glossary

Carnot cycle
a cyclical process that uses only reversible processes.

Carnot engine
a heat engine that uses a Carnot cycle

Carnot efficiency
the maximum theoretical efficiency for a heat engine

This page titled 6.7.2: Carnot's Perfect Heat Engine- The Second Law of Thermodynamics Restated is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.11: Carnot's Perfect Heat Engine- The Second Law of Thermodynamics Restated** by OpenStax is licensed CC BY 4.0.

6.8: Entropy and the Second Law of Thermodynamics

Learning Objectives

- Define entropy and calculate the increase of entropy in a system with reversible and irreversible processes.
- Calculate the increasing disorder of a system.



Figure 6.8.1: The ice in this drink is slowly melting. Eventually the liquid will reach thermal equilibrium, as predicted by the second law of thermodynamics. (credit: Jon Sullivan, PDPhoto.org)

There is yet another way of expressing the second law of thermodynamics. This version relates to a concept called **entropy**. By examining it, we shall see that the directions associated with the second law—heat transfer from hot to cold, for example—are related to the tendency in nature for systems to become disordered and for less energy to be available for use as work. The entropy of a system can in fact be shown to be a measure of its disorder and of the unavailability of energy to do work. We will first present the physical way in which this can be measured, and then discuss additional aspects of it in the subsections. In the first subsection we will discuss why entropy works the way it does. In the second subsection we will discuss some of the implications of this for us, the world we live in, and the universe as a whole.

MAKING CONNECTIONS: ENTROPY, ENERGY, AND WORK

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

A value for entropy can be derived from the Carnot Cycle:

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

where Q is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin (J/K).

The definition of ΔS is strictly valid only for reversible processes, such as used in a Carnot engine. However, we can find ΔS precisely even for real, irreversible processes. The reason is that the entropy S of a system, like internal energy U, depends only on the state of the system and not how it reached that condition. Entropy is a property of state. Thus, the change in entropy ΔS of a system between state 1 and state 2 is the same no matter how the change occurs. We just need to find or imagine a reversible process that takes us from state 1 to state 2 and calculate ΔS for that process. That will be the change in entropy for any process going from state 1 to state 2. (See Figure 6.8.2.)

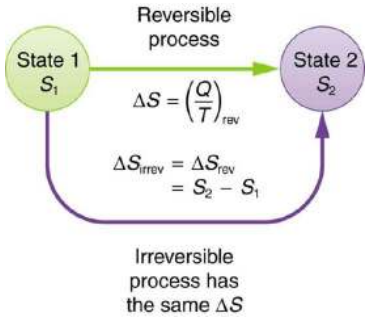


Figure 6.8.2: When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS, whether a hypothetical reversible path is followed or a real irreversible path is taken.

We can also say that *the total change in entropy for a system in any reversible process is zero*. The entropy of various parts of the system may change, but the total change is zero. Furthermore, the system does not affect the entropy of its surroundings, since heat transfer between them does not occur. Thus, the reversible process changes neither the total entropy of the system nor the entropy of its surroundings. Sometimes this is stated as follows: *Reversible processes do not affect the total entropy of the universe.* Real processes are not reversible, though, and they do change total entropy.

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

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

With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process.

THE SECOND LAW OF THERMODYNAMICS (FOURTH EXPRESSION)

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

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

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

Order to Disorder

Entropy is related not only to the unavailability of energy to do work—it can also be thought of as a measure of disorder. (Perhaps more accurately phrased as dispersion.) This notion was initially postulated by Ludwig Boltzmann in the 1800s, and will be examined in more detail in the following subsection. For example, melting a block of ice means taking a highly structured and orderly system of water molecules and converting it into a disorderly liquid in which molecules have no fixed positions. (See Figure 6.8.4.) There is a large increase in entropy in the process, as seen in the following example.

Example 6.8.2: Entropy Associated with Disorder

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

Strategy

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

Solution

The change in entropy is defined as:

$$\Delta S = \frac{Q}{T}.$$

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

$$Q = mL_f,$$

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

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

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

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

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

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

Discussion

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

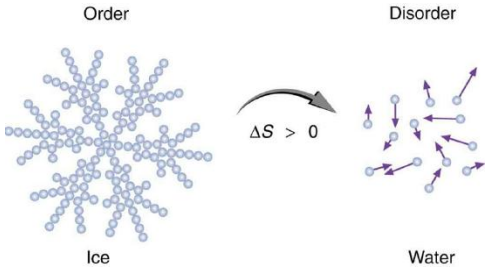


Figure 6.8.4: When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without fixed locations or orientations. Its entropy increases because heat transfer occurs into it. Entropy is a measure of disorder.

Section Summary

- Entropy is the loss of energy available to do work.
- Another form of the second law of thermodynamics states that the total entropy of a system either increases or remains constant; it never decreases.
- Change of entropy is zero in a reversible process; it increases in an irreversible process.
- Entropy is also associated with the tendency toward disorder in an *isolated* system.

Glossary

entropy

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

change in entropy

the ratio of heat transfer to temperature Q/T

second law of thermodynamics stated in terms of entropy

the total entropy of a system either increases or remains constant; it never decreases

This page titled 6.8: Entropy and the Second Law of Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.13: Entropy and the Second Law of Thermodynamics- Disorder and the Unavailability of Energy** by OpenStax is licensed CC BY 4.0.

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

Learning Objectives

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



Figure **6.8.1.1**: When you toss a coin a large number of times, heads and tails tend to come up in roughly equal numbers. Why does heads not come up 100, 90, or even 80% of the time? (credit: Jon Sullivan, PDPhoto.org)

The various ways of formulating the second law of thermodynamics tell what happens rather than why it happens. Why should heat transfer occur only from hot to cold? Why should energy become ever less available to do work? Why should the universe become increasingly disorderly? The answer is that it is a matter of overwhelming probability. Disorder is simply vastly more likely than order.

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

Coin Tosses

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

Table **6.8.1.1**: Possible Outcomes of 5 Coin Tosses

Combination of Heads and Tails	
5 heads	0 tails
4 heads	1 tails
3 heads	2 tails
2 heads	3 tails
1 heads	4 tails
0 heads	5 tails

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

Using this nomenclature, a system of 5 coins has the 6 possible macrostates just listed. Some macrostates are more likely to occur than others. For instance, there is only one way to get 5 heads, but there are several ways to get 3 heads and 2 tails, making the latter macrostate more probable. Table **6.8.1.2** lists of all the ways in which 5 coins can be tossed, taking into account the order in which heads and tails occur. Each sequence is called a **microstate**—a detailed description of every element of a system.

Table **6.8.1.2**: 5-Coin Toss

Combinations	Individual microstates	Number of microstates
5 heads, 0 tails	HHHHH	1
4 heads, 1 tail	HHHHT, HHHTH, HHTHH, HTHHH, THHHH	5
3 heads, 2 tails	HTHTH, THTHH, HTHHT, THHTH, THHHT HTHTH, THTHH, HTHHT, THHTH, THHHT	10
2 heads, 3 tails	TTHHH, TTHHT, THHTT, HHTTT, TTHTH, THTHT, HTHTT, THTTH, HTTHT, HTTTH	10
1 head, 4 tails	TTTTH, TTTHT, TTHTT, THTTT, HTTTT	5
0 heads, 5 tails	TTTTT	1
		Total: 32

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

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

Table **6.8.1.3**: 100-Coin Toss

Macrostate		Number of microstates
Heads	Tails	(W)
100	0	1
99	1	1.0×10^{2}
95	5	7.5×10^{7}
90	10	1.7×10^{13}
75	25	2.4×10^{23}
60	40	1.4×10^{28}
55	45	6.1×10^{28}
51	49	9.9×10^{28}
50	50	1.0×10^{29}
49	51	9.9×10^{28}
45	55	6.1×10^{28}
40	60	1.4×10^{28}
25	75	2.4×10^{23}
10	90	1.7×10^{13}
5	95	7.5×10^{7}
1	99	1.0×10^{2}
0	100	1
		Total: 1.27×10^{30}

This result becomes dramatic for larger systems. Consider what happens if you have 100 coins instead of just 5. The most orderly arrangements (most structured) are 100 heads or 100 tails. The least orderly (least structured) is that of 50 heads and 50 tails. There is only 1 way (1 microstate) to get the most orderly arrangement of 100 heads. There are 100 ways (100 microstates) to get the next most orderly arrangement of 99 heads and 1 tail (also 100 to get its reverse). And there are 1.0×10^{29} ways to get 50 heads and 50 tails, the least orderly arrangement. Table **6.8.1.3** is an abbreviated list of the various macrostates and the number of microstates for each macrostate. The total number of microstates—the total number of different ways 100 coins can be tossed—is an impressively large $1.27\times10^{30.}$ Now, if we start with an orderly macrostate like 100 heads and toss the coins, there is a virtual certainty that we will get a less orderly macrostate. If we keep tossing the coins, it is possible, but exceedingly unlikely, that we will ever get back to the most orderly macrostate. If you tossed the coins once each second, you could expect to get either 100 heads or 100 tails once in 2×10^{22} years! This period is 1 trillion (10^{12}) times longer than the age of the universe, and so the chances are essentially zero. In contrast, there is an 8% chance of getting 50 heads, a 73% chance of getting from 45 to 55 heads, and a 96% chance of getting from 40 to 60 heads. Disorder is highly likely.

Disorder in a Gas

The fantastic growth in the odds favoring disorder that we see in going from 5 to 100 coins continues as the number of entities in the system increases. Let us now imagine applying this approach to perhaps a small sample of gas. Because counting microstates and macrostates involves statistics, this is called **statistical analysis**. The macrostates of a gas correspond to its macroscopic properties, such as volume, temperature, and pressure; and its microstates correspond to the detailed description of the positions and velocities of its atoms. Even a small amount of gas has a huge number of atoms: 1.0 cm^3 of an ideal gas at 1.0 atm and 0° C has

2.7×10^{19} atoms. So each macrostate has an immense number of microstates. In plain language, this means that there are an immense number of ways in which the atoms in a gas can be arranged, while still having the same pressure, temperature, and so on.

The most likely conditions (or macrostates) for a gas are those we see all the time—a random distribution of atoms in space with a highly probable distribution of speeds in random directions, as predicted by kinetic theory. This is the most disorderly and least structured condition we can imagine. In contrast, one type of very orderly and structured macrostate has all of the atoms in one corner of a container with identical velocities. There are very few ways to accomplish this (very few microstates corresponding to it), and so it is exceedingly unlikely ever to occur. (See Figure **6.8.1.2**(b).) Indeed, it is so unlikely that we have a law saying that it is impossible, which has never been observed to be violated—the second law of thermodynamics.

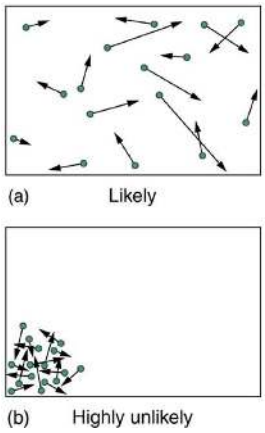


Figure **6.8.1.2**: (a) The ordinary state of gas in a container is a disorderly, random distribution of atoms or molecules with a Maxwell-Boltzmann distribution of speeds. It is so unlikely that these atoms or molecules would ever end up in one corner of the container that it might as well be impossible. (b) With energy transfer, the gas can be forced into one corner and its entropy greatly reduced. But left alone, it will spontaneously increase its entropy and return to the normal conditions, because they are immensely more likely.

The disordered condition is one of high entropy, and the ordered one has low entropy. With a transfer of energy from another system, we could force all of the atoms into one corner and have a local decrease in entropy, but at the cost of an overall increase in entropy of the universe. If the atoms start out in one corner, they will quickly disperse and become uniformly distributed and will never return to the orderly original state (Figure **6.8.1.2**(b)). Entropy will increase. With such a large sample of atoms, it is possible —but unimaginably unlikely—for entropy to decrease. Disorder is vastly more likely than order.

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

$$S = k \ln W,$$

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

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

Section Summary

- Disorder is far more likely than order, which can be seen statistically.
- The entropy of a system in a given state (a macrostate) can be written as

$$S = k \ln W,$$

where $k=1.38\times10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate.

Glossary

macrostate

an overall property of a system

microstate

each sequence within a larger macrostate

statistical analysis

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

---

This page titled 6.8.1: Statistical Interpretation of Entropy and the Second Law of Thermodynamics- The Underlying Explanation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.14: Statistical Interpretation of Entropy and the Second Law of Thermodynamics- The Underlying Explanation** by OpenStax is licensed CC BY 4.0.

6.8.2: Entropy, Life, the Universe, and Everything

Learning Objective

- Apply the Second Law of Thermodynamics to a variety of situations.

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

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

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

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

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

To summarize, we had several statements of the Second Law of Thermodynamics:

THE SECOND LAW OF THERMODYNAMICS (FIRST EXPRESSION)

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

THE SECOND LAW OF THERMODYNAMICS (SECOND EXPRESSION)

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

THE SECOND LAW OF THERMODYNAMICS (THIRD EXPRESSION)

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

THE SECOND LAW OF THERMODYNAMICS (FOURTH EXPRESSION)

The total entropy of the universe increases or remains constant in any process; it never decreases.

As we shall see with the ensuing examples, this fourth expression is the most generalizable, as it applies to systems which involve flows of energy beyond just heat.

Additionally, the reason why processes work in this way can be explained by the Boltzmann equation, which relates the number of possible arrangements of a thermodynamic system with its entropy. the Boltzmann equation is the mathematical description of the disorder that we understand entropy to be a measure of. These results are numerically equivalent for an ideal gas, and more advanced mathematical techniques can result in an equivalent relationship for more complicated systems.

BOLTZMANN EQUATION FOR RELATIONSHIP BETWEEN ENTROPY AND STATISTICAL ANALYSIS OF THE SYSTEM

$$S = k \ln W$$

We can see the connection between these different statements of The Second Law of Thermodynamics and our example of the mixed water at the start of this chapter.

It turns out that the Second Law of Thermodynamics is not just true for heat engines that we have manufactured, but for every process in the universe. However, it might be beneficial to start with another manufactured item before moving on to natural processes: computers. At their most basic level, computers are devices that organize information: whether that is a database or an aesthetically appealing video game, the information that we see on a computer screen is of low probability compared to all of the possible ways in which the pixels on the screen might be arranged. In order for your computer to organize this information, it needs a source of energy. It also must release some of this energy to the surrounding environment through the cooling system (generally a fan). We can think of your computer like a heat engine: energy in the form of electricity flows into it via the plug, performs work by organizing data (thereby lowering the entropy of those data), and then emitting heat to the environment (thereby increasing the entropy of the air surrounding the computer). (Strictly speaking, a computer is not a heat engine, because the electricity coming into it is of a much more useful form than heat, but we can see the same limits to what is possible for energy through this example.)

Turning to biological systems, the same ideas apply but with more complexity. As adult human beings, we are much more organized than the zygotes we existed as before birth. How is this possible if entropy is increasing? It is possible because we have eaten a lot of food and breathed a lot of air throughout our lives. The food we eat and the air we breathe contain energy that we extract from it via chemical reactions in our bodies. (We measure the amount of energy extracted in units of Calories, as labeled on our food containers.) These reactions provide the energy for us to grow. Part of this growth includes the organization of our body parts, including our brain. We also have a variety of waste matter that leaves our body. This waste matter does have significant potential energy stored within its chemical bonds, but it is a much lower amount than that stored in the bonds of the matter we take in through food and air. Although the energy entering and leaving our bodies is in the form of chemical energy stored in bonds, we see the same principle as we do with heat engines: some energy comes into a system and does useful work, but some amount of that energy leaves the system without doing any work. If we stopped eating, we would eventually die because the system of our body is much more ordered than our surroundings.

Finally, we can consider the Earth itself. Almost all of the energy on the Earth comes from the Sun. Most of this energy is eventually radiated back into space, but some of it powers chemical reactions which result in a more ordered environment. One of the most abundant such reactions is called photosynthesis, which turns low energy carbon dioxide into higher energy oxygen and carbon. The carbon becomes a part of the ordered structure of the organism which produced the photosynthesis, while the oxygen provides a readily available fuel for other organisms in the environment to use in the ordering of their own structures. If there were no sunlight, life on Earth would not be possible.

You might be wondering about all of the other planets where life does not exist (at least that we know of). Why does the application of the Second Law allow for life on Earth, but there is no life in those other places? The reason is because useful work was extracted on Earth but not in those other places. If we think of the example of mixing different temperature waters at the start of this section, that was another possibility for useful work to be extracted that didn't occur. If the situation had been a little bit different, useful work could have been extracted from it. The situations on Earth so happen to be those which allow for the extraction of useful work in the flow of energy and those conditions do not exist on those other planets. Just because it is possible to extract useful work does not mean that it always happens. (Consider all of the waste food in the United States, for example.) We are very fortunate to exist on a planet where life is possible!

Section Summary

- The Second Law of Thermodynamics is generalizable beyond what we have defined for heat engines.
- Some other applications include computers, biological life forms, and environments that support such life forms.

This page titled 6.8.2: Entropy, Life, the Universe, and Everything is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- **8.13: Entropy and the Second Law of Thermodynamics- Disorder and the Unavailability of Energy** by OpenStax is licensed CC BY 4.0.
- **8.10: Introduction to the Second Law of Thermodynamics- Heat Engines and Their Efficiency** by OpenStax is licensed CC BY 4.0.
- **8.11: Carnot's Perfect Heat Engine- The Second Law of Thermodynamics Restated** by OpenStax is licensed CC BY 4.0.

6.9: Energy Sources for Our Lives

Learning Objectives

- Know the diverse energy sources used in the United States.
- Apply the laws of thermodynamics in understanding how these energy sources function.

Ready access to energy is perhaps the greatest luxury of the modern world. It is also a limited resource which we have the necessary knowledge of the laws of nature to understand in context. Let us first consider where our energy comes from and where it might continue to come from going forward. Then we will consider how both the first and second laws of thermodynamics relate to our energy consumption.

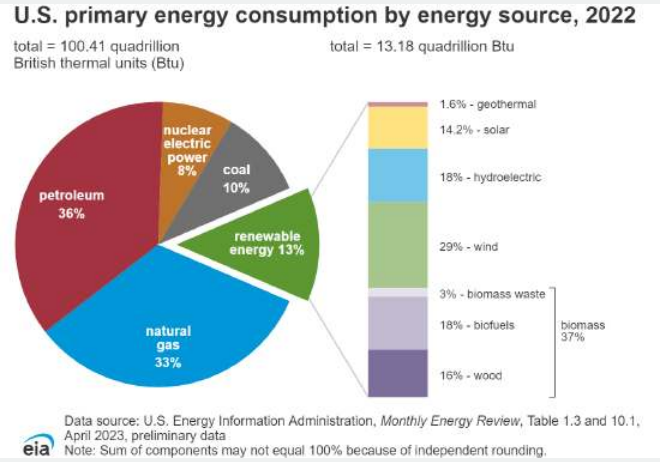


Figure **6.9.1**: The energy sources that were used for US primary energy consumption from 2022.

More than 90% of the energy use in the United States ultimately came from the Sun. (The two exceptions are nuclear electric power and geothermal energy. Nuclear electric power will be discussed more in a later chapter.) The three leading sources of energy (petroleum, natural gas, and coal, which are collectively referred to as fossil fuels) are all stored energy that initially came from the Sun. This energy was first stored via photosynthesis of living organisms, and remained stored in chemical bonds for hundreds of millions of years after those organisms died. When these resources are extracted from the Earth, it is an extraction of hundreds of millions of years of sunlight that will take an additional hundreds of millions of years to replenish.

Almost all of the renewable energy is also energy that ultimately comes from the Sun. The largest chunk of this energy is biomass, which is simply the renewable form of fossil fuels: energy stored in chemical bonds due to photosynthesis. In this case, that stored energy has not been buried within the earth for hundreds of millions of years. We could continue to rely on it without worrying that it will run out some day (although there are other considerations for this kind of energy that we will explore in a later chapter). Wind and hydroelectric power are also based on changes to our environment brought about by the Sun: there would be no wind or rain if it was not for the Sun.

One of the challenges with renewable energy sources is that the Sun does not always shine and the wind does not always blow. Because of this inconsistency, these energy sources have largely been used to supplement the existing fossil fuel-based energy economy. So long as we want to flip a switch to power a device at any time of day, we will need a consistent supply of energy.

One solution to this problem of inconsistent access to energy is finding renewable ways to store this energy when it is available for capture as either sunlight or wind and then use it at a later time when it is needed. There are a variety of technologies that have been created to aid in this process, but they are more expensive than the cheap fossil fuels we have become accustomed to. Such technologies include hydrogen electrolysis, several varieties of grid-scale batteries, pumped-storage hydropower, and solid mass

gravity batteries. Some of these technologies are currently being used at small scale, while others have been proposed or are under construction.

The largest challenge to cheaply available public power, however, is not the limited times when renewable energy is available, but in the high demand for energy throughout the residential sector at certain times of day. You may be familiar with time of use variations in charges on your energy bills. These charges are meant to incentivize people to space out their energy consumption where possible so as to avoid using more expensive energy sources during times with higher demand. If the demand becomes too high, such as when temperatures are high and people are running their air conditioners, there could be intermittent power outages throughout some of the area where this power is being supplied.

The First Law and Energy Sources in Our Lives

The conversation to this point has been about applying the First Law of Thermodynamics to our use of energy. We can think of the internal energy of the Earth changing as we extract resources from it and as the Sun provides additional energy. We have a net Q coming to us from the Sun. Some of this energy was stored within the Earth to increase the internal energy for hundreds of millions of years, but now those energy storage reservoirs are being depleted. Currently, this internal energy of stored sunlight is decreasing as we tap into it to do all the work we want whenever we power on a device.

Likewise, all of the ways in which we have found to harness renewable energy sources are based on an understanding of the First Law. Each of these methods is centered around the idea that all energy can be changed to other forms of energy, and in the case of humans we prefer electrical energy. Realizing that the gravitational potential energy of water behind a dam or the chemical potential energy within a biomass could be converted to electrical energy is necessary to create a functioning electrical grid. Likewise, conservation of energy principles are at play in determining electricity rates at different times of day based on the energy sources that are always available and that could be made available but at a higher cost. The energy demanded must be supplied, but the cost of the supply energy can vary quite a lot.

The Second Law and Energy Sources in our Lives

It turns out that the Second Law of Thermodynamics also plays a role in supplying us with the electricity we have become accustomed to. As was mentioned earlier, the Second Law was discovered during the Industrial Revolution as researchers tried to maximize the amount of work that could be done due to heating an object. From this scientific research we know that there is a limit to the amount of useful work that can be obtained by heat flowing between objects at different temperatures. In fact, the best way to maximize the efficiency is by making the highest possible temperature for the hot reservoir and the lowest possible temperature for the cold reservoir.

If we apply this idea to the ways in which our power is generated and used, we can see why some options are preferable to others. One application is the massive energy needs of data servers. There are more and more people using the internet, and the growing demand for AI creates a need for increased server capacity. Although the increasing demands on servers have been massive, rapid improvements on energy efficiency within these servers has limited the impact. One of these improvements has been the recent increase in servers housed within very cold regions of the world. Keeping the servers cold maximizes the efficiency of the energy transfer.

Another application is in where we create energy by burning fossil fuels (or any other fuel, for that matter). Most power plants attempt to maximize the temperature of the burning substance in order to maximize the efficiency of the power production process. These power plants often operate at well in excess of 1000°C. Some of the energy generated is also stored by increasing the pressure of the steam into which this energy flows before it flows on to the electrical generator. But in the internal combustion engine of an automobile, the maximum temperature must be less than 900°C due to the limits of the engine materials. Additionally, the engine coolant system transfers much of this energy away from the engine in a way that does not result in any increase in the work being done by the car. (In contrast to the coolant system within a power plant, which is directly linked to the power generation process.) Because of these differences, a power plant is more than twice as efficient as an internal combustion engine in generating useful work.

The value of electric vehicles can be seen as relates to both the first and second law discussions. Electric vehicles can be charged at times of day that help to stabilize the load on the electric grid. Additionally, the electricity created to power an electric vehicle was done so much more efficiently at a power plant than it would be inside of an internal combustion engine. So although the battery of

an electric vehicle is not itself a source of energy, the energy that led to the powering of the electric vehicle was created more efficiently than it would have been within an internal combustion engine. One of the benefits of electric vehicles is that they will decrease the depletion of stored energy from the Sun, even if they ultimately derive their energy from a power plant that burns fossil fuels.

Long term we want to find ways to become more reliant on renewable energy sources and less so on the limited supply of fossil fuels. In later chapters we will learn about additional challenges related to the use of fossil fuels.

Section Summary

- Fossil fuels consist of petroleum, coal, and natural gas, and consist of energy in chemical bonds which has been stored for hundreds of millions of years.
- The most prevalent sources of renewable energy include biomass, solar, wind, and hydroelectric.
- All energy sources except nuclear power and geothermal energy ultimately derive from the Sun.
- The United States relies mostly on fossil fuels as an energy source, but there are also a limited number of renewable energy sources.
- More efficient systems can be designed by having bigger differences in temperature between the hot and cold reservoirs and by lowering the temperature of the cold reservoir.

This page titled 6.9: Energy Sources for Our Lives is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

CHAPTER OVERVIEW

7: Fundamentals and Applications of Electricity

7.1: Introduction to Electricity
7.2: Static Electricity and Charge- Conservation of Charge
7.3: Coulomb's Law
7.4: Representing Electric Fields
7.5: Electric Potential and Potential Energy
7.6: Conductors and Current
7.7: Ohm's Law
7.8: Circuits
7.9: Applications- Power and Safety
7.9.1: Electric Power and Energy
7.9.2: Electric Hazards and the Human Body

This page titled 7: Fundamentals and Applications of Electricity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

7.1: Introduction to Electricity



Figure 7.1.1: Static electricity from this plastic slide causes the child's hair to stand on end. The sliding motion stripped electrons away from the child's body, leaving an excess of positive charges, which repel each other along each strand of hair. (credit: Ken Bosma/Wikimedia Commons)

The image of American politician and scientist Benjamin Franklin (1706–1790) flying a kite in a thunderstorm is familiar to every schoolchild. (See Figure.) In this experiment, Franklin demonstrated a connection between lightning and **static electricity**. Sparks were drawn from a key hung on a kite string during an electrical storm. These sparks were like those produced by static electricity, such as the spark that jumps from your finger to a metal doorknob after you walk across a wool carpet. What Franklin demonstrated in his dangerous experiment was a connection between phenomena on two different scales: one the grand power of an electrical storm, the other an effect of more human proportions. Connections like this one reveal the underlying unity of the laws of nature, an aspect we humans find particularly appealing.

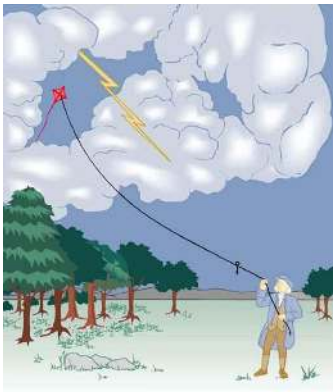


Figure 7.1.2: When Benjamin Franklin demonstrated that lightning was related to static electricity, he made a connection that is now part of the evidence that all directly experienced forces except the gravitational force are manifestations of the electromagnetic force.

Our understanding of electricity and the range of phenomena that are electrical in nature has vastly expanded since those early days. For example, atomic and molecular interactions, such as the forces of friction, cohesion, and adhesion, are now known to be manifestations of the **electromagnetic force**. Static electricity is just one aspect of the electromagnetic force, which also includes moving electricity and magnetism. Furthermore, all the macroscopic forces that we experience directly, such as the sensations of touch and the tension in a rope, are due to the electromagnetic force, one of the four fundamental forces in nature. The gravitational force, another fundamental force, is actually sensed through the electromagnetic interaction of molecules, such as between those in our feet and those on the top of a bathroom scale. (The other two fundamental forces, the strong nuclear force and the weak nuclear force, cannot be sensed on the human scale.)

In this chapter, we will develop the concepts of electric fields and electric potential (also known as *voltage*), which are used to describe electric force and electrical energy. Consider, for example, great amounts of electrical energy stored in batteries or transmitted cross-country through power lines, or consider electrical signals sent in our nervous systems at molecular levels, with

ions crossing cell membranes and transferring information, or the role of electricity in our household appliances and devices, including your laptop and smartphones. We will look at the laws that help us describe electric circuits and begin to explore some of the many applications of electricity.

Glossary

static electricity

a buildup of electric charge on the surface of an object

electromagnetic force

one of the four fundamental forces of nature; the electromagnetic force consists of static electricity, moving electricity and magnetism

---

This page titled 7.1: Introduction to Electricity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.1: Introduction to Electricity** by OpenStax is licensed CC BY 4.0.

7.2: Static Electricity and Charge- Conservation of Charge

Learning Objectives

- Define electric charge, and describe how the two types of charge interact.
- Describe three common situations that generate static electricity.
- State the law of conservation of charge.



Figure **7.2.1**: Borneo amber was mined in Sabah, Malaysia, from shale-sandstone-mudstone veins. When a piece of amber is rubbed with a piece of silk, the amber gains more electrons, giving it a net negative charge. At the same time, the silk, having lost electrons, becomes positively charged. (credit: Sebakoamber, Wikimedia Commons)

What makes plastic wrap cling? Static electricity. Not only are applications of static electricity common these days, its existence has been known since ancient times. The first record of its effects dates to ancient Greeks who noted more than 500 years B.C. that polishing amber temporarily enabled it to attract bits of straw (see Figure **7.2.1**). The very word *electric* derives from the Greek word for amber (*electron*).

Many of the characteristics of static electricity can be explored by rubbing things together. Rubbing creates the spark you get from walking across a wool carpet, for example. Static cling generated in a clothes dryer and the attraction of straw to recently polished amber also result from rubbing. Similarly, lightning results from air movements under certain weather conditions. You can also rub a balloon on your hair, and the static electricity created can then make the balloon cling to a wall. We also have to be cautious of static electricity, especially in dry climates. When we pump gasoline, we are warned to discharge ourselves (after sliding across the seat) on a metal surface before grabbing the gas nozzle. Attendants in hospital operating rooms must wear booties with aluminum foil on the bottoms to avoid creating sparks which may ignite the oxygen being used.

Some of the most basic characteristics of static electricity include:

- The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge.
- There are only two types of charge, one called positive and the other called negative.
- Like charges repel, whereas unlike charges attract.
- The force between charges decreases with distance.

How do we know there are two types of **electric charge**? When various materials are rubbed together in controlled ways, certain combinations of materials always produce one type of charge on one material and the opposite type on the other. By convention, we call one type of charge “positive”, and the other type “negative.” For example, when glass is rubbed with silk, the glass becomes positively charged and the silk negatively charged. Since the glass and silk have opposite charges, they attract one another like clothes that have rubbed together in a dryer. Two glass rods rubbed with silk in this manner will repel one another, since each rod has positive charge on it. Similarly, two silk cloths so rubbed will repel, since both cloths have negative charge. Figure **7.2.2** shows how these simple materials can be used to explore the nature of the force between charges.

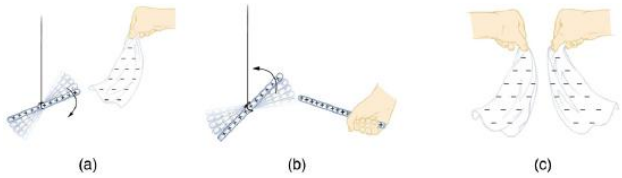


Figure **7.2.2**: A glass rod becomes positively charged when rubbed with silk, while the silk becomes negatively charged. (a) The glass rod is attracted to the silk because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

More sophisticated questions arise. Where do these charges come from? Can you create or destroy charge? Is there a smallest unit of charge? Exactly how does the force depend on the amount of charge and the distance between charges? Such questions obviously occurred to Benjamin Franklin and other early researchers, and they interest us even today.

Franklin wrote in his letters and books that he could see the effects of electric charge but did not understand what caused the phenomenon. Today we have the advantage of knowing that normal matter is made of atoms, and that atoms contain positive and negative charges, usually in equal amounts. We will learn more about atoms later in this text, but for now we will state that this negative charges come from something we call electrons. It's possible to have particles which carry either positive or negative charges, but for most of our observations of charge in the regular world, the charges we observe are due to variable amounts of electrons. A negatively charged object occurs because it has extra electrons. A positively charged object occurs because it does not have enough electrons. And an uncharged object (which is most objects) has the correct number of electrons. In each of the examples we looked at, the charge on an object occurred because electrons were transferred from one object to another, as we shall see next.

Separation of Charge in Atoms

Charges in matter can be separated—for example, by rubbing materials together. Some forms of matter have a greater affinity for electrons than others and will become negatively charged by close contact in rubbing, leaving the other material positively charged. (See Figure **7.2.6**.) Positive charge can similarly be induced by rubbing. Methods other than rubbing can also separate charges. Batteries, for example, use combinations of substances that interact in such a way as to separate charges, as we will see later in this chapter.

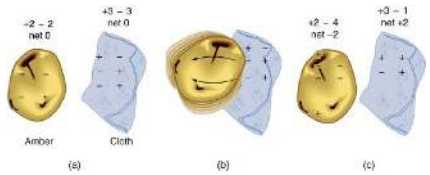


Figure **7.2.3**: When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

No charge is actually created or destroyed when charges are separated as we have been discussing. Rather, existing charges are moved about. In fact, in all situations the total amount of charge is always constant. This universally obeyed law of nature is called the **law of conservation of charge**.

LAW OF CONSERVATION OF CHARGE

Total charge is constant in any process.

The law of conservation of charge is absolute—it has never been observed to be violated. Charge, then, is a special physical quantity, joining a very short list of other quantities in nature that are always conserved. Other conserved quantities include energy, momentum, and angular momentum.

MAKING CONNECTIONS: CONSERVATION LAWS

Only a limited number of physical quantities are universally conserved. Charge is one—energy, momentum, and angular momentum are others. Because they are conserved, these physical quantities are used to explain more phenomena and form more connections than other, less basic quantities. We find that conserved quantities give us great insight into the rules followed by nature and hints to the organization of nature. Discoveries of conservation laws have led to further discoveries, such as the weak nuclear force and the quark substructure of protons and other particles.

Section Summary

- There are only two types of charge, which we call positive and negative.
- Like charges repel, unlike charges attract, and the force between charges decreases with the square of the distance.
- Most of our observations of charge involve electrons, which carry a negative charge.
- The SI unit for charge is the coulomb (C), with protons and electrons having charges of opposite sign but equal magnitude; the magnitude of this basic charge $|q_e|$ is

$$|q_e| = 1.60 \times 10^{-19}\text{C}.$$

- Whenever charge is created or destroyed, equal amounts of positive and negative are involved.
- Most often, existing charges are separated from neutral objects to obtain some net charge.
- Both positive and negative charges exist in neutral objects and can be separated by rubbing one object with another. For macroscopic objects, negatively charged means an excess of electrons and positively charged means a depletion of electrons.
- The law of conservation of charge ensures that whenever a charge is created, an equal charge of the opposite sign is created at the same time.

Glossary

electric charge

a physical property of an object that causes it to be attracted toward or repelled from another charged object; each charged object generates and is influenced by a force called an electromagnetic force

law of conservation of charge

states that whenever a charge is created, an equal amount of charge with the opposite sign is created simultaneously

electron

a particle inside of an atom that carries the smallest unit of negative charge

This page titled 7.2: Static Electricity and Charge- Conservation of Charge is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.2: Static Electricity and Charge- Conservation of Charge** by OpenStax is licensed CC BY 4.0.

7.3: Coulomb's Law

Learning Objectives

- State Coulomb's law in terms of how the electrostatic force changes with the distance between two objects.
- Calculate the electrostatic force between two charged point forces, such as electrons or protons.
- Compare the electrostatic force to the gravitational attraction for a proton and an electron; for a human and the Earth.



Figure **7.3.1**: This NASA image of Arp 87 shows the result of a strong gravitational attraction between two galaxies. In contrast, at the subatomic level, the electrostatic attraction between two objects, such as an electron and a proton, is far greater than their mutual attraction due to gravity. (credit: NASA/HST)

Through the work of scientists in the late 18th century, the main features of the **electrostatic force**—the existence of two types of charge, the observation that like charges repel, unlike charges attract, and the decrease of force with distance—were eventually refined, and expressed as a mathematical formula. The mathematical formula for the electrostatic force is called **Coulomb's law** after the French physicist Charles Coulomb (1736–1806), who performed experiments and first proposed a formula to calculate it.

Definition: COULOMB'S LAW

$$F = k\frac{|q_1 q_2|}{r^2}.$$

Coulomb's law calculates the magnitude of the force F between two point charges, q_1 and q_2, separated by a distance r. In SI units, the constant k is equal to

$$k = 8.988 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \approx 8.99 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}.$$

The electrostatic force is a vector quantity and is expressed in units of newtons. The force is understood to be along the line joining the two charges. (See Figure **7.3.2**.)

Although the formula for Coulomb's law is simple, it was no mean task to prove it. The experiments Coulomb did, with the primitive equipment then available, were difficult. Modern experiments have verified Coulomb's law to great precision. For example, it has been shown that the force is inversely proportional to distance between two objects squared ($F \propto 1/r^2$) to an accuracy of 1 part in 10^{16}. No exceptions have ever been found, even at the small distances within the atom.

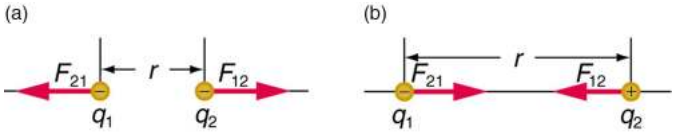


Figure **7.3.2**: The magnitude of the electrostatic force F between point charges q_1 and q_2 separated by a distance r is given by Coulomb's law. Note that Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2. (a) Like charges. (b) Unlike charges.

MISCONCEPTION ALERT

Although we have previously used the simple d for distance, Coulomb's Law uses the symbol r. This is similar to the way in which r is used in the equation for universal gravity. Note the similarities between these two equations. The r in each case is an abbreviation for radius. As we will soon discover, this distance is different than the distance associated with displacement.

Example 7.3.1: How Strong is the Coulomb Force Relative to the Gravitational Force?

Compare the electrostatic force between an electron and proton separated by 0.530×10^{-10} m with the gravitational force between them. This distance is their average separation in a hydrogen atom.

Strategy

To compare the two forces, we first compute the electrostatic force using Coulomb's law, $F = k\frac{|q_1 q_2|}{r^2}$. We then calculate the gravitational force using Newton's universal law of gravitation. Finally, we take a ratio to see how the forces compare in magnitude.

Solution

Entering the given and known information about the charges and separation of the electron and proton into the expression of Coulomb's law yields

$$F = k\frac{|q_1 q_2|}{r^2}$$
$$= \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right) \times \frac{\left(1.60 \times 10^{-19}\text{C}\right)\left(1.60 \times 10^{-19}\text{C}\right)}{\left(0.530\times10^{-10} \text{ m}\right)^2}$$

Thus, the Coulomb force is

$$F = 8.19 \times 10^{-8} \text{ N}.$$

The charges are opposite in sign, so this is an attractive force. This is a very large force for an electron—it would cause an acceleration of 8.99×10^{22} m/s^2(verification is left as an end-of-section problem).The gravitational force is given by Newton's law of gravitation as:

$$F_G = G\frac{mM}{r^2},$$

where $G = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$. Here, m and M represent the electron and proton masses, which can be found in the appendices. Entering values for the knowns yields

$$F_G = \left(6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2\right) \times \frac{\left(9.11 \times 10^{-31} \text{ kg}\right)\left(1.67 \times 10^{-27} \text{ kg}\right)}{\left(0.530\times10^{-10} \text{ m}\right)^2} = 3.61 \times 10^{-47} \text{ N}$$

This is also an attractive force, although it is traditionally shown as positive since gravitational force is always attractive. The ratio of the magnitude of the electrostatic force to gravitational force in this case is, thus,

$$\frac{F}{F_G} = 2.27 \times 10^{39}.$$

Discussion

This is a remarkably large ratio! Note that this will be the ratio of electrostatic force to gravitational force for an electron and a proton at any distance (taking the ratio before entering numerical values shows that the distance cancels). This ratio gives some indication of just how much larger the Coulomb force is than the gravitational force between two of the most common particles in nature.

As the example implies, gravitational force is completely negligible on a small scale, where the interactions of individual charged particles are important. On a large scale, such as between the Earth and a person, the reverse is true. Most objects are nearly electrically neutral, and so attractive and repulsive **Coulomb forces** nearly cancel. Gravitational force on a large scale dominates interactions between large objects because it is always attractive, while Coulomb forces tend to cancel.

Section Summary

- Frenchman Charles Coulomb was the first to publish the mathematical equation that describes the electrostatic force between two objects.
- Coulomb's law gives the magnitude of the force between point charges. It is

$$F = k\frac{|q_1 q_2|}{r^2},$$

where q_1 and q_1 are two point charges separated by a distance r, and $k \approx 8.99 \times 10^9\ \mathrm{N \cdot m^2/C^2}$

- This Coulomb force is extremely basic, since most charges are due to point-like particles. It is responsible for all electrostatic effects and underlies most macroscopic forces.
- The Coulomb force is extraordinarily strong compared with the gravitational force, another basic force—but unlike gravitational force it can cancel, since it can be either attractive or repulsive.
- The electrostatic force between two subatomic particles is far greater than the gravitational force between the same two particles.

Glossary

Coulomb's law

the mathematical equation calculating the electrostatic force vector between two charged particles

Coulomb force

another term for the electrostatic force

electrostatic force

the amount and direction of attraction or repulsion between two charged bodies

---

This page titled 7.3: Coulomb's Law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.3: Coulomb's Law** by OpenStax is licensed CC BY 4.0.

7.4: Representing Electric Fields

Learning Objectives

- Describe a force field and calculate the strength of an electric field due to a point charge.
- Calculate the total force (magnitude and direction) exerted on a test charge from more than one charge.
- Describe an electric field diagram of a positive point charge; of a negative point charge with twice the magnitude of positive charge.
- Draw the electric field lines between two points of the same charge; between two points of opposite charge.

Contact forces, such as between a baseball and a bat, are explained on the small scale by the interaction of the charges in atoms and molecules in close proximity. They interact through forces that include the **Coulomb force**. Action at a distance is a force between objects that are not close enough for their atoms to "touch." That is, they are separated by more than a few atomic diameters.

For example, a charged rubber comb attracts neutral bits of paper from a distance via the Coulomb force. It is very useful to think of an object being surrounded in space by a **force field**. The force field carries the force to another object (called a test object) some distance away.

Concept of a Field

A field is a way of conceptualizing and mapping the force that surrounds any object and acts on another object at a distance without apparent physical connection. For example, the gravitational field surrounding the earth (and all other masses) represents the gravitational force that would be experienced if another mass were placed at a given point within the field.

In the same way, the Coulomb force field surrounding any charge extends throughout space. Using Coulomb's law, $F = k|q_1q_2|/r^2$, its magnitude is given by the equation $F = k|qQ|/r^2$, for a **point charge** (a particle having a charge Q) acting on a **test charge** q at a distance r (see Figure 7.4.1). Both the magnitude and direction of the Coulomb force field depend on Q and the test charge q.

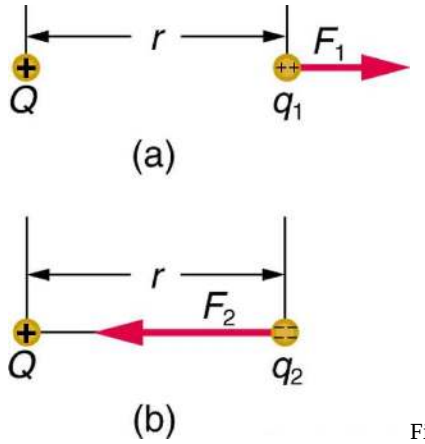


Figure 7.4.1: The Coulomb force field due to a positive charge Q is shown acting on two different charges. Both charges are the same distance from Q. (a) Since q_1 is positive, the force F_1 acting on it is repulsive. (b) The charge q_2 is negative and greater in magnitude than q_1, and so the force F_2 acting on it is attractive and stronger than F_1. The Coulomb force field is thus not unique at any point in space, because it depends on the test charges q_1 and q_2 as well as the charge Q.

Drawings using lines to represent **electric fields** around charged objects are very useful in visualizing field strength and direction. Since the electric field has both magnitude and direction, it is a vector. Like all **vectors**, the electric field can be represented by an arrow that has length proportional to its magnitude and that points in the correct direction. (We have used arrows extensively to represent force vectors, for example.)

Figure 7.4.1 shows two pictorial representations of the same electric field created by a positive point charge Q. Figure 7.4.1(b) shows the standard representation using continuous lines. Figure 7.4.1(a) shows numerous individual arrows with each arrow representing the force on a test charge q.

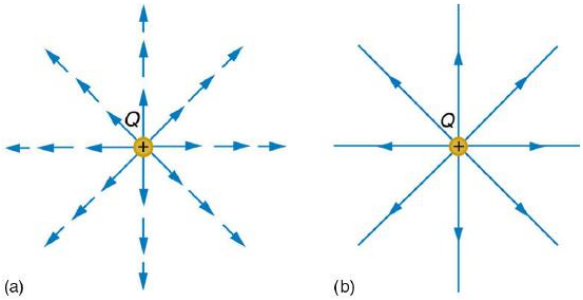


Figure 7.4.2: Two equivalent representations of the electric field due to a positive charge Q. (a) Arrows representing the electric field's magnitude and direction. (b) In the standard representation, the arrows are replaced by continuous field lines having the same direction at any point as the electric field. The closeness of the lines is directly related to the strength of the electric field. A test charge placed anywhere will feel a force in the direction of the field line; this force will have a strength proportional to the density of the lines (being greater near the charge, for example).

Note that the electric field is defined for a positive test charge q, so that the field lines point away from a positive charge and toward a negative charge. (See Figure 7.4.2.) The electric field strength is exactly proportional to the number of field lines per unit area, since the magnitude of the electric field for a point charge is $E = k|Q|/r^2$ and area is proportional to r^2. This pictorial representation, in which field lines represent the direction and their closeness (that is, their areal density or the number of lines crossing a unit area) represents strength, is used for all fields: electrostatic, gravitational, magnetic, and others.

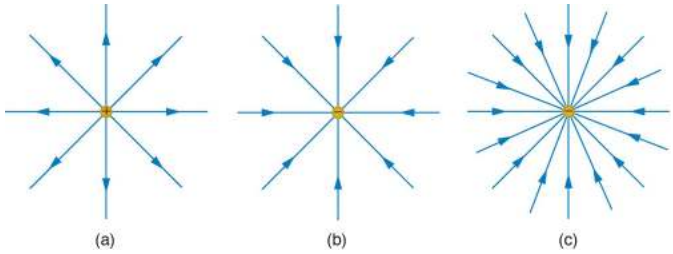


Figure 7.4.3: The electric field surrounding three different point charges. (a) A positive charge. (b) A negative charge of equal magnitude. (c) A larger negative charge.

In many situations, there are multiple charges. The total electric field created by multiple charges is the vector sum of the individual fields created by each charge.

Figure 7.4.3 shows how the electric field from two point charges can be drawn by finding the total field at representative points and drawing electric field lines consistent with those points. While the electric fields from multiple charges are more complex than those of single charges, some simple features are easily noticed.

For example, the field is weaker between like charges, as shown by the lines being farther apart in that region. (This is because the fields from each charge exert opposing forces on any charge placed between them.) (See Figure 7.4.3 and Figure 7.4.4(a).) Furthermore, at a great distance from two like charges, the field becomes identical to the field from a single, larger charge.

Figure 7.4.4(b) shows the electric field of two unlike charges. The field is stronger between the charges. In that region, the fields from each charge are in the same direction, and so their strengths add. The field of two unlike charges is weak at large distances, because the fields of the individual charges are in opposite directions and so their strengths subtract. At very large distances, the field of two unlike charges looks like that of a smaller single charge.

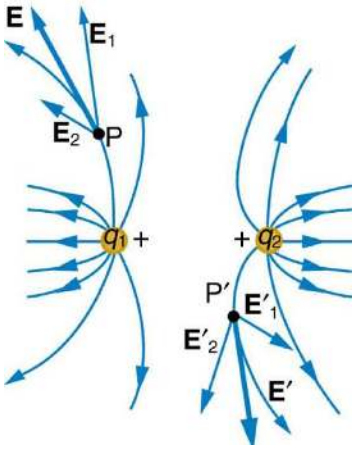


Figure 7.4.4: Two positive point charges q_1 and q_2 produce the resultant electric field shown. The field is calculated at representative points and then smooth field lines drawn following the rules outlined in the text.

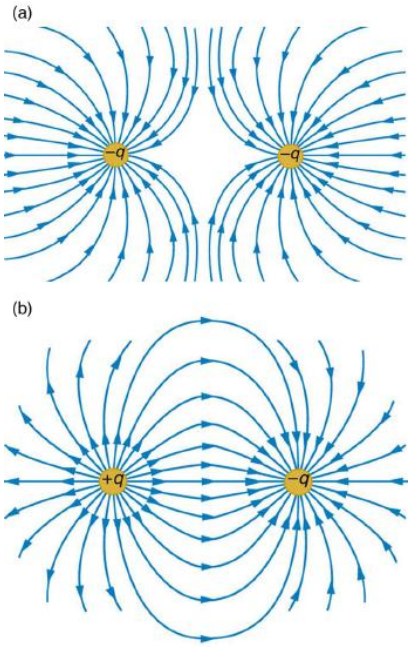


Figure 7.4.5: (a) Two negative charges produce the fields shown. It is very similar to the field produced by two positive charges, except that the directions are reversed. The field is clearly weaker between the charges. The individual forces on a test charge in that region are in opposite directions. (b) Two opposite charges produce the field shown, which is stronger in the region between the charges.

We use electric field lines to visualize and analyze electric fields (the lines are a pictorial tool, not a physical entity in themselves). The properties of electric field lines for any charge distribution can be summarized as follows:

1. Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
2. The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
3. The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
4. The direction of the electric field is tangent to the field line at any point in space.
5. Field lines can never cross.

The last property means that the field is unique at any point. The field line represents the direction of the field; so if they crossed, the field would have two directions at that location (an impossibility if the field is unique).

Section Summary

- Drawings of electric field lines are useful visual tools. The properties of electric field lines for any charge distribution are that:
  - Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
  - The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
  - The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
  - The direction of the electric field is tangent to the field line at any point in space.
  - Field lines can never cross.

Glossary

electric field
a three-dimensional map of the electric force extended out into space from a point charge

electric field lines
a series of lines drawn from a point charge representing the magnitude and direction of force exerted by that charge

vector
a quantity with both magnitude and direction

vector addition
mathematical combination of two or more vectors, including their magnitudes, directions, and positions

---

This page titled 7.4: Representing Electric Fields is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.5: Electric Field Lines** by OpenStax is licensed CC BY 4.0.
- **9.4: Electric Field- Concept of a Field Revisited** by OpenStax is licensed CC BY 4.0.

7.5: Electric Potential and Potential Energy

Learning Objectives

- Define electric potential and electric potential energy.
- Describe the relationship between electric potential difference and electric field.
- Describe the relationship between electric potential and electrical potential energy.

When a free positive charge q is accelerated by an electric field, such as shown in Figure 7.5.1, it is given kinetic energy. The process is analogous to an object being accelerated by a gravitational field. It is as if the charge is going down an electrical hill where its electric potential energy is converted to kinetic energy. Let us explore the work done on a charge q by the electric field in this process, so that we may develop a definition of electric potential energy.

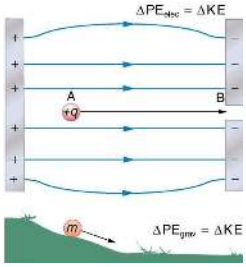


Figure 7.5.1: A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases potential energy is converted to another form. Work is done by a force, but since this force is conservative, we can write $W = -\Delta PE$.

The electrostatic or Coulomb force is conservative, which means that the work done on q is independent of the path taken. This is exactly analogous to the gravitational force in the absence of dissipative forces such as friction. When a force is conservative, it is possible to define a potential energy associated with the force, and it is usually easier to deal with the potential energy (because it depends only on position) than to calculate the work done directly from force $W = Fd$, where d is displacement and F is the force).

We use the letters PE to denote electric potential energy, which has units of joules (J). The change in potential energy, ΔPE, is crucial, since the work done by a conservative force is the negative of the change in potential energy; that is, $W = -\Delta PE$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Definition: POTENTIAL ENERGY

$W = -\Delta PE$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of voltage (related to electric potential energy) than to deal with the electric field (related to Coulomb force) directly.

Given some conservative force F and displacement d under the force, the work done and the change in potential energy can be calculated as, $W = \langle F, d \rangle$ and $\Delta PE = -W = \langle -F, d \rangle$. For electric force, the force is given by the product of electric charge and the electric field, $F = qE$, where q is the charge experiencing the force and E is the electric field at the location of the charge. So the potential energy change due to work done by electric force is $\Delta PE = q(\langle -E, d \rangle)$. If we define change in **electric potential** V as $\Delta V = \langle -E, d \rangle$, then the electric potential energy PE is simply expressed in terms of electric potential, $PE = qV$, or,

$$V = \frac{PE}{q},$$

electric potential energy per charge.

Definition: ELECTRIC POTENTIAL

Electric potential is the electric potential energy per unit charge.

$$V = \frac{\text{PE}}{q}$$

With potential energy, the case often is that its value at a single point has no significant meaning but what is important is the difference in potential energy. From the difference in potential energy, we are able to calculate other quantities, such as change in kinetic energy (if no force other than the conservative force acts) or work needing to be done by other forces (if other forces act). So likewise, rather than the electric potential itself, we are often interested in difference in electric potential ΔV between two points, where,

$$\Delta V = V_{\text{B}} - V_{\text{A}} = \frac{\Delta\text{PE}}{q}.$$

The **potential difference** between points A and B, $V_{\text{B}} - V_{\text{A}}$, is thus defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

$$1\ \text{V} = 1\frac{\text{J}}{\text{C}}$$

Definition: POTENTIAL DIFFERENCE

The potential difference between points A and B, $V_{\text{B}} - V_{\text{A}}$, is defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

$$1\ \text{V} = 1\frac{\text{J}}{\text{C}}$$

The familiar term **voltage** is the common name for potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor.

In summary, the relationship between potential difference (or voltage) and electrical potential energy is given by

$$\Delta V = \frac{\Delta\text{PE}}{q} \text{ and } \Delta\text{PE} = q\Delta V.$$

POTENTIAL DIFFERENCE AND ELECTRICAL POTENTIAL ENERGY

The relationship between potential difference (or voltage) and electrical potential energy is given by

$$\Delta V = \frac{\Delta\text{PE}}{q} \text{ and } \Delta\text{PE} = q\Delta V.$$

The second equation is equivalent to the first.

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus, a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other since $\Delta\text{PE} = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12 V batteries.

Example 7.5.1: Calculating Energy

Suppose you have a 12.0 V motorcycle battery that can move 5000 C of charge, and a 12.0 V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)

Strategy

To say we have a 12.0 V battery means that its terminals have a 12.0 V potential difference. When such a battery moves charge, it puts the charge through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta PE = q\Delta V$.

So to find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000\ \text{C}$ and $\Delta V = 12.0\ \text{V}$. The total energy delivered by the motorcycle battery is

$$\begin{aligned} \Delta \text{PE}_{\text{cycle}} &= (5000\text{C})(12.0\ \text{V}) \\ &= (5000\text{C})(12.0\ \text{J/C}) \\ &= 6.00 \times 10^4\ \text{J}. \end{aligned}$$

Similarly, for the car battery, $q = 60,000\ \text{C}$ and

$$\begin{aligned} \Delta \text{PE}_{\text{car}} &= (60,000\text{C})(12.0\ \text{V}) \\ &= 7.20 \times 10^5\ \text{J}. \end{aligned}$$

Discussion

While voltage and energy are related, they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a low car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (A) through whatever circuitry is involved and attract them to their positive terminals (B) as shown in Figure 7.5.2. The change in potential is $\Delta V = V_B - V_A = +12\ \text{V}$ and the charge q is negative, so that $\Delta PE = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from A to B.

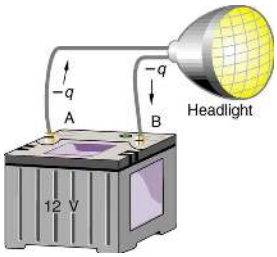


Figure 7.5.2: A battery moves negative charge from its negative terminal through a headlight to its positive terminal. Appropriate combinations of chemicals in the battery separate charges so that the negative terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative. Inside the battery, both positive and negative charges move.

Section Summary

- Electric potential is potential energy per unit charge.
- The potential difference between points A and B, $V_B - V_A$, defined to be the change in potential energy of a charge q moved from A to B, is equal to the change in potential energy divided by the charge, Potential difference is commonly called voltage, represented by the symbol ΔV.

$$.\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V$$

Glossary

electric potential

potential energy per unit charge

potential difference (or voltage)

change in potential energy of a charge moved from one point to another, divided by the charge; units of potential difference are joules per coulomb, known as volt

---

This page titled 7.5: Electric Potential and Potential Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.6: Electric Potential and Potential Energy** by OpenStax is licensed CC BY 4.0.

7.6: Conductors and Current

Learning Objectives

- Define electric current and ampere.
- Describe the direction of charge flow in conventional current.

Conductors

Some materials will more easily allow the passage of an electrical charge than others. Objects which have this property are called electrical **conductors**. (And those that do not are called electrical insulators.) Hopefully you see similarities to the concept of conductors and insulators of heat flow. We can also consider the comparison to the ball rolling down the hill. A ball will roll more quickly down the hill if the path is smoother than if the path is rough. (A smoother path might be along a plastic slide. A rough path might be unmown grass.) Regardless of how quickly the ball rolls, the potential for it to move down the hill remains the same due to the gravitational field in which it exists. An object which is a conductor allows relatively easy passage of electrical current, similar to the smooth plastic slide.

Electric Current

Electric current is defined to be the rate at which charge flows. A large current, such as that used to start a truck engine, moves a large amount of charge in a small time, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge over a long period of time. In equation form, **electric current** I is defined to be

$$I = \frac{\Delta Q}{\Delta t},$$

where ΔQ is the amount of charge passing through a given area in time Δt. (As in previous chapters, initial time is often taken to be zero, in which case $\Delta t = t$.) (See Figure 7.6.1.) The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). Since $I = \Delta Q/\Delta t$, we see that an ampere is one coulomb per second:

$$1 \text{ A} = 1 \text{ C/s}$$

Not only are fuses and circuit breakers rated in amperes (or amps), so are many electrical appliances.

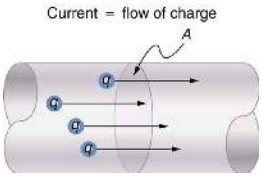


Figure 7.6.1: The rate of flow of charge is current. An ampere is the flow of one coulomb through an area in one second.

Example 7.6.1: Calculating Currents: Current in a Truck Battery and a Handheld Calculator

(a) What is the current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine? (b) How long does it take 1.00 C of charge to flow through a handheld calculator if a 0.300-mA current is flowing?

Strategy

We can use the definition of current in the equation $I = \Delta Q/\Delta t$ to find the current in part (a), since charge and time are given. In part (b), we rearrange the definition of current and use the given values of charge and current to find the time required.

Solution for (a)

Entering the given values for charge and time into the definition of current gives

$$I = \frac{\Delta Q}{\Delta t} = \frac{720\text{C}}{4.00\text{ s}} = 180\text{C/s} = 180\text{ A}.$$

Discussion for (a)

This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these "starter motors" are fairly large because large frictional forces need to be overcome when setting something in motion.

Solution for (b)

Solving the relationship $I = \Delta Q/\Delta t$ for time Δt, and entering the known values for charge and current gives

$$\Delta t = \frac{\Delta Q}{I} = \frac{1.00\text{C}}{0.300 \times 10^{-3}\text{C/s}}$$
$$= 3.33 \times 10^{3}\ \text{s}.$$

Discussion for (b)

This time is slightly less than an hour. The small current used by the hand-held calculator takes a much longer time to move a smaller charge than the large current of the truck starter. So why can we operate our calculators only seconds after turning them on? It is because calculators require very little energy. Such small current and energy demands allow handheld calculators to operate from solar cells or to get many hours of use out of small batteries. Remember, calculators do not have moving parts in the same way that a truck engine has with cylinders and pistons, so the technology requires smaller currents.

Figure 7.6.2 shows a simple circuit and the standard schematic representation of a battery, conducting path, and load (a resistor). Schematics are very useful in visualizing the main features of a circuit. A single schematic can represent a wide variety of situations. The schematic in Figure 7.6.2(b), for example, can represent anything from a truck battery connected to a headlight lighting the street in front of the truck to a small battery connected to a penlight lighting a keyhole in a door. Such schematics are useful because the analysis is the same for a wide variety of situations. We need to understand a few schematics to apply the concepts and analysis to many more situations.

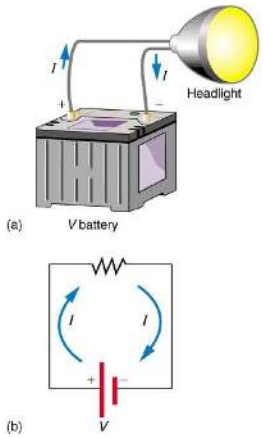


Figure 7.6.2: (a) A simple electric circuit. A closed path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by the two parallel red lines, conducting wires are shown as straight lines, and the zigzag represents the load. The schematic represents a wide variety of similar circuits.

Note that the direction of current flow in Figure 7.6.2 is from positive to negative. *The direction of conventional current is the direction that positive charge would flow*. Depending on the situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator used for nuclear research can produce a current of pure positive charges, such as protons. Figure 7.6.3 illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American politician and scientist Benjamin Franklin in the 1700s. He named the type of charge associated with electrons negative, long before they were known to carry current in so many situations. Franklin, in fact, was totally unaware of the small-scale structure of electricity.

It is important to realize that there is an electric field in conductors responsible for producing the current, as illustrated in Figure 7.6.3. Unlike static electricity, where a conductor in equilibrium cannot have an electric field in it, conductors carrying a current have an electric field and are not in static equilibrium. An electric field is needed to supply energy to move the charges.

MAKING CONNECTIONS: TAKE-HOME INVESTIGATION—ELECTRIC CURRENT ILLUSTRATION

Find a straw and little peas that can move freely in the straw. Place the straw flat on a table and fill the straw with peas. When you pop one pea in at one end, a different pea should pop out the other end. This demonstration is an analogy for an electric current. Identify what compares to the electrons and what compares to the supply of energy. What other analogies can you find for an electric current?

Note that the flow of peas is based on the peas physically bumping into each other; electrons flow due to mutually repulsive electrostatic forces.

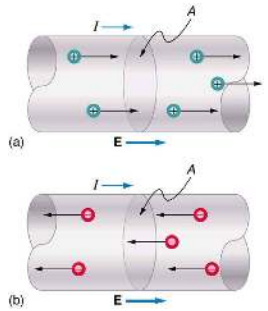


Figure **7.6.3**: Current I is the rate at which charge moves through an area A, such as the cross-section of a wire. Conventional current is defined to move in the direction of the electric field. (a) Positive charges move in the direction of the electric field and the same direction as conventional current. (b) Negative charges move in the direction opposite to the electric field. Conventional current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

Section Summary

- Electric current I is the rate at which charge flows.
- The direction of conventional current is taken as the direction in which positive charge moves.
- The SI unit for current is the ampere (A), where $1\ \text{A} = 1\ \text{C/s}$.
- Current is the flow of free charges, such as electrons and ions.

Glossary

conductor

a material which allows the easy passage of electric current.

electric current

the rate at which charge flows, $I = \Delta Q/\Delta t$

ampere

(amp) the SI unit for current; 1 A = 1 C/s

This page titled 7.6: Conductors and Current is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.8: Current** by OpenStax is licensed CC BY 4.0.

7.7: Ohm's Law

Learning Objectives

- Explain the origin of Ohm's law.
- Calculate voltages, currents, or resistances with Ohm's law.
- Explain what an ohmic material is.
- Describe a simple circuit.

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—which are necessary to maintain a current. All such devices create a potential difference and are loosely referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electric field. The electric field in turn exerts force on charges, causing current.

Ohm's Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:

$$I \propto V.$$

This important relationship is known as **Ohm's law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction—an experimentally observed phenomenon. Such a linear relationship does not always occur. Materials for which it is true are called ohmic materials.

Resistance and Simple Circuits

If voltage drives current, what impedes it? The electric property that impedes current (crudely similar to friction and air resistance) is called **resistance** R. Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or

$$I \propto \frac{1}{R}.$$

Thus, for example, current is cut in half if resistance doubles. Combining the relationships of current to voltage and current to resistance gives

$$I = \frac{V}{R}.$$

This relationship is also called Ohm's law. Ohm's law in this form really defines resistance for certain materials. Ohm's law (like Hooke's law) is not universally valid. The many substances for which Ohm's law holds are called **ohmic**. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I. An object that has simple resistance is called a *resistor*, even if its resistance is small. The unit for resistance is an **ohm** and is given the symbol Ω (upper case Greek omega). Rearranging $I = V/R$ gives $R = V/I$, and so the units of resistance are 1 ohm = 1 volt per ampere:

$$1\ \Omega = 1\frac{V}{A}.$$

Figure **7.7.1** shows the schematic for a simple circuit. A **simple circuit** has a single voltage source and a single resistor. The wires connecting the voltage source to the resistor can be assumed to have negligible resistance, or their resistance can be included in R.

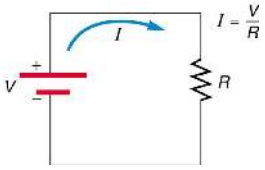


Figure 7.7.1: A simple electric circuit in which a closed path for current to flow is supplied by conductors (usually metal wires) connecting a load to the terminals of a battery, represented by the red parallel lines. The zigzag symbol represents the single resistor and includes any resistance in the connections to the voltage source.

Example 7.7.1: Calculating Resistance: An Automobile Headlight

What is the resistance of an automobile headlight through which 2.50 A flows when 12.0 V is applied to it?

Strategy

We can rearrange Ohm's law as stated by $I = V/R$ and use it to find the resistance.

Solution

Rearranging $I = V/R$ and substituting known values gives

$$R = \frac{V}{I} = \frac{12.0\ \text{V}}{2.50\ \text{A}} = 4.80\Omega.$$

Discussion

This is a relatively small resistance, but it is larger than the cold resistance of the headlight. Resistance usually increases with temperature, and so the bulb has a lower resistance when it is first switched on and will draw considerably more current during its brief warm-up period.

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12}\ \Omega$ or more. A dry person may have a hand-to-foot resistance of $10^{5}\ \Omega$, whereas the resistance of the human heart is about $10^{3}\ \Omega$. A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5}\ \Omega$, and superconductors have no resistance at all (they are non-ohmic). Resistance is related to the shape of an object and the material of which it is composed.

Additional insight is gained by solving $I = V/R$ for V yielding

$$V = IR.$$

This expression for V can be interpreted as the *voltage drop across a resistor produced by the flow of current* I. The phrase IR *drop* is often used for this voltage. For instance, the headlight in Example 7.7.1 has an IR drop of 12.0 V. If voltage is measured at various points in a circuit, it will be seen to increase at the voltage source and decrease at the resistor. Voltage is similar to fluid pressure. The voltage source is like a pump, creating a pressure difference, causing current—the flow of charge. The resistor is like a pipe that reduces pressure and limits flow because of its resistance. Conservation of energy has important consequences here. The voltage source supplies energy (causing an electric field and a current), and the resistor converts it to another form (such as thermal energy). In a simple circuit (one with a single simple resistor), the voltage supplied by the source equals the voltage drop across the resistor, since $\text{PE} = q\Delta V$, and the same q flows through each. Thus, the energy supplied by the voltage source and the energy converted by the resistor are equal. (See Figure 7.7.2.)

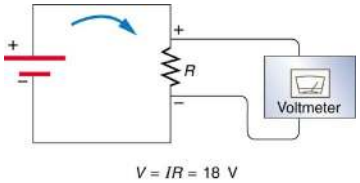


Figure 7.7.2: The voltage drop across a resistor in a simple circuit equals the voltage output of the battery.

MAKING CONNECTIONS: CONSERVATION OF ENERGY

In a simple electrical circuit, the sole resistor converts energy supplied by the source into another form. Conservation of energy is evidenced here by the fact that all of the energy supplied by the source is converted to another form by the resistor alone. We will find that conservation of energy has other important applications in circuits and is a powerful tool in circuit analysis.

Section Summary

- A simple circuit *is* one in which there is a single voltage source and a single resistance.
- One statement of Ohm's law gives the relationship between current I, voltage V, and resistance R in a simple circuit to be $I = \frac{V}{R}$
- Resistance has units of ohms (Ω), related to volts and amperes by $1\ \Omega = 1\ \mathrm{V/A}$.
- There is a voltage or IR drop across a resistor, caused by the current flowing through it, given by $V = IR$.

Glossary

Ohm's law

an empirical relation stating that the current I is proportional to the potential difference V. It is often written as $I = V/R$, where R is the resistance

resistance

the electric property that impedes current; for ohmic materials, it is the ratio of voltage to current, $R = V/I$

ohm

the unit of resistance, given by $1\ \Omega = 1\ \mathrm{V/A}$

ohmic

a type of a material for which Ohm's law is valid

simple circuit

a circuit with a single voltage source and a single resistor

This page titled 7.7: Ohm's Law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.9: Ohm's Law- Resistance and Simple Circuits** by OpenStax is licensed CC BY 4.0.

7.8: Circuits

Learning Objectives

- Draw a circuit with resistors in parallel and in series.
- Explain why total resistance of a parallel circuit is less than the smallest resistance of any of the resistors in that circuit.

Most circuits have more than one component, called a **resistor** that limits the flow of charge in the circuit. A measure of this limit on charge flow is called **resistance**. The simplest combinations of resistors are the series and parallel connections illustrated in Figure 7.8.1. The total resistance of a combination of resistors depends on both their individual values and how they are connected.

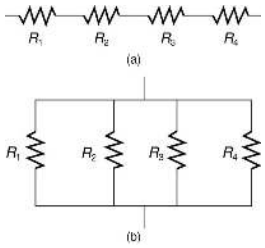


Figure 7.8.1: (a) A series connection of resistors. (b) A parallel connection of resistors.

Resistors in Series

When are resistors in **series**? Resistors are in series whenever the flow of charge, called the **current**, must flow through devices sequentially. For example, if current flows through a person holding a screwdriver and into the Earth, then R_1 in Figure 7.8.1(a) could be the resistance of the screwdriver's shaft, R_2 the resistance of its handle, R_3 the person's body resistance, and R_4 the resistance of her shoes.

Figure 7.8.2 shows resistors in series connected to a **voltage** source. It seems reasonable that the total resistance is the sum of the individual resistances, considering that the current has to pass through each resistor in sequence. (This fact would be an advantage to a person wishing to avoid an electrical shock, who could reduce the current by wearing high-resistance rubber-soled shoes. It could be a disadvantage if one of the resistances were a faulty high-resistance cord to an appliance that would reduce the operating current.)

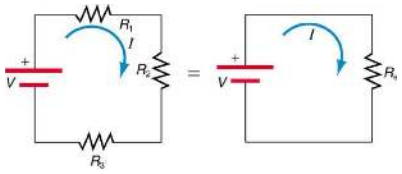


Figure 7.8.2: Three resistors connected in series to a battery (left) and the equivalent single or series resistance (right).

MAJOR FEATURES OF RESISTORS IN SERIES

1. Series resistances add: $R_s = R_1 + R_2 + R_3 + ...$
2. The same current flows through each resistor in series.
3. Individual resistors in series do not get the total source voltage, but divide it.

Resistors in Parallel

Figure 7.8.3 shows resistors in **parallel**, wired to a voltage source. Resistors are in parallel when each resistor is connected directly to the voltage source by connecting wires having negligible resistance. Each resistor thus has the full voltage of the source applied to it.

Each resistor draws the same current it would if it alone were connected to the voltage source (provided the voltage source is not overloaded). For example, an automobile's headlights, radio, and so on, are wired in parallel, so that they utilize the full voltage of the source and can operate completely independently. The same is true in your house, or any building. (See Figure 7.8.3(b).)

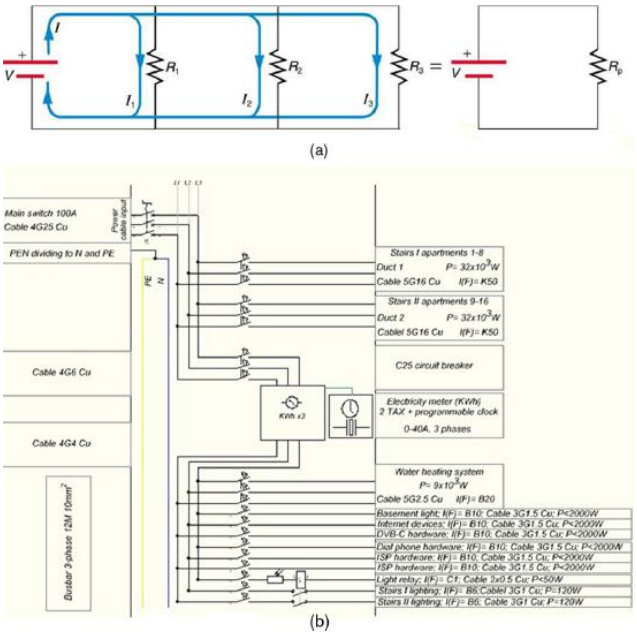


Figure **7.8.3**: (a) Three resistors connected in parallel to a battery and the equivalent single or parallel resistance. (b) Electrical power setup in a house. (credit: Dmitry G, Wikimedia Commons)

MAJOR FEATURES OF RESISTORS IN PARALLEL

1. Parallel resistance is found from $\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + ...$, and it is smaller than any individual resistance in the combination.
2. Each resistor in parallel has the same full voltage of the source applied to it. (Power distribution systems most often use parallel connections to supply the myriad devices served with the same voltage and to allow them to operate independently.)
3. Parallel resistors do not each get the total current; they divide it.

Combinations of Series and Parallel

More complex connections of resistors are sometimes just combinations of series and parallel. These are commonly encountered, especially when wire resistance is considered. In that case, wire resistance is in series with other resistances that are in parallel.

Combinations of series and parallel can be reduced to a single equivalent resistance using the technique illustrated in Figure **7.8.4**. Various parts are identified as either series or parallel, reduced to their equivalents, and further reduced until a single resistance is left. The process is more time consuming than difficult.

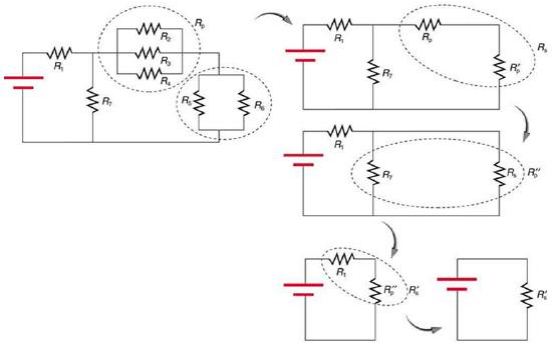


Figure **7.8.4**: This combination of seven resistors has both series and parallel parts. Each is identified and reduced to an equivalent resistance, and these are further reduced until a single equivalent resistance is reached.

The simplest combination of series and parallel resistance, shown in Figure **7.8.5**, is also the most instructive, since it is found in many applications. For example, R_1 could be the resistance of wires from a car battery to its electrical devices, which are in

parallel. R_2 and R_3 could be the starter motor and a passenger compartment light. We have previously assumed that wire resistance is negligible, but, when it is not, it has important effects, as the next example indicates.

CONNECTIONS: CONSERVATION LAWS

The derivations of the expressions for series and parallel resistance are based on the laws of conservation of energy and conservation of charge, which state that total charge and total energy are constant in any process. These two laws are directly involved in all electrical phenomena and will be invoked repeatedly to explain both specific effects and the general behavior of electricity.

Practical Implications

One implication of this last example is that resistance in wires reduces the current and power delivered to a resistor. If wire resistance is relatively large, as in a worn (or a very long) extension cord, then this loss can be significant. If a large current is drawn, the IR drop in the wires can also be significant.

For example, when you are rummaging in the refrigerator and the motor comes on, the refrigerator light dims momentarily. Similarly, you can see the passenger compartment light dim when you start the engine of your car (although this may be due to resistance inside the battery itself).

What is happening in these high-current situations is illustrated in Figure **7.8.6**. The device represented by R_3 has a very low resistance, and so when it is switched on, a large current flows. This increased current causes a larger IR drop in the wires represented by R_1, reducing the voltage across the light bulb (which is R_2), which then dims noticeably.

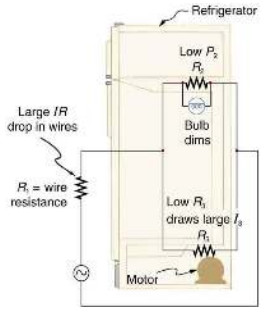


Figure **7.8.6**: Why do lights dim when a large appliance is switched on? The answer is that the large current the appliance motor draws causes a significant IR drop in the wires and reduces the voltage across the light.

Whether something is wired in parallel or in series will also affect the performance of the devices involved, as we will see in a couple of examples. You are probably familiar with the custom of decorating with strands of light for winter holidays (in fact, these are sometimes called Christmas lights even if they are used for a different purpose). These are typically wired in series, as the number of wires necessary to wiring them in parallel would be prohibitive to their function. Because they are wired in series, this means that the current has to pass through each bulb before going to the next bulb. If one of the bulbs stops functioning, then current no longer can flow through any of them, and none of the lights will function. Modern string lights are often designed with a bypass system so that they can continue to function even if one of them goes out.

Another important application to understand is the power strip. These are useful tools to have when there are lots of lower power devices that need to be plugged in but not enough plugs in the area of need. However, they can be quite dangerous if used incorrectly. Because a power strip is wired in parallel for each of the outlets, they can each draw the full voltage of the line the power strip is plugged into. If the current to each of the plugs is low, then it is safe. But if there is a high current, perhaps because the power is being used to heat something, then there might be more power passing through the power strip than it can reasonably handle. If used unsafely, this can result in a fire.

Additional safety issues for other electrical applications are addressed later in this chapter.

Section Summary

- The total resistance of an electrical circuit with resistors wired in a series is the sum of the individual resistances.
- Each resistor in a series circuit has the same amount of current flowing through it.

- The voltage drop, or power dissipation, across each individual resistor in a series is different, and their combined total adds up to the power source input.
- The total resistance of an electrical circuit with resistors wired in parallel is less than the lowest resistance of any of the components.
- Each resistor in a parallel circuit has the same full voltage of the source applied to it.
- The current flowing through each resistor in a parallel circuit is different, depending on the resistance.
- If a more complex connection of resistors is a combination of series and parallel, it can be reduced to a single equivalent resistance by identifying its various parts as series or parallel, reducing each to its equivalent, and continuing until a single resistance is eventually reached.

Glossary

series

a sequence of resistors or other components wired into a circuit one after the other

resistor

a component that provides resistance to the current flowing through an electrical circuit

resistance

causing a loss of electrical power in a circuit

Ohm's law

the relationship between current, voltage, and resistance within an electrical circuit: $V = IR$

voltage

the electrical potential energy per unit charge; electric pressure created by a power source, such as a battery

voltage drop

the loss of electrical power as a current travels through a resistor, wire or other component

current

the flow of charge through an electric circuit past a given point of measurement

Joule's law

the relationship between potential electrical power, voltage, and resistance in an electrical circuit, given by: $P_e = IV$

parallel

the wiring of resistors or other components in an electrical circuit such that each component receives an equal voltage from the power source; often pictured in a ladder-shaped diagram, with each component on a rung of the ladder

---

This page titled 7.8: Circuits is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.11: Resistors in Series and Parallel** by OpenStax is licensed CC BY 4.0.

7.9: Applications- Power and Safety

When you hear the word electricity, you probably think of one of three things: something very dangerous, something that can power your devices, or something that you have to pay for. In the following subsections we will discuss the dangers of electricity as well as the energy costs associated with living in a home.

This page titled 7.9: Applications- Power and Safety is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

7.9.1: Electric Power and Energy

Learning Objectives

- Calculate the power dissipated by a resistor and power supplied by a power supply.
- Calculate the cost of electricity under various circumstances.

Power in Electric Circuits

Power is associated by many people with electricity. Knowing that power is the rate of energy use or energy conversion, what is the expression for **electric power**? Power transmission lines might come to mind. We also think of lightbulbs in terms of their power ratings in watts. Let us compare a 25-W bulb with a 60-W bulb. (See Figure 7.9.1.1(a).) Since both operate on the same voltage, the 60-W bulb must draw more current to have a greater power rating. Thus, the 60-W bulb's resistance must be lower than that of a 25-W bulb. If we increase voltage, we also increase power. For example, when a 25-W bulb that is designed to operate on 120 V is connected to 240 V, it briefly glows very brightly and then burns out. Precisely how are voltage, current, and resistance related to electric power?

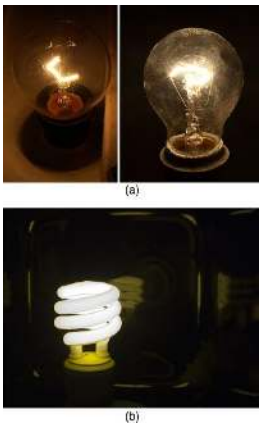


Figure 7.9.1.1: (a) Which of these lightbulbs, the 25-W bulb (upper left) or the 60-W bulb (upper right), has the higher resistance? Which draws more current? Which uses the most energy? Can you tell from the color that the 25-W filament is cooler? Is the brighter bulb a different color and if so why? (credits: Dickbauch, Wikimedia Commons; Greg Westfall, Flickr) (b) This compact fluorescent light (CFL) puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit: dbgg1979, Flickr)

Electric energy depends on both the voltage involved and the charge moved. This is expressed most simply as $\mathbf{PE} = qV$, where q is the charge moved and V is the voltage (or more precisely, the potential difference the charge moves through). Power is the rate at which energy is moved, and so electric power is

$$P = \frac{PE}{t} = \frac{qV}{t}.$$

Recognizing that current is $I = q/t$ (note that $\Delta t = t$ here), the expression for power becomes

$$P = IV.$$

Electric power (P) is simply the product of current times voltage. Power has familiar units of watts. Since the SI unit for potential energy (PE) is the joule, power has units of joules per second, or watts. Thus, $1\ \mathbf{A} \cdot \mathbf{V} = 1\ \mathbf{W}$. For example, cars often have one or more auxiliary power outlets with which you can charge a cell phone or other electronic devices. These outlets may be rated at 20 A, so that the circuit can deliver a maximum power $P = IV = (20\ \mathbf{A})(12\ \mathbf{V}) = 240\ \mathbf{W}$. In some applications, electric power may be expressed as volt-amperes or even kilovolt-amperes ($1\mathbf{kA} \cdot \mathbf{V} = 1\ \mathbf{kW}$).

To see the relationship of power to resistance, we combine Ohm's law with $P = IV$. Substituting $I = V/R$ gives $P = (V/R)V = V^2/R$. Similarly, substituting $V = IR$ gives $P = I(IR) = I^2R$. Three expressions for electric power are listed together here for convenience:

$$P = IV$$

$$P = \frac{V^2}{R}$$

$$P = I^2R.$$

Note that the first equation is always valid, whereas the other two can be used only for resistors. In a simple circuit, with one voltage source and a single resistor, the power supplied by the voltage source and that dissipated by the resistor are identical. (In more complicated circuits, P can be the power dissipated by a single device and not the total power in the circuit.)

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature its resistance is higher, too.

Example 7.9.1.1: Calculating Power Dissipation and Current

Consider the example given in "Ohm's Law: Resistance and Simple Circuits." Then, find the power dissipated by the car headlight.

Strategy

For the headlight, we know voltage and current, so we can use $P = IV$ to find the power.

Solution

Entering the known values of current and voltage for the hot headlight, we obtain

$$P = IV = (2.50\ \text{A})(12.0\ \text{V}) = 30.0\ \text{W}.$$

Discussion

The 30 W dissipated by the hot headlight is typical.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P = E/t$, we see that

$$E = Pt$$

is the energy used by a device using power P for a time interval t. For example, the more lightbulbs burning, the greater P used; the longer they are on, the greater t is. The energy unit on electric bills is the kilowatt-hour ($\text{kW} \cdot \text{h}$), consistent with the relationship $E = Pt$. It is easy to estimate the cost of operating electric appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted to joules. You can prove to yourself that $1\ \text{kW} \cdot \text{h} = 3.6 \times 10^6\ \text{J}$.

The electrical energy (E) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This will not only reduce the cost, but it will also result in a reduced impact on the environment. Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, while the number for commercial establishments is closer to 40%. Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFL). (See Figure 7.9.1.1(b).) Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.) The heat transfer from these CFLs is less, and they last up to 10 times longer. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last 5 times longer than CFLs. However, their cost is still high.

MAKING CONNECTIONS: ENERGY, POWER, AND TIME

The relationship $E = Pt$ is one that you will find useful in many different contexts. The energy your body uses in exercise is related to the power level and duration of your activity, for example. The amount of heating by a power source is related to the power level and time it is applied. Even the radiation dose of an X-ray image is related to the power and time of exposure.

Example 7.9.1.2: Calculating the Cost Effectiveness of Compact Fluorescent Lights (CFL)

If the cost of electricity in your area is 12 cents per kWh, what is the total cost (capital plus operation) of using a 60-W incandescent bulb for 1000 hours (the lifetime of that bulb) if the bulb cost 25 cents? (b) If we replace this bulb with a compact fluorescent light that provides the same light output, but at one-quarter the wattage, and which costs $1.50 but lasts 10 times longer (10,000 hours), what will that total cost be?

Strategy

To find the operating cost, we first find the energy used in kilowatt-hours and then multiply by the cost per kilowatt-hour.

Solution for (a)

The energy used in kilowatt-hours is found by entering the power and time into the expression for energy:

$$E = Pt = (60\ \text{W})(1000\ \text{h}) = 60,000\ \text{W} \cdot \text{h}.$$

In kilowatt-hours, this is

$$E = 60.0\ \text{kW} \cdot \text{h}.$$

Now the electricity cost is

$$\text{cost} = (60.0\ \text{kW} \cdot \text{h})(\$0.12/\text{kW} \cdot \text{h}) = \$7.20.$$

The total cost will be $7.20 for 1000 hours (about one-half year at 5 hours per day).

Solution for (b)

Since the CFL uses only 15 W and not 60 W, the electricity cost will be $7.20/4 = $1.80. The CFL will last 10 times longer than the incandescent, so that the investment cost will be 1/10 of the bulb cost for that time period of use, or 0.1($1.50) = $0.15. Therefore, the total cost will be $1.95 for 1000 hours.

Discussion

Therefore, it is much cheaper to use the CFLs, even though the initial investment is higher. The increased cost of labor that a business must include for replacing the incandescent bulbs more often has not been figured in here.

MAKING CONNECTIONS: TAKE-HOME EXPERIMENT—ELECTRICAL ENERGY USE INVENTORY

1) Make a list of the power ratings on a range of appliances in your home or room. Explain why something like a toaster has a higher rating than a digital clock. Estimate the energy consumed by these appliances in an average day (by estimating their time of use). Some appliances might only state the operating current. If the household voltage is 120 V, then use $P = IV$. 2) Check out the total wattage used in the rest rooms of your school's floor or building. (You might need to assume the long fluorescent lights in use are rated at 32 W.) Suppose that the building was closed all weekend and that these lights were left on from 6 p.m. Friday until 8 a.m. Monday. What would this oversight cost? How about for an entire year of weekends?

Section Summary

- Electric power P is the rate (in watts) that energy is supplied by a source or dissipated by a device.
- Three expressions for electrical power are

$$P = IV,$$

$$P = \frac{V^2}{R},$$

and

$$P = I^2R.$$

- The energy used by a device with a power P over a time t is $E = Pt$.

Glossary

electric power

the rate at which electrical energy is supplied by a source or dissipated by a device; it is the product of current times voltage

---

This page titled 7.9.1: Electric Power and Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.10: Electric Power and Energy** by OpenStax is licensed CC BY 4.0.

7.9.2: Electric Hazards and the Human Body

Learning Objectives

- Define thermal hazard, shock hazard, and short circuit.
- Explain what effects various levels of current have on the human body.

There are two known hazards of electricity—thermal and shock. A **thermal hazard** is one where excessive electric power causes undesired thermal effects, such as starting a fire in the wall of a house. A **shock hazard** occurs when electric current passes through a person. Shocks range in severity from painful, but otherwise harmless, to heart-stopping lethality. This section considers these hazards and the various factors affecting them in a quantitative manner.

Thermal Hazards

Electric power causes undesired heating effects whenever electric energy is converted to thermal energy at a rate faster than it can be safely dissipated. A classic example of this is the **short circuit**, a low-resistance path between terminals of a voltage source. An example of a short circuit is shown in Figure **7.9.2.1**. Insulation on wires leading to an appliance has worn through, allowing the two wires to come into contact. Such an undesired contact with a high voltage is called a *short*. Since the resistance of the short, r, is very small, the power dissipated in the short, $P = V^2/r$, is very large. For example, if V is 120 V and r is $0.100\ \Omega$, then the power is 144 kW, *much* greater than that used by a typical household appliance. Thermal energy delivered at this rate will very quickly raise the temperature of surrounding materials, melting or perhaps igniting them.

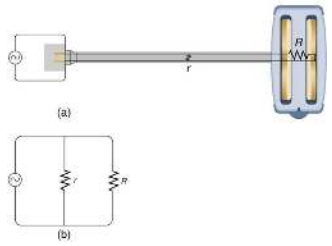


Figure **7.9.2.1**: A short circuit is an undesired low-resistance path across a voltage source. (a) Worn insulation on the wires of a toaster allow them to come into contact with a low resistance r. Since $P = V^2/r$, thermal power is created so rapidly that the cord melts or burns. (b) A schematic of the short circuit.

One particularly insidious aspect of a short circuit is that its resistance may actually be decreased due to the increase in temperature. This can happen if the short creates ionization. These charged atoms and molecules are free to move and, thus, lower the resistance r. Since $P = V^2/r$, the power dissipated in the short rises, possibly causing more ionization, more power, and so on. High voltages, such as the 480-V AC used in some industrial applications, lend themselves to this hazard, because higher voltages create higher initial power production in a short.

Another serious, but less dramatic, thermal hazard occurs when wires supplying power to a user are overloaded with too great a current. As discussed in the previous section, the power dissipated in the supply wires is $P = I^2R_w$, where R_w is the resistance of the wires and I the current flowing through them. If either I or R_w is too large, the wires overheat. For example, a worn appliance cord (with some of its braided wires broken) may have $R_w = 2.00\ \Omega$ rather than the $0.100\ \Omega$ it should be. If 10.0 A of current passes through the cord, then $P = I^2R_w = 200\ \text{W}$ is dissipated in the cord—much more than is safe. Similarly, if a wire with a $0.100 - \Omega$ resistance is meant to carry a few amps, but is instead carrying 100 A, it will severely overheat. The power dissipated in the wire will in that case be $P = 1000\ \text{W}$. Fuses and circuit breakers are used to limit excessive currents. (See Figure **7.9.2.2** and Figure **7.9.2.3**.) Each device opens the circuit automatically when a sustained current exceeds safe limits.

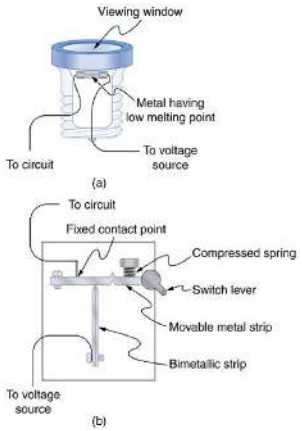


Figure 7.9.2.2: (a) A fuse has a metal strip with a low melting point that, when overheated by an excessive current, permanently breaks the connection of a circuit to a voltage source. (b) A circuit breaker is an automatic but restorable electric switch. The one shown here has a bimetallic strip that bends to the right and into the notch if overheated. The spring then forces the metal strip downward, breaking the electrical connection at the points.

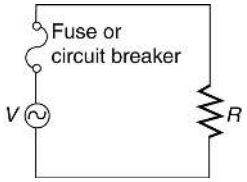


Figure 7.9.2.3: Schematic of a circuit with a fuse or circuit breaker in it. Fuses and circuit breakers act like automatic switches that open when sustained current exceeds desired limits.

Fuses and circuit breakers for typical household voltages and currents are relatively simple to produce, but those for large voltages and currents experience special problems. For example, when a circuit breaker tries to interrupt the flow of high-voltage electricity, a spark can jump across its points that ionizes the air in the gap and allows the current to continue flowing. Large circuit breakers found in power-distribution systems employ insulating gas and even use jets of gas to blow out such sparks. Here AC is safer than DC, since AC current goes through zero 120 times per second, giving a quick opportunity to extinguish these arcs.

Shock Hazards

Electrical currents through people produce tremendously varied effects. An electrical current can be used to block back pain. The possibility of using electrical current to stimulate muscle action in paralyzed limbs, perhaps allowing paraplegics to walk, is under study. TV dramatizations in which electrical shocks are used to bring a heart attack victim out of ventricular fibrillation (a massively irregular, often fatal, beating of the heart) are more than common. Yet most electrical shock fatalities occur because a current put the heart into fibrillation. A pacemaker uses electrical shocks to stimulate the heart to beat properly. Some fatal shocks do not produce burns, but warts can be safely burned off with electric current (though freezing using liquid nitrogen is now more common). Of course, there are consistent explanations for these disparate effects. The major factors upon which the effects of electrical shock depend are

1. The amount of current I.
2. The path taken by the current.
3. The duration of the shock.
4. The frequency f of the current ($f = 0$ for DC).

Table 7.9.2.1 gives the effects of electrical shocks as a function of current for a typical accidental shock. The effects are for a shock that passes through the trunk of the body, has a duration of 1 s, and is caused by 60-Hz power.

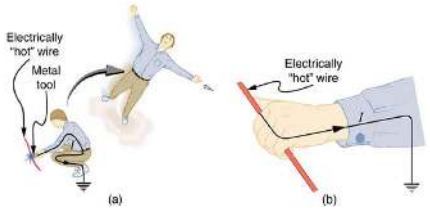


Figure **7.9.2.4**: An electric current can cause muscular contractions with varying effects. (a) The victim is "thrown" backward by involuntary muscle contractions that extend the legs and torso. (b) The victim can not let go of the wire that is stimulating all the muscles in the hand. Those that close the fingers are stronger than those that open them.

Table 7.9.2.1: Effects of Electrical Shock as a Function of Current[1]

Current (mA)	Effect
1	Threshold of sensation
5	Maximum harmless current
10–20	Onset of sustained muscular contraction; cannot let go for duration of shock; contraction of chest muscles may stop breathing during shock
50	Onset of pain
100–300+	Ventricular fibrillation possible; often fatal
300	Onset of burns depending on concentration of current
6000 (6 A)	Onset of sustained ventricular contraction and respiratory paralysis; both cease when shock ends; heartbeat may return to normal; used to defibrillate the heart

Our bodies are relatively good conductors due to the water in our bodies. Given that larger currents will flow through sections with lower resistance (to be further discussed in the next chapter), electric currents preferentially flow through paths in the human body that have a minimum resistance in a direct path to earth. The earth is a natural electron sink. Wearing insulating shoes, a requirement in many professions, prohibits a pathway for electrons by providing a large resistance in that path. Whenever working with high-power tools (drills), or in risky situations, ensure that you do not provide a pathway for current flow (especially through the heart).

Very small currents pass harmlessly and unfelt through the body. This happens to you regularly without your knowledge. The threshold of sensation is only 1 mA and, although unpleasant, shocks are apparently harmless for currents less than 5 mA. A great number of safety rules take the 5-mA value for the maximum allowed shock. At 10 to 20 mA and above, the current can stimulate sustained muscular contractions much as regular nerve impulses do. People sometimes say they were knocked across the room by a shock, but what really happened was that certain muscles contracted, propelling them in a manner not of their own choosing. (See Figure **7.9.2.4**(a).) More frightening, and potentially more dangerous, is the "can't let go" effect illustrated in Figure **7.9.2.4**(b). The muscles that close the fingers are stronger than those that open them, so the hand closes involuntarily on the wire shocking it. This can prolong the shock indefinitely. It can also be a danger to a person trying to rescue the victim, because the rescuer's hand may close about the victim's wrist. Usually the best way to help the victim is to give the fist a hard knock/blow/jar with an insulator or to throw an insulator at the fist. Modern electric fences, used in animal enclosures, are now pulsed on and off to allow people who touch them to get free, rendering them less lethal than in the past.

Greater currents may affect the heart. Its electrical patterns can be disrupted, so that it beats irregularly and ineffectively in a condition called "ventricular fibrillation." This condition often lingers after the shock and is fatal due to a lack of blood circulation. The threshold for ventricular fibrillation is between 100 and 300 mA. At about 300 mA and above, the shock can cause burns, depending on the concentration of current—the more concentrated, the greater the likelihood of burns.

Very large currents cause the heart and diaphragm to contract for the duration of the shock. Both the heart and breathing stop. Interestingly, both often return to normal following the shock. The electrical patterns on the heart are completely erased in a manner that the heart can start afresh with normal beating, as opposed to the permanent disruption caused by smaller currents that can put the heart into ventricular fibrillation. The latter is something like scribbling on a blackboard, whereas the former completely erases it. TV dramatizations of electric shock used to bring a heart attack victim out of ventricular fibrillation also show large paddles. These are used to spread out current passed through the victim to reduce the likelihood of burns.

Current is the major factor determining shock severity (given that other conditions such as path, duration, and frequency are fixed, such as in the table and preceding discussion). A larger voltage is more hazardous, but since $I = V/R$, the severity of the shock depends on the combination of voltage and resistance. For example, a person with dry skin has a resistance of about $200\ \text{k}\Omega$. If he comes into contact with 120-V AC, a current $I = (120\ \text{V})/(200\ \text{k}\Omega) = 0.6\ \text{mA}$ passes harmlessly through him. The same person soaking wet may have a resistance of $10.0\ \text{k}\Omega$ and the same 120 V will produce a current of 12 mA—above the "can't let go" threshold and potentially dangerous.

Most of the body's resistance is in its dry skin. When wet, salts go into ion form, lowering the resistance significantly. The interior of the body has a much lower resistance than dry skin because of all the ionic solutions and fluids it contains. If skin resistance is bypassed, such as by an intravenous infusion, a catheter, or exposed pacemaker leads, a person is rendered **microshock sensitive**. In this condition, currents about 1/1000 those listed in Table 7.9.2.1 produce similar effects. During open-heart surgery, currents as small as $20\ \mu\text{A}$ can be used to still the heart. Stringent electrical safety requirements in hospitals, particularly in surgery and intensive care, are related to the doubly disadvantaged microshock-sensitive patient. The break in the skin has reduced his resistance, and so the same voltage causes a greater current, and a much smaller current has a greater effect.

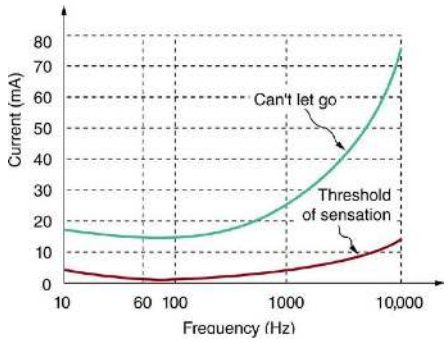


Figure 7.9.2.5: Graph of average values for the threshold of sensation and the "can't let go" current as a function of frequency. The lower the value, the more sensitive the body is at that frequency.

Factors other than current that affect the severity of a shock are its path, duration, and AC frequency. Path has obvious consequences. For example, the heart is unaffected by an electric shock through the brain, such as may be used to treat manic depression. And it is a general truth that the longer the duration of a shock, the greater its effects. Figure 7.9.2.5 presents a graph that illustrates the effects of frequency on a shock. The curves show the minimum current for two different effects, as a function of frequency. The lower the current needed, the more sensitive the body is at that frequency. Ironically, the body is most sensitive to frequencies near the 50- or 60-Hz frequencies in common use. The body is slightly less sensitive for DC ($f = 0$), mildly confirming Edison's claims that AC presents a greater hazard. At higher and higher frequencies, the body becomes progressively less sensitive to any effects that involve nerves. This is related to the maximum rates at which nerves can fire or be stimulated. At very high frequencies, electrical current travels only on the surface of a person. Thus, a wart can be burned off with very high frequency current without causing the heart to stop. (Do not try this at home with 60-Hz AC!) Some of the spectacular demonstrations of electricity, in which high-voltage arcs are passed through the air and over people's bodies, employ high frequencies and low currents. (See Figure 7.9.2.6.)

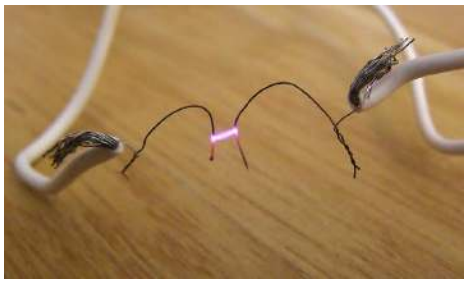


Figure **7.9.2.6**: Is this electric arc dangerous? The answer depends on the AC frequency and the power involved. (credit: Khimich Alex, Wikimedia Commons)

Section Summary

- The two types of electric hazards are thermal (excessive power) and shock (current through a person).
- Shock severity is determined by current, path, duration, and AC frequency.
- Table **7.9.2.1** lists shock hazards as a function of current.
- Figure **7.9.2.5** graphs the threshold current for two hazards as a function of frequency.

Footnotes

- 1 For an average male shocked through trunk of body for 1 s by 60-Hz AC. Values for females are 60–80% of those listed.

Glossary

thermal hazard

a hazard in which electric current causes undesired thermal effects

shock hazard

when electric current passes through a person

short circuit

also known as a “short,” a low-resistance path between terminals of a voltage source

microshock sensitive

a condition in which a person’s skin resistance is bypassed, possibly by a medical procedure, rendering the person vulnerable to electrical shock at currents about 1/1000 the normally required level

---

This page titled 7.9.2: Electric Hazards and the Human Body is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.12: Electric Hazards and the Human Body** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

8: The Amazing Connections Between Magnets and Light Switches

8.1: Introduction to Magnetism
8.2: Magnets
8.3: Natural Magnetism
8.4: The Source of Magnetism
8.5: Magnetic Fields and Magnetic Field Lines
8.6: The relationship between electric current, magnetic fields, and magnetic force
8.7: Force from Electricity and Magnetism- Motors and Meters
8.8: Faraday's Law- Electricity from Magnetism and Force
8.8.1: Applications of Electromagnetic Induction
8.9: Magnetism from Electricity
8.10: Maxwell’s Equations- Summarizing the Connections Between Electricity and Magnetism
8.11: Sources of Voltage and the Electrification of America
8.11.1: Batteries
8.11.2: Induced Voltage and Electrical Generators
8.11.3: Alternating Current versus Direct Current
8.11.4: Transformers

This page titled 8: The Amazing Connections Between Magnets and Light Switches is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

8.1: Introduction to Magnetism



Figure **8.1.1**: The magnificent spectacle of the Aurora Borealis, or northern lights, glows in the northern sky above Bear Lake near Eielson Air Force Base, Alaska. Shaped by the Earth's magnetic field, this light is produced by radiation spewed from solar storms. (credit: Senior Airman Joshua Strang, via Flickr)

One evening, an Alaskan sticks a note to his refrigerator with a small magnet. Through the kitchen window, the Aurora Borealis glows in the night sky. This grand spectacle is shaped by the same force that holds the note to the refrigerator.

People have been aware of magnets and magnetism for thousands of years. The earliest records date to around 500 BCE, in a region of Asia Minor called Magnesia (the name of this region is the source of words like *magnetic*). Magnetic rocks found in Magnesia, which is now part of western Turkey, stimulated interest during ancient times. A practical application for magnets was found later, when they were employed as navigational compasses. The use of magnets in compasses resulted not only in improved long-distance sailing, but also in the names of "north" and "south" being given to the two types of magnetic poles.

Today magnetism plays many important roles in our lives. Physicists' understanding of magnetism has enabled the development of technologies that affect our everyday lives. The iphone in your purse or backpack, for example, would not have been possible without the applications of magnetism and electricity on a small scale.

The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology. The 2007 Nobel Prize in Physics went to Albert Fert from France and Peter Grunberg from Germany for this discovery of *giant magnetoresistance* and its applications to computer memory.

All electric motors, with uses as diverse as powering refrigerators, starting cars, and moving elevators, contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Hundreds of millions of dollars are spent annually on magnetic containment of fusion as a future energy source. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is used to explain atomic energy levels, cosmic rays, and charged particles trapped in the Van Allen belts. Once again, we will find all these disparate phenomena are linked by a small number of underlying physical principles.

There are a surprising number of similarities between electricity and magnetism. There are also some important differences. Pay attention to these similarities and differences as you move through the text. Towards the end we will connect them all together.



Figure **8.1.2**: Engineering of technology like iPods would not be possible without a deep understanding magnetism. (credit: Jesse! S?, Flickr)

This page titled 8.1: Introduction to Magnetism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.1: Introduction to Magnetism** by OpenStax is licensed CC BY 4.0.

8.2: Magnets

Learning Objectives

- Describe the difference between the north and south poles of a magnet.
- Describe how magnetic poles interact with each other.



Figure **8.2.1**: Magnets come in various shapes, sizes, and strengths. All have both a north pole and a south pole. There is never an isolated pole (a monopole).

All magnets attract iron, such as that in a refrigerator door. However, magnets may attract or repel other magnets. Experimentation shows that all magnets have two poles. If freely suspended, one pole will point toward the north. The two poles are thus named the **north magnetic pole** and the **south magnetic pole** (or more properly, north-seeking and south-seeking poles, for the attractions in those directions).

UNIVERSAL CHARACTERISTICS OF MAGNETS AND MAGNETIC POLES

Every magnet has a north pole and a south pole and it *is impossible to separate them*. (Note the difference in how + and - charges can be separated.)

It is a universal characteristic of all magnets that *like poles repel and unlike poles attract*. (Note the similarity with electrostatics: unlike charges attract and like charges repel.)

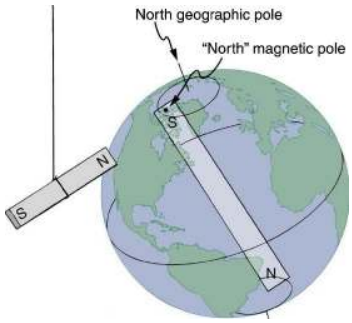


Figure **8.2.2**: One end of a bar magnet is suspended from a thread that points toward north. The magnet's two poles are labeled N and S for north-seeking and south-seeking poles, respectively.

MISCONCEPTION ALERT: EARTH'S GEOGRAPHIC NORTH POLE HIDES AN S

The Earth acts like a very large bar magnet with its south-seeking pole near the geographic North Pole. That is why the north pole of your compass is attracted toward the geographic north pole of the Earth—because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole! Confusion arises because the geographic term "North Pole" has come to be used (incorrectly) for the magnetic pole that is near the North Pole. Thus, "North magnetic pole" is actually a misnomer—it should be called the South magnetic pole.

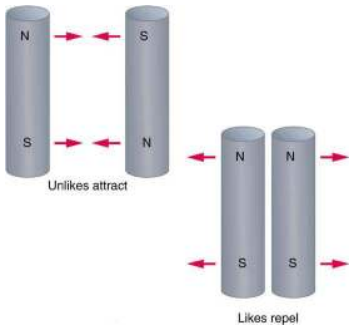


Figure **8.2.3**: Unlike poles attract, whereas like poles repel.

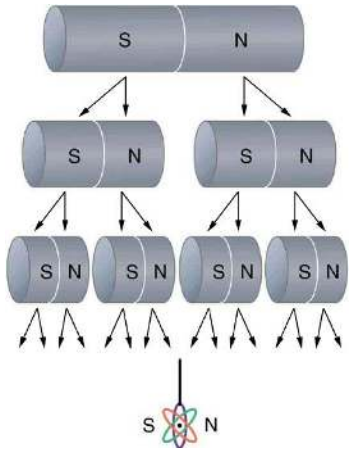


Figure **8.2.4**: North and south poles always occur in pairs. Attempts to separate them result in more pairs of poles. If we continue to split the magnet, we will eventually get down to an iron atom with a north pole and a south pole—these, too, cannot be separated.

The fact that magnetic poles always occur in pairs of north and south is true from the very large scale—for example, sunspots always occur in pairs that are north and south magnetic poles—all the way down to the very small scale. Magnetic atoms have both a north pole and a south pole, as do many types of subatomic particles, such as electrons, protons, and neutrons.

MAKING CONNECTIONS: TAKE-HOME EXPERIMENT—REFRIGERATOR MAGNETS

We know that like magnetic poles repel and unlike poles attract. See if you can show this for two refrigerator magnets. Will the magnets stick if you turn them over? Why do they stick to the door anyway? What can you say about the magnetic properties of the door next to the magnet? Do refrigerator magnets stick to metal or plastic spoons? Do they stick to all types of metal?

Section Summary

- Magnetism is a subject that includes the properties of magnets, the effect of the magnetic force on moving charges and currents, and the creation of magnetic fields by currents.
- There are two types of magnetic poles, called the north magnetic pole and south magnetic pole.
- North magnetic poles are those that are attracted toward the Earth’s geographic north pole.
- Like poles repel and unlike poles attract.
- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.

Glossary

north magnetic pole

the end or the side of a magnet that is attracted toward Earth’s geographic north pole

south magnetic pole

the end or the side of a magnet that is attracted toward Earth’s geographic south pole

This page titled 8.2: Magnets is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.2: Magnets** by OpenStax is licensed CC BY 4.0.

8.3: Natural Magnetism

Learning Objectives

- Define diamagnetism, paramagnetism, and ferromagnetism.
- Describe the role of magnetic domains in magnetization.
- Explain the significance of the Curie temperature.

All matter exhibits some sort of magnetism, as it turns out. There are three types of magnetism possible for matter: diamagnetism, paramagnetism, and ferromagnetism. Some materials can also be made into permanent magnets. The reason for these magnetic properties is due to the arrangement of the electrons within the substances that they are composed of.

Diamagnetism

As we will learn later in this text, all substances are made of atoms and all atoms contain electrons. Those electrons may exist in pairs within a certain portion of the atom, or they might be the only electron found within that portion of the atom. Most electrons in any atom will exist in pairs. Electrons which are paired have the property of diamagnetism. Diamagnetism is a weak repulsive force to an applied magnetic field. If all of the electrons in all of the atoms of a substance are paired, that substance is said to be diamagnetic. If some of them are not paired, then another type of magnetism will come into play which is much stronger.

Paramagnetism

If some of the electrons of an atom are unpaired, the magnetic properties of that substance will be completely different. Instead of being repelled by an external magnetic field, the substance will be attracted to that magnetic field. This phenomena is called paramagnetism. Most substances in their elemental states are paramagnetic, and hence attracted to external magnetic fields. There are some unexpected substances such as molecular oxygen which are paramagnetic. In fact, the electronic structure of oxygen was deduced partially on the experimental evidence that it was paramagnetic.



Figure **8.3.1**: Liquid oxygen suspended between the poles of a strong magnet. By Bob Burk, work supported by the National Science Foundation under grant numbers: 1246120, 1525057, and 1413739 - [1], frame at 4:26, CC BY 3.0, (Source)

Ferromagnetism

Only certain materials exhibit strong magnetic effects. Such materials are called **ferromagnetic**, after the Latin word for iron, *ferrum*. Ferromagnetism is more complicated and more rare than the other two types of magnetism. It is this rarity that makes it such a novelty when we first discover it! Ferromagnetic materials have the properties they do because of both the behavior of the electrons within an atom and the ways the the atoms align with each other. Because this property partly depends on the alignment of the atoms with each other, it is possible to use heating, cooling, or applying an external magnetic field to change the properties of a ferromagnetic substance to become either more magnetic or less magnetic.

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. A group of materials made from the alloys of the rare earth elements are also used as strong and permanent magnets; a popular one is neodymium. Other materials exhibit the weak magnetic effects we have already discussed, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets (the way iron is attracted to magnets), they can also be **magnetized** themselves—that is, they can be induced to be magnetic or made into permanent magnets.

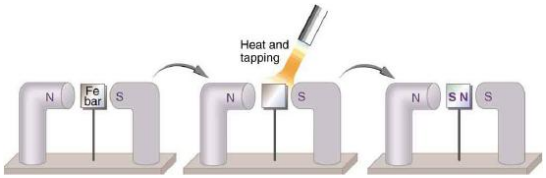


Figure 8.3.2: An unmagnetized piece of iron is placed between two magnets, heated, and then cooled, or simply tapped when cold. The iron becomes a permanent magnet with the poles aligned as shown: its south pole is adjacent to the north pole of the original magnet, and its north pole is adjacent to the south pole of the original magnet. Note that there are attractive forces between the magnets.

When a magnet is brought near a previously unmagnetized ferromagnetic material, it causes local magnetization of the material with unlike poles closest, as in Figure 8.3.2. (This results in the attraction of the previously unmagnetized material to the magnet.) What happens on a microscopic scale is illustrated in Figure 8.3.3. The regions within the material called **domains** act like small bar magnets. Within domains, the poles of individual atoms are aligned. Each atom acts like a tiny bar magnet. Domains are small and randomly oriented in an unmagnetized ferromagnetic object. In response to an external magnetic field, the domains may grow to millimeter size, aligning themselves as shown in Figure 8.3.2(b). This induced magnetization can be made permanent if the material is heated and then cooled, or simply tapped in the presence of other magnets.

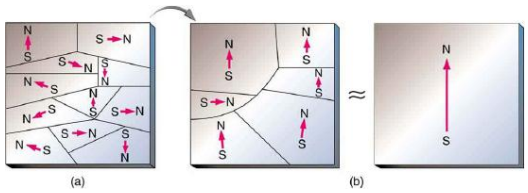


Figure 8.3.3: (a) An unmagnetized piece of iron (or other ferromagnetic material) has randomly oriented domains. (b) When magnetized by an external field, the domains show greater alignment, and some grow at the expense of others. Individual atoms are aligned within domains; each atom acts like a tiny bar magnet.

Conversely, a permanent magnet can be demagnetized by hard blows or by heating it in the absence of another magnet. Increased thermal motion at higher temperature can disrupt and randomize the orientation and the size of the domains. There is a well-defined temperature for ferromagnetic materials, which is called the **Curie temperature**, above which they cannot be magnetized. The Curie temperature for iron is 1043 K (770°C), which is well above room temperature. There are several elements and alloys that have Curie temperatures much lower than room temperature and are ferromagnetic only below those temperatures.

Section Summary

- Diamagnetic materials are repelled by magnetic fields and Paramagnetic materials are weakly attracted to them.
- Ferromagnetic materials, such as iron, are those that exhibit strong magnetic effects.
- The atoms in ferromagnetic materials act like small magnets (due to currents within the atoms) and can be aligned, usually in millimeter-sized regions called domains.
- Domains can grow and align on a larger scale, producing permanent magnets. Such a material is magnetized, or induced to be magnetic.
- Above a material's Curie temperature, thermal agitation destroys the alignment of atoms, and ferromagnetism disappears.

Glossary

diamagnetic
materials with only unpaired electrons that exhibit a weak magnetic repulsion to a magnetic field

paramagnetic
materials with unpaired electrons that exhibit weak magnetic attraction to a magnetic field

ferromagnetic
materials, such as iron, cobalt, nickel, and gadolinium, that exhibit strong magnetic effects

magnetized
to be turned into a magnet; to be induced to be magnetic

domains

regions within a material that behave like small bar magnets

Curie temperature

the temperature above which a ferromagnetic material cannot be magnetized

magnetic monopoles

an isolated magnetic pole; a south pole without a north pole, or vice versa (no magnetic monopole has ever been observed)

This page titled 8.3: Natural Magnetism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.3: Ferromagnets and Electromagnets** by OpenStax is licensed CC BY 4.0.

8.4: The Source of Magnetism

Learning Objectives

- Describe the relationship between electricity and magnetism.

Based on the previous discussion, you might wonder if electricity and magnetism are related. Because electrons have a charge and the behavior of electrons in an atom in part determines what type of magnetic properties a substance will have, there is definitely reason to be curious about this. It turns out that there is a relationship, and it explains some of the properties of magnets we have already discussed. (and many more that we will discuss later in this chapter.) But the discovery of this relationship actually involved artificially created elctromagnets.

Electromagnets

Early in the 19th century, it was discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of charges had any connection with magnets. **Electromagnetism** is the use of electric current to make magnets. These temporarily induced magnets are called **electromagnets**. Electromagnets are employed for everything from a wrecking yard crane that lifts scrapped cars to controlling the beam of a 90-km-circumference particle accelerator to the magnets in medical imaging machines (See Figure **8.4.3**).



Figure **8.4.3**: Instrument for magnetic resonance imaging (MRI). The device uses a superconducting cylindrical coil for the main magnetic field. The patient goes into this “tunnel” on the gurney. (credit: Bill McChesney, Flickr)

Figure **8.4.4** shows that the response of iron filings to a current-carrying coil and to a permanent bar magnet. The patterns are similar. In fact, electromagnets and ferromagnets have the same basic characteristics—for example, they have north and south poles that cannot be separated and for which like poles repel and unlike poles attract.

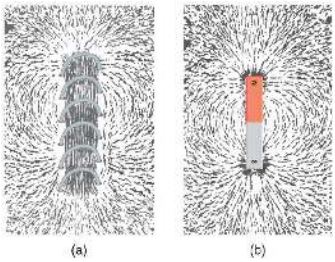


Figure **8.4.4**: Iron filings near (a) a current-carrying coil and (b) a magnet act like tiny compass needles, showing the shape of their fields. Their response to a current-carrying coil and a permanent magnet is seen to be very similar, especially near the ends of the coil and the magnet.

Combining a ferromagnet with an electromagnet can produce particularly strong magnetic effects. (See Figure **8.4.5**.) Whenever strong magnetic effects are needed, such as lifting scrap metal, or in particle accelerators, electromagnets are enhanced by ferromagnetic materials. Limits to how strong the magnets can be made are imposed by coil resistance (it will overheat and melt at sufficiently high current), and so superconducting magnets may be employed. These are still limited, because superconducting properties are destroyed by too great a magnetic field.



Figure 8.4.5: An electromagnet with a ferromagnetic core can produce very strong magnetic effects. Alignment of domains in the core produces a magnet, the poles of which are aligned with the electromagnet.

Figure 8.4.6 shows a few uses of combinations of electromagnets and ferromagnets. Ferromagnetic materials can act as memory devices, because the orientation of the magnetic fields of small domains can be reversed or erased. Magnetic information storage on videotapes and computer hard drives are among the most common applications. This property is vital in our digital world.

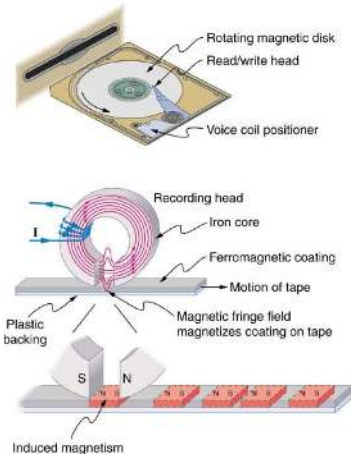


Figure 8.4.6: An electromagnet induces regions of permanent magnetism on a floppy disk coated with a ferromagnetic material. The information stored here is digital (a region is either magnetic or not); in other applications, it can be analog (with a varying strength), such as on audiotapes.

Current: The Source of All Magnetism

An electromagnet creates magnetism with an electric current. In later sections we explore this in more detail. We have already discussed the relationship between naturally occuring magnetic properties arising from the behavior of electrons. Figure 8.4.7 shows models of how electric currents create magnetism at the submicroscopic level. (Note that we cannot directly observe the paths of individual electrons about atoms, and so a model or visual image, consistent with all direct observations, is made. We can directly observe the electron's orbital angular momentum, its spin momentum, and subsequent magnetic moments, all of which are explained with electric-current-creating subatomic magnetism.) Electric currents, including those associated with other submicroscopic particles like protons, allow us to explain ferromagnetism and all other magnetic effects. Magnetism is caused by the motion of an electric charge.

Crucial to the statement that electric current is the source of all magnetism is the fact that it is impossible to separate north and south magnetic poles. (This is far different from the case of positive and negative charges, which are easily separated.) A current loop always produces a magnetic dipole—that is, a magnetic field that acts like a north pole and south pole pair. Since isolated north and south magnetic poles, called **magnetic monopoles**, are not observed, currents are used to explain all magnetic effects. If magnetic monopoles did exist, then we would have to modify this underlying connection that all magnetism is due to electrical current.

ELECTRIC CURRENTS AND MAGNETISM

Electric current is the source of all magnetism.

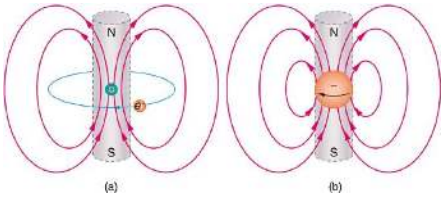


Figure 8.4.7: (a) In the planetary model of the atom, an electron orbits a nucleus, forming a closed-current loop and producing a magnetic field with a north pole and a south pole. (b) Electrons have spin and can be crudely pictured as rotating charge, forming a current that produces a magnetic field with a north pole and a south pole. Neither the planetary model nor the image of a spinning electron is completely consistent with modern physics. However, they do provide a useful way of understanding phenomena.

Section Summary

- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.
- All magnetism is created by electric current.
- Electromagnets employ electric currents to make magnetic fields, often aided by induced fields in ferromagnetic materials.

Glossary

domains

regions within a material that behave like small bar magnets

electromagnetism

the use of electrical currents to induce magnetism

electromagnet

an object that is temporarily magnetic when an electrical current is passed through it

magnetic monopoles

an isolated magnetic pole; a south pole without a north pole, or vice versa (no magnetic monopole has ever been observed)

This page titled 8.4: The Source of Magnetism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.3: Ferromagnets and Electromagnets** by OpenStax is licensed CC BY 4.0.

8.5: Magnetic Fields and Magnetic Field Lines

Learning Objectives

- Define magnetic field and describe the magnetic field lines of various magnetic fields.

Einstein is said to have been fascinated by a compass as a child, perhaps musing on how the needle felt a force without direct physical contact. His ability to think deeply and clearly about action at a distance, particularly for gravitational, electric, and magnetic forces, later enabled him to create his revolutionary theory of relativity. Since magnetic forces act at a distance, we define a **magnetic field** to represent magnetic forces. The pictorial representation of **magnetic field lines** is very useful in visualizing the strength and direction of the magnetic field. As shown in Figure 8.5.1, the **direction of magnetic field lines** is defined to be the direction in which the north end of a compass needle points. The magnetic field is traditionally called the ***B*-field**. Contrast the look of the magnetic field lines with what we saw for electric field lines in the previous chapter.

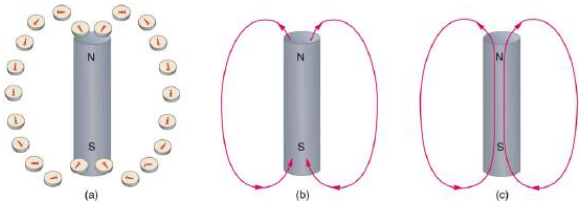


Figure 8.5.1: Magnetic field lines are defined to have the direction that a small compass points when placed at a location. (a) If small compasses are used to map the magnetic field around a bar magnet, they will point in the directions shown: away from the north pole of the magnet, toward the south pole of the magnet. (Recall that the Earth's north magnetic pole is really a south pole in terms of definitions of poles on a bar magnet.) (b) Connecting the arrows gives continuous magnetic field lines. The strength of the field is proportional to the closeness (or density) of the lines. (c) If the interior of the magnet could be probed, the field lines would be found to form continuous closed loops.

Small compasses used to test a magnetic field will not disturb it. (This is analogous to the way we tested electric fields with a small test charge. In both cases, the fields represent only the object creating them and not the probe testing them.) Figure 8.5.2 shows how the magnetic field appears for a loop and a long straight wire, each of which is conducting an electrical current. A small compass placed in these fields will align itself parallel to the field line at its location, with its north pole pointing in the direction of *B*. Note the symbols used for field into and out of the paper.

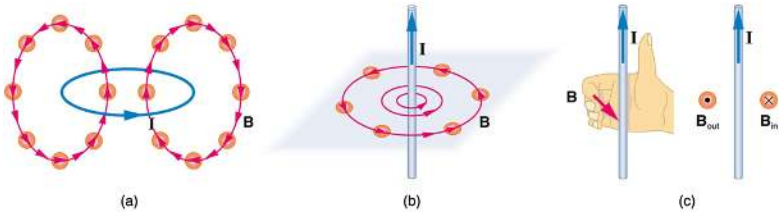


Figure 8.5.2: Small compasses could be used to map the fields shown here. (a) The magnetic field of a circular current loop is similar to that of a bar magnet. (b) A long and straight wire creates a field with magnetic field lines forming circular loops. (c) When the wire is in the plane of the paper, the field is perpendicular to the paper. Note that the symbols used for the field pointing inward (like the tail of an arrow) and the field pointing outward (like the tip of an arrow).

MAKING CONNECTIONS: CONCEPT OF A FIELD

A field is a way of mapping forces surrounding any object that can act on another object at a distance without apparent physical connection. The field represents the object generating it. Gravitational fields map gravitational forces, electric fields map electrical forces, and magnetic fields map magnetic forces.

Extensive exploration of magnetic fields has revealed a number of hard-and-fast rules. We use magnetic field lines to represent the field (the lines are a pictorial tool, not a physical entity in and of themselves). The properties of magnetic field lines can be summarized by these rules:

1. The direction of the magnetic field is tangent to the magnetic field line at any point in space. A small compass will point in the direction of the field line.

2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without beginning or end. They go from the north pole to the south pole.

The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which begin and end on the positive and negative charges. If magnetic monopoles existed, then magnetic field lines would begin and end on them.

Section Summary

- Magnetic fields can be pictorially represented by magnetic field lines, the properties of which are as follows:

1. The field is tangent to the magnetic field line.
2. Field strength is proportional to the line density.
3. Field lines cannot cross.
4. Field lines are continuous loops.

Glossary

magnetic field
the representation of magnetic forces

***B*-field**
another term for magnetic field

magnetic field lines
the pictorial representation of the strength and the direction of a magnetic field

direction of magnetic field lines
the direction that the north end of a compass needle points

This page titled 8.5: Magnetic Fields and Magnetic Field Lines is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.4: Magnetic Fields and Magnetic Field Lines** by OpenStax is licensed CC BY 4.0.

8.6: The relationship between electric current, magnetic fields, and magnetic force

Learning Objectives

- Describe the effects of magnetic fields on moving charges.
- Use the right hand rule 1 to determine the relationship between the velocity of a charge, the direction of the magnetic field, and the direction of the magnetic force on a moving charge.

What is the mechanism by which one magnet exerts a force on another? The answer is related to the fact that all magnetism is caused by current, the flow of charge. *Magnetic fields exert forces on moving charges*, and so they exert forces on other magnets, all of which have moving charges. However, unlike most forces that we have examined in this course, the relationship between the direction of the moving charge and the magnetic force applied by it are not intuitive. We will examine this connection briefly, but not in a great deal of detail. In a more advanced physics text you would be using the rules for vectors to perform calculations to find the relationship between the moving charge and the applied force. For this text we only want to understanding a little bit about how the direction is important in order to understand the ensuing phenomena.

MAKING CONNECTIONS: CHARGES AND MAGNETS

There is no magnetic force on static charges. However, there is a magnetic force on moving charges. When charges are stationary, their electric fields do not affect magnets. But, when charges move, they produce magnetic fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic fields emerges—each affects the other.

Direction of Force: Right Hand Rule 1

The *direction* of the magnetic force $\mathbf{F}$ is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$, as determined by the **right hand rule 1** (or RHR-1), which is illustrated in Figure 8.6.1. For the purposes of this text, what you should understand about this is that the magnetic force applied by the moving charge is always perpendicular to both the direction of the moving charge and the magnetic field. The direction of the moving charge is not always perpendicular to the magnetic field, but is strongest when it is perpendicular. When the direction is parallel, the magnetic force is zero. The variation in the angle between the moving charge and the magnetic field can have some dramatic effects, as we shall soon seen. Whenever you see the RHR-1 notation or accompanying graphics in this text, simply consider the three dimensional relationship between the direction of the moving charge, the direction of the magnetic field, and the direction of the magnetic force.

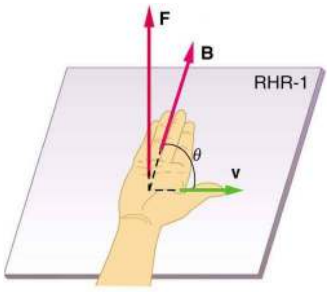


Figure 8.6.1: Magnetic fields exert forces on moving charges. This force is one of the most basic known. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$ and follows right hand rule–1 (RHR-1) as shown. The magnitude of the force is proportional to q, v, B, and depends on the angle between $\mathbf{v}$ and $\mathbf{B}$.

Section Summary

- The maximum force a magnetic field can exert on a moving charge is

$$F = qvB$$

- The *direction* of the force on a moving charge is given by right hand rule 1 (RHR-1): Point the thumb of the right hand in the direction of v, the fingers in the direction of B, and a perpendicular to the palm points in the direction of F.

- The force is perpendicular to the plane formed by $\mathbf{V}$ and $\mathbf{B}$. Since the force is zero if $\mathbf{V}$ is parallel to $\mathbf{B}$, charged particles often follow magnetic field lines rather than cross them.

Glossary

right hand rule 1 (RHR-1)

the rule to determine the direction of the magnetic force on a positive moving charge: when the thumb of the right hand points in the direction of the charge's velocity $\mathbf{v}$ and the fingers point in the direction of the magnetic field $\mathbf{B}$, then the force on the charge is perpendicular and away from the palm; the force on a negative charge is perpendicular and into the palm

tesla

T, the SI unit of the magnetic field strength; $1\ \mathrm{T} = \frac{1\ \mathrm{N}}{\mathrm{A \cdot m}}$

magnetic force

the force on a charge produced by its motion through a magnetic field

gauss

G, the unit of the magnetic field strength; $1\mathrm{G} = 10^{-4}\ \mathrm{T}$

This page titled 8.6: The relationship between electric current, magnetic fields, and magnetic force is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.5: Magnetic Field Strength- Force on a Moving Charge in a Magnetic Field** by OpenStax is licensed CC BY 4.0.

8.7: Force from Electricity and Magnetism- Motors and Meters

Learning Objectives

- Describe how motors and meters work in terms of force on a current loop.

Because charges ordinarily cannot escape a conductor, the magnetic force on charges moving in a conductor is transmitted to the conductor itself. This force is sufficient to move the wire that the charge is traveling through, which we will see can be put to great effect. And as we will also see, the manipulation of magnets can itself create a charge. We will explore each of these ideas.

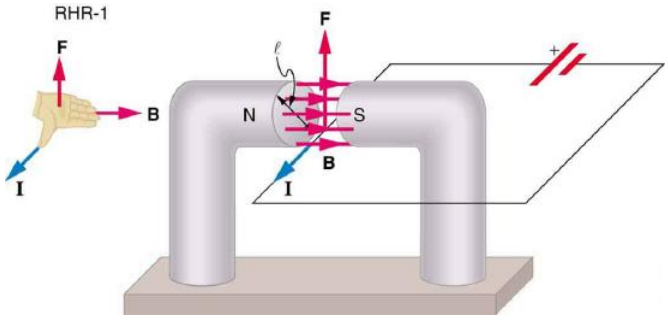


Figure 8.7.1: The magnetic field exerts a force on a current-carrying wire in a direction given by the right hand rule 1 (the same direction as that on the individual moving charges). This force can easily be large enough to move the wire, since typical currents consist of very large numbers of moving charges.

Motors

Motors are the most common application of magnetic force on current-carrying wires. Motors have loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts force on the loops, which rotates a shaft. Electrical energy is converted to mechanical work in the process. (See Figure 8.7.2.)

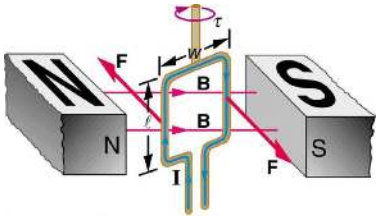


Figure 8.7.2: Force on a current loop. A current-carrying loop of wire attached to a vertically rotating shaft feels magnetic forces that produce a clockwise rotation as viewed from above.

The coil rotates because the forces exerted on the current carrying wire by the magnetic field are in opposite directions, which pull the wire towards an equilibrium position that is perpendicular to the direction of the magnetic field. If nothing else were done, this would be the end of the process. However, if the direction of the electric current changes right as this equilibrium position is reached, the wire loop will continue to rotate as it approaches the next equilibrium position where the process can repeat itself. The direction of the electrical current must be switched twice in one rotation in order to maintain the process. To get the coil to continue rotating in the same direction, we can reverse the current as it passes through $\theta = 0$ with automatic switches called *brushes*. (See Figure 8.7.3.)

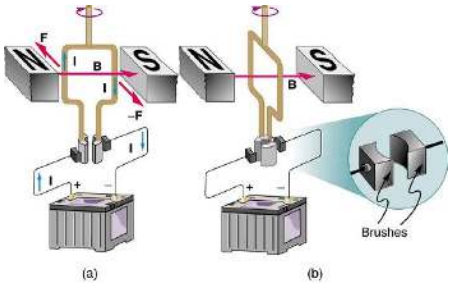


Figure **8.7.3**: (a) As the momentum of the coil carries it through $\theta = 0$, the brushes reverse the current to keep the motion clockwise. (b) The coil will rotate continuously in the clockwise direction, with the current reversing each half revolution to maintain the motion.

Meters

Meters, such as those in analog fuel gauges on a car, are another common application of magnetic force on a current-carrying loop. Figure **8.7.4** shows that a meter is very similar in construction to a motor. The meter in the figure has its magnets shaped to limit the effect of θ by making B perpendicular to the loop over a large angular range. A linear spring exerts a counter-force that balances the current-produced force. This makes the needle deflection proportional to I. If an exact proportionality cannot be achieved, the gauge reading can be calibrated. To produce a galvanometer for use in analog voltmeters and ammeters that have a low resistance and respond to small currents, we use a large loop area A, high magnetic field B, and low-resistance coils.

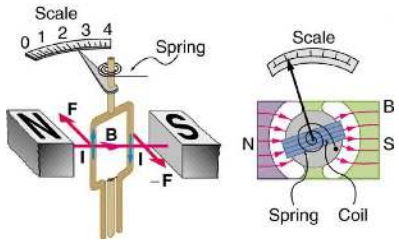


Figure **8.7.4**: Meters are very similar to motors but only rotate through a part of a revolution. The magnetic poles of this meter are shaped to keep the component of B perpendicular to the loop constant, so that the force does not depend on θ and the deflection against the return spring is proportional only to the current I.

Glossary

motor

loop of wire in a magnetic field; when current is passed through the loops, the magnetic field exerts force on the loops, which rotates a shaft; electrical energy is converted to mechanical work in the process

meter

common application of magnetic force on a current-carrying loop that is very similar in construction to a motor; by design, the force is proportional to I and not θ, so the needle deflection is proportional to the current

---

This page titled 8.7: Force from Electricity and Magnetism- Motors and Meters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.7: Motors and Meters** by OpenStax is licensed CC BY 4.0.
- **10.6: Magnetic Force on a Current-Carrying Conductor** by OpenStax is licensed CC BY 4.0.

8.8: Faraday's Law- Electricity from Magnetism and Force

Learning Objectives

- Describe methods to produce an induced voltage with a magnetic field or magnet and a loop of wire.
- Explain Faraday's Law of Induction.

The apparatus to demonstrate that magnetic fields can create currents is illustrated in Figure **8.8.1**. When the switch is closed, a magnetic field is produced in the coil on the top part of the iron ring and transmitted to the coil on the bottom part of the ring. The galvanometer is used to detect any current induced in the coil on the bottom. It was found that each time the switch is closed, the galvanometer detects a current in one direction in the coil on the bottom. Each time the switch is opened, the galvanometer detects a current in the opposite direction. Interestingly, if the switch remains closed or open for any length of time, there is no current through the galvanometer. *Closing and opening the switch* induces the current. It is the *change* in magnetic field that creates the current in the following way: *a changing magnetic field induces an electric field*, which results in the induced voltage. When this induced voltage occurs over a conducting path, as in this example, the induced voltage causes a current to flow. As a shorthand, we call the resulting current **induced current**; the changing magnetic field does not induce the current directly but through the induced voltage and an application of Ohm's law.

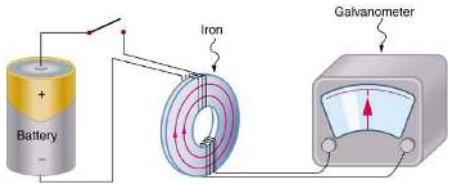


Figure **8.8.1**: Faraday's apparatus for demonstrating that a magnetic field can produce a current. A change in the field produced by the top coil induces an voltage and, hence, a current in the bottom coil. When the switch is opened and closed, the galvanometer registers currents in opposite directions. No current flows through the galvanometer when the switch remains closed or open.

An experiment easily performed and often done in physics labs is illustrated in Figure **8.8.2**. A voltage is induced in the coil when a bar magnet is pushed in and out of it. Voltages of opposite signs are produced by motion in opposite directions, and the voltages are also reversed by reversing poles. The same results are produced if the coil is moved rather than the magnet—it is the relative motion that is important. The faster the motion, the greater the voltage, and there is no voltage when the magnet is stationary relative to the coil.

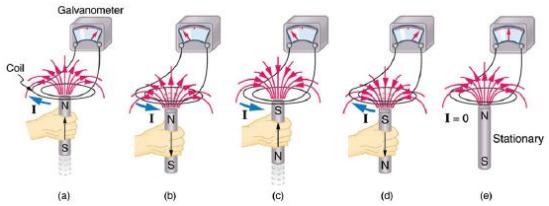


Figure **8.8.2**: Movement of a magnet relative to a coil produces voltage as shown. The same voltages are produced if the coil is moved relative to the magnet. The greater the speed, the greater the magnitude of the voltage, and the voltage is zero when there is no motion.

The method of inducing a voltage used in most electric generators is shown in Figure **8.8.3**. A coil is rotated in a magnetic field, producing an alternating voltage (and current), which depends on rotation rate and other factors that will be explored in later sections. Note that the generator is remarkably similar in construction to a motor.

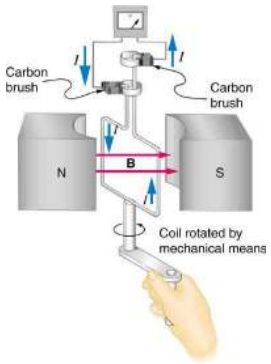


Figure 8.8.3: Rotation of a coil in a magnetic field produces a voltage. This is the basic construction of a generator, where work done to turn the coil is converted to electric energy. Note the generator is very similar in construction to a motor.

So we see that changing the magnitude or direction of a magnetic field produces a voltage. Experiments revealed that there is a crucial quantity called the **magnetic flux**, Φ, given by

$$\Phi = B_\perp A,$$

where B is the magnetic field strength over an area A, at an angle θ with the perpendicular to the area as shown in Figure 8.8.4. We need advanced mathematics to formally describe flux, but we can visualize the idea with concepts we have already introduced in this chapter. The more magnetic field lines that pass through a surface, the higher the magnetic flux. **Any change in magnetic flux Φ induces a voltage.** This process is defined to be **electromagnetic induction**. Units of magnetic flux Φ are $\mathrm{T \cdot m^2}$.

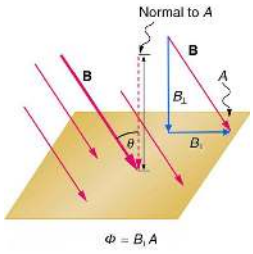


Figure 8.8.4: Magnetic flux Φ is related to the magnetic field and the area A over which it exists. Only the portion of the magnetic field that is perpendicular to the area ($B_\perp$) contributes to the flux. The flux $\Phi = B_\perp A$ is related to induction; any change in Φ induces a voltage.

All induction, including the examples given so far, arises from some change in magnetic flux Φ. For example, Faraday changed B (and, hence, Φ) when opening and closing the switch in his apparatus (shown in Figure 8.8.1). This is also true for the bar magnet and coil shown in Figure 8.8.2. When rotating the coil of a generator, the angle θ (and, hence, Φ) is changed.

Faraday's experiments showed that the voltage induced by a change in magnetic flux depends on only a few factors. First, voltage is directly proportional to the change in flux $\Delta\Phi$. Second, voltage is greatest when the change in time Δt is smallest—that is, voltage is inversely proportional to Δt. Finally, if a coil has N turns, a voltage will be produced that is N times greater than for a single coil, so that voltage is directly proportional to N. The equation for the voltage induced by a change in magnetic flux is referred to as Faraday's law of induction.

FARADAY'S LAW OF INDUCTION

$$V = -N\frac{\Delta\Phi}{\Delta t}.$$

The minus sign in Faraday's law of induction is very important. The minus means that *the induced voltage creates a current I and magnetic field B that oppose the change in flux $\Delta\Phi$—this is known as Lenz's law*. Faraday was aware of the direction, but Lenz stated it so clearly that he is credited for its discovery. (See Figure 8.8.1.)

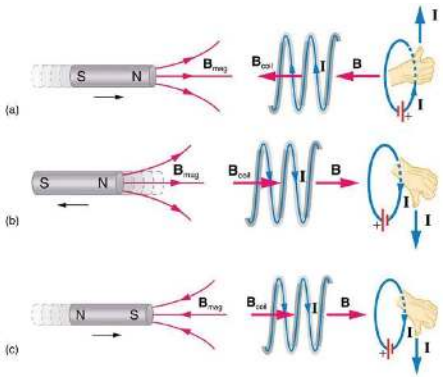


Figure 8.8.5: (a) When this bar magnet is thrust into the coil, the strength of the magnetic field increases in the coil. The current induced in the coil creates another field, in the opposite direction of the bar magnet's to oppose the increase. This is one aspect of *Lenz's law—induction opposes any change in flux*. (b) and (c) are two other situations showing similar trends.

MAKING CONNECTIONS: CONSERVATION OF ENERGY

Lenz's law is a manifestation of the conservation of energy. The induced voltage produces a current that opposes the change in flux, because a change in flux means a change in energy. Energy can enter or leave, but not instantaneously. Lenz's law is a consequence. As the change begins, the law says induction opposes and, thus, slows the change. In fact, if the induced voltage were in the same direction as the change in flux, there would be a positive feedback that would give us free energy from no apparent source—conservation of energy would be violated.

Section Summary

- The crucial quantity in induction is magnetic flux Φ, defined to be $\Phi = B_\perp A$, where $B_\perp$ is the magnetic field strength perpendicular to the area A.
- Units of magnetic flux Φ are $\mathrm{T \cdot m^2}$.
- Any change in magnetic flux Φ induces a voltage—the process is defined to be electromagnetic induction.
- Faraday's law of induction states that the voltage induced by a change in magnetic flux is

$$V = -N\frac{\Delta\Phi}{\Delta t}$$

  when flux changes by $\Delta\Phi$ in a time Δt.

- If voltage is induced in a coil, N is its number of turns.
- The minus sign means that the induced voltage creates a current I and magnetic field B that *oppose the change in flux* $\Delta\Phi$ — this opposition is known as Lenz's law.

Glossary

induced current

the current created by a changing magnetic field through voltage induced over a conducting path

magnetic flux

the amount of magnetic field going through a particular area, calculated with $\Phi = B_\perp A$, where $B_\perp$ is the magnetic field strength perpendicular to the area A

electromagnetic induction

the process of inducing a voltage with a change in magnetic flux

Faraday's law of induction

the means of calculating the voltage in a coil due to changing magnetic flux, given by $V = -N\frac{\Delta\Phi}{\Delta t}$

This page titled 8.8: Faraday's Law- Electricity from Magnetism and Force is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.9: Induced Voltage and Magnetic Flux** by OpenStax is licensed CC BY 4.0.
- **10.10: Faraday's Law of Induction- Lenz's Law** by OpenStax is licensed CC BY 4.0.

8.8.1: Applications of Electromagnetic Induction

Learning Objectives

- Understand how Faraday's Law of Induction has led to several new technologies.

There are many applications of Faraday's Law of induction, as we will explore in this chapter and others. At this juncture, let us mention several that have to do with data storage and magnetic fields. A very important application has to do with audio and video *recording tapes*. A plastic tape, coated with iron oxide, moves past a recording head. This recording head is basically a round iron ring about which is wrapped a coil of wire—an electromagnet (Figure 8.8.1.2). A signal in the form of a varying input current from a microphone or camera goes to the recording head. These signals (which are a function of the signal amplitude and frequency) produce varying magnetic fields at the recording head. As the tape moves past the recording head, the magnetic field orientations of the iron oxide molecules on the tape are changed, thus recording the signal. In the playback mode, the magnetized tape is run past another head, similar in structure to the recording head. The different magnetic field orientations of the iron oxide molecules on the tape induces a voltage in the coil of wire in the playback head. This signal then is sent to a loudspeaker or video player.



Figure 8.8.1.2: Recording and playback heads used with audio and video magnetic tapes. (credit: Steve Jurvetson)

Similar principles apply to computer hard drives, except at a much faster rate. Here recordings are on a coated, spinning disk. Read heads historically were made to work on the principle of induction. However, the input information is carried in digital rather than analog form – a series of 0's or 1's are written upon the spinning hard drive. Today, most hard drive readout devices do not work on the principle of induction, but use a technique known as *giant magnetoresistance*. (The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology.) Another application of induction is found on the magnetic stripe on the back of your personal credit card as used at the grocery store or the ATM machine. This works on the same principle as the audio or video tape mentioned in the last paragraph in which a head reads personal information from your card.

Another application of electromagnetic induction is when electrical signals need to be transmitted across a barrier. Some people use cochlear implants as shown in Figure 8.8.1.3. Sound is picked up by a microphone on the outside of the skull and is used to set up a varying magnetic field. A current is induced in a receiver secured in the bone beneath the skin and transmitted to electrodes in the inner ear. Electromagnetic induction can be used in other instances where electric signals need to be conveyed across various media.



Figure 8.8.1.3: Electromagnetic induction used in transmitting electric currents across mediums. The device on the baby's head induces an electrical current in a receiver secured in the bone beneath the skin. (credit: Bjorn Knetsch)

Another contemporary area of research in which electromagnetic induction is being successfully implemented (and with substantial potential) is transcranial magnetic simulation. A host of disorders, including depression and hallucinations can be traced to irregular localized electrical activity in the brain. In *transcranial magnetic stimulation*, a rapidly varying and very localized magnetic field is placed close to certain sites identified in the brain. Weak electric currents are induced in the identified sites and can result in recovery of electrical functioning in the brain tissue.

Sleep apnea ("the cessation of breath") affects both adults and infants (especially premature babies and it may be a cause of sudden infant deaths [SID]). In such individuals, breath can stop repeatedly during their sleep. A cessation of more than 20 seconds can be very dangerous. Stroke, heart failure, and tiredness are just some of the possible consequences for a person having sleep apnea. The

concern in infants is the stopping of breath for these longer times. One type of monitor to alert parents when a child is not breathing uses electromagnetic induction. A wire wrapped around the infant's chest has an alternating current running through it. The expansion and contraction of the infant's chest as the infant breathes changes the area through the coil. A pickup coil located nearby has an alternating current induced in it due to the changing magnetic field of the initial wire. If the child stops breathing, there will be a change in the induced current, and so a parent can be alerted.

As we shall soon see, the most impactful application of electromagnetic induction is in the process that provides most of our electricity.

Section Summary

- There are many applications of Faraday's Law in technology developed and used in our modern world.

This page titled 8.8.1: Applications of Electromagnetic Induction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.10: Faraday's Law of Induction- Lenz's Law** by OpenStax is licensed CC BY 4.0.

8.9: Magnetism from Electricity

Learning Objectives

- Calculate current that produces a magnetic field.
- Use the right hand rule 2 to determine the direction of current or the direction of magnetic field loops.

How much current is needed to produce a significant magnetic field, perhaps as strong as the Earth's field? Surveyors will tell you that overhead electric power lines create magnetic fields that interfere with their compass readings. Indeed, when Oersted discovered in 1820 that a current in a wire affected a compass needle, he was not dealing with extremely large currents. How does the shape of wires carrying current affect the shape of the magnetic field created? We noted earlier that a current loop created a magnetic field similar to that of a bar magnet, but what about a straight wire or a toroid (doughnut)? How is the direction of a current-created field related to the direction of the current? Answers to these questions are explored in this section, together with a brief discussion of the law governing the fields created by currents.

Magnetic Field Created by a Long Straight Current-Carrying Wire

If we shift our perspective to consider the effect that the moving charge has on the magnetic field, as in **Figure 8.9.1**, we can see how electrical current can produce a magnetic effect. Please note, this is the exact same relationship we were looking at previously, we are simply shifting the perspective to the relationship between electrical current and magnetic field instead of electrical current and magnetic force.

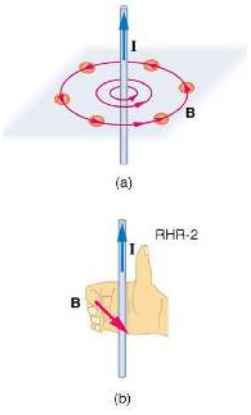


Figure **8.9.1**: (a) Compasses placed near a long straight current-carrying wire indicate that field lines form circular loops centered on the wire. (b) Right hand rule 2 states that, if the right hand thumb points in the direction of the current, the fingers curl in the direction of the field. This rule is consistent with the field mapped for the long straight wire and is valid for any current segment.

Ampere's Law and Others

The magnetic field of a long straight wire has more implications than you might at first suspect. *Each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment.* The formal statement of the direction and magnitude of the field due to each segment is called the **Biot-Savart law**. Using advanced mathematical techniques that are beyond the scope of this course results in a more complete law, **Ampere's law.** This law relates magnetic field and current in a general for a variety of shapes for both the wire and the field. Leaving aside the math, Ampere's Law can be stated as:

Ampere's Law

The magnetic field created by an electrical current is proportional to this electrical current.

More generally it can be stated as magnetic effects can be produced by electrical phenomenon.

Magnetic Field Produced by a Current-Carrying Circular Loop

The magnetic field near a current-carrying loop of wire is shown in Figure 8.9.2. Consider how the direction of the electrical current and the magnetic field are related.

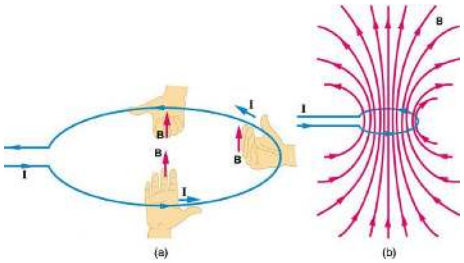


Figure 8.9.2: (a) RHR-2 gives the direction of the magnetic field inside and outside a current-carrying loop. (b) More detailed mapping with compasses or with a Hall probe completes the picture. The field is similar to that of a bar magnet.

Magnetic Field Produced by a Current-Carrying Solenoid

A **solenoid** is a long coil of wire (with many turns or loops, as opposed to a flat loop). Because of its shape, the field inside a solenoid can be very uniform, and also very strong. The field just outside the coils is nearly zero. Figure 8.9.3 shows how the field looks. The magnetic field inside of a current-carrying solenoid is very uniform in direction and magnitude. Only near the ends does it begin to weaken and change direction. The field outside has similar complexities to flat loops and bar magnets, but the **magnetic field strength inside a solenoid** is related to the number of loops and the current in the wire.

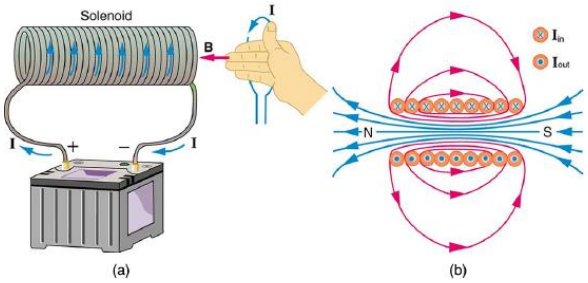


Figure 8.9.3: (a) Because of its shape, the field inside a solenoid of length l is remarkably uniform in magnitude and direction, as indicated by the straight and uniformly spaced field lines. The field outside the coils is nearly zero. (b) This cutaway shows the magnetic field generated by the current in the solenoid.

There are interesting variations of the flat coil and solenoid. For example, the toroidal coil used to confine the reactive particles in tokamaks is much like a solenoid bent into a circle. The field inside a toroid is very strong but circular. Charged particles travel in circles, following the field lines, and collide with one another, perhaps inducing fusion. But the charged particles do not cross field lines and escape the toroid. A whole range of coil shapes are used to produce all sorts of magnetic field shapes. Adding ferromagnetic materials produces greater field strengths and can have a significant effect on the shape of the field. Ferromagnetic materials tend to trap magnetic fields (the field lines bend into the ferromagnetic material, leaving weaker fields outside it) and are used as shields for devices that are adversely affected by magnetic fields, including the Earth's magnetic field.

Section Summary

- The magnetic field is perpendicular to the direction of the electric current which generated it.
- The magnetic field created by current following any path is the sum (or integral) of the fields due to segments along the path (magnitude and direction as for a straight wire), resulting in a general relationship between current and field known as Ampere's law.
- The magnetic field strength created by an electrical current is proportional to this electrical current.
- The magnetic field strength inside a solenoid is proportional the the current and the number of loops of wire. The field inside is very uniform in magnitude and direction.

Glossary

solenoid
: a thin wire wound into a coil that produces a magnetic field when an electric current is passed through it

Biot-Savart law
: a physical law that describes the magnetic field generated by an electric current in terms of a specific equation

Ampere's law
: the physical law that states that the magnetic field around an electric current is proportional to the current; each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment

---

This page titled 8.9: Magnetism from Electricity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.8: Magnetic Fields Produced by Currents- Ampere's Law** by OpenStax is licensed CC BY 4.0.

8.10: Maxwell’s Equations- Summarizing the Connections Between Electricity and Magnetism

Learning Objectives

- Restate Maxwell’s equations.

The Scotsman James Clerk Maxwell (1831–1879) is regarded as the greatest theoretical physicist of the 19th century. (See Figure 8.10.1). Although he died young, Maxwell not only formulated a complete electromagnetic theory, represented by **Maxwell’s equations**, he also developed the kinetic theory of gases and made significant contributions to the understanding of color vision and the nature of Saturn’s rings.



Figure 8.10.1: James Clerk Maxwell, a 19th-century physicist, developed a theory that explained the relationship between electricity and magnetism and correctly predicted that visible light is caused by electromagnetic waves. (credit: G. J. Stodart)

Maxwell brought together all the work that had been done by brilliant physicists such as Oersted, Coulomb, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. Let us first summarize some of the previous observations from this chapter so we can see their connections. Maxwell's equations are summarizing the results of many different types of work in electricity and magnetism and state the same general ideas as we have just summarized, but do so with mathematical rigor.

The first law we looked at for electricity was Coulomb's Law. This law was applied to point charges, but we could extend the concept to look at electric fields as well. This would result in a similar mathematical concept known as Gauss's Law. (Similar to what we saw with magnetism, it is difficult to express this law in equation form without getting into advanced mathematics, so we will not do so here.) As we extended our discussion of electricity into the idea of an electric field, we saw some characteristics of an electric field, namely that the fields originate at positive charges and move towards negative charges.

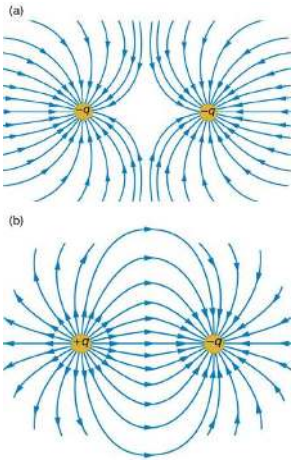


Figure 8.10.2: Characteristic behavior of electric fields. The field lines start on the positive charge and move towards the negative charge.

When we considered magnetic fields, we saw a somewhat different behavior. Instead of starting on a charge and extending to another charge, as we saw with electric fields, magnetic fields form loops that move through the magnet and back again. Whereas

we can have discreet electrical charges, this was not observed for magnetic fields. In fact, one of the phenomena we introduced related to magnetism was that magnetic monopoles do not exist.

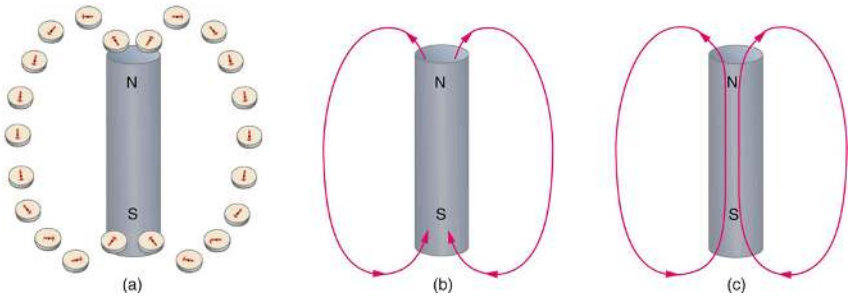


Figure 8.10.3: Characteristic behavior of magnetic fields. The field lines form loops and there are no monopoles.

The behavior of electricity and magnetism are distinctly different, but also mirrors of each other. Electrical fields extend out from electrical charges to other electrical charges. They have a beginning and an end. Magnetic fields do look similar to electric fields, but they do not extend from discreet points, but form loops. The lack of monopoles prevents the possibility of these discreet points from existing. The beauty of Maxwell's equations are the elegant way in which the mathematics shows these connected, but very distinct ideas.

Maxwell also incorporated the ways in which these fields interacted with each other into his analysis. Faraday had found that a changing magnetic field induces an electric field. And the work of Oersted and Ampere led to the analagous conclusion by Maxwell that moving charges result in magnetic fields. The two phenomenon, electricity and magnetism, were inextricably linked together regardless of how different they seemed to be. Maxwell's equations also linked these ideas together.

The results of Maxwell’s equations are paraphrased here in words because their mathematical statement is beyond the level of this text. However, the equations illustrate how apparently simple mathematical statements can elegantly unite and express a multitude of concepts—why mathematics is the language of science.

THE RESULTS OF MAXWELL’S EQUATIONS

1. Electric field lines originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge.
2. Magnetic field lines are continuous, having no beginning or end. No magnetic monopoles are known to exist.
3. A changing magnetic field induces an electric field.
4. Magnetic fields are generated by moving charges or by changing electric fields.

Maxwell’s equations encompass the major laws of electricity and magnetism. What is not so apparent is the symmetry that Maxwell introduced in his mathematical framework. Especially important is his addition of the hypothesis that changing electric fields create magnetic fields. This is exactly analogous (and symmetric) to Faraday’s law of induction and had been suspected for some time, but fits beautifully into Maxwell’s equations. Maxwell's equations showed that electric and magnetic forces are not separate, but different manifestations of the same thing—the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces.

Section Summary

- Maxwell’s prediction of electromagnetic waves resulted from his formulation of a complete and symmetric theory of electricity and magnetism, known as Maxwell’s equations.
- These four equations are paraphrased in this text, rather than presented numerically, and encompass the major laws of electricity and magnetism. First is Gauss’s law for electricity, second is Gauss’s law for magnetism, third is Faraday’s law of induction, including Lenz’s law, and fourth is Ampere’s law in a symmetric formulation that adds another source of magnetism—changing electric fields.

Glossary

Maxwell’s equations

a set of four equations that comprise a complete, overarching theory of electromagnetism

electric field lines

a pattern of imaginary lines that extend between an electric source and charged objects in the surrounding area, with arrows pointed away from positively charged objects and toward negatively charged objects. The more lines in the pattern, the stronger the electric field in that region

magnetic field lines

a pattern of continuous, imaginary lines that emerge from and enter into opposite magnetic poles. The density of the lines indicates the magnitude of the magnetic field

---

This page titled 8.10: Maxwell’s Equations- Summarizing the Connections Between Electricity and Magnetism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.2: Maxwell’s Equations- Electromagnetic Waves Predicted and Observed** by OpenStax is licensed CC BY 4.0.

8.11: Sources of Voltage and the Electrification of America

Almost all electrical power in the United States is created in one of two ways: batteries or generators. (The major exception being solar power, in which the solar panels convert the radiant energy directly into electricity.) In these subsections we will explore each of these ideas, and then look at the history of how we have arrived at the type of power sources we currently rely on the most.

This page titled 8.11: Sources of Voltage and the Electrification of America is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

8.11.1: Batteries

Learning Objectives

- Know the components of a battery: anode, cathode, salt bridge
- Know where electrons are flowing as energy is transferred from a battery.

As we discussed in the previous chapter, a charged particle will move through an electric field similar to a ball rolling down a hill. We can see the ball moving and the hill and understand how it works, as we have seen similar phenomena several times. But why does an electric potential work the way that it does? Potential differences can be caused by a variety of factors, some of which will be explored later in this text. For now we will discuss the functioning of a battery.

There are some chemical reactions which involve the transfer of electrons in order for the reaction to take place. Remember that energy is stored in chemical bonds and is often released from those bonds as chemical reactions take place. It turns out that when electrons are transferred in a chemical reaction, there is an enormous amount of energy that is also released in these reactions. (We will explore chemical reactions in more detail later in this text, but for now consider a combustion reaction as an example of a reaction in which electrons are transferred as the reaction takes place.)

One way to create power from a chemical reaction is to extract useful work from the heat produced as was discussed earlier in this text. But what if we could take those individual electrons that are being transferred and make them do some work in the process? It turns out that we can do this, and in fact this is exactly what occurs inside of a battery. A battery is simply a chemical reaction divided into sections: one section where the electrons are lost and another section where the electrons are gained. These sections are commonly referred to as half-cells. These half-cells are then separated from each other. If a material that allows the flow of electrical energy is used to connect these two half cells, some of that electrical energy may be transferred to do useful work. Once the reaction is complete, there is no longer a potential difference to drive the flow of electrons and the battery ceases to function.

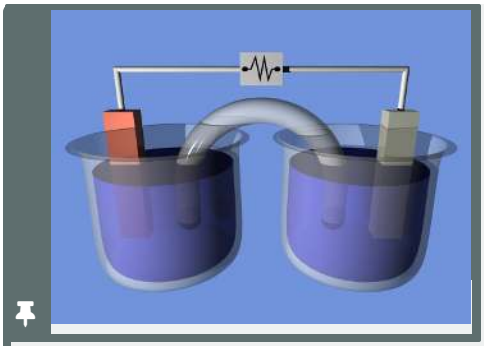


Figure **8.11.1.1**: A voltaic cell for demonstration purposes. By Alksub at the English Wikipedia, CC BY-SA 3.0, (Source)

Figure **8.11.1.1** illustrates a schematic illustration of the function of a battery (though this specific type of battery would not be very functional for any application in which it might move around). Most batteries function because of electrodes, which allow for the flow of electrical energy from the battery to do work and also sometimes function as reactants or products of the reaction. Each electrode is placed in a solution within a half cell. The electrode may react by dissolving into the solution and releasing electrical energy in the process. Or the electrical energy transferred into an electrode might cause a reaction within the solution. The electrode that the electrons flow away from within a battery is called the **anode**. The electrode that the electrons flow to on the other side of the battery is called the **cathode**. In addition to the two half cells, each battery also contains a salt bridge which connects the half cells. The function of the salt bridge it to prevent the accumulation of charge at either of the electrodes within the battery.

MISCONCEPTION ALERT: CONVENTIONAL CURRENT AND FLOW OF ELECTRONS

When we discuss the function of the anode and cathode related to the flow of electricity, we have been considering these from the perspective of the flow of electrons. It turns out that flowing electricity was understood prior to the knowledge of electrons and a convention was created by Benjamin Franklin opposite of what we now understand. As we discussed previously, the convention in studying electricity is to consider the current as flowing from the positive charge. We have seen this with the

diagrams of electric field lines. Conventional Current is the flow of positive charge. However, the electrons are flowing in the opposite direction because they have a negative charge.

Of course the batteries that we use do not have beakers filled with solutions. These would be impractical, as any movement would cause them to spill. However, they do function on the same principle: there are two half reactions which are separated from each other, and the electrons that flow from one of these reactions to the other are available to do work. These batteries can be different sizes and shapes, and in some cases a battery is multiple batteries strung together. But in all cases, batteries function on a similar principle.

Section Summary

- A battery is a chemical reaction separated into two parts.
- Electricity can flow from one part of the battery to the other in order to complete this reaction.
- Electrons flow from the anode to the cathode while conventional current flows from the cathode to the anode.

Glossary

anode

The electrode that electrons flow away from within a battery.

cathode

The electrode into which electrons flow when they return to a battery.

conventional current

The flow of positive charge.

electrode

A portion of a battery capable of transferring an electrical charge to a device outside of the battery, or accepting an electrical charge from a device outside of the battery. It is also sometimes engage in a chemical reaction related to the function of the battery.

salt bridge

connects the anode and cathode half-cells and provides charge neutrality within the battery.

This page titled 8.11.1: Batteries is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **9.7: Conductors and Applications of Electrostatics** by OpenStax is licensed CC BY 4.0.

8.11.2: Induced Voltage and Electrical Generators

Learning Objectives

- Describe how an electric generator produceds an electric current.

We started this section by saying that most of the power that we use comes to us either from a battery or a generator. But previously we stated that there were many sources of energy: fossil fuels, wind, hydroelectric, and others. How can both of these be true at the same time? The reality is that every power plant relies on a generator.

In our discussion on electromagnetic induction, we learned that it was possible to create an electrical charge by spinning a wire through a magnetic field. We also learned that the more loops of wire, the greater the charge of the induced electricity, but also the harder it is to move the wire through the magnetic field. If it was possible to find an energy source that could move the loop of wire through the magnetic field we would be able to transform that kinetic energy into electrical energy. This is essentially what is done with our power sources. wind and hydroelectric power use a turbine to convert the movement of the requisite fluid into rotational energy which can then rotate loops of wire through a magnetic field in order to create an electric current. For fossil fuels, biomass, or nuclear power the production of heat is used to cause high pressure steam to move against a turbine to produce similar results.

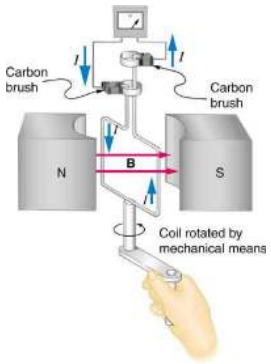


Figure **8.11.2.3**: Rotation of a coil in a magnetic field produces a voltage. This is the basic construction of a generator, where work done to turn the coil is converted to electric energy. Note the generator is very similar in construction to a motor.

One thing to note with the function of a generator is that as the wire loop moves through the magnetic field, the current generated will fluctuate. At the points where the electric current is parallel to the magnetic field, the generated current will be zero. At the points where the direction of the electric current is perpendicular to the magnetic field, the generated current will be the strongest. But also, the current will change the direction that it flows in from one maxima to the next. This is different than the type of current that flows from a battery, and the contrast between them will be discussed in greater detail in the next subsection.

Section Summary

- The movement of an electric wire through a magnetic field will result in an electric current, and can provide a sustained electric current if there is a way to sustain the movement of the electric wire.
- Various energy sources are able to power electric generators.
- Electric generators result in a current whose flow fluctuates as the wire loops move through the magnetic field.

Glossary

generator

a device that produces electric current by turning an electric wire through a magnetic field.

turbine

a device that converts the kinetic energy of a flowing fluid into rotational energy.

electromagnetic induction

the process of inducing a voltage with a change in magnetic flux

This page titled 8.11.2: Induced Voltage and Electrical Generators is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.9: Induced Voltage and Magnetic Flux** by OpenStax is licensed CC BY 4.0.

8.11.3: Alternating Current versus Direct Current

Learning Objectives

- Explain the differences and similarities between AC and DC current.
- Describe rms voltage, current, and average power.
- Explain why AC current is used for power transmission.

Alternating Current

Most of the examples in electric circuits, and particularly those utilizing batteries, have constant voltage sources. Once the current is established, it is thus also a constant. **Direct current** (DC) is the flow of electric charge in only one direction. It is the steady state of a constant-voltage circuit. Many well-known applications, however, use a time-varying voltage source. **Alternating current** (AC) is the flow of electric charge that periodically reverses direction. If the source varies periodically, particularly sinusoidally, the circuit is known as an alternating current circuit. Examples include the commercial and residential power that serves so many of our needs. Figure **8.11.3.1** shows graphs of voltage and current versus time for typical DC and AC power. The AC voltages and frequencies commonly used in homes and businesses vary around the world. The AC voltages range from 100 V to 240 V; the frequencies range from 50 Hz to 60 Hz.

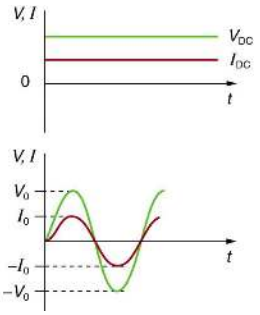


Figure **8.11.3.1**: (a) DC voltage and current are constant in time, once the current is established. (b) A graph of voltage and current versus time for 60-Hz AC power. The voltage and current are sinusoidal and are in phase for a simple resistance circuit. The frequencies and peak voltages of AC sources differ greatly.

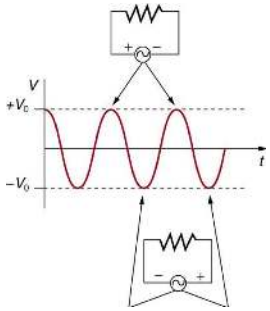


Figure **8.11.3.2**: The potential difference V between the terminals of an AC voltage source fluctuates as shown.

Figure **8.11.3.2** shows a schematic of a simple circuit with an AC voltage source. The voltage between the terminals fluctuates as shown. For this example, the voltage and current are said to be in phase, as seen in Figure **8.11.3.1**(b).

Current in the resistor alternates back and forth just like the driving voltage, since $I = V/R$. If the resistor is a fluorescent light bulb, for example, it brightens and dims 120 times per second as the current repeatedly goes through zero. A 120-Hz flicker is too rapid for your eyes to detect, but if you wave your hand back and forth between your face and a fluorescent light, you will see a stroboscopic effect evidencing AC. The fact that the light output fluctuates means that the power is fluctuating. The power supplied is $P = IV$.

MAKING CONNECTIONS: TAKE-HOME EXPERIMENT—AC/DC LIGHTS

Wave your hand back and forth between your face and a fluorescent light bulb. Do you observe the same thing with the headlights on your car? Explain what you observe. *Warning: Do not look directly at very bright light.*

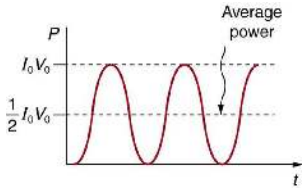


Figure **8.11.3.3**: AC power as a function of time.

We are most often concerned with average power rather than its fluctuations—that 60-W light bulb in your desk lamp has an average power consumption of 60 W, for example. As illustrated in Figure **8.11.3.3**. One common way to express an average is "root-mean-square," or "**rms**." For example, rms voltage of an AC voltage source is found by first squaring the voltage ("square"), taking an average of this value over one period of oscillation ("mean"), and taking the square root ("root").

It is standard practice to quote I_{rms}, V_{rms}, and P_{ave} rather than the peak values. For example, most household electricity is 120 V AC, which means that V_{rms} is 120 V. The common 10-A circuit breaker will interrupt a sustained I_{rms} greater than 10 A. Your 1.0-kW microwave oven consumes $P_{\text{ave}} = 1.0\ \text{kW}$, and so on. You can think of these rms and average values as the equivalent DC values for a simple resistive circuit.

To summarize, when dealing with AC, Ohm's law and the equations for power are completely analogous to those for DC, but rms and average values are used for AC. Thus, for AC, Ohm's law is written

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{R}.$$

The various expressions for AC power P_{ave} are

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}},$$

$$P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R},$$

and

$$P_{\text{ave}} = I_{\text{rms}}^2 R.$$

Why Use AC for Power Distribution?

Most large power-distribution systems are AC. Moreover, the power is transmitted at much higher voltages than the 120-V AC (240 V in most parts of the world) we use in homes and on the job. Economies of scale make it cheaper to build a few very large electric power-generation plants than to build numerous small ones. This necessitates sending power long distances, and it is obviously important that energy losses en route be minimized. High voltages can be transmitted with much smaller power losses than low voltages, as we shall show below. (See Figure **8.11.3.4**.) For safety reasons, the voltage at the user is reduced to familiar values. The crucial factor is that AC voltages can be increased and decreased efficiently with transformers (which uses electromagnetic induction to produce time-varying voltages), while it is more difficult to change DC voltages without power losses. So AC is used in most large power distribution systems.



Figure **8.11.3.4**: Power is distributed over large distances at high voltage to reduce power loss in the transmission lines. The voltages generated at the power plant are stepped up by passive devices called transformers to 330,000 volts (or more in some places worldwide). At the point of use, the transformers reduce the voltage transmitted for safe residential and commercial use. (Credit: GeorgHH, Wikimedia Commons)

Example 8.11.3.1: Power Losses Are Less for High-Voltage Transmission

(a) What current is needed to transmit 100 MW of power at 200 kV? (b) What is the power dissipated by the transmission lines if they have a resistance of $1.00\ \Omega$? (c) What percentage of the power is lost in the transmission lines?

Strategy

We are given $P_{\text{ave}} = 100\ \text{MW}$, $V_{\text{rms}} = 200\ \text{kV}$, and the resistance of the lines is $R = 1.00\ \Omega$. Using these givens, we can find the current flowing (from $P = IV$) and then the power dissipated in the lines ($P = I^2R$), and we take the ratio to the total power transmitted.

Solution

To find the current, we rearrange the relationship $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$ and substitute known values. This gives

$$I_{\text{rms}} = \frac{P_{\text{ave}}}{V_{\text{rms}}} = \frac{100 \times 10^6\ \text{W}}{200 \times 10^3\ \text{V}} = 500\ \text{A}.$$

Solution

Knowing the current and given the resistance of the lines, the power dissipated in them is found from $P_{\text{ave}} = I_{\text{rms}}^2 R$. Substituting the known values gives

$$P_{\text{ave}} = I_{\text{rms}}^2 R = (500\ \text{A})^2(1.00\Omega) = 250\ \text{kW}.$$

Solution

The percent loss is the ratio of this lost power to the total or input power, multiplied by 100:

$$\%\ \text{loss} = \frac{250\ \text{kW}}{100\ \text{MW}} \times 100 = 0.250\ \%.$$

Discussion

One-fourth of a percent is an acceptable loss. Note that if 100 MW of power had been transmitted at 25 kV, then a current of 4000 A would have been needed. This would result in a power loss in the lines of 16.0 MW, or 16.0% rather than 0.250%. The lower the voltage, the more current is needed, and the greater the power loss in the fixed-resistance transmission lines. Of course, lower-resistance lines can be built, but this requires larger and more expensive wires. If superconducting lines could be economically produced, there would be no loss in the transmission lines at all. But, as we shall see in a later chapter, there is a limit to current in superconductors, too. In short, high voltages are more economical for transmitting power, and AC voltage is much easier to raise and lower, so that AC is used in most large-scale power distribution systems.

Section Summary

- Ohm's law for AC is $I_{\text{rms}} = \frac{V_{\text{rms}}}{R}$.
- Expressions for the average power of an AC circuit are $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$, $P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R}$, and $P_{\text{ave}} = I_{\text{rms}}^2 R$, analogous to the expressions for DC circuits.

Glossary

direct current

(DC) the flow of electric charge in only one direction

alternating current

(AC) the flow of electric charge that periodically reverses direction

AC voltage

voltage that fluctuates sinusoidally with time.

AC current

current that fluctuates sinusoidally with time.

rms

a type of average taken for a time-varying quantity by squaring it, taking the mean of the square, and then taking the square-root of the mean.

This page titled 8.11.3: Alternating Current versus Direct Current is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.12: Alternating Current versus Direct Current** by OpenStax is licensed CC BY 4.0.

8.11.4: Transformers

Learning Objectives

- Explain how a transformer works.
- Calculate voltage, current, and/or number of turns given the other quantities.

Transformers do what their name implies—they transform voltages from one value to another. For example, many cell phones, laptops, video games, and power tools and small appliances have a transformer built into their plug-in unit (like that in Figure 8.11.4.1) that changes 120 V or 240 V AC into whatever voltage the device uses. Transformers are also used at several points in the power distribution systems, such as illustrated in Figure 8.11.4.2. Power is sent long distances at high voltages, because less current is required for a given amount of power, and this means less line loss, as was discussed previously. But high voltages pose greater hazards, so that transformers are employed to produce lower voltage at the user's location.



Figure 8.11.4.1: The plug-in transformer has become increasingly familiar with the proliferation of electronic devices that operate on voltages other than common 120 V AC. Most are in the 3 to 12 V range. (credit: Shop Xtreme)

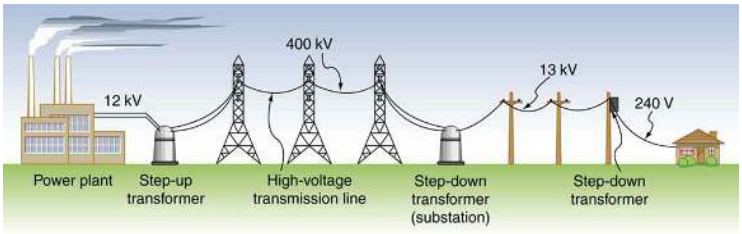


Figure 8.11.4.2: Transformers change voltages at several points in a power distribution system. Electric power is usually generated at greater than 10 kV, and transmitted long distances at voltages over 200 kV—sometimes as great as 700 kV—to limit energy losses. Local power distribution to neighborhoods or industries goes through a substation and is sent short distances at voltages ranging from 5 to 13 kV. This is reduced to 120, 240, or 480 V for safety at the individual user site.

The type of transformer considered in this text—see Figure 8.11.4.3 —is based on Faraday's law of induction and is very similar in construction to the apparatus Faraday used to demonstrate magnetic fields could cause currents. The two coils are called the *primary* and *secondary coils*. In normal use, the input voltage is placed on the primary, and the secondary produces the transformed output voltage. Not only does the iron core trap the magnetic field created by the primary coil, its magnetization increases the field strength. Since the input voltage is AC, a time-varying magnetic flux is sent to the secondary, inducing its AC output voltage.

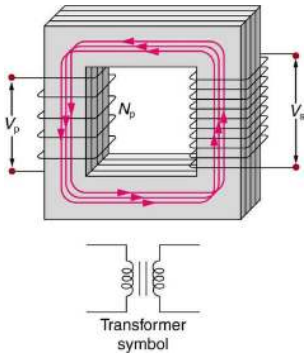


Figure **8.11.4.3**: A typical construction of a simple transformer has two coils wound on a ferromagnetic core that is laminated to minimize eddy currents. The magnetic field created by the primary is mostly confined to and increased by the core, which transmits it to the secondary coil. Any change in current in the primary induces a current in the secondary.

For the simple transformer shown in Figure **8.11.4.3**, the output voltage V_s depends almost entirely on the input voltage V_p and the ratio of the number of loops in the primary and secondary coils. Faraday's law of induction for the secondary coil gives its induced output voltage V_s to be

$$V_s = -N_s \frac{\Delta \Phi}{\Delta t},$$

where N_s is the number of loops in the secondary coil and $\Delta\Phi/\Delta t$ is the rate of change of magnetic flux. The cross-sectional area of the coils is the same on either side, as is the magnetic field strength, and so $\Delta\Phi/\Delta t$ is the same on either side. The input primary voltage V_p is also related to changing flux by

$$V_p = -N_p \frac{\Delta \Phi}{\Delta t}.$$

The reason for this is a little more subtle. Lenz's law tells us that the primary coil opposes the change in flux caused by the input voltage V_p, hence the minus sign. Assuming negligible coil resistance, Kirchhoff's loop rule tells us that the induced voltage exactly equals the input voltage. Taking the ratio of these last two equations yields a useful relationship:

$$\frac{V_s}{V_p} = \frac{N_s}{N_p}.$$

This is known as the **transformer equation**, and it simply states that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of loops in their coils.

The output voltage of a transformer can be less than, greater than, or equal to the input voltage, depending on the ratio of the number of loops in their coils. Some transformers even provide a variable output by allowing connection to be made at different points on the secondary coil. A **step-up transformer** is one that increases voltage, whereas a **step-down transformer** decreases voltage. Assuming, as we have, that resistance is negligible, the electrical power output of a transformer equals its input. This is nearly true in practice—transformer efficiency often exceeds 99%. Equating the power input and output,

$$P_p = I_p V_p = I_s V_s = P_s.$$

Rearranging terms gives

$$\frac{V_s}{V_p} = \frac{I_p}{I_s}.$$

Combining this with $\frac{V_s}{V_p} = \frac{N_s}{N_p}$, we find that

$$\frac{I_s}{I_p} = \frac{N_p}{N_s}$$

is the relationship between the output and input currents of a transformer. So if voltage increases, current decreases. Conversely, if voltage decreases, current increases.

Example 8.11.4.1: Calculating Characteristics of a Step-Up Transformer

A portable x-ray unit has a step-up transformer, the 120 V input of which is transformed to the 100 kV output needed by the x-ray tube. The primary has 50 loops and draws a current of 10.00 A when in use. (a) What is the number of loops in the secondary? (b) Find the current output of the secondary.

Strategy and Solution for (a)

We solve $\frac{V_s}{V_p} = \frac{N_s}{N_p}$ for N_s, the number of loops in the secondary, and enter the known values. This gives

$$\begin{aligned} N_s &= N_p \frac{V_s}{V_p} \\ &= (50)\frac{100,000\ \text{V}}{120\ \text{V}} = 4.17 \times 10^4. \end{aligned}$$

Discussion for (a)

A large number of loops in the secondary (compared with the primary) is required to produce such a large voltage. This would be true for neon sign transformers and those supplying high voltage inside TVs and CRTs.

Strategy and Solution for (b)

We can similarly find the output current of the secondary by solving $\frac{I_s}{I_p} = \frac{N_p}{N_s}$ for I_s and entering known values. This gives

$$\begin{aligned} I_s &= I_p \frac{N_p}{N_s} \\ &= (10.00\ \text{A})\frac{50}{4.17 \times 10^4} = 12.0\ \text{mA}. \end{aligned}$$

Discussion for (b)

As expected, the current output is significantly less than the input. In certain spectacular demonstrations, very large voltages are used to produce long arcs, but they are relatively safe because the transformer output does not supply a large current. Note that the power input here is $P_p = I_p V_p = (10.00\ \text{A})(120\ \text{V}) = 1.20\ \text{kW}$. This equals the power output $P_p = I_s V_s = (12.0\ \text{mA})(100\text{kV}) = 1.20\ \text{kW}$, as we assumed in the derivation of the equations used.

Example 8.11.4.2: Calculating Characteristics of a Step-Down Transformer

A battery charger meant for a series connection of ten nickel-cadmium batteries needs to have a 15.0 V output to charge the batteries. It uses a step-down transformer with a 200-loop primary and a 120 V input. (a) How many loops should there be in the secondary coil? (b) If the charging current is 16.0 A, what is the input current?

Strategy and Solution for (a)

You would expect the secondary to have a small number of loops. Solving $\frac{V_s}{V_p} = \frac{N_s}{N_p}$ for N_s and entering known values gives

$$\begin{aligned} N_s &= N_p \frac{V_s}{V_p} \\ &= (200)\frac{15.0\ \text{V}}{120\ \text{V}} = 25. \end{aligned}$$

Strategy and Solution for (b)

The current input can be obtained by solving $\frac{I_s}{I_p} = \frac{N_p}{N_s}$ for I_p and entering known values. This gives

$$\begin{aligned} I_p &= I_s \frac{N_s}{N_p} \\ &= (16.0\ \text{A})\frac{25}{200} = 2.00\ \text{A}. \end{aligned}$$

Discussion

The number of loops in the secondary is small, as expected for a step-down transformer. We also see that a small input current produces a larger output current in a step-down transformer. When transformers are used to operate large magnets, they sometimes have a small number of very heavy loops in the secondary. This allows the secondary to have low internal resistance and produce large currents. Note again that this solution is based on the assumption of 100% efficiency—or power out equals

power in ($P_p = P_s$)—reasonable for good transformers. In this case the primary and secondary power is 240 W. (Verify this for yourself as a consistency check.) Note that the Ni-Cd batteries need to be charged from a DC power source (as would a 12 V battery). So the AC output of the secondary coil needs to be converted into DC. This is done using something called a rectifier, which uses devices called diodes that allow only a one-way flow of current.

Section Summary

- Transformers use induction to transform voltages from one value to another.
- For a transformer, the voltages across the primary and secondary coils are related by

$$\frac{V_s}{V_p} = \frac{N_s}{N_p},$$

  where V_p and V_s are the voltages across primary and secondary coils having N_p and N_s turns.
- The currents I_p and I_s in the primary and secondary coils are related by $\frac{I_s}{I_p} = \frac{N_p}{N_s}$.
- A step-up transformer increases voltage and decreases current, whereas a step-down transformer decreases voltage and increases current.

Glossary

transformer
a device that transforms voltages from one value to another using induction

transformer equation
the equation showing that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of loops in their coils; $\frac{V_s}{V_p} = \frac{N_s}{N_p}$

step-up transformer
a transformer that increases voltage

step-down transformer
a transformer that decreases voltage

This page titled 8.11.4: Transformers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **10.11: Transformers** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

9: Electromagnetic Radiation and Optics

9.1: Introduction to Light
9.2: Maxwell’s Equations- Electromagnetic Waves Predicted and Observed
9.3: Producing Electromagnetic Radiation
9.4: The Electromagnetic Spectrum- an Overview
9.5: The Electromagnetic Spectrum- Application Notes
9.6: Interaction of Light with Matter
9.6.1: Reflection
9.6.2: Refraction
9.6.3: Dispersion- The Rainbow and Prisms
9.7: Our Extraordinary World of Colors
9.8: Lenses, Mirrors, and Polarization
9.8.1: Image Formation by Lenses
9.8.2: Image Formation by Mirrors
9.8.3: Polarization
9.9: The Greenhouse Effect and Climate Change

This page titled 9: Electromagnetic Radiation and Optics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

9.1: Introduction to Light

Light from this page or screen is formed into an image by the lens of your eye, much as the lens of the camera that made this photograph. Mirrors, like lenses, can also form images that in turn are captured by your eye.



Figure **9.1.1**: Image seen as a result of reflection of light on a plane smooth surface. (credit: NASA Goddard Photo and Video, via Flickr)

Our lives are filled with light. Through vision, the most valued of our senses, light can evoke spiritual emotions, such as when we view a magnificent sunset or glimpse a rainbow breaking through the clouds. Light can also simply amuse us in a theater, or warn us to stop at an intersection. It has innumerable uses beyond vision. Light can carry telephone signals through glass fibers or cook a meal in a solar oven. Life itself could not exist without light's energy. From photosynthesis in plants to the sun warming a cold-blooded animal, its supply of energy is vital.



Figure **9.1.2**: Double Rainbow over the bay of Pocitos in Montevideo, Uruguay. (credit: Madrax, Wikimedia Commons)

We will start our discussion of visible light as a type of electromagnetic wave. This knowledge will help us answer questions regarding the nature of light and vision. What is color, and how do our eyes detect it? Why do diamonds sparkle? How does light travel? How do lenses and mirrors form images? These are but a few of the questions that are answered by the study of optics. Optics is the branch of physics that deals with the behavior of visible light and other electromagnetic waves.

It is convenient to divide optics into two major parts based on the features we are interested in. The wave characteristics of light, such as frequency and wavelength, relates to the colors we perceive and to how we characterize different types of electromagnetic wave along the electromagnetic spectrum. The wave nature of light is also responsible for phenomena such as diffraction and interference. We call this part of optics "wave optics" or "physical optics." But when light interacts with an object that is several times as large as the light's wavelength, its observable behavior is like that of a ray; it does not prominently display its wave characteristics, and we can look at primarily how it refracts and reflects. We call this part of optics "geometric optics" or "ray optics."

This page titled 9.1: Introduction to Light is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.1: Introduction to Light** by OpenStax is licensed CC BY 4.0.

9.2: Maxwell’s Equations- Electromagnetic Waves Predicted and Observed

Learning Objectives

- Know the speed of light, and the significance of it related to Maxwell's equations.
- Restate Maxwell’s equations.

At the end of the last chapter we introduced the results of Maxwell's equations. Maxwell made some remarkable and surprising (even to himself!) predictions from these equations. Before looking at these predictions, we must first backtrack to some other information that was known at the time of Maxwell's work: the speed of light.

The Speed of Light

Early attempts to measure the speed of light, such as those made by Galileo, determined that light moved extremely fast, perhaps instantaneously. The first real evidence that light traveled at a finite speed came from the Danish astronomer Ole Roemer in the late 17th century. Roemer had noted that the average orbital period of one of Jupiter’s moons, as measured from Earth, varied depending on whether Earth was moving toward or away from Jupiter. He correctly concluded that the apparent change in period was due to the change in distance between Earth and Jupiter and the time it took light to travel this distance. From his 1676 data, a value of the speed of light was calculated to be $2.26 \times 10^8\ \mathrm{m/s}$ (only 25% different from today’s accepted value). In more recent times, physicists have measured the speed of light in numerous ways and with increasing accuracy. One particularly direct method, used in 1887 by the American physicist Albert Michelson (1852–1931), is illustrated in Figure 9.2.2. Light reflected from a rotating set of mirrors was reflected from a stationary mirror 35 km away and returned to the rotating mirrors. The time for the light to travel can be determined by how fast the mirrors must rotate for the light to be returned to the observer’s eye.

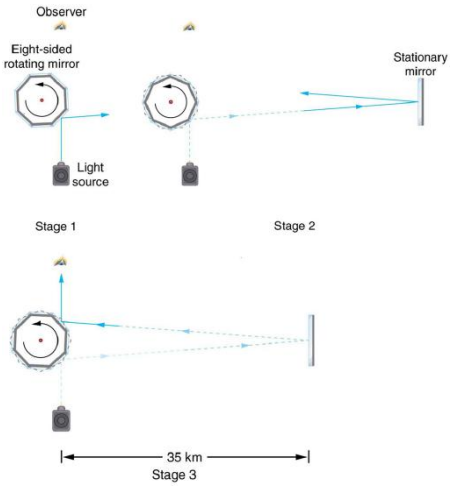


Figure 9.2.2: A schematic of early apparatus used by Michelson and others to determine the speed of light. As the mirrors rotate, the reflected ray is only briefly directed at the stationary mirror. The returning ray will be reflected into the observer's eye only if the next mirror has rotated into the correct position just as the ray returns. By measuring the correct rotation rate, the time for the round trip can be measured and the speed of light calculated. Michelson’s calculated value of the speed of light was only 0.04% different from the value used today.

The speed of light is now known to great precision. In fact, the speed of light in a vacuum c is so important that it is accepted as one of the basic physical quantities and has the fixed value

$$c = 2.9972458 \times 10^8\ \mathrm{m/s} \approx 3.00 \times 10^8\ \mathrm{m/s},$$

where the approximate value of $3.00 \times 10^8\ \mathrm{m/s}$ is used whenever three-digit accuracy is sufficient.

VALUE OF THE SPEED OF LIGHT

$$c = 2.9972458 \times 10^8\ \mathrm{m/s} \approx 3.00 \times 10^8\ \mathrm{m/s}$$

Maxwell's Prediction

Before looking at Maxwell's prediction, let us restate again the results of the equations.

THE RESULTS OF MAXWELL'S EQUATIONS

1. Electric field lines originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge.
2. Magnetic field lines are continuous, having no beginning or end. No magnetic monopoles are known to exist.
3. A changing magnetic field induces an electric field.
4. Magnetic fields are generated by moving charges or by changing electric fields.

Maxwell's equations encompass the major laws of electricity and magnetism. What is not so apparent is the symmetry that Maxwell introduced in his mathematical framework. Especially important is his addition of the hypothesis that changing electric fields create magnetic fields. This is exactly analogous (and symmetric) to Faraday's law of induction and had been suspected for some time, but fits beautifully into Maxwell's equations.

MAKING CONNECTIONS: UNIFICATION OF FORCES

Maxwell's complete and symmetric theory showed that electric and magnetic forces are not separate, but different manifestations of the same thing—the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces.

Since changing electric fields create relatively weak magnetic fields, they could not be easily detected at the time of Maxwell's hypothesis. Maxwell realized, however, that oscillating charges, like those in AC circuits, produce changing electric fields. He predicted that these changing fields would propagate from the source like waves generated on a lake by a jumping fish. The waves predicted from this equation would move at a constant velocity, determined only by empirical constants such as the one present in Coulomb's Law. When Maxwell performed this calculation, the results were a very peculiar number:

$$3.00 \times 10^8 \text{ m/s}.$$

The most peculiar thing about it was that people already knew this was the speed of light! It wasn't likely a coincidence, these calculations indicated that light itself was a form of this electromagnetic radiation that Maxwell had been hypothesizing.

Other wavelengths should exist—it remained to be seen if they did. If so, Maxwell's theory and remarkable predictions would be verified, the greatest triumph of physics since Newton. Experimental verification came within a few years, but not before Maxwell's death.

Hertz's Observations

The German physicist Heinrich Hertz (1857–1894) was the first to generate and detect certain types of electromagnetic waves in the laboratory. Starting in 1887, he performed a series of experiments that not only confirmed the existence of electromagnetic waves, but also verified that they travel at the speed of light.

Hertz used an AC circuit that resonates at a known frequency and connected it to a loop of wire as shown in Figure **9.2.2**. High voltages induced across the gap in the loop produced sparks that were visible evidence of the current in the circuit and that helped generate electromagnetic waves.

Across the laboratory, Hertz had another loop attached to another circuit, which could be tuned (as the dial on a radio) to the same resonant frequency as the first and could, thus, be made to receive electromagnetic waves. This loop also had a gap across which sparks were generated, giving solid evidence that electromagnetic waves had been received.

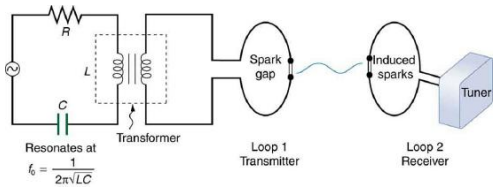


Figure **9.2.2**: The apparatus used by Hertz in 1887 to generate and detect electromagnetic waves. An AC circuit connected to the first loop caused sparks across a gap in the wire loop and generated electromagnetic waves. Sparks across a gap in the second loop located across the laboratory gave evidence that the waves had been received.

Hertz also studied the reflection, refraction, and interference patterns of the electromagnetic waves he generated, verifying their wave character. He was able to determine wavelength from the interference patterns, and knowing their frequency, he could calculate the propagation speed using the equation $v = f\lambda$ (velocity—or speed—equals frequency times wavelength). Hertz was thus able to prove that electromagnetic waves travel at the speed of light. The SI unit for frequency, the hertz ($1\ \text{Hz} = 1\ \text{cycle/sec}$), is named in his honor.

Section Summary

- The speed of light was determined fairly accurately prior to the formulation of Maxwell's equations.
- Electromagnetic waves consist of oscillating electric and magnetic fields and propagate at the speed of light *c*. They were predicted by Maxwell.
- Maxwell's prediction of electromagnetic waves resulted from his formulation of a complete and symmetric theory of electricity and magnetism, known as Maxwell's equations.
- These four equations are paraphrased in this text, rather than presented numerically, and encompass the major laws of electricity and magnetism. First is Gauss's law for electricity, second is Gauss's law for magnetism, third is Faraday's law of induction, including Lenz's law, and fourth is Ampere's law in a symmetric formulation that adds another source of magnetism—changing electric fields.
- Hertz verified Maxwell's equations by creating and detecting that radio waves which also moved at the speed of light.

Glossary

electromagnetic waves
radiation in the form of waves of electric and magnetic energy

Maxwell's equations
a set of four equations that comprise a complete, overarching theory of electromagnetism

hertz
an SI unit denoting the frequency of an electromagnetic wave, in cycles per second

speed of light
in a vacuum, such as space, the speed of light is a constant 3 x 10^8 m/s

electric field lines
a pattern of imaginary lines that extend between an electric source and charged objects in the surrounding area, with arrows pointed away from positively charged objects and toward negatively charged objects. The more lines in the pattern, the stronger the electric field in that region

magnetic field lines
a pattern of continuous, imaginary lines that emerge from and enter into opposite magnetic poles. The density of the lines indicates the magnitude of the magnetic field

This page titled 9.2: Maxwell's Equations- Electromagnetic Waves Predicted and Observed is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.2: Maxwell's Equations- Electromagnetic Waves Predicted and Observed** by OpenStax is licensed CC BY 4.0.

- **11.7: Refraction** by OpenStax is licensed CC BY 4.0.

9.3: Producing Electromagnetic Radiation

Learning Objectives

- Describe the electric and magnetic waves as they move out from a source, such as an AC generator.
- Describe the characterstics of how blackbody radiation works.

We can get a good understanding of **electromagnetic waves** (EM) by considering how they are produced. Whenever a current varies, associated electric and magnetic fields vary, moving out from the source like waves. Perhaps the easiest situation to visualize is a varying current in a long straight wire, produced by an AC generator at its center, as illustrated in Figure 9.3.1.

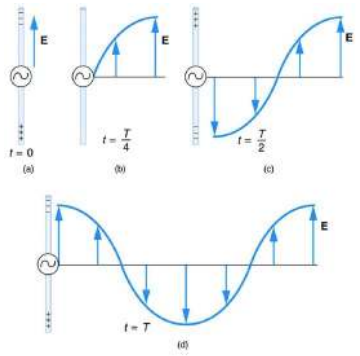


Figure 9.3.1: This long straight gray wire with an AC generator at its center becomes a broadcast antenna for electromagnetic waves. Shown here are the charge distributions at four different times. The electric field (**E**) propagates away from the antenna at the speed of light, forming part of an electromagnetic wave.

The **electric field** (**E**) shown surrounding the wire is produced by the charge distribution on the wire. Both the **E** and the charge distribution vary as the current changes. The changing field propagates outward at the speed of light.

There is an associated **magnetic field** (**B**) which propagates outward as well (see Figure 9.3.2). The electric and magnetic fields are closely related and propagate as an electromagnetic wave. This is what happens in broadcast antennae such as those in radio and TV stations.

Closer examination of the one complete cycle shown in Figure 9.3.1 reveals the periodic nature of the generator-driven charges oscillating up and down in the antenna and the electric field produced. At time $t = 0$, there is the maximum separation of charge, with negative charges at the top and positive charges at the bottom, producing the maximum magnitude of the electric field (or E-field) in the upward direction. One-fourth of a cycle later, there is no charge separation and the field next to the antenna is zero, while the maximum E-field has moved away at speed c.

As the process continues, the charge separation reverses and the field reaches its maximum downward value, returns to zero, and rises to its maximum upward value at the end of one complete cycle. The outgoing wave has an **amplitude** proportional to the maximum separation of charge. Its **wavelength** (λ) is proportional to the period of the oscillation and, hence, is smaller for short periods or high frequencies. (As usual, wavelength and **frequency** (f) are inversely proportional.)

Electric and Magnetic Waves: Moving Together

Based on Maxwell's equations, current in the antenna produces a magnetic field, as shown in Figure 9.3.2. The relationship between **E** and **B** is shown at one instant in Figure 9.3.2(a). As the current varies, the magnetic field varies in magnitude and direction.

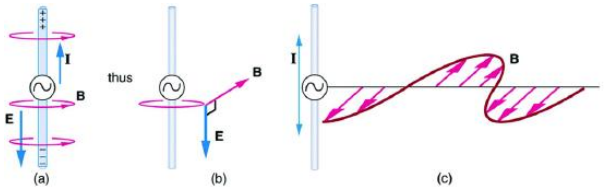


Figure **9.3.2**: (a) The current in the antenna produces the circular magnetic field lines. The current (***I***) produces the separation of charge along the wire, which in turn creates the electric field as shown. (b) The electric and magnetic fields (**E** and **B**) near the wire are perpendicular; they are shown here for one point in space. (c) The magnetic field varies with current and propagates away from the antenna at the speed of light.

The magnetic field lines also propagate away from the antenna at the speed of light, forming the other part of the electromagnetic wave, as seen in Figure **9.3.2**(b). The magnetic part of the wave has the same period and wavelength as the electric part, since they are both produced by the same movement and separation of charges in the antenna.

The electric and magnetic waves are shown together at one instant in time in Figure **9.3.3**. The electric and magnetic fields produced by a long straight wire antenna are exactly in phase. Note that they are perpendicular to one another and to the direction of propagation, making this a **transverse wave**.

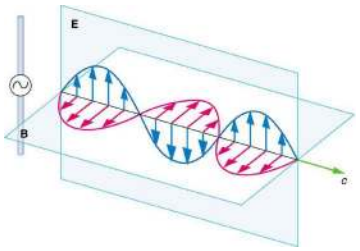


Figure **9.3.3**: A part of the electromagnetic wave sent out from the antenna at one instant in time. The electric and magnetic fields (**E** and **B**) are in phase, and they are perpendicular to one another and the direction of propagation. For clarity, the waves are shown only along one direction, but they propagate out in other directions too.

Electromagnetic waves generally propagate out from a source in all directions, sometimes forming a complex radiation pattern. A linear antenna like this one will not radiate parallel to its length, for example. The wave is shown in one direction from the antenna in Figure **9.3.3** to illustrate its basic characteristics.

Instead of the AC generator, the antenna can also be driven by an AC circuit. In fact, charges radiate whenever they are accelerated. But while a current in a circuit needs a complete path, an antenna has a varying charge distribution forming a **standing wave**, driven by the AC. The dimensions of the antenna are critical for determining the frequency of the radiated electromagnetic waves. This is a **resonant** phenomenon and when we tune radios or TV, we vary electrical properties to achieve appropriate resonant conditions in the antenna.

Receiving Electromagnetic Waves

Electromagnetic waves carry energy away from their source, similar to a sound wave carrying energy away from a standing wave on a guitar string. An antenna for receiving EM signals works in reverse. And like antennas that produce EM waves, receiver antennas are specially designed to resonate at particular frequencies.

An incoming electromagnetic wave accelerates electrons in the antenna, setting up a standing wave. If the radio or TV is switched on, electrical components pick up and amplify the signal formed by the accelerating electrons. The signal is then converted to audio and/or video format. Sometimes big receiver dishes are used to focus the signal onto an antenna.

In fact, charges radiate whenever they are accelerated. When designing circuits, we often assume that energy does not quickly escape AC circuits, and mostly this is true. A broadcast antenna is specially designed to enhance the rate of electromagnetic radiation, and shielding is necessary to keep the radiation close to zero. Some familiar phenomena are based on the production of electromagnetic waves by varying currents. Your microwave oven, for example, sends electromagnetic waves, called microwaves, from a concealed antenna that has an oscillating current imposed on it.

Producing Electromagnetic Radiation

There are many ways in which electromagnetic radiation can be produced. The device which Hertz created was just one artificial method that produced a very low energy electromagnetic wave. Other processes with much higher frequencies and energies are produced in nature. One of the most important processes is related to how we receive most of our light: high temperature objects radiating energy as discussed previously in this text. (Though this is not the only way of emitting light, more efficient fluorescent and light emitting diode technologies now allow us to only emit the wavelengths of light that are useful for illumination.)

For processes which radiate electromagnetic energy as a function of temperature, something called a **blackbody** is a useful model for first approximations. A blackbody is a convenient, ideal emitter that approximates the behavior of many materials when heated. It is "ideal" in the sense that it ignores more complicated interactions in order to simplify the math. A good approximation of a blackbody that can be used to observe blackbody radiation is a metal oven that can be heated to very high temperatures. The oven has a small hole allowing for the light being emitted within the oven to be observed with a spectrometer so that the wavelengths and their intensities can be measured. Both the Sun and the tungsten filament from a standard light bulb emit light according to the characteristics of blackbody radiation. If a blackbody is heated, we will first see it begin to glow a red color, and as it increases in temperature it might move though orange into yellow into white. The color we see is a combination of colors of various intensities being emitted from the blackbody, as shown in Figure **9.3.4**. Notice that even though the light emitted exists along a spectrum, there is a maximum intensity that is associated with the temperature at which the radiation occurs. We can know the different wavelengths at which light is emitted by using devices to create a spectrograph which separate the light. We will discuss the methods necessary to create a spectrograph later in the chapter.

If a glowing object is observed to have a color other than red, orange, yellow, or white, it is not a blackbody but emitting light in a different way which we will discuss later in this text. (There are some extraterrestrial exceptions to this, but it requires temperatures above 7000 K, hotter than the Sun, and definitely hotter than anything we encounter on the Earth.)

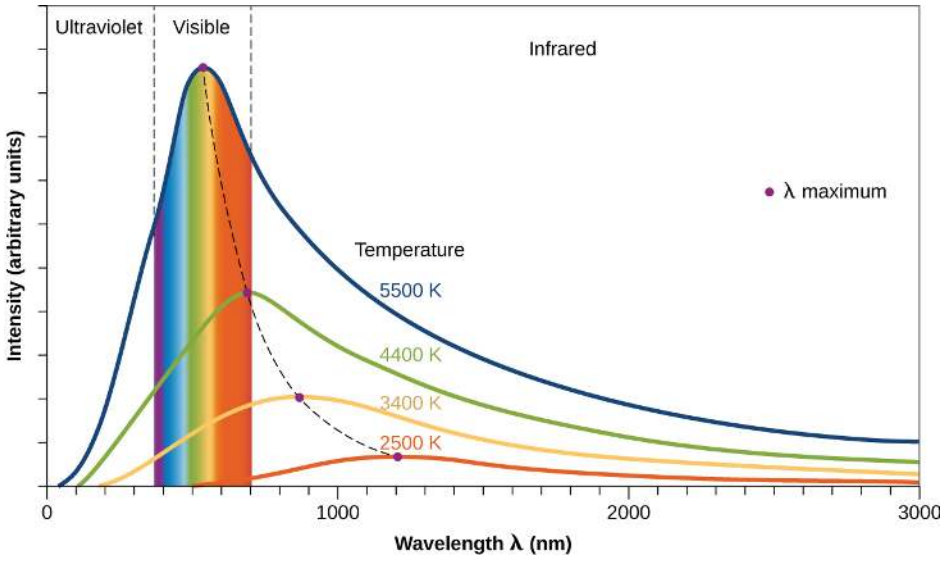


Figure **9.3.4**: Blackbody spectral distribution curves are shown for some representative temperatures.

Section Summary

- Electromagnetic waves are created by oscillating charges (which radiate whenever accelerated) and have the same frequency as the oscillation.
- Since the electric and magnetic fields in most electromagnetic waves are perpendicular to the direction in which the wave moves, it is ordinarily a transverse wave.
- Blackbody radiation is a model for electromagnetic radiation emitted from objects based on their absolute temperature, and follows characteristic patterns.

Glossary

electric field
a vector quantity (**E**); the lines of electric force per unit charge, moving radially outward from a positive charge and in toward a negative charge

electric field strength
the magnitude of the electric field, denoted *E*-field

magnetic field
a vector quantity (**B**); can be used to determine the magnetic force on a moving charged particle

magnetic field strength
the magnitude of the magnetic field, denoted *B*-field

transverse wave
a wave, such as an electromagnetic wave, which oscillates perpendicular to the axis along the line of travel

standing wave
a wave that oscillates in place, with nodes where no motion happens

wavelength
the distance from one peak to the next in a wave

amplitude
the height, or magnitude, of an electromagnetic wave

frequency
the number of complete wave cycles (up-down-up) passing a given point within one second (cycles/second)

resonant
a system that displays enhanced oscillation when subjected to a periodic disturbance of the same frequency as its natural frequency

oscillate
to fluctuate back and forth in a steady beat

blackbody
a convenient, ideal emitter that approximates the behavior of many materials when heated.

This page titled 9.3: Producing Electromagnetic Radiation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.3: Production and Properties of Electromagnetic Waves** by OpenStax is licensed CC BY 4.0.

9.4: The Electromagnetic Spectrum- an Overview

Learning Objectives

- List three “rules of thumb” that apply to the different frequencies along the electromagnetic spectrum.
- Explain why the higher the frequency, the shorter the wavelength of an electromagnetic wave.
- Explain why energy increases with increasing frequency of electromagnetic waves.
- Draw a simplified electromagnetic spectrum, indicating the relative positions, frequencies, and spacing of the different types of radiation bands.

In this module we examine how electromagnetic waves are classified into categories such as radio, infrared, ultraviolet, and so on, so that we can understand some of their similarities as well as some of their differences. We will also find that there are many connections with previously discussed topics, such as wavelength and resonance. A brief overview of the production and utilization of electromagnetic waves is found in Table 9.4.1. Note that the vast majority of the different types of electromagnetic waves originate from atomic and/or molecular electron transitions—that is, from electrons changing their energy levels within atoms or molecules.

Table 9.4.1: Electromagnetic Waves

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Radio and TV	Accelerating charges	Communications, Remote controls	MRI	Requires controls for band use
Microwaves	Accelerating charges and thermal agitation	Communications, Ovens, Radar	Deep heating	Cell phone use
Infrared	Thermal agitations and atomic/molecular electron transitions	Thermal imaging, Heating	Absorbed by atmosphere	Greenhouse effect
Visible light	Thermal agitations and atomic/molecular electron transitions	All pervasive	Photosynthesis, Human vision	
Ultraviolet	Thermal agitations and atomic/molecular electron transitions	Sterilization, Cancer control	Vitamin D production	Ozone depletion, Cancer causing
X-rays	Inner atomic electron transitions and fast collisions	Medical, Security	Medical diagnosis, Cancer therapy	Cancer causing
Gamma rays	Nuclear decay	Nuclear medicine, Security	Medical diagnosis, Cancer therapy	Cancer causing, Radiation damage

CONNECTIONS: WAVES

There are many types of waves, such as water waves and even earthquakes. Among the many shared attributes of waves are propagation speed, frequency, and wavelength. These are always related by the expression $v_W = f\lambda$. This module concentrates on EM waves, but other modules contain examples of all of these characteristics for sound waves and submicroscopic particles.

As noted before, an electromagnetic wave has a frequency and a wavelength associated with it and travels at the speed of light, or c. The relationship among these wave characteristics can be described by $v_W = f\lambda$, where v_W is the propagation speed of the wave, f is the frequency, and λ is the wavelength. Here $v_W = c$, so that for all electromagnetic waves,

$$c = f\lambda.$$

Thus, for all electromagnetic waves, the greater the frequency, the smaller the wavelength.

It is also true that the energy carried by a single photon of an electromagnetic wave is proportional to its frequency,

$$E = hf$$

where *h* is a constant.

Figure **9.4.1** shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies—that is, it shows the electromagnetic spectrum. Many of the characteristics of the various types of electromagnetic waves are related to their frequencies and wavelengths, as we shall see.

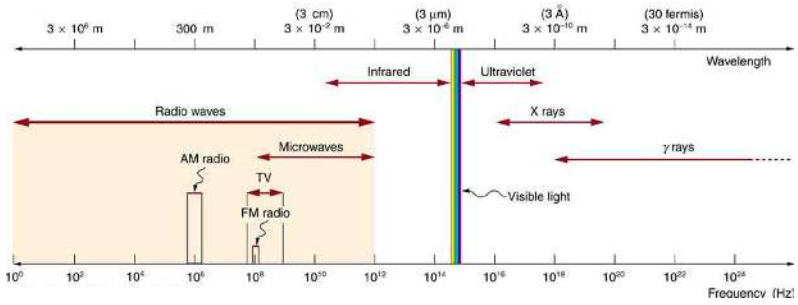


Figure **9.4.1**: The electromagnetic spectrum, showing the major categories of electromagnetic waves. The range of frequencies and wavelengths is remarkable. The dividing line between some categories is distinct, whereas other categories overlap.

ELECTROMAGNETIC SPECTRUM: RULES OF THUMB

Three rules that apply to electromagnetic waves in general are as follows:

- High-frequency electromagnetic waves are more able to penetrate than low-frequency waves.
- High-frequency electromagnetic waves can carry more information per unit time than low-frequency waves.
- The shorter the wavelength of any electromagnetic wave probing a material, the smaller the detail it is possible to resolve.

Note that there are exceptions to these rules of thumb.

Section Summary

- The relationship among the speed of propagation, wavelength, and frequency for any wave is given by $v_W = f\lambda$, so that for electromagnetic waves,

  $$c = f\lambda,$$

  where f is the frequency, λ is the wavelength, and c is the speed of light.
- The relationship between the energy carried by light and its frequency is given by

  $$E = hf$$

- The electromagnetic spectrum is separated into many categories and subcategories, based on the frequency and wavelength, source, and uses of the electromagnetic waves.

Glossary

electromagnetic spectrum
the full range of wavelengths or frequencies of electromagnetic radiation

radio waves
electromagnetic waves with wavelengths in the range from 1 mm to 100 km; they are produced by currents in wires and circuits and by astronomical phenomena

microwaves
electromagnetic waves with wavelengths in the range from 1 mm to 1 m; they can be produced by currents in macroscopic circuits and devices

infrared radiation (IR)

a region of the electromagnetic spectrum with a frequency range that extends from just below the red region of the visible light spectrum up to the microwave region, or from 0.74 μm to 300 μm

ultraviolet radiation (UV)

electromagnetic radiation in the range extending upward in frequency from violet light and overlapping with the lowest X-ray frequencies, with wavelengths from 400 nm down to about 10 nm

visible light

the narrow segment of the electromagnetic spectrum to which the normal human eye responds

X-ray

invisible, penetrating form of very high frequency electromagnetic radiation, overlapping both the ultraviolet range and the γ-ray range

gamma ray

(γ ray); extremely high frequency electromagnetic radiation emitted by the nucleus of an atom, either from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ-ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation

---

This page titled 9.4: The Electromagnetic Spectrum- an Overview is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.4: The Electromagnetic Spectrum- an Overview** by OpenStax is licensed CC BY 4.0.

9.5: The Electromagnetic Spectrum- Application Notes

Learning Objectives

- List and explain the characteristics and applications of different parts of the electromagnetic spectrum.

In this module, we look at the properties of different types of electromagnetic waves. Again, Figure **9.5.1** shows the electromagnetic spectrum. The characteristics of the various types of electromagnetic waves you will read about below are related to their frequencies and wavelengths.

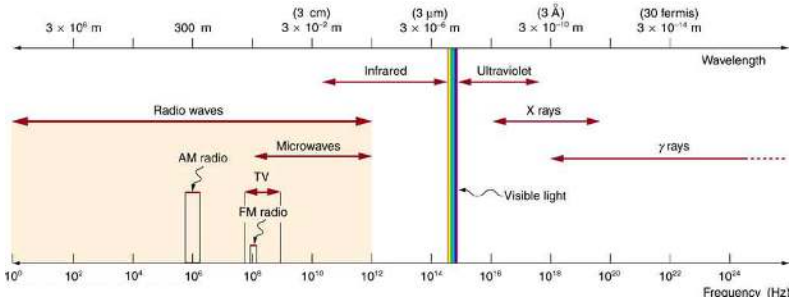


Figure **9.5.1**: The electromagnetic spectrum, showing the major categories of electromagnetic waves. The range of frequencies and wavelengths is remarkable. The dividing line between some categories is distinct, whereas other categories overlap.

Radio and TV Waves

The broad category of **radio waves** is defined to contain any electromagnetic wave produced by currents in wires and circuits. Its name derives from their most common use as a carrier of audio information (i.e., radio). The name is applied to electromagnetic waves of similar frequencies regardless of source. Radio waves from outer space, for example, do not come from alien radio stations. They are created by many astronomical phenomena, and their study has revealed much about nature on the largest scales.

There are many uses for radio waves, and so the category is divided into many subcategories, including microwaves and those electromagnetic waves used for AM and FM radio, cellular telephones, and TV.

The lowest commonly encountered radio frequencies are produced by high-voltage AC power transmission lines at frequencies of 50 or 60 Hz. (See Figure **9.5.2**.) These extremely long wavelength electromagnetic waves (about 6000 km!) are one means of energy loss in long-distance power transmission.



Figure **9.5.2**: This high-voltage traction power line running to Eutingen Railway Substation in Germany radiates electromagnetic waves with very long wavelengths. (credit: Zonk43, Wikimedia Commons)

There is an ongoing controversy regarding potential health hazards associated with exposure to these electromagnetic fields (***E***-fields). Some people suspect that living near such transmission lines may cause a variety of illnesses, including cancer. But demographic data are either inconclusive or simply do not support the hazard theory. Recent reports that have looked at many European and American epidemiological studies have found no increase in risk for cancer due to exposure to ***E***-fields.

Extremely low frequency (ELF) radio waves of about 1 kHz are used to communicate with submerged submarines. The ability of radio waves to penetrate salt water is related to their wavelength (much like ultrasound penetrating tissue)—the longer the wavelength, the farther they penetrate. Since salt water is a good conductor, radio waves are strongly absorbed by it, and very long wavelengths are needed to reach a submarine under the surface. (See Figure **9.5.3**.)

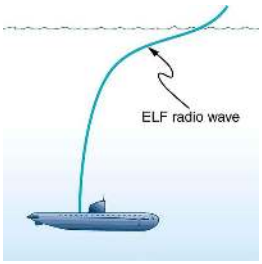


Figure 9.5.3: Very long wavelength radio waves are needed to reach this submarine, requiring extremely low frequency signals (ELF). Shorter wavelengths do not penetrate to any significant depth.

AM radio waves are used to carry commercial radio signals in the frequency range from 540 to 1600 kHz. The abbreviation AM stands for **amplitude modulation**, which is the method for placing information on these waves. (See Figure 9.5.4.) A **carrier wave** having the basic frequency of the radio station, say 1530 kHz, is varied or modulated in amplitude by an audio signal. The resulting wave has a constant frequency, but a varying amplitude.

A radio receiver tuned to have the same resonant frequency as the carrier wave can pick up the signal, while rejecting the many other frequencies impinging on its antenna. The receiver's circuitry is designed to respond to variations in amplitude of the carrier wave to replicate the original audio signal. That audio signal is amplified to drive a speaker or perhaps to be recorded.

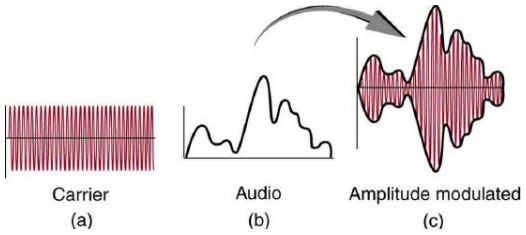


Figure 9.5.4: Amplitude modulation for AM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The amplitude of the carrier is modulated by the audio signal without changing its basic frequency.

FM Radio Waves

FM radio waves are also used for commercial radio transmission, but in the frequency range of 88 to 108 MHz. FM stands for **frequency modulation**, another method of carrying information. (See Figure 9.5.5.) Here a carrier wave having the basic frequency of the radio station, perhaps 105.1 MHz, is modulated in frequency by the audio signal, producing a wave of constant amplitude but varying frequency.

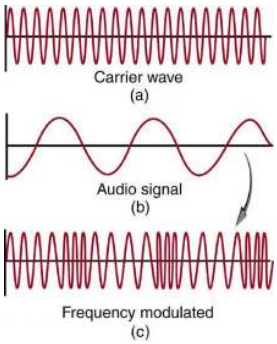


Figure 9.5.5: Frequency modulation for FM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The frequency of the carrier is modulated by the audio signal without changing its amplitude.

Since audible frequencies range up to 20 kHz (or 0.020 MHz) at most, the frequency of the FM radio wave can vary from the carrier by as much as 0.020 MHz. Thus the carrier frequencies of two different radio stations cannot be closer than 0.020 MHz. An FM receiver is tuned to resonate at the carrier frequency and has circuitry that responds to variations in frequency, reproducing the audio information.

FM radio is inherently less subject to noise from stray radio sources than AM radio. The reason is that amplitudes of waves add. So an AM receiver would interpret noise added onto the amplitude of its carrier wave as part of the information. An FM receiver can

be made to reject amplitudes other than that of the basic carrier wave and only look for variations in frequency. It is thus easier to reject noise from FM, since noise produces a variation in amplitude.

Television is also broadcast on electromagnetic waves. Since the waves must carry a great deal of visual as well as audio information, each channel requires a larger range of frequencies than simple radio transmission. TV channels utilize frequencies in the range of 54 to 88 MHz and 174 to 222 MHz. (The entire FM radio band lies between channels 88 MHz and 174 MHz.) These TV channels are called VHF (for **very high frequency**). Other channels called UHF (for **ultra high frequency**) utilize an even higher frequency range of 470 to 1000 MHz.

The TV video signal is AM, while the TV audio is FM. Note that these frequencies are those of free transmission with the user utilizing an old-fashioned roof antenna. Satellite dishes and cable transmission of TV occurs at significantly higher frequencies and is rapidly evolving with the use of the high-definition or HD format.

Example 9.5.1: Calculating Wavelengths of Radio Waves

Calculate the wavelengths of a 1530-kHz AM radio signal, a 105.1-MHz FM radio signal, and a 1.90-GHz cell phone signal.

Strategy

The relationship between wavelength and frequency is $c = f\lambda$, where $c = 3.00 \times 10^8\ \text{m/s}$ is the speed of light (the speed of light is only very slightly smaller in air than it is in a vacuum). We can rearrange this equation to find the wavelength for all three frequencies.

Solution

Rearranging gives

$$\lambda = \frac{c}{f}.$$

(a) For the $f = 1530\ \text{kHz}$ AM radio signal, then,

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8\ \text{m/s}}{1530 \times 10^3 \text{cycles/s}} \\ &= 196\ \text{m}.\end{aligned}$$

(b) For the $f = 105.1\ \text{MHz}$ FM radio signal,

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8\ \text{m/s}}{105.1 \times 10^6 \text{cycles/s}} \\ &= 2.85\ \text{m}.\end{aligned}$$

(c) And for the $f = 1.90\ \text{GHz}$ cell phone,

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8\ \text{m/s}}{1.90 \times 10^9 \text{cycles/s}} \\ &= 0.158\ \text{m}.\end{aligned}$$

Discussion

These wavelengths are consistent with the spectrum in Figure 9.5.1. The wavelengths are also related to other properties of these electromagnetic waves, as we shall see.

The wavelengths found in the preceding example are representative of AM, FM, and cell phones, and account for some of the differences in how they are broadcast and how well they travel. The most efficient length for a linear antenna, such as discussed in "Production of Electromagnetic Waves", is $\lambda/2$, half the wavelength of the electromagnetic wave. Thus a very large antenna is needed to efficiently broadcast typical AM radio with its carrier wavelengths on the order of hundreds of meters.

One benefit to these long AM wavelengths is that they can go over and around rather large obstacles (like buildings and hills), just as ocean waves can go around large rocks. FM and TV are best received when there is a line of sight between the broadcast antenna and receiver, and they are often sent from very tall structures. FM, TV, and mobile phone antennas themselves are much smaller than those used for AM, but they are elevated to achieve an unobstructed line of sight. (See Figure 9.5.6.)

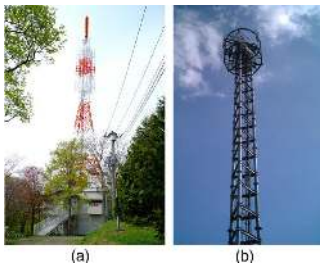


Figure 9.5.6: (a) A large tower is used to broadcast TV signals. The actual antennas are small structures on top of the tower—they are placed at great heights to have a clear line of sight over a large broadcast area. (credit: Ozizo, Wikimedia Commons) (b) The NTT Dokomo mobile phone tower at Tokorozawa City, Japan. (credit: tokoroten, Wikimedia Commons)

Radio Wave Interference

Astronomers and astrophysicists collect signals from outer space using electromagnetic waves. A common problem for astrophysicists is the "pollution" from electromagnetic radiation pervading our surroundings from communication systems in general. Even everyday gadgets like our car keys having the facility to lock car doors remotely and being able to turn TVs on and off using remotes involve radio-wave frequencies. In order to prevent interference between all these electromagnetic signals, strict regulations are drawn up for different organizations to utilize different radio frequency bands.

One reason why we are sometimes asked to switch off our mobile phones (operating in the range of 1.9 GHz) on airplanes and in hospitals is that important communications or medical equipment often uses similar radio frequencies and their operation can be affected by frequencies used in the communication devices.

For example, radio waves used in magnetic resonance imaging (MRI) have frequencies on the order of 100 MHz, although this varies significantly depending on the strength of the magnetic field used and the nuclear type being scanned. MRI is an important medical imaging and research tool, producing highly detailed two- and three-dimensional images. Radio waves are broadcast, absorbed, and reemitted in a resonance process that is sensitive to the density of nuclei (usually protons or hydrogen nuclei).

The wavelength of 100-MHz radio waves is 3 m, yet using the sensitivity of the resonant frequency to the magnetic field strength, details smaller than a millimeter can be imaged. This is a good example of an exception to a rule of thumb (in this case, the rubric that details much smaller than the probe's wavelength cannot be detected). The intensity of the radio waves used in MRI presents little or no hazard to human health.

Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about 10^9 Hz to the highest practical LC resonance at nearly 10^{12} Hz. Since they have high frequencies, their wavelengths are short compared with those of other radio waves—hence the name "microwave."

Microwaves can also be produced by atoms and molecules. They are, for example, a component of electromagnetic radiation generated by **thermal agitation**. The thermal motion of atoms and molecules in any object at a temperature above absolute zero causes them to emit and absorb radiation.

Since it is possible to carry more information per unit time on high frequencies, microwaves are quite suitable for communications. Most satellite-transmitted information is carried on microwaves, as are land-based long-distance transmissions. A clear line of sight between transmitter and receiver is needed because of the short wavelengths involved.

Radar is a common application of microwaves that was first developed in World War II. By detecting and timing microwave echoes, radar systems can determine the distance to objects as diverse as clouds and aircraft. A Doppler shift in the radar echo can be used to determine the speed of a car or the intensity of a rainstorm. Sophisticated radar systems are used to map the Earth and other planets, with a resolution limited by wavelength. (See Figure 9.5.7.) The shorter the wavelength of any probe, the smaller the detail it is possible to observe.

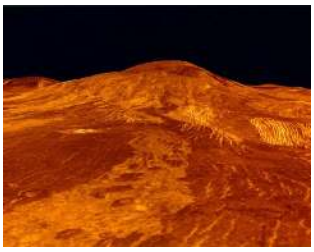


Figure **9.5.7**: An image of Sif Mons with lava flows on Venus, based on Magellan synthetic aperture radar data combined with radar altimetry to produce a three-dimensional map of the surface. The Venusian atmosphere is opaque to visible light, but not to the microwaves that were used to create this image. (credit: NSSDC, NASA/JPL)

Heating with Microwaves

How does the ubiquitous microwave oven produce microwaves electronically, and why does food absorb them preferentially? Microwaves at a frequency of 2.45 GHz are produced by accelerating electrons. The microwaves are then used to induce an alternating electric field in the oven.

Water and some other constituents of food have a slightly negative charge at one end and a slightly positive charge at one end (called polar molecules, which we will cover in more detail later in this text.). The range of microwave frequencies is specially selected so that the polar molecules, in trying to keep orienting themselves with the electric field, absorb these energies and increase their temperatures—called dielectric heating.

The energy thereby absorbed results in thermal agitation heating food and not the plate, which does not contain water. Hot spots in the food are related to constructive and destructive interference patterns. Rotating antennas and food turntables help spread out the hot spots.

Another use of microwaves for heating is within the human body. Microwaves will penetrate more than shorter wavelengths into tissue and so can accomplish “deep heating” (called microwave diathermy). This is used for treating muscular pains, spasms, tendonitis, and rheumatoid arthritis.

MAKING CONNECTIONS: TAKE-HOME EXPERIMENT—MICROWAVE OVENS

1. Look at the door of a microwave oven. Describe the structure of the door. Why is there a metal grid on the door? How does the size of the holes in the grid compare with the wavelengths of microwaves used in microwave ovens? What is this wavelength?
2. Place a glass of water (about 250 ml) in the microwave and heat it for 30 seconds. Measure the temperature gain (the ΔT). Assuming that the power output of the oven is 1000 W, calculate the efficiency of the heat-transfer process.
3. Remove the rotating turntable or moving plate and place a cup of water in several places along a line parallel with the opening. Heat for 30 seconds and measure the ΔT for each position. Do you see cases of destructive interference?

Microwaves generated by atoms and molecules far away in time and space can be received and detected by electronic circuits. Deep space acts like a blackbody with a 2.7 K temperature, radiating most of its energy in the microwave frequency range. In 1964, Penzias and Wilson detected this radiation and eventually recognized that it was the radiation of the Big Bang’s cooled remnants.

Infrared Radiation

The microwave and infrared regions of the electromagnetic spectrum overlap (see Figure 9.5.1). **Infrared radiation** is generally produced by thermal motion and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce infrared radiation.

The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means “below red.” Frequencies at its upper limit are too high to be produced by accelerating electrons in circuits, but small systems, such as atoms and molecules, can vibrate fast enough to produce these waves.

Water molecules rotate and vibrate particularly well at infrared frequencies. Night-vision scopes can detect the infrared emitted by various warm objects, including humans, and convert it to visible light.

We can examine radiant heat transfer from a house by using a camera capable of detecting infrared radiation. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, some of which are called quartz heaters, to preferentially warm us because we absorb infrared better than our surroundings.

The Sun radiates like a nearly perfect blackbody (that is, it has $e = 1$), with a 6000 K surface temperature. About half of the solar energy arriving at the Earth is in the infrared region, with most of the rest in the visible part of the spectrum, and a relatively small amount in the ultraviolet. On average, 50 percent of the incident solar energy is absorbed by the Earth.

The relatively constant temperature of the Earth is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Some of the infrared radiation emitted from the Earth is absorbed by CO_2 and H_2O in the atmosphere and then radiated back to Earth or into outer space. This phenomena will be explored in greater detail later in this chapter.

Visible Light

Visible light is the narrow segment of the electromagnetic spectrum to which the normal human eye responds. Visible light is produced by vibrations and rotations of atoms and molecules, as well as by electronic transitions within atoms and molecules. The receivers or detectors of light largely utilize electronic transitions. We say the atoms and molecules are excited when they absorb and relax when they emit through electronic transitions.

Figure 9.5.8 shows this part of the spectrum, together with the colors associated with particular pure wavelengths. We usually refer to visible light as having wavelengths of between 400 nm and 750 nm. (The retina of the eye actually responds to the lowest ultraviolet frequencies, but these do not normally reach the retina because they are absorbed by the cornea and lens of the eye.)

Red light has the lowest frequencies and longest wavelengths, while violet has the highest frequencies and shortest wavelengths. Blackbody radiation from the Sun peaks in the visible part of the spectrum but is more intense in the red than in the violet, making the Sun yellowish in appearance.

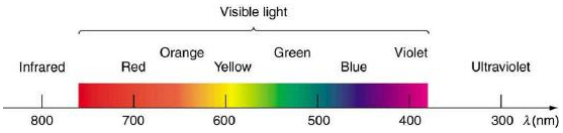


Figure 9.5.8: A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly distinct, nor are those between the seven rainbow colors.

Living things—plants and animals—have evolved to utilize and respond to parts of the electromagnetic spectrum they are embedded in. Visible light is the most predominant and we enjoy the beauty of nature through visible light. Plants are more selective. Photosynthesis makes use of parts of the visible spectrum to make sugars.

Optics is the study of the behavior of visible light and other forms of electromagnetic waves. Optics falls into two distinct categories. When electromagnetic radiation, such as visible light, interacts with objects that are large compared with its wavelength, its motion can be represented by straight lines like rays. Ray optics is the study of such situations and includes lenses and mirrors.

When electromagnetic radiation interacts with objects about the same size as the wavelength or smaller, its wave nature becomes apparent. For example, observable detail is limited by the wavelength, and so visible light can never detect individual atoms, because they are so much smaller than its wavelength. Physical or wave optics is the study of such situations and includes all wave characteristics.

TAKE-HOME EXPERIMENT: COLORS THAT MATCH

When you light a match you see largely orange light; when you light a gas stove you see blue light. Why are the colors different? What other colors are present in these?

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. Ultraviolet is also produced by atomic and molecular motions and electronic transitions. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies, which overlap with the lowest X-

ray frequencies. It was recognized as early as 1801 by Johann Ritter that the solar spectrum had an invisible component beyond the violet range.

Solar UV radiation is broadly subdivided into three regions: UV-A (320–400 nm), UV-B (290–320 nm), and UV-C (220–290 nm), ranked from long to shorter wavelengths (from smaller to larger energies). Most UV-B and all UV-C is absorbed by ozone (O_3) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching the Earth's surface is UV-A.

Human Exposure to UV Radiation

It is largely exposure to UV-B that causes skin cancer. It is estimated that as many as 20% of adults will develop skin cancer over the course of their lifetime. Again, treatment is often successful if caught early. Despite very little UV-B reaching the Earth's surface, there are substantial increases in skin-cancer rates in countries such as Australia, indicating how important it is that UV-B and UV-C continue to be absorbed by the upper atmosphere.

All UV radiation can damage collagen fibers, resulting in an acceleration of the aging process of skin and the formation of wrinkles. Because there is so little UV-B and UV-C reaching the Earth's surface, sunburn is caused by large exposures, and skin cancer from repeated exposure. Some studies indicate a link between overexposure to the Sun when young and melanoma later in life.

The tanning response is a defense mechanism in which the body produces pigments to absorb future exposures in inert skin layers above living cells. Basically UV-B radiation excites DNA molecules, distorting the DNA helix, leading to mutations and the possible formation of cancerous cells.

Repeated exposure to UV-B may also lead to the formation of cataracts in the eyes—a cause of blindness among people living in the equatorial belt where medical treatment is limited. Cataracts, clouding in the eye's lens and a loss of vision, are age related; 60% of those between the ages of 65 and 74 will develop cataracts. However, treatment is easy and successful, as one replaces the lens of the eye with a plastic lens. Prevention is important. Eye protection from UV is more effective with plastic sunglasses than those made of glass.

A major acute effect of extreme UV exposure is the suppression of the immune system, both locally and throughout the body.

Low-intensity ultraviolet is used to sterilize various types of equipment, implying that the energy associated with ultraviolet is deposited in a manner different from lower-frequency electromagnetic waves. (Actually this is true for all electromagnetic waves with frequencies greater than visible light.)

Flash photography is generally not allowed of precious artworks and colored prints because the UV radiation from the flash can cause photo-degradation in the artworks. Often artworks will have an extra-thick layer of glass in front of them, which is especially designed to absorb UV radiation.

UV Light and the Ozone Layer

If all of the Sun's ultraviolet radiation reached the Earth's surface, there would be extremely grave effects on the biosphere from the severe cell damage it causes. However, the layer of ozone (O_3) in our upper atmosphere (10 to 50 km above the Earth) protects life by absorbing most of the dangerous UV radiation.

Unfortunately, today we are observing a depletion in ozone concentrations in the upper atmosphere. This depletion has led to the formation of an "ozone hole" in the upper atmosphere. The hole is more centered over the southern hemisphere, and changes with the seasons, being largest in the spring. This depletion is attributed to the breakdown of ozone molecules by refrigerant gases called chlorofluorocarbons (CFCs).

The UV radiation helps dissociate the CFC's, releasing highly reactive chlorine (Cl) atoms, which catalyze the destruction of the ozone layer. As we will learn later in this text, atoms that act as a catalyst can lead to multiple cycles of a reaction taking place. A single chlorine atom could destroy ozone molecules for up to two years before being transported down to the surface. The CFCs are relatively stable and will contribute to ozone depletion for years to come. CFCs are found in refrigerants, air conditioning systems, foams, and aerosols.

International concern over this problem led to the establishment of the "Montreal Protocol" agreement (1987) to phase out CFC production in most countries. However, developing-country participation is needed if worldwide production and elimination of

CFCs is to be achieved. Probably the largest contributor to CFC emissions today is India. But the protocol seems to be working, as there are signs of an ozone recovery. (See Figure 9.5.9.)

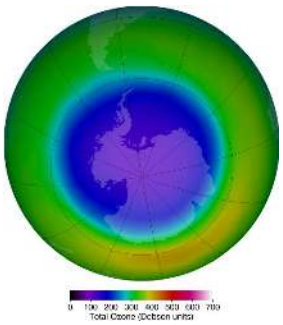


Figure 9.5.9: This map of ozone concentration over Antarctica in October 2011 shows severe depletion suspected to be caused by CFCs. Less dramatic but more general depletion has been observed over northern latitudes, suggesting the effect is global. With less ozone, more ultraviolet radiation from the Sun reaches the surface, causing more damage. (credit: NASA Ozone Watch)

Benefits of UV Light

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin (epidermis) results from exposure to UVB radiation, generally from sunlight. A number of studies indicate lack of vitamin D can result in the development of a range of cancers (prostate, breast, colon), so a certain amount of UV exposure is helpful. Lack of vitamin D is also linked to osteoporosis. Exposures (with no sunscreen) of 10 minutes a day to arms, face, and legs might be sufficient to provide the accepted dietary level. However, in the winter time north of about **37°** latitude, most UVB gets blocked by the atmosphere.

UV radiation is used in the treatment of infantile jaundice and in some skin conditions. It is also used in sterilizing workspaces and tools, and killing germs in a wide range of applications. It is also used as an analytical tool to identify substances.

When exposed to ultraviolet, some substances, such as minerals, glow in characteristic visible wavelengths, a process called fluorescence. So-called black lights emit ultraviolet to cause posters and clothing to fluoresce in the visible.The reason why fluorescence works is because the substance that absorbs the UV light will then emit a lower energy form of light through an internal process. This will often result in the emission of visible light.

Ultraviolet is also used in special microscopes to detect details smaller than those observable with longer-wavelength visible-light microscopes.

X-Rays

In the 1850s, scientists (such as Faraday) began experimenting with high-voltage electrical discharges in tubes filled with rarefied gases. It was later found that these discharges created an invisible, penetrating form of very high frequency electromagnetic radiation. This radiation was called an **X-ray**, because its identity and nature were unknown.

THINGS GREAT AND SMALL: A SUBMICROSCOPIC VIEW OF X-RAY PRODUCTION

X-rays can be created in a high-voltage discharge. They are emitted in the material struck by electrons in the discharge current. There are two mechanisms by which the electrons create X-rays.

The first method is illustrated in Figure 9.5.10. An electron is accelerated in an evacuated tube by a high positive voltage. The electron strikes a metal plate (e.g., copper) and produces X-rays. Since this is a high-voltage discharge, the electron gains sufficient energy to ionize the atom.

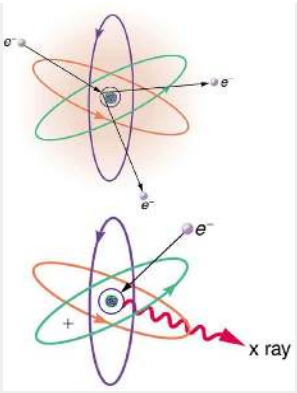


Figure **9.5.10**: Artist's conception of an electron ionizing an atom followed by the recapture of an electron and emission of an X-ray. An energetic electron strikes an atom and knocks an electron out of one of the orbits closest to the nucleus. Later, the atom captures another electron, and the energy released by its fall into a low orbit generates a high-energy EM wave called an X-ray.

In the case shown, an inner-shell electron (one in an orbit relatively close to and tightly bound to the nucleus) is ejected. A short time later, another electron is captured and falls into the orbit in a single great plunge. The energy released by this fall is given to an EM wave known as an X-ray. Since the orbits of the atom are unique to the type of atom, the energy of the X-ray is characteristic of the atom, hence the name characteristic X-ray.

The second method by which an energetic electron creates an X-ray when it strikes a material is illustrated in Figure **9.5.11**. The electron interacts with charges in the material as it penetrates. These collisions transfer kinetic energy from the electron to the electrons and atoms in the material.

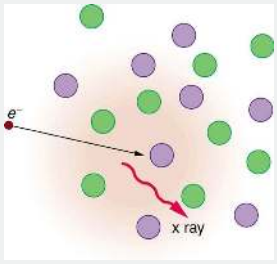


Figure **9.5.11**: Artist's conception of an electron being slowed by collisions in a material and emitting X-ray radiation. This energetic electron makes numerous collisions with electrons and atoms in a material it penetrates. An accelerated charge radiates EM waves, a second method by which X-rays are created.

A loss of kinetic energy implies an acceleration, in this case decreasing the electron's velocity. Whenever a charge is accelerated, it radiates EM waves. Given the high energy of the electron, these EM waves can have high energy. We call them X-rays. Since the process is random, a broad spectrum of X-ray energy is emitted that is more characteristic of the electron energy than the type of material the electron encounters. Such EM radiation is called "bremsstrahlung" (German for "braking radiation").

As described above, there are two methods by which X-rays are created—both are submicroscopic processes and can be caused by high-voltage discharges. While the low-frequency end of the X-ray range overlaps with the ultraviolet, X-rays extend to much higher frequencies (and energies).

X-rays have adverse effects on living cells similar to those of ultraviolet radiation, and they have the additional liability of being more penetrating, affecting more than the surface layers of cells. Cancer and genetic defects can be induced by exposure to X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained. However, questions have risen in recent years as to accidental overexposure of some people during CT scans—a mistake at least in part due to poor monitoring of radiation dose.

The ability of X-rays to penetrate matter depends on density, and so an X-ray image can reveal very detailed density information. Figure **9.5.12** shows an example of the simplest type of X-ray image, an X-ray shadow on film. The amount of information in a simple X-ray image is impressive, but more sophisticated techniques, such as CT scans, can reveal three-dimensional information with details smaller than a millimeter.



Figure **9.5.12**: This shadow X-ray image shows many interesting features, such as artificial heart valves, a pacemaker, and the wires used to close the sternum. (credit: P. P. Urone)

The use of X-ray technology in medicine is called radiology—an established and relatively cheap tool in comparison to more sophisticated technologies. Consequently, X-rays are widely available and used extensively in medical diagnostics. During World War I, mobile X-ray units, advocated by Madame Marie Curie, were used to diagnose soldiers.

Because they can have wavelengths less than 0.01 nm, X-rays can be scattered (a process called X-ray diffraction) to detect the shape of molecules and the structure of crystals. X-ray diffraction was crucial to Crick, Watson, and Wilkins in the determination of the shape of the double-helix DNA molecule.

X-rays are also used as a precise tool for trace-metal analysis in X-ray induced fluorescence, in which the energy of the X-ray emissions are related to the specific types of elements and amounts of materials present.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted. The most penetrating nuclear radiation was called a **gamma ray (γ ray)** (again a name given because its identity and character were unknown), and it was later found to be an extremely high frequency electromagnetic wave.

In fact, γ rays are any electromagnetic radiation emitted by the nucleus of an atom, as we shall learn later in this text. This can be from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ-ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation.

Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. At higher frequencies, γ rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

Figure **9.5.13** shows a medical image based on γ rays. Food spoilage can be greatly inhibited by exposing it to large doses of γ radiation, thereby obliterating responsible microorganisms. Damage to food cells through irradiation occurs as well, and the long-term hazards of consuming radiation-preserved food are unknown and controversial for some groups. Both X-ray and γ-ray technologies are also used in scanning luggage at airports.



Figure **9.5.13**: This is an image of the γ rays emitted by nuclei in a compound that is concentrated in the bones and eliminated through the kidneys. Bone cancer is evidenced by nonuniform concentration in similar structures. For example, some ribs are darker than others. (credit: P. P. Urone)

Detecting Electromagnetic Waves from Space

A final note on star gazing. The entire electromagnetic spectrum is used by researchers for investigating stars, space, and time. As noted earlier, Penzias and Wilson detected microwaves to identify the background radiation originating from the Big Bang. Radio telescopes such as the Arecibo Radio Telescope in Puerto Rico and Parkes Observatory in Australia were designed to detect radio waves.

Infrared telescopes need to have their detectors cooled by liquid nitrogen to be able to gather useful signals. Since infrared radiation is predominantly from thermal agitation, if the detectors were not cooled, the vibrations of the molecules in the antenna would be stronger than the signal being collected.

The most famous of these infrared sensitive telescopes is the James Clerk Maxwell Telescope in Hawaii. The earliest telescopes, developed in the seventeenth century, were optical telescopes, collecting visible light. Telescopes in the ultraviolet, X-ray, and γ-ray regions are placed outside the atmosphere on satellites orbiting the Earth.

The Hubble Space Telescope (launched in 1990) gathers ultraviolet radiation as well as visible light. In the X-ray region, there is the Chandra X-ray Observatory (launched in 1999), and in the γ-ray region, there is the new Fermi Gamma-ray Space Telescope (launched in 2008—taking the place of the Compton Gamma Ray Observatory, 1991–2000.).

Section Summary

- The electromagnetic spectrum is separated into many categories and subcategories, based on the frequency and wavelength, source, and uses of the electromagnetic waves.
- Any electromagnetic wave produced by currents in wires is classified as a radio wave, the lowest frequency electromagnetic waves. Radio waves are divided into many types, depending on their applications, ranging up to microwaves at their highest frequencies.
- Infrared radiation lies below visible light in frequency and is produced by thermal motion and the vibration and rotation of atoms and molecules. Infrared's lower frequencies overlap with the highest-frequency microwaves.
- Visible light is largely produced by electronic transitions in atoms and molecules, and is defined as being detectable by the human eye. Its colors vary with frequency, from red at the lowest to violet at the highest.
- Ultraviolet radiation starts with frequencies just above violet in the visible range and is produced primarily by electronic transitions in atoms and molecules.

- X-rays are created in high-voltage discharges and by electron bombardment of metal targets. Their lowest frequencies overlap the ultraviolet range but extend to much higher values, overlapping at the high end with gamma rays.
- Gamma rays are nuclear in origin and are defined to include the highest-frequency electromagnetic radiation of any type.

Glossary

electromagnetic spectrum
the full range of wavelengths or frequencies of electromagnetic radiation

radio waves
electromagnetic waves with wavelengths in the range from 1 mm to 100 km; they are produced by currents in wires and circuits and by astronomical phenomena

microwaves
electromagnetic waves with wavelengths in the range from 1 mm to 1 m; they can be produced by currents in macroscopic circuits and devices

thermal agitation
the thermal motion of atoms and molecules in any object at a temperature above absolute zero, which causes them to emit and absorb radiation

radar
a common application of microwaves. Radar can determine the distance to objects as diverse as clouds and aircraft, as well as determine the speed of a car or the intensity of a rainstorm

infrared radiation (IR)
a region of the electromagnetic spectrum with a frequency range that extends from just below the red region of the visible light spectrum up to the microwave region, or from 0.74 μm to 300 μm

ultraviolet radiation (UV)
electromagnetic radiation in the range extending upward in frequency from violet light and overlapping with the lowest X-ray frequencies, with wavelengths from 400 nm down to about 10 nm

visible light
the narrow segment of the electromagnetic spectrum to which the normal human eye responds

amplitude modulation (AM)
a method for placing information on electromagnetic waves by modulating the amplitude of a carrier wave with an audio signal, resulting in a wave with constant frequency but varying amplitude

extremely low frequency (ELF)
electromagnetic radiation with wavelengths usually in the range of 0 to 300 Hz, but also about 1kHz

carrier wave
an electromagnetic wave that carries a signal by modulation of its amplitude or frequency

frequency modulation (FM)
a method of placing information on electromagnetic waves by modulating the frequency of a carrier wave with an audio signal, producing a wave of constant amplitude but varying frequency

TV
video and audio signals broadcast on electromagnetic waves

very high frequency (VHF)

TV channels utilizing frequencies in the two ranges of 54 to 88 MHz and 174 to 222 MHz

ultra-high frequency (UHF)

TV channels in an even higher frequency range than VHF, of 470 to 1000 MHz

X-ray

invisible, penetrating form of very high frequency electromagnetic radiation, overlapping both the ultraviolet range and the γ-ray range

gamma ray

(γ ray); extremely high frequency electromagnetic radiation emitted by the nucleus of an atom, either from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ-ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation

---

This page titled 9.5: The Electromagnetic Spectrum- Application Notes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.5: The Electromagnetic Spectrum- Application Notes** by OpenStax is licensed CC BY 4.0.

9.6: Interaction of Light with Matter

Learning Objectives

- Determine if light is absorbed, transmitted, or emitted in a given interaction with matter.

What happens when an electromagnetic wave impinges on a material? There are a variety of possibilities depending on the characteristics of that material as well as the frequency of the electromagnetic wave. As we will see, the material might not interact at all, it might completely absorb the light of the wave, or it might interact in a way that is somewhere between those two extremes.

Most of the world around us is **opaque** to visible light. We know because we can see it! When a material is opaque that means it **absorbs** the light. This means that the energy stored in that object has increased by the amount of the energy of that form of radiation. None of the absorbed radiation leaves the material after impinging upon it. Materials can be opaque to other forms of radiation as well. For example, our atmosphere is opaque to most forms of ultraviolet light. (If it were not we would have even greater concerns about sunburns than we already do.) Matter can also be opaque to some forms of visible light but not others, resulting in colors, as we will see.

Materials that allow electromagnetic radiation to pass through them with little or no absorption are said to be **transparent**. Examples of some common materials transparent to visible light include most molecules in the atmosphere, glass, and water. Just as with absorption, materials can be transparent to any form of electromagnetic radiation, but are hardly ever transparent to all forms of radiation. For example, even though water is transparent to visible light, it is not transparent to microwaves or infrared light. When something is transparent, the radiation that enters the material will eventually exit the material with the same frequency that it entered. In the meantime, it might interact with the material in ways we will discuss later, but it eventually exits with the same energy that it entered with. When radiation impinges on matter that is transparent to it, we say that the radiation is **transmitted.**

Sometimes materials will scatter the light that they transmit. Almost all transmitted light is slowed down somewhat when passing through a material, and this can result in bending the direction of the light. For transparent materials this bending is relatively small and uniform. This kind of bending is referred to as refraction. But some materials will result in a greater change in direction for the light. This is referred to scattering of light. This scattering can either be diffuse scattering when it occurs in many directions, or reflection when it occurs in a specific direction. We will explore each of these three possibilities in more detail in the following subsections.

Finally, we should mention the process by which a material releases energy in the form of radiation. This is referred to as **emission.** We will explore some types of emission in more detail later in this text, but it is worth discussing a little bit here so that we have this knowledge when it comes to our discussion of color. If an object is emitting visible light, we would see it as glowing. We could even identify the color that it is glowing and measure the specific wavelength of this light with some simple devices. This is very different than the color that we see for an object because it is absorbing light. When an object absorbs light, we don't see the absorbed light we see the transmitted light. The object is absorbing the colors we do not see and transmitting the colors that we do see. When an object emits light, the light we see is the color that is emitted. We will discuss color in more detail later in this text.

Section Summary

- Matter that is opaque to a form of radiation absorbs that radiation.
- Matter that is transparent to a form of radiation trasmits that form of radiation.
- Matter that releases energy in the form of radiation emits that radiation.

Glossary

opaque

a material that absorbs a particular type of radiation

transparent

a material that allows a form of radiation to pass through it

transmit
when radiation impinges on matter that is transparent to it and the matter passes along radiation of the same frequency.

emission
when matter releases energy in the form of radiation.

This page titled 9.6: Interaction of Light with Matter is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.4: The Electromagnetic Spectrum- an Overview** by OpenStax is licensed CC BY 4.0.
- **11.5: The Electromagnetic Spectrum- Application Notes** by OpenStax is licensed CC BY 4.0.

9.6.1: Reflection

Learning Objectives

- Explain reflection of light from polished and rough surfaces.

Whenever we look into a mirror, or squint at sunlight glinting from a lake, we are seeing a reflection. When you look at a printed page, too, you are seeing light reflected from it. Large telescopes use reflection to form an image of stars and other astronomical objects.

The law of reflection is illustrated in Figure **9.6.1.1**, which also shows how the angles are measured relative to the perpendicular to the surface at the point where the light ray strikes. We expect to see reflections from smooth surfaces, but Figure **9.6.1.2** illustrates how a rough surface reflects light. Since the light strikes different parts of the surface at different angles, it is reflected in many different directions, or diffused. Diffused light is what allows us to see a sheet of paper from any angle, as illustrated in Figure **9.6.1.3**. Many objects, such as people, clothing, leaves, and walls, have rough surfaces and can be seen from all sides. A mirror, on the other hand, has a smooth surface (compared with the wavelength of light) and reflects light at specific angles, as illustrated in Figure **9.6.1.4**. When the moon reflects from a lake, as shown in Figure **9.6.1.5**, a combination of these effects takes place.

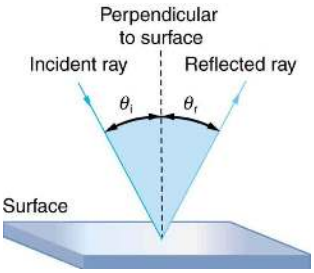


Figure **9.6.1.1**: The law of reflection states that the angle of reflection equals the angle of incidence— $\theta_r = \theta_i$. The angles are measured relative to the perpendicular to the surface at the point where the ray strikes the surface.

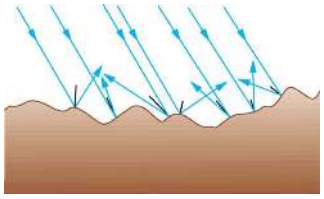


Figure **9.6.1.2**: Light is diffused when it reflects from a rough surface. Here, many parallel rays are incident, but they are reflected at many different angles since the surface is rough.

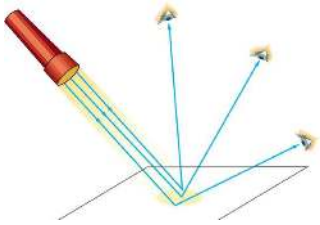


Figure **9.6.1.3**: When a sheet of paper is illuminated with many parallel incident rays, it can be seen at many different angles, because its surface is rough and diffuses the light.

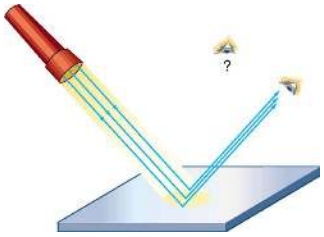


Figure **9.6.1.4**: A mirror illuminated by many parallel rays reflects them in only one direction, since its surface is very smooth. Only the observer at a particular angle will see the reflected light.



Figure 9.6.1.5: Moonlight is spread out when it is reflected by the lake, since the surface is shiny but uneven. (credit: Diego Torres Silvestre, Flickr)

The law of reflection is very simple: The angle of reflection equals the angle of incidence.

THE LAW OF REFLECTION

The angle of reflection equals the angle of incidence.

When we see ourselves in a mirror, it appears that our image is actually behind the mirror. This is illustrated in Figure 9.6.1.6. We see the light coming from a direction determined by the law of reflection. The angles are such that our image is exactly the same distance behind the mirror as we stand away from the mirror. If the mirror is on the wall of a room, the images in it are all behind the mirror, which can make the room seem bigger. Although these mirror images make objects appear to be where they cannot be (like behind a solid wall), the images are not figments of our imagination. Mirror images can be photographed and videotaped by instruments and look just as they do with our eyes (optical instruments themselves). The precise manner in which images are formed by mirrors and lenses will be treated in later sections of this chapter.

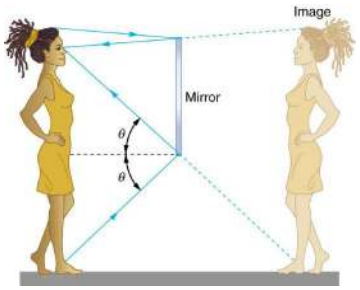


Figure 9.6.1.6: Our image in a mirror is behind the mirror. The two rays shown are those that strike the mirror at just the correct angles to be reflected into the eyes of the person. The image appears to be in the direction the rays are coming from when they enter the eyes.

TAKE-HOME EXPERIMENT: LAW OF REFLECTION

Take a piece of paper and shine a flashlight at an angle at the paper, as shown in Figure 9.6.1.3. Now shine the flashlight at a mirror at an angle. Do your observations confirm the predictions in Figure 9.6.1.3 and Figure 9.6.1.4? Shine the flashlight on various surfaces and determine whether the reflected light is diffuse or not. You can choose a shiny metallic lid of a pot or your skin. Using the mirror and flashlight, can you confirm the law of reflection? You will need to draw lines on a piece of paper showing the incident and reflected rays. (This part works even better if you use a laser pencil.)

Section Summary

- The angle of reflection equals the angle of incidence.
- A mirror has a smooth surface and reflects light at specific angles.
- Light is diffused when it reflects from a rough surface.
- Mirror images can be photographed and videotaped by instruments.

Glossary

mirror

smooth surface that reflects light at specific angles, forming an image of the person or object in front of it

law of reflection

angle of reflection equals the angle of incidence

This page titled 9.6.1: Reflection is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.6: Reflection** by OpenStax is licensed CC BY 4.0.

9.6.2: Refraction

Learning Objectives

- Determine the index of refraction, given the speed of light in a medium.

It is easy to notice some odd things when looking into a fish tank. For example, you may see the same fish appearing to be in two different places. (See Figure 9.6.2.1.) This is because light coming from the fish to us changes direction when it leaves the tank, and in this case, it can travel two different paths to get to our eyes. The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called **refraction**. Refraction is responsible for a tremendous range of optical phenomena, from the action of lenses to voice transmission through optical fibers.

Definition: REFRACTION

The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called refraction.

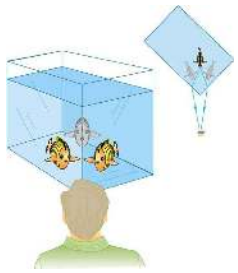


Figure 9.6.2.1: Looking at the fish tank as shown, we can see the same fish in two different locations, because light changes directions when it passes from water to air. In this case, the light can reach the observer by two different paths, and so the fish seems to be in two different places. This bending of light is called refraction and is responsible for many optical phenomena.

Why does light change direction when passing from one material (medium) to another? It is because light changes speed when going from one material to another. As stated earlier in this chapter, the speed of light is now known to great precision. In fact, the speed of light in a vacuum c is so important that it is accepted as one of the basic physical quantities and has the fixed value. The speed of light depends strongly on the type of material, since its interaction with different atoms, crystal lattices, and other substructures varies. We define the **index of refraction** n of a material to be

$$n = \frac{c}{v},$$

where v is the observed speed of light in the material. Since the speed of light is always less than c in matter and equals c only in a vacuum, the index of refraction is always greater than or equal to one.

Definition: INDEX OF REFRACTION

$$n = \frac{c}{v}$$

That is, $n \geq 1$. Table 9.6.2.1 gives the indices of refraction for some representative substances. The values are listed for a particular wavelength of light, because they vary slightly with wavelength. (This can have important effects, such as colors produced by a prism.) Note that for gases, n is close to 1.0. This seems reasonable, since atoms in gases are widely separated and light travels at c in the vacuum between atoms. It is common to take $n = 1$ for gases unless great precision is needed. Although the speed of light v in a medium varies considerably from its value c in a vacuum, it is still a large speed.

Table 9.6.2.1: Index of Refraction in Various Media

Medium	n
Gases at 0ºC, 1 atm	
Air	1.000293
Carbon dioxide	1.00045
Hydrogen	1.000139
Oxygen	1.000271
Liquids at 20ºC	
Benzene	1.501
Carbon disulfide	1.628
Carbon tetrachloride	1.461
Ethanol	1.361
Glycerine	1.473
Water, fresh	1.333
Solids at 20ºC	
Diamond	2.419
Fluorite	1.434
Glass, crown	1.52
Glass, flint	1.66
Ice at 20ºC	1.309
Polystyrene	1.49
Plexiglas	1.51
Quartz, crystalline	1.544
Quartz, fused	1.458
Sodium chloride	1.544
Zircon	1.923

Example **9.6.2.1**: Speed of Light in Matter

Calculate the speed of light in zircon, a material used in jewelry to imitate diamond.

Strategy

The speed of light in a material, v, can be calculated from the index of refraction n of the material using the equation $n = c/v$.

Solution

The equation for index of refraction states that $n = c/v$. Rearranging this to determine v gives

$$v = \frac{c}{n}.$$

The index of refraction for zircon is given as 1.923 in Table **9.6.2.1**, and c is given in the equation for speed of light. Entering these values in the last expression gives

$$v = \frac{3.00 \times 10^8 \text{ m/s}}{1.923}$$
$$= 1.56 \times 10^8 \text{ m/s}.$$

Discussion

This speed is slightly larger than half the speed of light in a vacuum and is still high compared with speeds we normally experience. The only substance listed in Table 9.6.2.1 that has a greater index of refraction than zircon is diamond. We shall see later that the large index of refraction for zircon makes it sparkle more than glass, but less than diamond.

Law of Refraction

Figure 9.6.2.3 shows how a ray of light changes direction when it passes from one medium to another. As before, the angles are measured relative to a perpendicular to the surface at the point where the light ray crosses it. (Some of the incident light will be reflected from the surface, but for now we will concentrate on the light that is transmitted.) The change in direction of the light ray depends on how the speed of light changes. The change in the speed of light is related to the indices of refraction of the media involved. In the situations shown in Figure 9.6.2.3, medium 2 has a greater index of refraction than medium 1. This means that the speed of light is less in medium 2 than in medium 1. Note that as shown in Figure 9.6.2.3(a), the direction of the ray moves closer to the perpendicular when it slows down. Conversely, as shown in Figure 9.6.2.3(b), the direction of the ray moves away from the perpendicular when it speeds up. The path is exactly reversible. In both cases, you can imagine what happens by thinking about pushing a lawn mower from a footpath onto grass, and vice versa. Going from the footpath to grass, the front wheels are slowed and pulled to the side as shown. This is the same change in direction as for light when it goes from a fast medium to a slow one. When going from the grass to the footpath, the front wheels can move faster and the mower changes direction as shown. This, too, is the same change in direction as for light going from slow to fast.

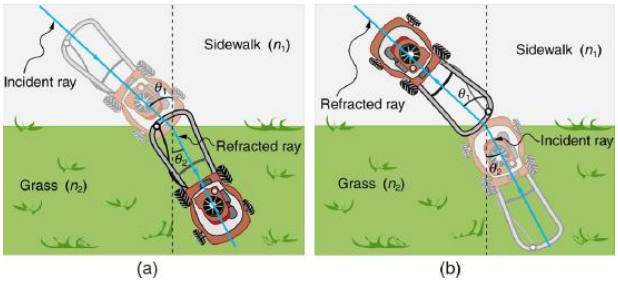


Figure 9.6.2.3: The change in direction of a light ray depends on how the speed of light changes when it crosses from one medium to another. The speed of light is greater in medium 1 than in medium 2 in the situations shown here. (a) A ray of light moves closer to the perpendicular when it slows down. This is analogous to what happens when a lawn mower goes from a footpath to grass. (b) A ray of light moves away from the perpendicular when it speeds up. This is analogous to what happens when a lawn mower goes from grass to footpath. The paths are exactly reversible.

The amount that a light ray changes its direction depends both on the incident angle and the amount that the speed changes. For a ray at a given incident angle, a large change in speed causes a large change in direction, and thus a large change in angle.

TAKE-HOME EXPERIMENT: A BROKEN PENCIL

A classic observation of refraction occurs when a pencil is placed in a glass half filled with water. Do this and observe the shape of the pencil when you look at the pencil sideways, that is, through air, glass, water. Explain your observations. Draw ray diagrams for the situation.

Section Summary

- The changing of a light ray's direction when it passes through variations in matter is called refraction.
- The speed of light in vacuum $c = 2.9972458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$.
- Index of refraction $n = \frac{c}{v}$, where v is the speed of light in the material, c is the speed of light in vacuum, and n is the index of refraction.

Glossary

refraction

changing of a light ray’s direction when it passes through variations in matter

index of refraction

for a material, the ratio of the speed of light in vacuum to that in the material

---

This page titled 9.6.2: Refraction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.7: Refraction** by OpenStax is licensed CC BY 4.0.

9.6.3: Dispersion- The Rainbow and Prisms

Learning Objectives

- Explain the phenomenon of dispersion and discuss its advantages and disadvantages.

Everyone enjoys the spectacle of a rainbow glimmering against a dark stormy sky. How does sunlight falling on clear drops of rain get broken into the rainbow of colors we see? The same process causes white light to be broken into colors by a clear glass prism or a diamond. (See Figure **9.6.3.1**.)

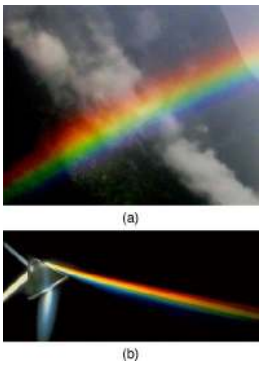


Figure **9.6.3.1**: The colors of the rainbow (a) and those produced by a prism (b) are identical. (credit: Alfredo55, Wikimedia Commons; NASA)

We see about six colors in a rainbow—red, orange, yellow, green, blue, and violet; sometimes indigo is listed, too. Those colors are associated with different wavelengths of light, as shown in Figure **9.6.3.2**. When our eye receives pure-wavelength light, we tend to see only one of the six colors, depending on wavelength. The thousands of other hues we can sense in other situations are our eye's response to various mixtures of wavelengths. White light, in particular, is a fairly uniform mixture of all visible wavelengths. Sunlight, considered to be white, actually appears to be a bit yellow because of its mixture of wavelengths, but it does contain all visible wavelengths. The sequence of colors in rainbows is the same sequence as the colors plotted versus wavelength in Figure **9.6.3.2**. What this implies is that white light is spread out according to wavelength in a rainbow. **Dispersion** is defined as the spreading of white light into its full spectrum of wavelengths. More technically, dispersion occurs whenever there is a process that changes the direction of light in a manner that depends on wavelength. Dispersion, as a general phenomenon, can occur for any type of wave and always involves wavelength-dependent processes.

Definition: DISPERSION

Dispersion is defined to be the spreading of white light into its full spectrum of wavelengths.

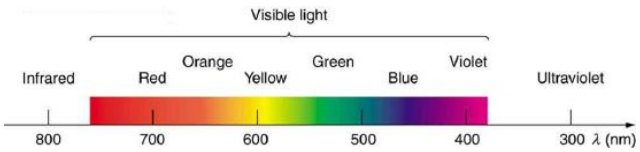


Figure **9.6.3.2**: Even though rainbows are associated with seven colors, the rainbow is a continuous distribution of colors according to wavelengths.

Refraction is responsible for dispersion in rainbows and many other situations. The angle of refraction depends on the index of refraction. We know that the index of refraction n depends on the medium. But for a given medium, n also depends on wavelength. (See Table **9.6.3.1**. Note that, for a given medium, n increases as wavelength decreases and is greatest for violet light. Thus, violet light is bent more than red light, as shown for a prism in Figure **9.6.3.3**(b), and the light is dispersed into the same sequence of wavelengths as seen in Figure **9.6.3.1** and Figure **9.6.3.2**.

MAKING CONNECTIONS: DISPERSION

Any type of wave can exhibit dispersion. Sound waves, all types of electromagnetic waves, and water waves can be dispersed according to wavelength. Dispersion occurs whenever the speed of propagation depends on wavelength, thus separating and spreading out various wavelengths. Dispersion may require special circumstances and can result in spectacular displays such as in the production of a rainbow. This is also true for sound, since all frequencies ordinarily travel at the same speed. If you listen to sound through a long tube, such as a vacuum cleaner hose, you can easily hear it is dispersed by interaction with the tube. Dispersion, in fact, can reveal a great deal about what the wave has encountered that disperses its wavelengths. The dispersion of electromagnetic radiation from outer space, for example, has revealed much about what exists between the stars—the so-called empty space.

Table **9.6.3.1**: Index of Refraction *n* in Selected Media at Various Wavelengths

Medium	Red (660 nm)	Orange (610 nm)	Yellow (580 nm)	Green (550 nm)	Blue (470 nm)	Violet (410 nm)
Water	1.331	1.332	1.333	1.335	1.338	1.342
Diamond	2.410	2.415	2.417	2.426	2.444	2.458
Glass, crown	1.512	1.514	1.518	1.519	1.524	1.530
Glass, flint	1.662	1.665	1.667	1.674	1.684	1.698
Polystyrene	1.488	1.490	1.492	1.493	1.499	1.506
Quartz, fused	1.455	1.456	1.458	1.459	1.462	1.468

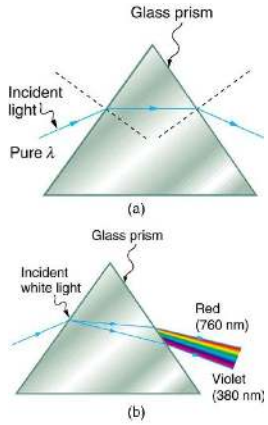


Figure **9.6.3.3**: (a) A pure wavelength of light falls onto a prism and is refracted at both surfaces. (b) White light is dispersed by the prism (shown exaggerated). Since the index of refraction varies with wavelength, the angles of refraction vary with wavelength. A sequence of red to violet is produced, because the index of refraction increases steadily with decreasing wavelength.

Rainbows are produced by a combination of refraction and reflection. You may have noticed that you see a rainbow only when you look away from the sun. Light enters a drop of water and is reflected from the back of the drop, as shown in Figure **9.6.3.4**. The light is refracted both as it enters and as it leaves the drop. Since the index of refraction of water varies with wavelength, the light is dispersed, and a rainbow is observed, as shown in Figure **9.6.3.5**(a). (There is no dispersion caused by reflection at the back surface, since the law of reflection does not depend on wavelength.) The actual rainbow of colors seen by an observer depends on the myriad of rays being refracted and reflected toward the observer's eyes from numerous drops of water. The effect is most spectacular when the background is dark, as in stormy weather, but can also be observed in waterfalls and lawn sprinklers. The arc of a rainbow comes from the need to be looking at a specific angle relative to the direction of the sun, as illustrated in Figure **9.6.3.5** (b). (If there are two reflections of light within the water drop, another "secondary" rainbow is produced. This rare event produces an arc that lies above the primary rainbow arc—see Figure **9.6.3.5**(c).)

Definition: RAINBOWS

Rainbows are produced by a combination of refraction and reflection.

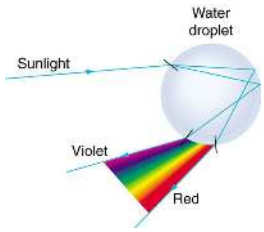


Figure **9.6.3.4**: Part of the light falling on this water drop enters and is reflected from the back of the drop. This light is refracted and dispersed both as it enters and as it leaves the drop.

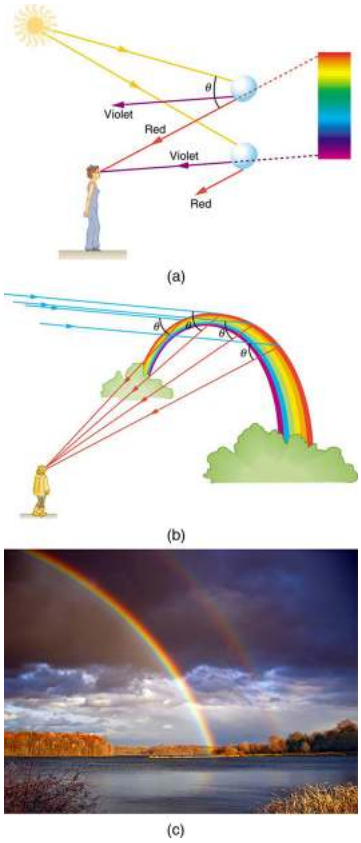


Figure **9.6.3.5**: (a) Different colors emerge in different directions, and so you must look at different locations to see the various colors of a rainbow. (b) The arc of a rainbow results from the fact that a line between the observer and any point on the arc must make the correct angle with the parallel rays of sunlight to receive the refracted rays. (c) Double rainbow. (credit: Nicholas, Wikimedia Commons)

Dispersion may produce beautiful rainbows, but it can cause problems in optical systems. White light used to transmit messages in a fiber is dispersed, spreading out in time and eventually overlapping with other messages. Since a laser produces a nearly pure wavelength, its light experiences little dispersion, an advantage over white light for transmission of information. In contrast, dispersion of electromagnetic waves coming to us from outer space can be used to determine the amount of matter they pass through. As with many phenomena, dispersion can be useful or a nuisance, depending on the situation and our human goals.

Section Summary

- The spreading of white light into its full spectrum of wavelengths is called dispersion.
- Rainbows are produced by a combination of refraction and reflection and involve the dispersion of sunlight into a continuous distribution of colors.
- Dispersion produces beautiful rainbows but also causes problems in certain optical systems.

Glossary

dispersion

spreading of white light into its full spectrum of wavelengths

rainbow

dispersion of sunlight into a continuous distribution of colors according to wavelength, produced by the refraction and reflection of sunlight by water droplets in the sky

---

This page titled 9.6.3: Dispersion- The Rainbow and Prisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.8: Dispersion- The Rainbow and Prisms** by OpenStax is licensed CC BY 4.0.

9.7: Our Extraordinary World of Colors

Learning Objectives

- Know how the following factors contribute to our sense of color: absorption, transmission, emission, additive mixing, and subtractive mixing.
- Know how the structure of our eyes affects our interpretations of color.
- Know how these factors affect our perceptions of natural phenomena such as the color of the ocean, the sky and sunsets.

We have said a lot about how light can and does interact with matter, but how does that result in the colorful world that we live in? Why is the sky blue? Why are sunsets red? We now have the background knowledge to answer these kinds of questions that could be very important to people we meet in our lives.

The visible light portion of the electromagnetic spectrum has wavelengths in approximately (because everyone's vision is a little bit different) the 400-700 nanometer range (a nanometer, “nm,” is 10^{-9} meters). Each specific wavelength corresponds to a different color (Figure **9.7.1**), and when all these wavelengths are present, it appears to us as white light. It turns out that for the light from the Sun and the light from the lamps in your house, the number of wavelengths emitted is essentially infinite. This is why it is called a spectrum. Where does red end and orange begin? From a scientific perspective, we can measure a specific wavelength and agree on a convention for the naming of that color. As we shall see later in this text, sometimes knowing the specifics of a wavelength of light can be very important. But most of our interactions with light are not scientific, they are aesthetic. When is something green or blue or cyan or teal? It is a very subjective decision, and there have probably been many arguments about it. Different cultures throughout time have defined color differently based on what was important to them. At the time that much of the work on optics was being done the number seven and the color indigo were very important. Hence we have the conventional colors of red orange yellow green blue indigo violet. And depending on the source, sometimes either the indigo or violet are left out.



Figure **9.7.1**: A linear representation of the visible light spectrum. (Public Domain; Gringer via Wikipedia)

Absorption and Transmission and Color

Color is perceived in two ways, through additive mixing, where different colors are made by combining different colors of light, and through subtractive mixing, where different wavelengths of light are taken out so that the light is no longer pure white. As shown in Figure **9.7.2**, the idea behind subtractive mixing is that white light (which is made from all the colors mixed together) interacts with an object. The object absorbs some of the light, and then reflects or transmits (or both, depending on the object) the rest of the light, which contacts the eye. The object is perceived as whichever color is not absorbed. In Figure **9.7.2**, white light (simplified as green, red, and blue bands) is shone through a solution. The solution absorbs the red and green wavelengths; however, the blue light is reflected and passes through, so the solution appears blue. (This is why large bodies of water, such as the ocean, appear blue. The bluish color of water is too faint to observe in the smaller quantities of water we see in our drinking glasses, for example.) This procedure takes place whenever an object displays visible color. If none of the light is absorbed, and all is reflected back off, the object appears white; if all of the light is absorbed, and there is none left to reflect or transmit through, the object appears black. This is why white objects do not get as hot when exposed to sunlight as black objects.

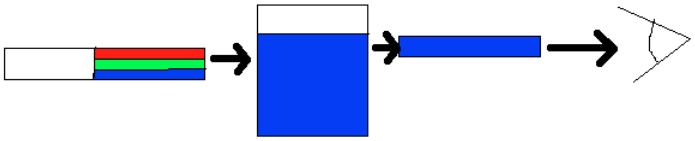


Figure **9.7.2**: When white light (simplified as green, red, and blue bands) is shone on an object, the object absorbs the green and red wavelengths. The blue wavelength is not absorbed, but transmitted. So we see the object as being the color blue. Image created by Libretexts user Deyu Wang (UCD).

Complimentary Colors

Additive mixing is when colors are mixed together to produce a new color. If we combine red, green, and blue, as in our last example, we get white light. There are other combinations that will produce other colors, as you may have learned in art class. It turns out that for the eyes of humans we have receptors which are specifically geared towards red, green, and blue colors. We interpret other colors as mixtures of these three based on how our eyes work. Those other colors do still exist in the world: we know this from the wavelengths of the spectrum.

A color wheel shows the complimentary color for any other color. The complimentary color is the color that will make white light if mixed together. A color wheel is shown in figure **9.7.3**. The color directly across from a given color is its complimentary color. Another way of understanding the color we see for a substance related to absorption of light is that we will see it as the complimentary color to the color of light that is most strongly absorbed. Most objects will absorb light of a variety of different wavelengths, giving rise to a variety of different colors.

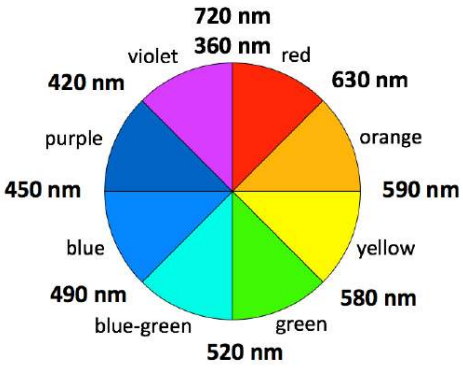


Figure **9.7.3**: Color wheel with wavelengths marked (CC BY-SA 4.0 International; Tem5psu via source)

Factors Affecting What We See in the World

In addition to the properties of light, the properties of our eyes will determine what we are able to see. We already mentioned that our eyes have receptors to red, green, and blue and interpret other colors as combinations of how these receptors interact with the light we see. It also turns out that our eyes are most receptive to yellow light, which happens to be the highest intensity wavelength of light coming from the Sun. As we move away from yellow towards the other colors in the visible range, our ability to see those colors decreases.

Sometimes matter will absorb some forms of radiation but be transparent to others. When there are some wavelengths of visible light that are absorbed and others that are not, this affects the color that we perceive when we look at the object. But if no light is there to impinge on the object, then it is invisible to us. If an object is emitting light, such as the Sun or a lamp, the color of the wavelengths emitted by that object is the color that we see. If it is not emitting that light, then we see the color that it transmits. And there are even more complicated interactions, such as a light that emits only certain wavelengths impinging on an object.

Another factor that we need to consider related to the color of some objects is the scattering of light. As we discussed previously, different wavelengths of light are bent to different extents when they interact with various forms of matter. This gives rise to the rainbows we see when light from the sun interacts with a drop of water, as we had previously discussed. But what happens when there are no raindrops to interact with? In this case, the light from the Sun interacts with what will always be present in our

atmosphere: the atoms of the air. These atoms absorb the light from the Sun and re-emit it. Just as when light travels through any substance, the amount each of the wavelengths are bent will vary. It turns out that violet light is bent the most and red light is bent the least. This means that the violet light is scattered the most. So why does the sky not appear violet? This is mostly because our eyes do not see violet as well. Blue is the color where the degree of scattering and the degree to which our eyes can properly interpret the light match up. The shading of the blue is also affected by how much scattering is occurring in the air. If there is a lot of moisture or particles in the air, it will be a lighter blue, as other interactions mean there are more colors mixed in with it. If the air is very dry, it will be a darker blue, perhaps some people can even see some of it as violet.

Then why are sunsets red? Red is the color that is the least scattered by the atoms of the atmosphere. The more atmosphere that the sunlight has to travel through, the more each of the other colors will be scattered. Sunlight has to travel through the greatest amount of atmosphere at sunrise and sunset. In the middle of the day, the sun is directly overhead and it has the fewest number of atoms for the sunlight to interact with, but at the beginning and end of the day there are more atoms, hence the reddish color of the sky around the sun at both sunrise and sunset.

Section Summary

- The visible portion of the spectrum occurs for electromagnetic radiation with a wavelength between 400 and 700 nm, with colors closer to 400 appearing more violet and transitioning through all of the other visible colors to red before reaching 700.
- When a substance absorbs some wavelengths of light but not others, we see that substance as the colors which were not absorbed, but transmitted. This is called subtractive mixing.
- Additive mixing is when different wavelengths of light are mixed and appear as a different color. Complimentary colors are the ones which create white when they are mixed additively.
- Our eyesight is most receptive to yellow light and is made of receptors that respond to blue, green, and red light. We have difficulty perceiving violet.
- The blue color of the sky and the red color of sunrises and sunsets are observed due to the scattering of light by the atoms in the atmosphere.

Contributors and Attributions

- Jamie MacArthur
- Deyu Wang (UCD)

This page titled 9.7: Our Extraordinary World of Colors is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- **Colors of Coordination Complexes** is licensed CC BY 4.0.

9.8: Lenses, Mirrors, and Polarization

When we understand how light works, we can apply our understanding to create new technologies such as lenses, mirrors, and polarization. We will explore each of these topics in the subsections that follow.

This page titled 9.8: Lenses, Mirrors, and Polarization is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

9.8.1: Image Formation by Lenses

Learning Objectives

- List the rules for ray tracking for thin lenses.
- Illustrate the formation of images using the technique of ray tracking.
- Determine the power of a lens given the focal length.

Lenses are found in a huge array of optical instruments, ranging from a simple magnifying glass to the eye to a camera's zoom lens. In this section, we will use the law of refraction to explore the properties of lenses and how they form images.

The word *lens* derives from the Latin word for a lentil bean, the shape of which is similar to the convex lens in Figure **9.8.1.1**. The convex lens shown has been shaped so that all light rays that enter it parallel to its axis cross one another at a single point on the opposite side of the lens. (The axis is defined to be a line normal to the lens at its center, as shown in Figure **9.8.1.1**.) Such a lens is called a **converging (or convex) lens** for the converging effect it has on light rays. An expanded view of the path of one ray through the lens is shown, to illustrate how the ray changes direction both as it enters and as it leaves the lens. Since the index of refraction of the lens is greater than that of air, the ray moves towards the perpendicular as it enters and away from the perpendicular as it leaves. (This is in accordance with the law of refraction.) Due to the shape of the lens, light is thus bent toward the axis at both surfaces. The point at which the rays cross is defined to be the **focal point** F of the lens. The distance from the center of the lens to its focal point is defined to be the **focal length** f of the lens. Figure **9.8.1.2** shows how a converging lens, such as that in a magnifying glass, can converge the nearly parallel light rays from the sun to a small spot.

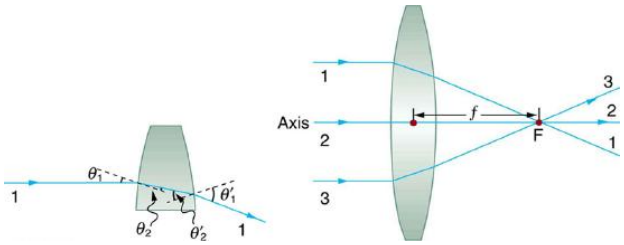


Figure **9.8.1.1**: Rays of light entering a converging lens parallel to its axis converge at its focal point F. (Ray 2 lies on the axis of the lens.) The distance from the center of the lens to the focal point is the focal length, *f*, of the lens. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Definition: CONVERGING OR CONVEX LENS

The lens in which light rays that enter it parallel to its axis cross one another at a single point on the opposite side with a converging effect is called converging lens.

Definition: FOCAL POINT F

The point at which the light rays cross is called the focal point F of the lens.

Definition: FOCAL LENGTH f

The distance from the center of the lens to its focal point is called focal length, f.



Figure **9.8.1.2**: Sunlight focused by a converging magnifying glass can burn paper. Light rays from the sun are nearly parallel and cross at the focal point of the lens. The more powerful the lens, the closer to the lens the rays will cross.

The greater effect a lens has on light rays, the more powerful it is said to be. For example, a powerful converging lens will focus parallel light rays closer to itself and will have a smaller focal length than a weak lens. The light will also focus into a smaller and more intense spot for a more powerful lens. The **power** P of a lens is defined to be the inverse of its focal length. In equation form, this is

$$P = \frac{1}{f}.$$

Definition: POWER P

The **power** P of a lens is defined to be the inverse of its focal length. In equation form, this is

$$P = \frac{1}{f}.$$

,

where f is the focal length of the lens, which must be given in meters (and not cm or mm). The power of a lens P has the unit diopters (D), provided that the focal length is given in meters. That is, $1\ \mathrm{D} = 1/\mathrm{m}$, or $1\ \mathrm{m}^{-1}$. (Note that this power (optical power, actually) is not the same as power in watts. It is a concept related to the effect of optical devices on light.) Optometrists prescribe common spectacles and contact lenses in units of diopters.

Figure 9.8.1.3 shows a concave lens and the effect it has on rays of light that enter it parallel to its axis (the path taken by ray 2 in the figure is the axis of the lens). The concave lens is a **diverging lens**, because it causes the light rays to bend away (diverge) from its axis. In this case, the lens has been shaped so that all light rays entering it parallel to its axis appear to originate from the same point, F, defined to be the focal point of a diverging lens. The distance from the center of the lens to the focal point is again called the focal length f of the lens. Note that the focal length and power of a diverging lens are defined to be negative. For example, if the distance to F in Figure 9.8.1.3 is 5.00 cm, then the focal length is $f = -5.00\ \mathrm{cm}$ and the power of the lens is $P = -20\ \mathrm{D}$. An expanded view of the path of one ray through the lens is shown in the figure to illustrate how the shape of the lens, together with the law of refraction, causes the ray to follow its particular path and be diverged.

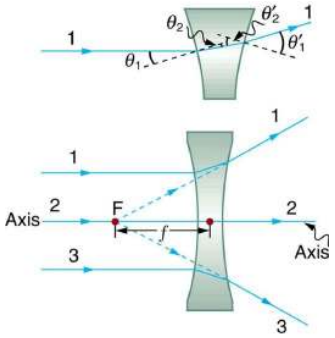


Figure 9.8.1.3: Rays of light entering a diverging lens parallel to its axis are diverged, and all appear to originate at its focal point F. The dashed lines are not rays—they indicate the directions from which the rays appear to come. The focal length f of a diverging lens is negative. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Definition: DIVERGING LENS

A lens that causes the light rays to bend away from its axis is called a diverging lens.

As noted in the initial discussion of the law of refraction, the paths of light rays are exactly reversible. This means that the direction of the arrows could be reversed for all of the rays in Figure 9.8.1.1 and Figure 9.8.1.3. For example, if a point light source is placed at the focal point of a convex lens, as shown in Figure 9.8.1.4, parallel light rays emerge from the other side.

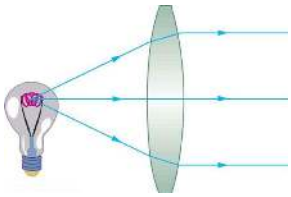


Figure 9.8.1.4: A small light source, like a light bulb filament, placed at the focal point of a convex lens, results in parallel rays of light emerging from the other side. The paths are exactly the reverse of those shown in Figure 9.8.1.1. This technique is used in lighthouses and sometimes in traffic lights to produce a directional beam of light from a source that emits light in all directions.

Ray Tracing and Thin Lenses

Ray tracing is the technique of determining or following (tracing) the paths that light rays take. For rays passing through matter, the law of refraction is used to trace the paths. Here we use ray tracing to help us understand the action of lenses in situations ranging from forming images on film to magnifying small print to correcting nearsightedness. While ray tracing for complicated lenses, such as those found in sophisticated cameras, may require computer techniques, there is a set of simple rules for tracing rays through thin lenses. A **thin lens** is defined to be one whose thickness allows rays to refract, as illustrated in Figure 9.8.1.1, but does not allow properties such as dispersion and aberrations. An ideal thin lens has two refracting surfaces but the lens is thin enough to assume that light rays bend only once. A thin symmetrical lens has two focal points, one on either side and both at the same distance from the lens. (See Figure 9.8.1.5.) Another important characteristic of a thin lens is that light rays through its center are deflected by a negligible amount, as seen in Figure 9.8.1.6.

Definition: THIN LENS

A thin lens is defined to be one whose thickness allows rays to refract but does not allow properties such as dispersion and aberrations.

TAKE-HOME EXPERIMENT: A VISIT TO THE OPTICIAN

Look through your eyeglasses (or those of a friend) backward and forward and comment on whether they act like thin lenses.

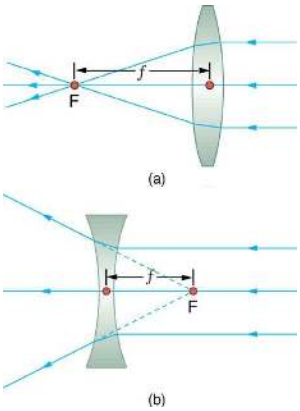


Figure 9.8.1.5: Thin lenses have the same focal length on either side. (a) Parallel light rays entering a converging lens from the right cross at its focal point on the left. (b) Parallel light rays entering a diverging lens from the right seem to come from the focal point on the right.

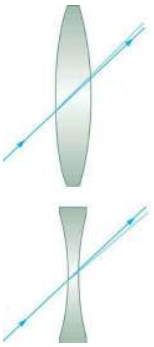


Figure 9.8.1.6: The light ray through the center of a thin lens is deflected by a negligible amount and is assumed to emerge parallel to its original path (shown as a shaded line).

Using paper, pencil, and a straight edge, ray tracing can accurately describe the operation of a lens. The rules for ray tracing for thin lenses are based on the illustrations already discussed:

- **Principal Ray 1**: a ray entering a converging lens parallel to its axis passes through the focal point F of the lens on the other side (see rays 1 and 3 in Figure 9.8.1.1); a ray entering a diverging lens parallel to its axis seems to come from the focal point F (see rays 1 and 3 in Figure 9.8.1.3)
- **Principal Ray 2**: a ray passing through the center of either a converging or a diverging lens does not change direction (see Figure 9.8.1.6, and see ray 2 in Figure 9.8.1.1 and Figure 9.8.1.3)
- **Principal Ray 3**: a ray entering a converging lens through its focal point exits parallel to its axis (the reverse of rays 1 and 3 in Figure 9.8.1.1); a ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis (the reverse of rays 1 and 3 in Figure 9.8.1.3). *The third principal ray is optional and may be used to verify the accuracy of image location.*

Image Formation by Thin Lenses

In some circumstances, a lens forms an obvious image, such as when a movie projector casts an image onto a screen. In other cases, the image is less obvious. Where, for example, is the image formed by eyeglasses? We will use ray tracing for thin lenses to illustrate how they form images, and we will develop equations to describe the image formation quantitatively.

Consider an object some distance away from a converging lens, as shown in Figure 9.8.1.7. To find the location and size of the image formed, we trace the paths of principal rays originating from one point on the object, in this case the top of the person's head. The figure shows the three principal rays from the top of the object as described above. (Note that there are many light rays leaving this point going in many directions, but we concentrate on the three principal rays which can be traced by simple rules.) The first ray is one that enters the lens parallel to its axis and passes through the focal point on the other side (Principal Ray 1). The second ray passes through the center of the lens without changing direction (Principal Ray 2). The third ray passes through the nearer focal point on its way into the lens and leaves the lens parallel to its axis (Principal Ray 3). If correctly drawn, all three rays cross at the same point on the other side of the lens. The image of the top of the person's head is located at this point. *All* rays, including those that are not principal rays, that come from the same point on the top of the person's head are refracted in such a way as to cross at the point shown; the principal rays are what we use to *find* this point. Rays from another point on the object, such as her belt buckle, will also cross at another common point, forming a complete image, as shown. Although three rays are traced in Figure 9.8.1.7, only two are necessary to locate the image. Before applying ray tracing to other situations, let us consider the example shown in Figure 9.8.1.7 in more detail.

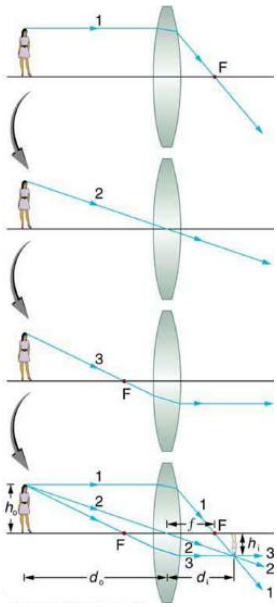


Figure **9.8.1.7**: Ray tracing is used to locate the image formed by a lens. Rays originating from the same point on the object are traced—the three chosen rays each follow one of the rules for ray tracing, so that their paths are easy to determine. The image is located at the point where the rays cross. In this case, a real image—one that can be projected on a screen—is formed.

The image formed in Figure **9.8.1.7** is a **real image**, meaning that it can be projected. That is, light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye, for example. Figure **9.8.1.8** shows how such an image would be projected onto film by a camera lens. This figure also shows how a real image is projected onto the retina by the lens of an eye. Note that the image is there whether it is projected onto a screen or not.

Definition: REAL IMAGE

The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.

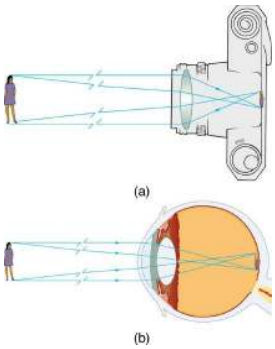


Figure **9.8.1.8**: Real images can be projected. (a) A real image of the person is projected onto film. (b) The converging nature of the multiple surfaces that make up the eye result in the projection of a real image on the retina.

Several important distances appear in Figure **9.8.1.7**. We define d_o to be the object distance, the distance of an object from the center of a lens. **Image distance** d_i is defined to be the distance of the image from the center of a lens. The height of the object and height of the image are given the symbols h_o and h_i, respectively. Images that appear upright relative to the object have heights that are positive and those that are inverted have negative heights. Using the rules of ray tracing and making a scale drawing with paper and pencil, like that in Figure **9.8.1.8**, we can accurately describe the location and size of an image. But the real benefit of ray tracing is in visualizing how images are formed in a variety of situations. To obtain numerical information, we use a pair of equations that can be derived from a geometric analysis of ray tracing for thin lenses. The **thin lens equations** are

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

and

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m.$$

We define the ratio of image height to object height (h_i/h_o) to be the **magnification** m. (The minus sign in the equation above will be discussed shortly.) The thin lens equations are broadly applicable to all situations involving thin lenses (and "thin" mirrors, as we will see later). We will explore many features of image formation in the following worked examples.

Definition: IMAGE DISTANCE

The distance of the image from the center of the lens is called image distance.

THIN LENS EQUATIONS AND MAGNIFICATION

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$$

Real images, such as the one considered in the previous example, are formed by converging lenses whenever an object is farther from the lens than its focal length. This is true for movie projectors, cameras, and the eye. We shall refer to these as *case 1* images. A case 1 image is formed when $d_0 > f$ and f is positive, as in Figure **9.8.1.10**(a). (A summary of the three cases or types of image formation appears at the end of this section.)

A different type of image is formed when an object, such as a person's face, is held close to a convex lens. The image is upright and larger than the object, as seen in Figure **9.8.1.10**(b), and so the lens is called a magnifier. If you slowly pull the magnifier away from the face, you will see that the magnification steadily increases until the image begins to blur. Pulling the magnifier even farther away produces an inverted image as seen in Figure **9.8.1.10**(a). The distance at which the image blurs, and beyond which it inverts, is the focal length of the lens. To use a convex lens as a magnifier, the object must be closer to the converging lens than its focal length. This is called a *case 2* image. A case 2 image is formed when $d_o < f$ and f is positive.

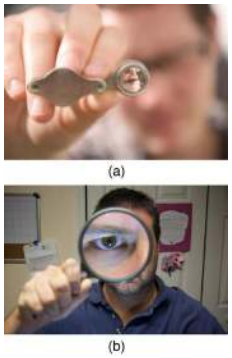


Figure **9.8.1.10**: (a) When a converging lens is held farther away from the face than the lens's focal length, an inverted image is formed. This is a case 1 image. Note that the image is in focus but the face is not, because the image is much closer to the camera taking this photograph than the face. (credit: DaMongMan, Flickr) (b) A magnified image of a face is produced by placing it closer to the converging lens than its focal length. This is a case 2 image. (credit: Casey Fleser, Flickr)

Figure **9.8.1.11** uses ray tracing to show how an image is formed when an object is held closer to a converging lens than its focal length. Rays coming from a common point on the object continue to diverge after passing through the lens, but all appear to originate from a point at the location of the image. The image is on the same side of the lens as the object and is farther away from the lens than the object. This image, like all case 2 images, cannot be projected and, hence, is called a **virtual image**. Light rays only appear to originate at a virtual image; they do not actually pass through that location in space. A screen placed at the location of a virtual image will receive only diffuse light from the object, not focused rays from the lens. Additionally, a screen placed on the opposite side of the lens will receive rays that are still diverging, and so no image will be projected on it. We can see the magnified image with our eyes, because the lens of the eye converges the rays into a real image projected on our retina. Finally, we note that a virtual image is upright and larger than the object, meaning that the magnification is positive and greater than 1.

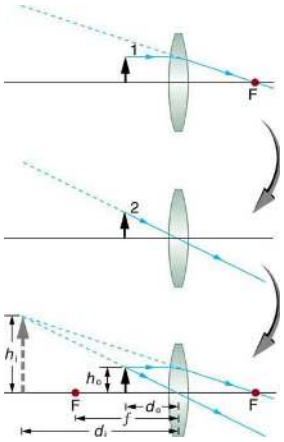


Figure **9.8.1.11**: Ray tracing predicts the image location and size for an object held closer to a converging lens than its focal length. Ray 1 enters parallel to the axis and exits through the focal point on the opposite side, while ray 2 passes through the center of the lens without changing path. The two rays continue to diverge on the other side of the lens, but both appear to come from a common point, locating the upright, magnified, virtual image. This is a case 2 image.

Definition: VIRTUAL IMAGE

An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

A third type of image is formed by a diverging or concave lens. Try looking through eyeglasses meant to correct nearsightedness. (See Figure **9.8.1.12**.) You will see an image that is upright but smaller than the object. This means that the magnification is positive but less than 1. The ray diagram in Figure **9.8.1.13** shows that the image is on the same side of the lens as the object and, hence, cannot be projected—it is a virtual image. Note that the image is closer to the lens than the object. This is a *case 3* image, formed for any object by a negative focal length or diverging lens.



Figure **9.8.1.12**: A car viewed through a concave or diverging lens looks upright. This is a case 3 image. (credit: Daniel Oines, Flickr)

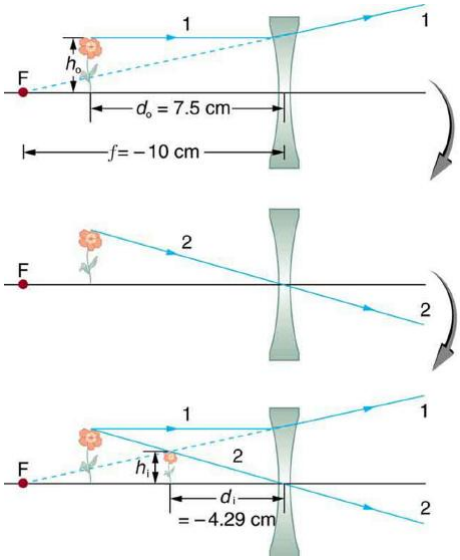


Figure **9.8.1.13**: Ray tracing predicts the image location and size for a concave or diverging lens. Ray 1 enters parallel to the axis and is bent so that it appears to originate from the focal point. Ray 2 passes through the center of the lens without changing path. The two rays appear to come from a common point, locating the upright image. This is a case 3 image, which is closer to the lens than the object and smaller in height.

Table 9.8.1.1 summarizes the three types of images formed by single thin lenses. These are referred to as case 1, 2, and 3 images. Convex (converging) lenses can form either real or virtual images (cases 1 and 2, respectively), whereas concave (diverging) lenses can form only virtual images (always case 3). Real images are always inverted, but they can be either larger or smaller than the object. For example, a slide projector forms an image larger than the slide, whereas a camera makes an image smaller than the object being photographed. Virtual images are always upright and cannot be projected. Virtual images are larger than the object only in case 2, where a convex lens is used. The virtual image produced by a concave lens is always smaller than the object—a case 3 image. We can see and photograph virtual images only by using an additional lens to form a real image.

Table 9.8.1.1: Three Types of Images Formed By Thin Lenses

Type	Formed when	Image type	d_i	m
Case 1	f positive, $d_o > f$	real	positive	negative
Case 2	f positive, $d_o < f$	virtual	negative	positive, $m > 1$
Case 3	f negative	virtual	negative	positive, $m < 1$

TAKE-HOME EXPERIMENT: CONCENTRATING SUNLIGHT

Find several lenses and determine whether they are converging or diverging. In general those that are thicker near the edges are diverging and those that are thicker near the center are converging. On a bright sunny day take the converging lenses outside and try focusing the sunlight onto a piece of paper. Determine the focal lengths of the lenses. Be careful because the paper may start to burn, depending on the type of lens you have selected.

Section Summary

- Light rays entering a converging lens parallel to its axis cross one another at a single point on the opposite side.
- For a converging lens, the focal point is the point at which converging light rays cross; for a diverging lens, the focal point is the point from which diverging light rays appear to originate.
- The distance from the center of the lens to its focal point is called the focal length f.
- Power P of a lens is defined to be the inverse of its focal length, $P = \frac{1}{f}$.
- A lens that causes the light rays to bend away from its axis is called a diverging lens.
- Ray tracing is the technique of graphically determining the paths that light rays take.
- The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.
- Thin lens equations are $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$ and $\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$ (magnification).
- The distance of the image from the center of the lens is called image distance.
- An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

Glossary

converging lens
a convex lens in which light rays that enter it parallel to its axis converge at a single point on the opposite side

diverging lens
a concave lens in which light rays that enter it parallel to its axis bend away (diverge) from its axis

focal point
for a converging lens or mirror, the point at which converging light rays cross; for a diverging lens or mirror, the point from which diverging light rays appear to originate

focal length
distance from the center of a lens or curved mirror to its focal point

magnification

ratio of image height to object height

power

inverse of focal length

real image

image that can be projected

virtual image

image that cannot be projected

This page titled 9.8.1: Image Formation by Lenses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.9: Image Formation by Lenses** by OpenStax is licensed CC BY 4.0.

9.8.2: Image Formation by Mirrors

Learning Objectives

- Illustrate image formation in a flat mirror.
- Explain with ray diagrams the formation of an image using spherical mirrors.
- Determine focal length and magnification given radius of curvature, distance of object and image.

We only have to look as far as the nearest bathroom to find an example of an image formed by a mirror. Images in flat mirrors are the same size as the object and are located behind the mirror. Like lenses, mirrors can form a variety of images. For example, dental mirrors may produce a magnified image, just as makeup mirrors do. Security mirrors in shops, on the other hand, form images that are smaller than the object. We will use the law of reflection to understand how mirrors form images, and we will find that mirror images are analogous to those formed by lenses.

Figure **9.8.2.1** helps illustrate how a flat mirror forms an image. Two rays are shown emerging from the same point, striking the mirror, and being reflected into the observer's eye. The rays can diverge slightly, and both still get into the eye. If the rays are extrapolated backward, they seem to originate from a common point behind the mirror, locating the image. (The paths of the reflected rays into the eye are the same as if they had come directly from that point behind the mirror.) Using the law of reflection —the angle of reflection equals the angle of incidence—we can see that the image and object are the same distance from the mirror. This is a virtual image, since it cannot be projected—the rays only appear to originate from a common point behind the mirror. Obviously, if you walk behind the mirror, you cannot see the image, since the rays do not go there. But in front of the mirror, the rays behave exactly as if they had come from behind the mirror, so that is where the image is situated.

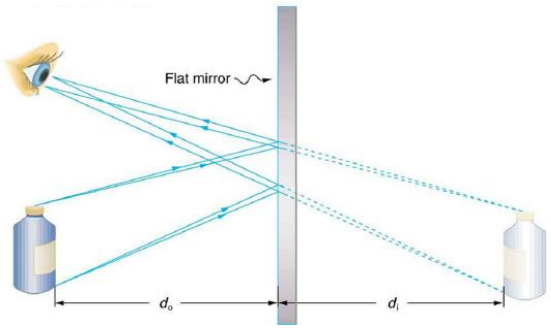


Figure **9.8.2.1**: Two sets of rays from common points on an object are reflected by a flat mirror into the eye of an observer. The reflected rays seem to originate from behind the mirror, locating the virtual image.

Now let us consider the focal length of a mirror—for example, the concave spherical mirrors in Figure **9.8.2.2**. Rays of light that strike the surface follow the law of reflection. For a mirror that is large compared with its radius of curvature, as in Figure **9.8.2.1** (a), we see that the reflected rays do not cross at the same point, and the mirror does not have a well-defined focal point. If the mirror had the shape of a parabola, the rays would all cross at a single point, and the mirror would have a well-defined focal point. But parabolic mirrors are much more expensive to make than spherical mirrors. The solution is to use a mirror that is small compared with its radius of curvature, as shown in Figure **9.8.2.2**(b). (This is the mirror equivalent of the thin lens approximation.) To a very good approximation, this mirror has a well-defined focal point at F that is the focal distance f from the center of the mirror. The focal length f of a concave mirror is positive, since it is a converging mirror.

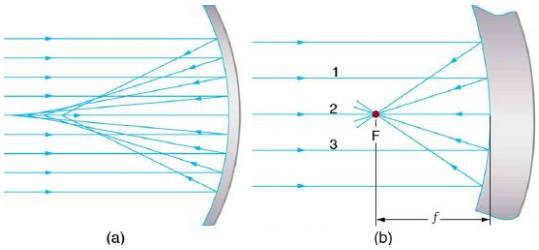


Figure 9.8.2.2: (a) Parallel rays reflected from a large spherical mirror do not all cross at a common point. (b) If a spherical mirror is small compared with its radius of curvature, parallel rays are focused to a common point. The distance of the focal point from the center of the mirror is its focal length f. Since this mirror is converging, it has a positive focal length.

Just as for lenses, the shorter the focal length, the more powerful the mirror; thus, $P = 1/f$ for a mirror, too. A more strongly curved mirror has a shorter focal length and a greater power. Using the law of reflection and some simple trigonometry, it can be shown that the focal length is half the radius of curvature, or

$$f = \frac{R}{2},$$

where R is the radius of curvature of a spherical mirror. The smaller the radius of curvature, the smaller the focal length and, thus, the more powerful the mirror.

The convex mirror shown in Figure 9.8.2.3 also has a focal point. Parallel rays of light reflected from the mirror seem to originate from the point F at the focal distance f behind the mirror. The focal length and power of a convex mirror are negative, since it is a diverging mirror.

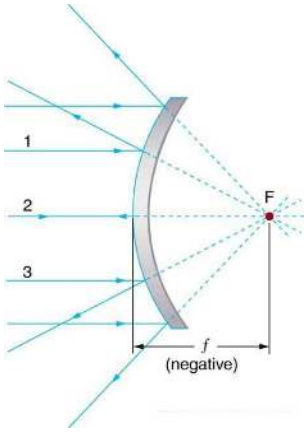


Figure 9.8.2.3: Parallel rays of light reflected from a convex spherical mirror (small in size compared with its radius of curvature) seem to originate from a well-defined focal point at the focal distance f behind the mirror. Convex mirrors diverge light rays and, thus, have a negative focal length.

Ray tracing is as useful for mirrors as for lenses. The rules for ray tracing for mirrors are based on the illustrations just discussed:

1. **Principal Ray 1**: a ray approaching a concave converging mirror parallel to its axis is reflected through the focal point F of the mirror on the same side. (See rays 1 and 3 in Figure 9.8.2.2(b).); a ray approaching a convex diverging mirror parallel to its axis is reflected so that it seems to come from the focal point F behind the mirror. (See rays 1 and 3 in Figure 9.8.2.3.)
2. **Principal Ray 2**: any ray striking the center of a mirror is followed by applying the law of reflection; it makes the same angle with the axis when leaving as when approaching. (See ray 2 in Figure 9.8.2.4.)
3. **Principal Ray 3**: a ray approaching a concave converging mirror through its focal point is reflected parallel to its axis. (The reverse of rays 1 and 3 in Figure 9.8.2.2.); a ray approaching a convex diverging mirror by heading toward its focal point on the opposite side is reflected parallel to the axis. (The reverse of rays 1 and 3 in Figure 9.8.2.3.) *The third principal ray is optional and may be used to verify the accuracy of image location.*

We will use ray tracing to illustrate how images are formed by mirrors, and we can use ray tracing quantitatively to obtain numerical information. But since we assume each mirror is small compared with its radius of curvature, we can use the thin lens equations for mirrors just as we did for lenses.

Consider the situation shown in Figure 9.8.2.4, concave spherical mirror reflection, in which an object is placed farther from a concave (converging) mirror than its focal length. That is, f is positive and $d_o > f$, so that we may expect an image similar to the case 1 real image formed by a converging lens. Ray tracing in Figure 9.8.2.4 shows that the rays from a common point on the object all cross at a point on the same side of the mirror as the object. Thus, a real image can be projected onto a screen placed at this location. The image distance is positive, and the image is inverted, so its magnification is negative. This is a *case 1 image for mirrors*. It differs from the case 1 image for lenses only in that the image is on the same side of the mirror as the object. It is otherwise identical.

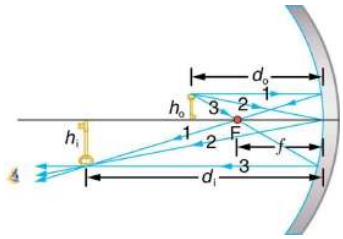


Figure 9.8.2.4: A case 1 image for a mirror. An object is farther from the converging mirror than its focal length. Rays from a common point on the object are traced using the rules in the text. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 goes through the focal point on the way toward the mirror. All three rays cross at the same point after being reflected, locating the inverted real image. Although three rays are shown, only two of the three are needed to locate the image and determine its height.



Figure 9.8.2.5: Parabolic trough collectors are used to generate electricity in southern California. An array of such pipes in the California desert can provide a thermal output of 250 MW on a sunny day, with fluids reaching temperatures as high as 400°C. (credit: kjkolb, Wikimedia Commons)

What happens if an object is closer to a concave mirror than its focal length? This is analogous to a case 2 image for lenses ($d_o < f$ and f positive), which is a magnifier. In fact, this is how makeup mirrors act as magnifiers. Figure 9.8.2.6(a) uses ray tracing to locate the image of an object placed close to a concave mirror. Rays from a common point on the object are reflected in such a manner that they appear to be coming from behind the mirror, meaning that the image is virtual and cannot be projected. As with a magnifying glass, the image is upright and larger than the object. This is a *case 2 image for mirrors* and is exactly analogous to that for lenses.

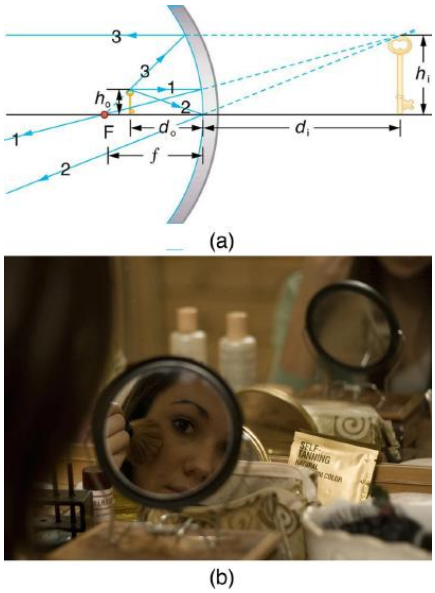


Figure **9.8.2.6**: (a) Case 2 images for mirrors are formed when a converging mirror has an object closer to it than its focal length. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 approaches the mirror as if it came from the focal point. (b) A magnifying mirror showing the reflection. (credit: Mike Melrose, Flickr)

All three rays appear to originate from the same point after being reflected, locating the upright virtual image behind the mirror and showing it to be larger than the object. (b) Makeup mirrors are perhaps the most common use of a concave mirror to produce a larger, upright image.

A convex mirror is a diverging mirror (***f*** is negative) and forms only one type of image. It is a *case 3* image—one that is upright and smaller than the object, just as for diverging lenses. Figure **9.8.2.7**(a) uses ray tracing to illustrate the location and size of the case 3 image for mirrors. Since the image is behind the mirror, it cannot be projected and is thus a virtual image. It is also seen to be smaller than the object.

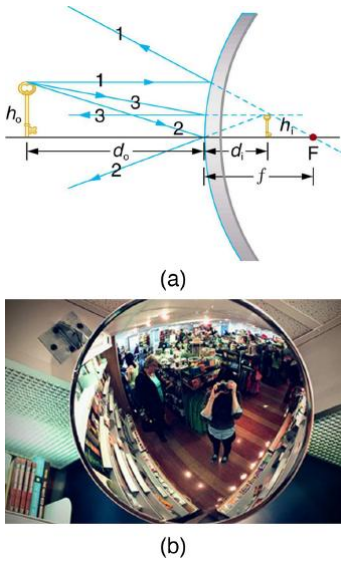


Figure **9.8.2.7**: Case 3 images for mirrors are formed by any convex mirror. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 approaches toward the focal point. All three rays appear to originate from the same point after being reflected, locating the upright virtual image behind the mirror and showing it to be smaller than the object. (b) Security mirrors are convex, producing a smaller, upright image. Because the image is smaller, a larger area is imaged compared to what would be observed for a flat mirror (and hence security is improved). (credit: Laura D'Alessandro, Flickr)

TAKE-HOME EXPERIMENT: CONCAVE MIRRORS CLOSE TO HOME

Find a flashlight and identify the curved mirror used in it. Find another flashlight and shine the first flashlight onto the second one, which is turned off. Estimate the focal length of the mirror. You might try shining a flashlight on the curved mirror behind the headlight of a car, keeping the headlight switched off, and determine its focal length.

Section Summary

- The characteristics of an image formed by a flat mirror are: (a) The image and object are the same distance from the mirror, (b) The image is a virtual image, and (c) The image is situated behind the mirror.
- Image length is half the radius of curvature.

$$f = \frac{R}{2}$$

- A convex mirror is a diverging mirror and forms only one type of image, namely a virtual image.
- The focal length and power of a convex mirror are negative.

Glossary

converging mirror

a concave mirror in which light rays that strike it parallel to its axis converge at one or more points along the axis

diverging mirror

a convex mirror in which light rays that strike it parallel to its axis bend away (diverge) from its axis

law of reflection

angle of reflection equals the angle of incidence

This page titled 9.8.2: Image Formation by Mirrors is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.10: Image Formation by Mirrors** by OpenStax is licensed CC BY 4.0.

9.8.3: Polarization

Learning Objectives

- Discuss the meaning of polarization.
- Discuss the property of optical activity of certain materials.

Polaroid sunglasses are familiar to most of us. They have a special ability to cut the glare of light reflected from water or glass (see Figure 9.8.3.1). Polaroids have this ability because of a wave characteristic of light called polarization. What is polarization? How is it produced? What are some of its uses? The answers to these questions are related to the wave character of light.



Figure 9.8.3.1: These two photographs of a river show the effect of a polarizing filter in reducing glare in light reflected from the surface of water. Part (b) of this figure was taken with a polarizing filter and part (a) was not. As a result, the reflection of clouds and sky observed in part (a) is not observed in part (b). Polarizing sunglasses are particularly useful on snow and water. (credit: Amithshs, Wikimedia Commons)

Light is one type of electromagnetic (EM) wave. As noted earlier, EM waves are *transverse waves* consisting of varying electric and magnetic fields that oscillate perpendicular to the direction of propagation (see Figure 9.8.3.2). There are specific directions for the oscillations of the electric and magnetic fields. **Polarization** is the attribute that a wave's oscillations have a definite direction relative to the direction of propagation of the wave. (This is not the same type of polarization as that discussed for the separation of charges.) Waves having such a direction are said to be **polarized**. For an EM wave, we define the **direction of polarization** to be the direction parallel to the electric field. Thus, we can think of the electric field arrows as showing the direction of polarization, as in Figure 9.8.3.2.

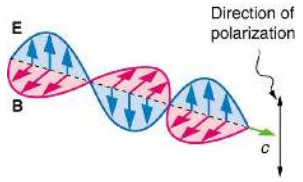


Figure 9.8.3.2: An EM wave, such as light, is a transverse wave. The electric and magnetic fields are perpendicular to the direction of propagation.

To examine this further, consider the transverse waves in the ropes shown in Figure 9.8.3.3. The oscillations in one rope are in a vertical plane and are said to be **vertically polarized**. Those in the other rope are in a horizontal plane and are **horizontally polarized**. If a vertical slit is placed on the first rope, the waves pass through. However, a vertical slit blocks the horizontally polarized waves. For EM waves, the direction of the electric field is analogous to the disturbances on the ropes.

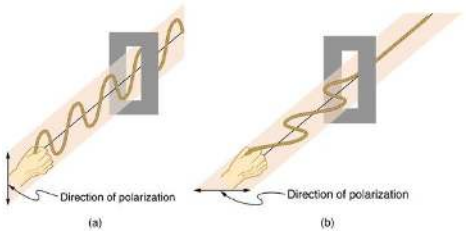


Figure 9.8.3.3: The transverse oscillations in one rope are in a vertical plane, and those in the other rope are in a horizontal plane. The first is said to be vertically polarized, and the other is said to be horizontally polarized. Vertical slits pass vertically polarized waves and block horizontally polarized waves.

The Sun and many other light sources produce waves that are randomly polarized (see Figure 9.8.3.4). Such light is said to be **unpolarized** because it is composed of many waves with all possible directions of polarization. Polaroid materials, invented by the founder of Polaroid Corporation, Edwin Land, act as a *polarizing* slit for light, allowing only polarization in one direction to pass through. Polarizing filters are composed of long molecules aligned in one direction. Thinking of the molecules as many slits, analogous to those for the oscillating ropes, we can understand why only light with a specific polarization can get through. The **axis of a polarizing filter** is the direction along which the filter passes the electric field of an EM wave (see Figure 9.8.3.5).

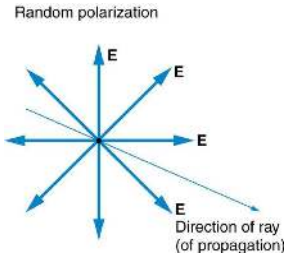


Figure 9.8.3.4: The slender arrow represents a ray of unpolarized light. The bold arrows represent the direction of polarization of the individual waves composing the ray. Since the light is unpolarized, the arrows point in all directions.

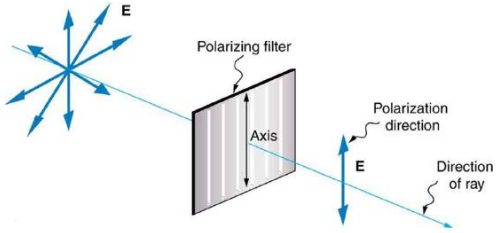


Figure 9.8.3.5: A polarizing filter has a polarization axis that acts as a slit passing through electric fields parallel to its direction. The direction of polarization of an EM wave is defined to be the direction of its electric field.

Figure 9.8.3.6 shows the effect of two polarizing filters on originally unpolarized light. The first filter polarizes the light along its axis. When the axes of the first and second filters are aligned (parallel), then all of the polarized light passed by the first filter is also passed by the second. If the second polarizing filter is rotated, only the component of the light parallel to the second filter's axis is passed. When the axes are perpendicular, no light is passed by the second.

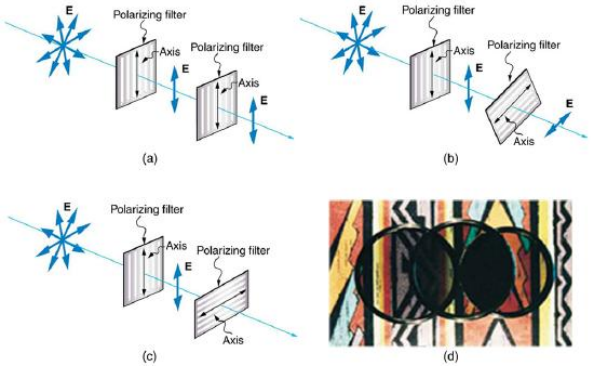


Figure 9.8.3.6: The effect of rotating two polarizing filters, where the first polarizes the light. (a) All of the polarized light is passed by the second polarizing filter, because its axis is parallel to the first. (b) As the second is rotated, only part of the light is passed. (c) When the second is perpendicular to the first, no light is passed. (d) In this photograph, a polarizing filter is placed above two others. Its axis is perpendicular to the filter on the right (dark area) and parallel to the filter on the left (lighter area). (credit: P.P. Urone)

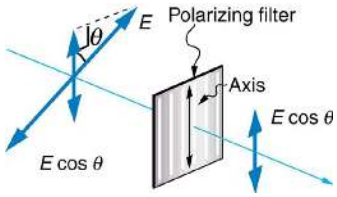


Figure 9.8.3.7: A polarizing filter transmits only the component of the wave parallel to its axis, $E\cos\theta$, reducing the intensity of any light not polarized parallel to its axis.

Polarization by Reflection

By now you can probably guess that Polaroid sunglasses cut the glare in reflected light because that light is polarized. You can check this for yourself by holding Polaroid sunglasses in front of you and rotating them while looking at light reflected from water or glass. As you rotate the sunglasses, you will notice the light gets bright and dim, but not completely black. This implies the reflected light is partially polarized and cannot be completely blocked by a polarizing filter.

Figure 9.8.3.8 illustrates what happens when unpolarized light is reflected from a surface. Vertically polarized light is preferentially refracted at the surface, so that *the reflected light is left more horizontally polarized*. The reasons for this phenomenon are beyond the scope of this text, but a convenient mnemonic for remembering this is to imagine the polarization direction to be like an arrow. Vertical polarization would be like an arrow perpendicular to the surface and would be more likely to stick and not be reflected. Horizontal polarization is like an arrow bouncing on its side and would be more likely to be reflected. Sunglasses with vertical axes would then block more reflected light than unpolarized light from other sources.

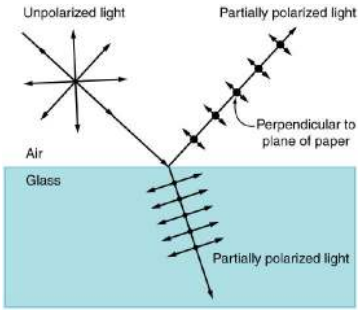


Figure 9.8.3.8: Polarization by reflection. Unpolarized light has equal amounts of vertical and horizontal polarization. After interaction with a surface, the vertical components are preferentially absorbed or refracted, leaving the reflected light more horizontally polarized. This is akin to arrows striking on their sides bouncing off, whereas arrows striking on their tips go into the surface.

THINGS GREAT AND SMALL: ATOMIC EXPLANATION OF POLARIZING FILTERS

Polarizing filters have a polarization axis that acts as a slit. This slit passes electromagnetic waves (often visible light) that have an electric field parallel to the axis. This is accomplished with long molecules aligned perpendicular to the axis as shown in Figure 9.8.3.9.

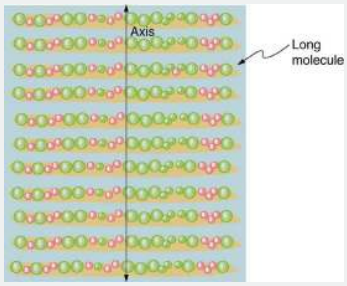


Figure 9.8.3.9: Long molecules are aligned perpendicular to the axis of a polarizing filter. The component of the electric field in an EM wave perpendicular to these molecules passes through the filter, while the component parallel to the molecules is absorbed.

Figure 9.8.3.10 illustrates how the component of the electric field parallel to the long molecules is absorbed. An electromagnetic wave is composed of oscillating electric and magnetic fields. The electric field is strong compared with the magnetic field and is more effective in exerting force on charges in the molecules. The most affected charged particles are the electrons in the molecules, since electron masses are small. If the electron is forced to oscillate, it can absorb energy from the EM wave. This reduces the fields in the wave and, hence, reduces its intensity. In long molecules, electrons can more easily oscillate parallel to the molecule than in the perpendicular direction. The electrons are bound to the molecule and are more restricted in their movement perpendicular to the molecule. Thus, the electrons can absorb EM waves that have a component of their electric field parallel to the molecule. The electrons are much less responsive to electric fields perpendicular to the molecule and will allow those fields to pass. Thus, the axis of the polarizing filter is perpendicular to the length of the molecule.

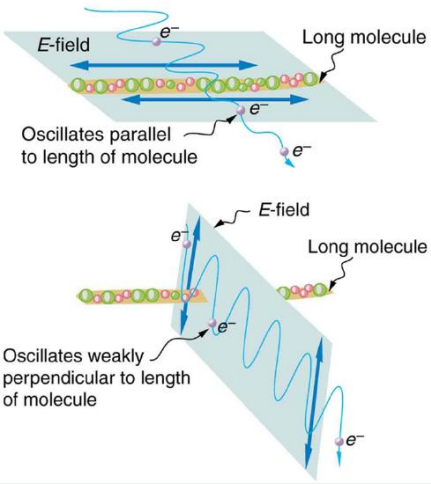


Figure 9.8.3.10: Artist's conception of an electron in a long molecule oscillating parallel to the molecule. The oscillation of the electron absorbs energy and reduces the intensity of the component of the EM wave that is parallel to the molecule.

Polarization by Scattering

If you hold your Polaroid sunglasses in front of you and rotate them while looking at blue sky, you will see the sky get bright and dim. This is a clear indication that light scattered by air is partially polarized. Figure 9.8.3.11 helps illustrate how this happens. Since light is a transverse EM wave, it vibrates the electrons of air molecules perpendicular to the direction it is traveling. The electrons then radiate like small antennae. Since they are oscillating perpendicular to the direction of the light ray, they produce EM radiation that is polarized perpendicular to the direction of the ray. When viewing the light along a line perpendicular to the original ray, as in Figure 9.8.3.11, there can be no polarization in the scattered light parallel to the original ray, because that would require the original ray to be a longitudinal wave. Along other directions, a component of the other polarization can be projected along the line of sight, and the scattered light will only be partially polarized. Furthermore, multiple scattering can bring light to your eyes from other directions and can contain different polarizations.

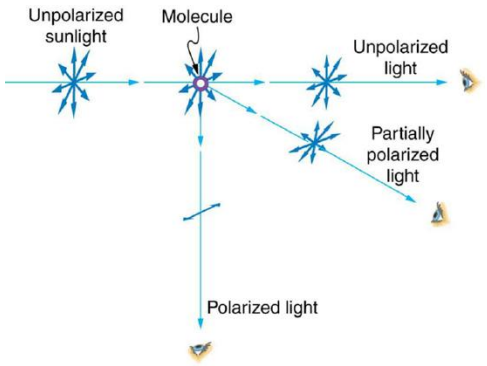


Figure **9.8.3.11**: Polarization by scattering. Unpolarized light scattering from air molecules shakes their electrons perpendicular to the direction of the original ray. The scattered light therefore has a polarization perpendicular to the original direction and none parallel to the original direction.

Photographs of the sky can be darkened by polarizing filters, a trick used by many photographers to make clouds brighter by contrast. Scattering from other particles, such as smoke or dust, can also polarize light. Detecting polarization in scattered EM waves can be a useful analytical tool in determining the scattering source.

There is a range of optical effects used in sunglasses. Besides being Polaroid, other sunglasses have colored pigments embedded in them, while others use non-reflective or even reflective coatings. A recent development is photochromic lenses, which darken in the sunlight and become clear indoors. Photochromic lenses are embedded with organic microcrystalline molecules that change their properties when exposed to UV in sunlight, but become clear in artificial lighting with no UV.

TAKE-HOME EXPERIMENT: POLARIZATION

Find Polaroid sunglasses and rotate one while holding the other still and look at different surfaces and objects. Explain your observations. What is the difference in angle from when you see a maximum intensity to when you see a minimum intensity? Find a reflective glass surface and do the same. At what angle does the glass need to be oriented to give minimum glare?

Liquid Crystals and Other Polarization Effects in Materials

While you are undoubtedly aware of liquid crystal displays (LCDs) found in watches, calculators, computer screens, cellphones, flat screen televisions, and other myriad places, you may not be aware that they are based on polarization. Liquid crystals are so named because their molecules can be aligned even though they are in a liquid. Liquid crystals have the property that they can rotate the polarization of light passing through them by **90°**. Furthermore, this property can be turned off by the application of a voltage, as illustrated in Figure **9.8.3.12**. It is possible to manipulate this characteristic quickly and in small well-defined regions to create the contrast patterns we see in so many LCD devices.

In flat screen LCD televisions, there is a large light at the back of the TV. The light travels to the front screen through millions of tiny units called pixels (picture elements). One of these is shown in Figure **9.8.3.1** (a) and (b). Each unit has three cells, with red, blue, or green filters, each controlled independently. When the voltage across a liquid crystal is switched off, the liquid crystal passes the light through the particular filter. One can vary the picture contrast by varying the strength of the voltage applied to the liquid crystal.

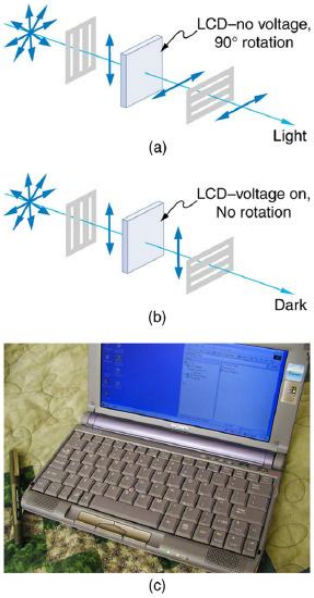


Figure 9.8.3.12: (a) Polarized light is rotated 90° by a liquid crystal and then passed by a polarizing filter that has its axis perpendicular to the original polarization direction. (b) When a voltage is applied to the liquid crystal, the polarized light is not rotated and is blocked by the filter, making the region dark in comparison with its surroundings. (c) LCDs can be made color specific, small, and fast enough to use in laptop computers and TVs. (credit: Jon Sullivan)

Many crystals and solutions rotate the plane of polarization of light passing through them. Such substances are said to be **optically active**. Examples include sugar water, insulin, and collagen (see Figure 9.8.3.13). In addition to depending on the type of substance, the amount and direction of rotation depends on a number of factors. Among these is the concentration of the substance, the distance the light travels through it, and the wavelength of light. Optical activity is due to the asymmetric shape of molecules in the substance, such as being helical. Measurements of the rotation of polarized light passing through substances can thus be used to measure concentrations, a standard technique for sugars. It can also give information on the shapes of molecules, such as proteins, and factors that affect their shapes, such as temperature and pH.

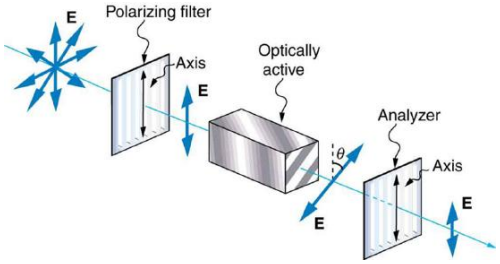


Figure 9.8.3.13: Optical activity is the ability of some substances to rotate the plane of polarization of light passing through them. The rotation is detected with a polarizing filter or analyzer.

Glass and plastic become optically active when stressed; the greater the stress, the greater the effect. Optical stress analysis on complicated shapes can be performed by making plastic models of them and observing them through crossed filters, as seen in Figure 9.8.3.14. It is apparent that the effect depends on wavelength as well as stress. The wavelength dependence is sometimes also used for artistic purposes.

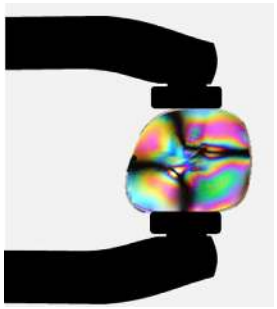


Figure **9.8.3.14**: Optical stress analysis of a plastic lens placed between crossed polarizers. (credit: Infopro, Wikimedia Commons)

Section Summary

- Polarization is the attribute that wave oscillations have a definite direction relative to the direction of propagation of the wave.
- EM waves are transverse waves that may be polarized.
- The direction of polarization is defined to be the direction parallel to the electric field of the EM wave.
- Unpolarized light is composed of many rays having random polarization directions.
- Light can be polarized by passing it through a polarizing filter or other polarizing material. The intensity of polarized light after passing through a polarizing filter depends on the direction of polarization and the axis of the filter.
- Polarization is also produced by reflection.
- Polarization can also be produced by scattering.
- There are a number of types of optically active substances that rotate the direction of polarization of light passing through them.

Glossary

axis of a polarizing filter
the direction along which the filter passes the electric field of an EM wave

direction of polarization
the direction parallel to the electric field for EM waves

horizontally polarized
the oscillations are in a horizontal plane

optically active
substances that rotate the plane of polarization of light passing through them

polarization
the attribute that wave oscillations have a definite direction relative to the direction of propagation of the wave

polarized
waves having the electric and magnetic field oscillations in a definite direction

unpolarized
waves that are randomly polarized

vertically polarized
the oscillations are in a vertical plane

This page titled 9.8.3: Polarization is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **11.11: Polarization** by OpenStax is licensed CC BY 4.0.

9.9: The Greenhouse Effect and Climate Change

Earth's Temperature is a Balancing Act

Earth's temperature depends on the balance between energy entering and leaving the planet. When incoming energy from the Sun is absorbed, Earth warms. When the Sun's energy is reflected back into space, Earth avoids warming. When energy is released from Earth into space, the planet cools. Many factors, both natural and human, can cause changes in Earth's energy balance, including:

- Changes in the greenhouse effect, which affects the amount of heat retained by Earth's atmosphere;
- Variations in the Sun's energy reaching Earth;
- Changes in the reflectivity of Earth's atmosphere and surface.

Scientists have pieced together a picture of Earth's climate, dating back hundreds of thousands of years, by analyzing a number of indirect measures of climate such as ice cores, tree rings, glacier size, pollen counts, and ocean sediments. Scientists have also studied changes in Earth's orbit around the sun and the activity of the sun itself.

The historical record shows that the climate varies naturally over a wide range of time scales. In general, climate changes prior to the Industrial Revolution in the 1700s can be explained by natural causes, such as changes in solar energy, volcanic eruptions, and natural changes in greenhouse gas (GHG) concentrations. (The details of these are beyond the scope of this text, but see the link in the attributions if you would like to know more about them.) Recent changes in climate**,** however, cannot be explained by natural causes alone. Research indicates that natural causes are very unlikely to explain most observed warming, especially warming since the mid-20th century. Rather, human activities, especially our combustion of fossil fuels, explains the current warming (figure **9.9.1**). The scientific consensus is clear: through alterations of the carbon cycle, humans are changing the global climate by increasing the effects of something known as the greenhouse effect.

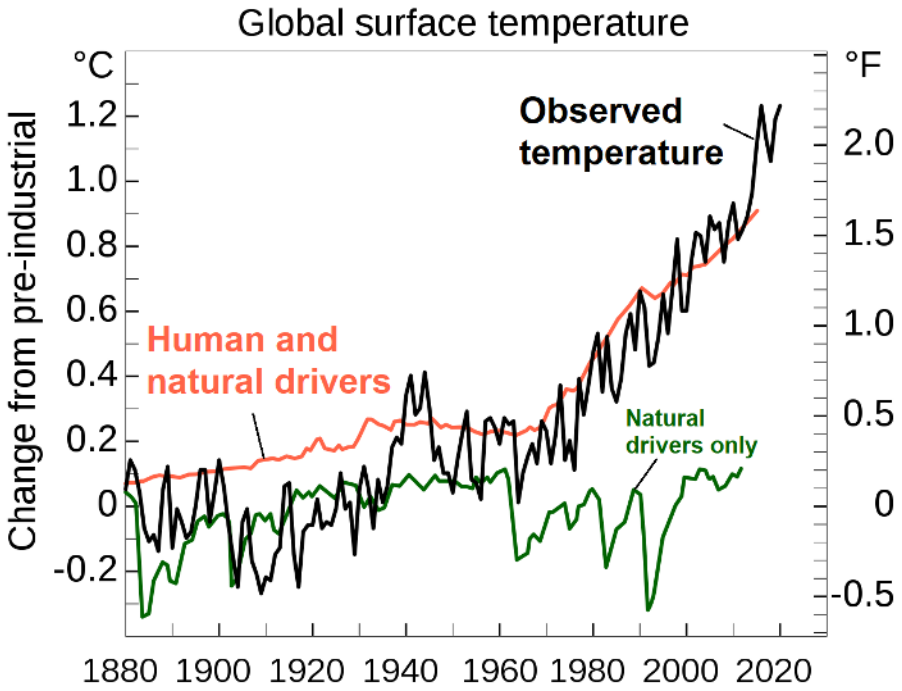


Figure **9.9.1**: This graph shows the predicted temperatures from two climate models and observed temperatures from 1880 to 2020. The first model considered only natural factors that could influence temperature and is represented by the green line (bottom). It shows some fluctuations in temperature by no overall increase or decrease. The second model considered both human and natural factors and is represented by the orange line (top). It shows an overall increase in temperature. Actual observations (black, jagged line; middle) more closely match the second model. Overall, temperature has increased about 1.2 degrees Celsius (2.1 degrees Fahrenheit) since pre-industrial times. Image relabeled from Efbrazil(opens in new window) (CC-BY-SA(opens in new window)).

The Greenhouse Effect Causes the Atmosphere to Retain Heat

Gardeners that live in moderate or cool environments use greenhouses because they trap heat and create an environment that is warmer than outside temperatures. This is great for plants that like heat, or are sensitive to cold temperatures, such as tomato and pepper plants. Greenhouses contain glass or plastic that is transparent to visible light from the Sun. This light is absorbed by plants, soil, and surfaces and heats them. Some of that heat energy is then radiated outwards in the form of infrared radiation. But the glass of the greenhouse is opaque to the infrared radiation, thereby trapping the heat energy, causing the temperature within the greenhouse to increase.

The same phenomenon happens inside a car on a sunny day. Have you ever noticed how much hotter a car can get compared to the outside temperature? Light energy from the Sun passes through the windows and is absorbed by the surfaces in the car such as seats and the dashboard. Those warm surfaces then radiate infrared radiation, which cannot pass through the glass. This trapped infrared energy causes the air temperatures in the car to increase. This process is commonly known as the **greenhouse effect**.

The video below made for kids, but provides a clear and simple introduction to the greenhouse effect.



The greenhouse effect also happens with the entire Earth. Of course, our planet is not surrounded by glass windows. Instead, the Earth is wrapped with an atmosphere that contains **greenhouse gases** (GHGs). Much like the glass in a greenhouse, GHGs are transparent to visible light from the Sun, but they are opaque to infrared radiation that is radiated from the Earth towards space (figure 9.9.2). In this way, they help trap heat energy that subsequently raises air temperature. Being a greenhouse gas is a physical property of certain types of gases; because of their molecular structure they absorb wavelengths of infrared radiation, but are transparent to visible light. Some notable greenhouse gases are water vapor (H_2O), carbon dioxide (CO_2), and methane (CH_4). GHGs act like a blanket, making Earth significantly warmer than it would otherwise be. Scientists estimate that average temperature on Earth would be -18° C without naturally-occurring GHGs.

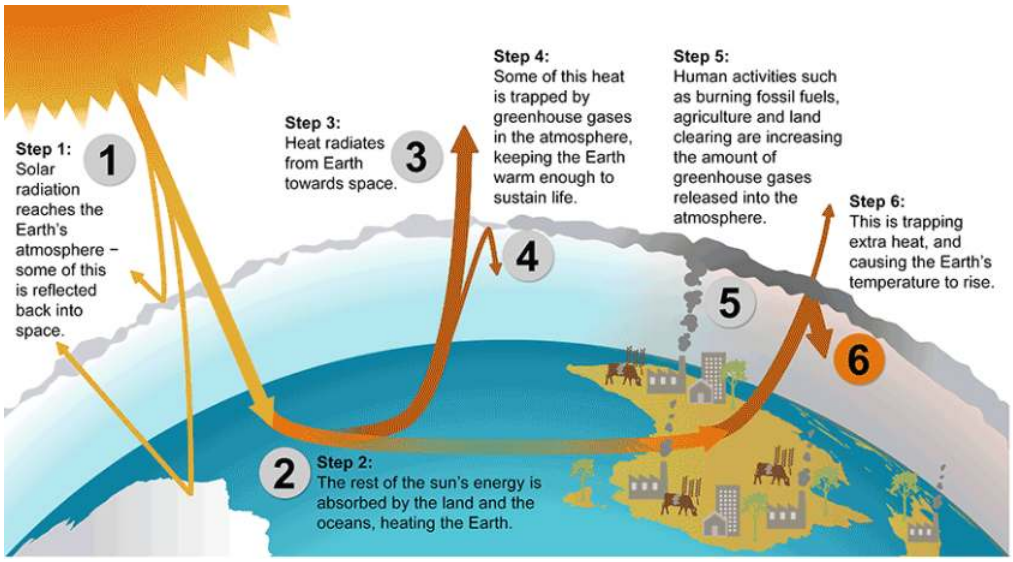


Figure **9.9.2**: The enhanced greenhouse effect. Step 1: Some solar radiation reaches the Earth's surface, and some is reflected back into space. Step 2: The rest of the sun's energy is absorbed by the land and the oceans, heating the Earth. Step 3: Heat radiates from Earth towards space. Step 4: Some of the heat is trapped by greenhouse gases in the atmosphere, warming the Earth. Step 5: Human activities such as burning fossil fuels, agriculture, and land clearing have increased the concentrations of greenhouse gases in the atmosphere. Step 6: This is trapping extra heat, causing the Earth's temperature to rise.

What is Global Warming?

Global warming refers to the recent and ongoing rise in global average temperature near Earth's surface. It is caused mostly by increasing concentrations of greenhouse gases in the atmosphere. Global warming is causing climate patterns to change. However, global warming itself represents only one aspect of climate change. **Climate change** refers to any significant change in the measures of climate lasting for an extended period of time. In other words, climate change includes major changes in temperature, precipitation, or wind patterns, among other effects, that occur over several decades or longer.

The Main Greenhouse Gasses

The most important GHGs directly emitted by humans include CO_2 and methane. **Carbon dioxide** (CO_2) is the primary greenhouse gas that is contributing to recent global climate change. CO_2 is a natural component of the carbon cycle, involved in such activities as photosynthesis, respiration, volcanic eruptions, and ocean-atmosphere exchange. Human activities, primarily the burning of fossil fuels and changes in land use, release very large amounts of CO_2 to the atmosphere, causing its concentration in the atmosphere to rise.

Atmospheric CO_2 concentrations have increased by 45% since pre-industrial times, from approximately 280 parts per million (ppm) in the 18th century to 409.8 ppm in 2019 (figure **9.9.*c***). The current CO_2 level is higher than it has been in at least 800,000 years, based on evidence from ice cores that preserve ancient atmospheric gases (figure **9.9.*d* – *f***). Human activities currently release over 30 billion tons of CO_2 into the atmosphere every year. While some volcanic eruptions released large quantities of CO_2 in the distant past, the U.S. Geological Survey (USGS) reports that human activities now emit more than 135 times as much CO_2 as volcanoes each year. This human-caused build-up of CO_2 in the atmosphere is like a tub filling with water, where more water flows from the faucet than the drain can take away.

Water vapor is the most abundant greenhouse gas and also the most important in terms of its contribution to the natural greenhouse effect, despite having a short atmospheric lifetime. Some human activities can influence local water vapor levels. However, on a global scale, the concentration of water vapor is controlled by temperature, which influences overall rates of evaporation and precipitation. Therefore, the global concentration of water vapor is not substantially affected by direct human emissions.

Methane is another GHG which has increased due to human activity, but its concentrations are much lower than CO_2.

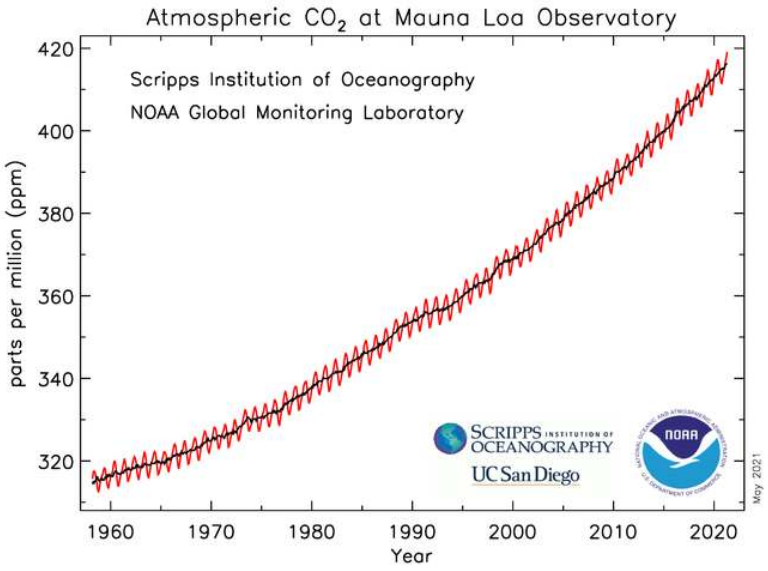


Figure **9.9.3**: Monthly mean carbon dioxide concentration in parts per million (ppm) measured at Mauna Loa Observatory, Hawaii. The carbon dioxide data on Mauna Loa constitute the longest record of direct measurements of CO_2 in the atmosphere. The fluctuating red line represents the monthly mean values, centered on the middle of each month. Carbon dioxide concentrations dip every summer due to increased photosynthesis. The smoother black line represents the same, after correction for the average seasonal cycle. Image and caption (modified) by NOAA(opens in new window) (public domain).

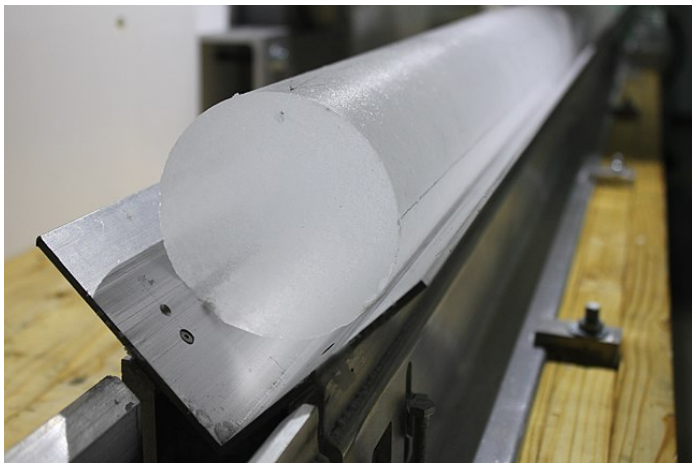


Figure **9.9.*d***: The EastGRIP ice core, which was drilled from the Greenland Ice Sheet. You can read more here. Image by Helle Astrid Kjær (CC-BY).

$$n\ CO_2 + n\ H_2O + \text{Energy} \longrightarrow C_nH_{2n}O_n + n\ O_2$$

Figure **9.9.4**: This graph, based on the comparison of atmospheric samples contained in ice cores and more recent direct measurements, provides evidence that atmospheric CO2 has increased since the Industrial Revolution. On the x-axis are the years before today (0 = 1950). It begins 400,000 years prior to 1950. On the y-axis is the carbon dioxide level in parts per million. Carbon dioxide levels have fluctuated over the years, but they never exceeded 300 parts per million until 1950. In 2018, carbon dioxide levels reached 409.8 ppm. (This graph is a few years old and shows the current CO_2 level at 400 ppm; Credit: Vostok ice core data/J.R. Petit et al.; NOAA Mauna Loa CO2 record.)

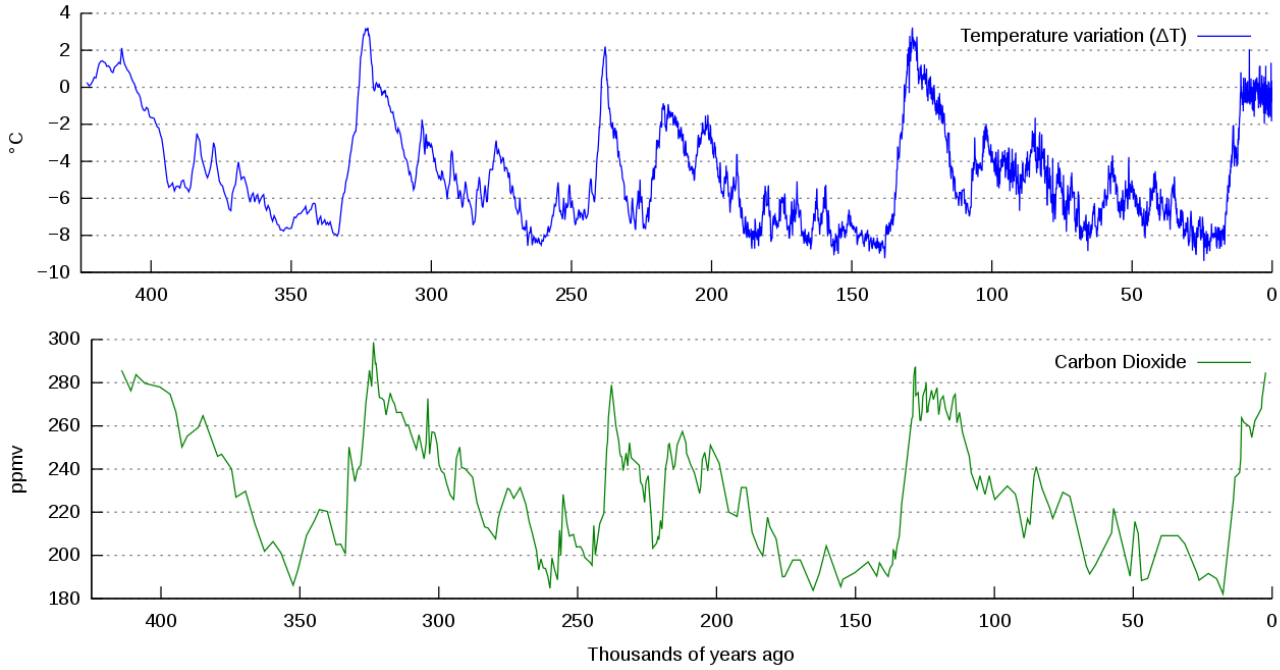


Figure 9.9.5: Graph of change in temperature in degrees Celsius (top, blue line) and carbon dioxide concentration in parts per million by volume (bottom, green line) measured from the Vostok, Antarctica ice core. These have been associated for over 400,000 years. As carbon dioxide concentration increased, so did temperature. As carbon dioxide concentration decreased, so did temperature. These data were collected in 1999. Since then, carbon dioxide concentrations have increased to 409.8 ppm (2019 average). Image and caption (modified) by NOAA/Autopilot(opens in new window) (CC-BY-SA(opens in new window))

Scientific Consensus: Global Climate Change is Real

The **Intergovernmental Panel on Climate Change** (IPCC) was created in 1988 by the United Nations Environment Programme and the World Meteorological Organization. It is charged with the task of evaluating and synthesizing the scientific evidence surrounding global climate change. The IPCC uses this information to evaluate current impacts and future risks, in addition to providing policymakers with assessments. These assessments are released about once every every six years. The most recent report, the 6th Assessment, was released in 2021. Hundreds of leading scientists from around the world are chosen to author these reports. Over the history of the IPCC, these scientists have reviewed thousands of peer-reviewed, publicly available studies. The scientific consensus is clear: global climate change is real and humans are very likely the cause for this change.

Additionally, the major scientific agencies of the United States, including the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA), also agree that climate change is occurring and that humans are driving it. In 2010, the US National Research Council concluded that “Climate change is occurring, is very likely caused by human activities, and poses significant risks for a broad range of human and natural systems”. Many independent scientific organizations have released similar statements, both in the United States and abroad. This does not necessarily mean that every scientist sees eye to eye on each component of the climate change problem, but broad agreement exists that climate change is happening and is primarily caused by excess greenhouse gases from human activities. Critics of climate change, driven by ideology instead of evidence, try to suggest to the public that there is no scientific consensus on global climate change. Such an assertion is patently false.

Current Status of Global Climate Change and Future Changes

Greenhouse gas concentrations in the atmosphere will continue to increase unless the billions of tons of anthropogenic emissions each year decrease substantially. Increased concentrations are expected to do the following:

- Increase Earth’s average temperature (figure 9.9.6).
- Influence the patterns and amounts of precipitation.
- Reduce ice and snow cover, as well as permafrost.
- Raise sea level.

- Increase the acidity of the oceans.

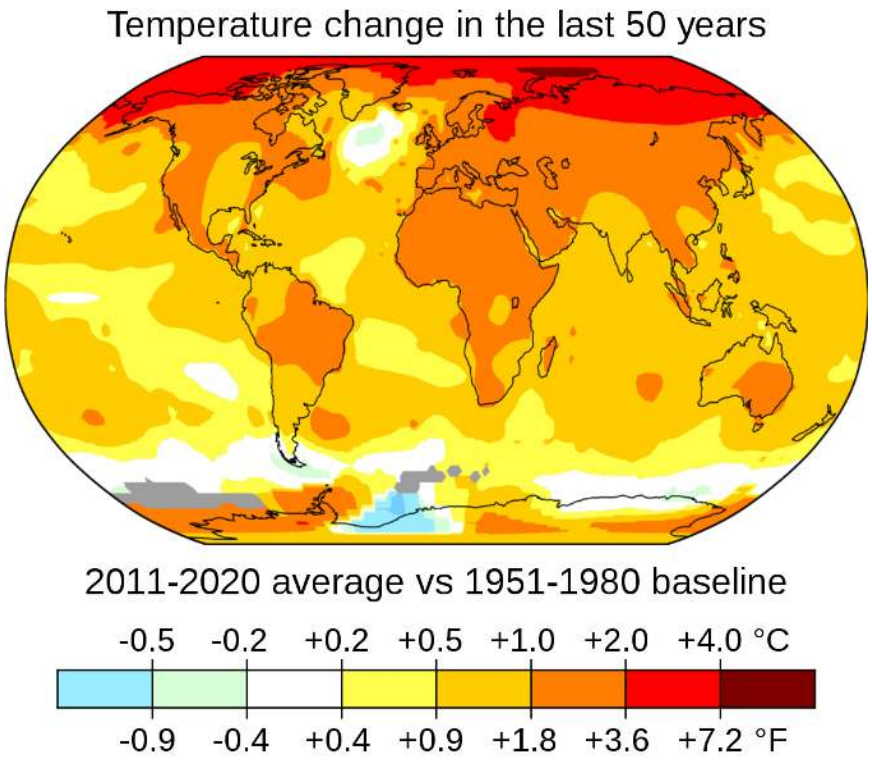


Figure **9.9.6**: Temperature increases have been most pronounced in northern latitudes and over land masses. The colors represent the temperature difference between the 2011-2020 average and the 1951-1980 baseline, with warmer colors (yellow, orange, red) representing increases, and cool colors (green, blue) representing decreases. The image uses longer term averages of at least a decade to smooth out climate variability due to factors such as El Niño. Grey areas in the image have insufficient data for rendering. Image and caption (modified) from NASA's Scientific Visualization Studio/Eric Fisk(opens in new window) (public domain).

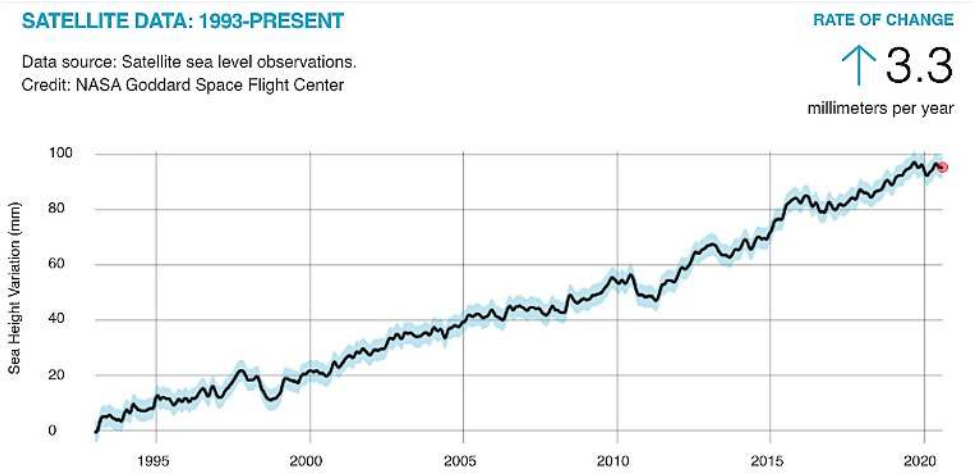


Figure 9.9.7: Sea height variation (mm) over time. Sea height has increased about 3.3 millimeters per year on average since 1993. Data is from satellite sea level observations by the NASA Goddard Space Flight Center. Image by NASA(opens in new window) (public domain).

These changes will impact our food supply, water resources, infrastructure, ecosystems, and even our own health. The magnitude and rate of future climate change will primarily depend on the following factors:

- The rate at which levels of greenhouse gas concentrations in our atmosphere continue to increase.

- How strongly features of the climate (e.g., temperature, precipitation, and sea level) respond to the expected increase in greenhouse gas concentrations.
- Natural influences on climate (e.g., from volcanic activity and changes in the sun's intensity) and natural processes within the climate system (e.g., changes in ocean circulation patterns).

These impacts will continue well into the future (please see the attribution link at the end of this page if you would like to read about the impacts and time frame in more detail.)

Climate Change Affects Everyone

Our lives are connected to the climate**.** Human societies have adapted to the relatively stable climate we have enjoyed since the last ice age which ended several thousand years ago. A warming climate will bring changes that can affect our water supplies, agriculture, power and transportation systems, the natural environment, and even our own health and safety.

Carbon dioxide can stay in the atmosphere for nearly a century, on average, so Earth will continue to warm in the coming decades. The warmer it gets, the greater the risk for more severe changes to the climate and Earth's system. Although it's difficult to predict the exact impacts of climate change, what is clear is that the climate we are accustomed to is no longer a reliable guide for what to expect in the future.

We can reduce the risks we will face from climate change**.** By making choices that reduce greenhouse gas pollution, and preparing for the changes that are already underway, we can reduce risks from climate change. Our decisions today will shape the world our children and grandchildren will live in.

You can take steps at home, on the road, and in your office to reduce greenhouse gas emissions and the risks associated with climate change. Many of these steps can save you money. Some, such as walking or biking to work, can even improve your health! You can also get involved on a local or state level to support energy efficiency, clean energy programs, or other climate programs.

Suggested Supplementary Reading

Intergovernmental Panel on Climate Change. 2021. 6th Assessment: Summary for Policymakers.

NASA. 2018. Global Climate Change: Vital Signs of the Planet. This website by NASA provides a multi-media smorgasbord of engaging content. Learn about climate change using data collected by NASA satellites and more.

Attributions

Remixed by Jamie MacArthur from the following source which was modified by Melissa Ha from the following sources:

- Climate and the Effects of Global Climate Change from *General Biology* by OpenStax (licensed under CC-BY(opens in new window))
- Climate Change from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY(opens in new window))
- Carbon Cycle from *Biology* by John W. Kimball (licensed under CC-BY(opens in new window))

This page titled 9.9: The Greenhouse Effect and Climate Change is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

This page titled 9.9: The Greenhouse Effect and Climate Change is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **21.1: The Greenhouse Effect and Climate Change** by Melissa Ha and Rachel Schleiger is licensed CC BY-SA 4.0.

CHAPTER OVERVIEW

10: Elements Are Made of Atoms

10.1: Introduction to Atoms and Elements
10.2: Modern Understandings of Elements
10.3: The Development of Atomic Theory
10.3.1: Law of Conservation of Mass
10.3.2: Law of Definite Proportions
10.3.3: Law of Multiple Proportions
10.3.4: Dalton's Atomic Theory
10.4: Organizing the Elements
10.4.1: Mendeleev's Periodic Table
10.4.2: Periodic Law
10.4.3: Modern Periodic Table- Periods and Groups
10.5: Understanding the Periodic Table
10.5.1: Metals, Nonmetals, and Metalloids.
10.5.2: Families of the Periodic Table
10.6: Inside the Atom
10.7: Electrons and Light
10.8: Quantum Mechanics
10.9: Electron Shells

This page titled 10: Elements Are Made of Atoms is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

10.1: Introduction to Atoms and Elements

In this chapter we will be exploring the concepts of atoms and elements. These two words make most people think of chemistry. They are related, but are not exactly the same thing. As we read through this chapter we will see the nuances of these two words. To introduce these ideas, we will first consider how they were thought about before people began to investigate them scientifically.



Figure **10.1.1** (Credit: Justin Norris (Flickr:JustinMN); Source: flickr)(opens in new window); License: CC by 2.0(opens in new window))

The Philosophers' Views on Atoms

People enjoy getting together to discuss things, whether it is how your favorite sports team is doing, what the best new movie is, the current politics, or any number of other topics. Often the question is raised about who is right and who is wrong. If the football game is to be played this coming weekend, all we can do is offer opinions as to its outcome. The game has not been played yet, so we do not know who will actually win.

The ancient Greek philosophers did a lot of discussing, with part of their conversations concerning the physical world and its composition. There were different opinions about what made up matter. Some felt one thing was true while others believed another set of ideas. Since these scholars did not have laboratories and had not developed the idea of the experiment, they were left to debate. Whoever could offer the best argument was considered right. However, often the best argument had little to do with reality.

One of the on-going debates had to do with sand. The question posed was: into how small of pieces can you divide a grain of sand? The prevailing thought at the time, pushed by Aristotle, was that the grain of sand could be divided indefinitely, that you could always get a smaller particle by dividing a larger one and there was no limit to how small the resulting particle could be.



Figure **10.1.2**: Into how small of pieces can you divide a grain of sand? (Credit: User:(WT-shared) Shoestring/WTS.Wikivoyage; Source: wikimedia; License: Public Domain)

Since Aristotle was such an influential philosopher, very few people disagreed with him. However, there were some philosophers who believed that there was a limit to how small a grain of sand could be divided. One of these philosophers was Democritus (~460-370 B.C.E.), often referred to as the "laughing philosopher" because of his emphasis on cheerfulness. He taught that there were substances called **atoms** and that these atoms made up all material things. The atoms were unchangeable, indestructible, and always existed.

The word "**atom**" comes from the Greek *ατομοσ* and means "indivisible". The atomists of the time (Democritus being one of the leading atomists) believed there were two realities that made up the physical world: atoms and void. There were an infinite number of atoms, but different types of atoms had different sizes and shapes. The void was the empty space in which the atoms moved and

collided with one another. When these atoms collided with one another, they might repel each other or they might connect in clusters, held together by tiny hooks and barbs on the surfaces of the atoms.

Aristotle disagreed with Democritus and offered his own idea of the composition of matter. According to Aristotle, everything was composed of four elements: earth, air, fire, and water. Democritus' theory better explained things, but Aristotle was more influential, so his ideas prevailed. It took almost two thousand years before scientists came around to seeing the atom as Democritus did.

It is very interesting that Democritus had the basic idea of atoms, even though he had no experimental evidence to support his thinking. We now know more about how atoms hold together in "clusters" (compounds), but the basic concept existed over two thousand years ago. We also know that atoms can be further subdivided, but there is still a lower limit to how small we can break up that grain of sand.

The Philosophers' Views on Elements

Greek philosophers proposed many ideas for what things were made of as well. Some proposed 5 elements, with distinct shapes: **octahedra**, **tetrahedra**, **cubes**, etc. For a long while, the four element model (earth, air, fire, water) was popular. This model, which Plato and Aristotle also used, suggested that all matter was composed of these four elements in different ratios. For instance, when wood is burned, you get smoke (air), ash (earth), pitch (a viscous liquid, here identified with water), and fire, so wood is made of all these things. Other models suggested everything was made of just water, or just air.

Despite the vast array of models for the elements that philosophers came up with, none of them got close to the current understanding of elements. It would take the work of science, not mere chance in a game of conjectures, to find that.

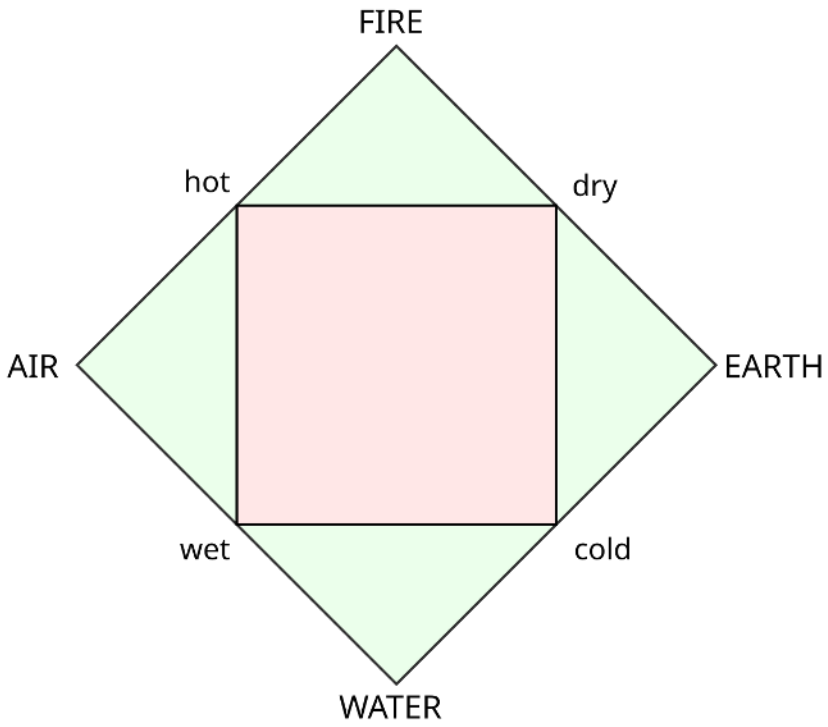


Figure 10.1.3: representation of an ancient idea of the elements. By en:User:Heron - Transwikied from en:. Originally uploaded by en:User:Heron. Recompressed with OptiPNG by Michael. Converted to SVG by tiZom, Public Domain, wikimedia

Contributors and Attributions

- Emily V Eames(opens in new window) (City College of San Francisco) wrote language adopted for the discussion of the ancient view of elements in this text.

This page titled 10.1: Introduction to Atoms and Elements is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **4.1: Democritus' Idea of the Atom** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **6.1: Early History of the Periodic Table** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.2: Modern Understandings of Elements

Learning Objectives

- Know the modern definition of an element.

The Modern Definition of an Element

Early chemical experimenters toyed with the many ideas of the ancients as possibilities for what was an element. Chemistry was a bit later than other sciences to experience a true scientific revolution, because much of the practice of chemical experimentation occurred in the field of alchemy, where there was much secrecy and linking of experimental practices to metaphysical interpretations.

It was Antoine Lavoisier (1743-1794) who finally came up with the modern concept of an element as something which can not be broken down into anything simpler by a chemical reaction. Lavoisier also proposed the Law of Conservation of Mass (which we shall discuss shortly). He also proposed many other important ideas in chemistry during his short time of experimentation. His work in defining elements and discovering this first law of chemistry paved the way for the work of Dalton and the others. He certainly had the most impact of any chemist of the 18th century, and in fact some consider him to be the first true chemist.

The idea that an element was something that could not be broken down any further was not necessarily new. What was new in the work of Lavoisier was that these substances were identified empirically instead of philosophically. Lavoisier proposed about 30 substances as elements which could not be broken down any further. Later experiments showed that some of these could be broken down further and others were not forms of matter at all. However, it was a huge step to list both lead and gold as elements which could not be broken down further when transmuting lead into gold had long been the goal of alchemists. The reliance on experimentation instead of reasoning was again a key to understanding the natural world.

Watch this video to learn more about the work of Lavoisier and some of the other earlier chemists:



Glossary

element

something which can't be broken down into anything simpler by a chemical reaction.

This page titled 10.2: Modern Understandings of Elements is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **6.1: Early History of the Periodic Table** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

10.3: The Development of Atomic Theory

The Greek and Roman philosophers debated, discussed, and sometimes even attacked one another. But the mode of discovery was talk. There was no experimentation—the idea had not yet been thought of. So science did not develop very far, and there was no reliable way to establish what was true and what was false.

John Dalton

While it must be assumed that many more scientists, philosophers, and others studied the composition of matter after Democritus, a major leap forward in our understanding of the composition of matter took place in the 1800s with the work of the British scientist John Dalton. He started teaching school at age twelve, and was primarily known as a teacher. In his twenties, he moved to the growing city of Manchester, where he was able to pursue some scientific studies. He did work in several areas of science including meteorology and the discovery of a gas law. He is most well known for atomic theory, which we will be discussing within this section. When he died, over 40,000 people in Manchester marched at his funeral.



Figure **10.3.2**: John Dalton. (Credit: Henry Roscoe (author), William Henry Worthington (engraver), and Joseph Allen (painter); Source:Wikimedia; License: Public Domain)

In the subsections that follow, we will explore various laws of nature which together along with Dalton's own observation allowed him to make a scientific theory for the existence of atoms that was, in fact, based on observations of the world around him.

This page titled 10.3: The Development of Atomic Theory is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **4.6: Dalton's Atomic Theory** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.3.1: Law of Conservation of Mass

Learning Objectives

- Recognize the law of conservation of mass.



Figure **10.3.1.1** (Credit: Douglas Heriot; Source:flickr; License: CC by 2.0(opens in new window))

Have you ever lost a screw?

The following situation happens all too often. You have taken apart a piece of equipment to clean it up. When you put the equipment back together, somehow you have an extra screw or two. Or you find out that a screw is missing that was a part of the original equipment. In either case, you know something is wrong. You expect to end up with the same amount of material that you started with, not with more or less than what you had originally.

Law of Conservation of Mass

By the late 1700s, chemists accepted the definition of an element as a substance that cannot be broken down into a simpler substance by ordinary chemical means. It was also clear that elements combine with one another to form more complex substances called compounds. The chemical and physical properties of these compounds are different than the properties of the elements from which they were formed. There were questions about the details of these processes.

In the 1790s, a greater emphasis began to be placed on the quantitative analysis of chemical reactions. Accurate and reproducible measurements of the masses of reacting elements and the compounds they form led to the formulation of several basic **laws**. One of these is called the law of **conservation of mass**, which states that during a chemical reaction, the total mass of the **products** must be equal to the total mass of the **reactants**. In other words, mass cannot be created or destroyed during a chemical reaction, but is always conserved. This is another of the rare conservation laws we are being introduced to.

LAW OF CONSERVATION OF MASS

During a chemical reaction, the total mass of all products must be equal to the total mass of all reactants.

As an example, consider the reaction between silver nitrate and sodium chloride. These two compounds will dissolve in water to form silver chloride and sodium nitrate. The silver chloride does not dissolve in water, so it forms a solid that we can filter off. When we evaporate the water, we can recover the sodium nitrate formed. If we react 58.5 grams of sodium chloride with 169.9 grams of silver nitrate, we start with 228.4 grams of materials. After the reaction is complete and the materials separated, we find that we have formed 143.4 grams of silver chloride and 85.0 grams of sodium nitrate, giving us a total mass of 228.4 grams for the products. So, the total mass of reactants equals the total mass of products, a proof of the law of conservation of mass.



Section Summary

- The law of conservation of mass states that, during a chemical reaction, the total mass of the products must be equal to the total mass of the reactants.

Glossary

reactants
: substances which react during a chemical reaction.

products
: substances which are formed during a chemical reaction.

law of conservation of mass
: mass cannot be created or destroyed during a chemical reaction, but is always conserved.

This page titled 10.3.1: Law of Conservation of Mass is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

10.3.2: Law of Definite Proportions

Learning Objectives

- Recognize the Law of Definite Proportions

The discovery that mass was always conserved in chemical reactions was soon followed by the **law of definite proportions**, which states that a given chemical compound always contains the same elements in the exact same proportions by mass. As an example, any sample of pure water contains 11.19% hydrogen and 88.81% oxygen by mass. It does not matter where the sample of water came from or how it was prepared. Its composition, like that of every other compound, is fixed.

LAW OF DEFINITE PROPORTIONS

Any sample of a chemical compound will always contain the same elements in the exact same proportion by mass.



Figure **10.3.2.1**: Water. (Credit: Jenn Durfey; Source: Flickr Photos, Water Faucet(opens in new window) [www.flickr.com]; License: CC by 2.0(opens in new window))

Another example is carbon dioxide. This gas is produced from a variety of reactions, often by the burning of materials. The structure of the gas consists of one atom of carbon and two atoms of oxygen. Carbon dioxide production is of interest in many areas, from the amount we breathe to the amount of the gas produced by burning wood or fossil fuels. By knowing the exact composition of carbon dioxide, we can make predictions as to the effects of different chemical processes.



Figure **10.3.2.2**: Carbon dioxide is produced during the burning process. (Credit: Courtesy of Patsy Lynch, FEMA; Source: Commons Wikimedia, FEMA Photograph by Patsy Lynch taken on 04-17-2006 in Missouri(opens in new window) [commons.wikimedia.org]; License: Public Domain)

Section Summary

- The law of definite proportions states that a given chemical compound always contains the same elements in the exact same proportions by mass.

This page titled 10.3.2: Law of Definite Proportions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **4.4: Law of Definite Proportions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.3.3: Law of Multiple Proportions

Learning Objectives

- Recognize the Law of Multiple Proportions

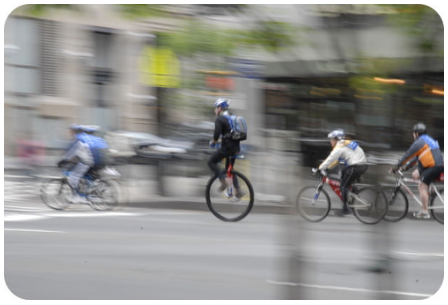


Figure **10.3.3.1** (Credit: Flickr: Sutibu; Source: Commons WIkimedia, Five Boro Bike Tour Unicycle(opens in new window) [commons.wikimedia.org]; License: CC by 2.0(opens in new window))

What are the similarities and differences between a unicycle and a bicycle?

Just from the words themselves, the astute Latin-speaking scholar can tell that, whatever it is made of, the unicycle has one of them (*uni* = "one") and the bicycle has two (*bi* = "two"). From pictures, we get additional information that helps us tell the two apart. The unicycle has one wheel and the bicycle has two. In particular, they are made up of the same materials, and the only significant difference is the number of wheels on the two vehicles. Now—how many wheels are on a tricycle?

Law of Multiple Proportions

Once the idea that elements combined in definite proportions to form compounds was established, experiments also began to demonstrate that the same pairs of certain elements could combine to form more than one compound. Consider the elements carbon and oxygen. Combined in one way, they form the familiar compound carbon dioxide. In every sample of carbon dioxide, there are $32.0\ \text{g}$ of oxygen present for every $12.0\ \text{g}$ of carbon. By dividing 32.0 by 12.0, this simplifies to a mass ratio of oxygen to carbon of 2.66 to 1. There is another compound that forms from the combination of carbon and oxygen called carbon monoxide. Every sample of carbon monoxide contains $16.0\ \text{g}$ of oxygen for every $12.0\ \text{g}$ of carbon. This is a mass ratio of oxygen to carbon of 1.33 to 1. In the carbon dioxide, there is exactly twice as much oxygen present as there is in the carbon monoxide. This example illustrates the **law of multiple proportions**: whenever the same two elements form more than one compound, the different masses of one element that combine with the same mass of the other element are in the ratio of small whole numbers.

LAW OF MULTIPLE PROPORTIONS

Whenever the same two elements form more than one compound, the different masses of one element that combine with the same mass of the other element are in the ratio of small whole numbers.

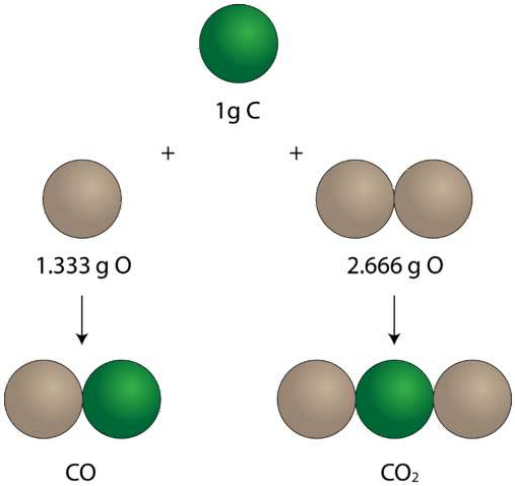


Figure **10.3.3.2**: Carbon can form two different compounds with oxygen. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))

In carbon monoxide, on the left, there is **1.333 g** of oxygen for every **1 g** of carbon. In carbon dioxide, on the right, there is **2.666 g** of oxygen for every gram of carbon. So the ratio of oxygen in the two compounds is 1:2, a small whole number ratio.

The difference between carbon monoxide and carbon dioxide is significant. Carbon monoxide is a deadly gas, formed from the incomplete combustion of some carbon-containing materials (such as wood and gasoline). This compound will attach to hemoglobin in the red blood cells and block the binding of oxygen to those cells. If oxygen does not bind, it cannot be carried to the cells of the body where it is needed, and death can occur. Carbon dioxide, on the other hand, is not toxic like carbon monoxide is. However, it can displace oxygen in systems since it is heavier. Carbon dioxide fire extinguishers cut off the flow of oxygen in a fire, putting out the fire.

Section Summary

- The law of multiple proportions states that whenever the same two elements form more than one compound, the different masses of one element that combine with the same mass of the other element are in the ratio of small whole numbers.

This page titled 10.3.3: Law of Multiple Proportions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **4.3: Law of Multiple Proportions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.3.4: Dalton's Atomic Theory

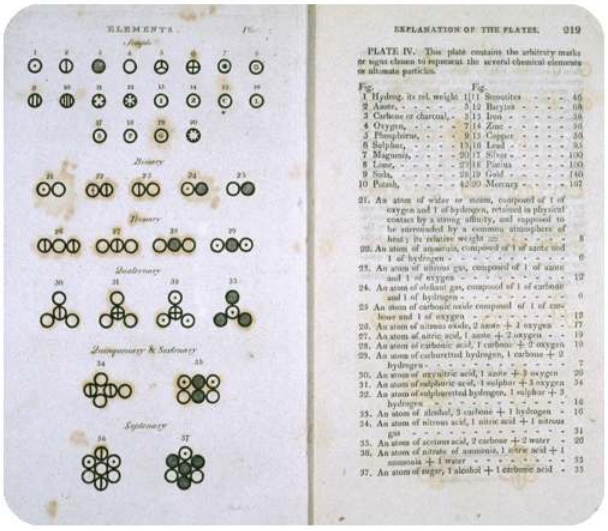
Learning Objectives

- Recognize the 5 tenants of Dalton's Atomic Theory.

Though we have illustrated our current knowledge of the behavior of atoms in our discussion of the laws of definite proportions and multiple proportions, this information was not available to Dalton when he formulated atomic theory. In fact, it was the information in the laws themselves which naturally led to the conclusion that atoms must exist. Using only the three laws we have introduced in this section, Dalton devised the **atomic theory** as an explanation for the observations which led to these laws. It was a scientific way of suggesting that atoms existed, as opposed to the non-scientific debates that the Greek philosophers engaged in. The general tenets of this theory are as follows:

- All matter is composed of extremely small particles called atoms.
- Atoms of a given element are identical in size, mass, and other properties. Atoms of different elements differ in size, mass, and other properties.
- Atoms cannot be subdivided, created, or destroyed.
- Atoms of different elements can combine in simple whole number ratios to form chemical compounds.
- In chemical reactions, atoms are combined, separated, or rearranged.

Dalton's atomic theory has been largely accepted by the scientific community, with the exception of three changes. We now know that (1) an atom can be further subdivided, (2) all atoms of an element are not identical in mass, and (3) using nuclear fission and fusion techniques, we can create or destroy atoms by changing them into other atoms. We will explore each of these ideas later in this text. However, for any chemical reaction the ideas held within the atomic theory would correctly predict the results of that reaction.


ELEMENTS. Pl.

EXPLANATION OF THE PLATES. 919

PLATE IV. This plate contains the arbitrary marks or signs chosen to represent the several chemical elements or ultimate particles.

Figure **10.3.4.3**: Dalton's symbols. (Credit: John Dalton; Source: wikimedia; License: Public Domain)



Summary

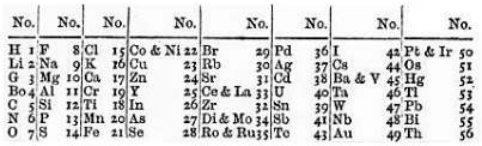
- The general tenets of Dalton's atomic theory are:
  - All matter is composed of extremely small particles called atoms
  - Atoms cannot be subdivided, created, or destroyed.
  - Atoms of a given element are identical in size, mass, and other properties. Atoms of different elements differ in size, mass, and other properties.
  - Atoms of different elements can combine in whole number ratios to form chemical compounds;
  - In chemical reaction, atoms are rearranged, combined, or separated.

This page titled 10.3.4: Dalton's Atomic Theory is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **4.6: Dalton's Atomic Theory** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.4: Organizing the Elements

As more elements were being discovered, many people began attempting to organize them in a variety of ways: Triads, Pentads, and Octaves were among these ideas. Many of these ideas were not taken very seriously. It wasn't until someone was able to find an organizational scheme that was able to make a prediction that the organization of the elements was taken seriously across the disciple. In these subsections we will discuss the discovery of this organization scheme, its implications, and introduce the current organization of the elements within the periodic table.



No.	No.	No.	No.	No.	No.	No.	No.
H 1	F 8	Cl 15	Co & Ni 22	Br 29	Pd 36	I 42	Pt & Ir 50
Li 2	Na 9	K 16	Cu 23	Rb 30	Ag 37	Cs 44	Os 51
G 3	Mg 10	Ca 17	Zn 24	Sr 31	Cd 38	Ba & V 45	Hg 52
Bo 4	Al 11	Cr 19	Y 25	Ce & La 33	U 40	Ta 46	Tl 53
C 5	Si 12	Ti 18	In 26	Zr 32	Sn 39	W 47	Pb 54
N 6	P 13	Mn 20	As 27	Di & Mo 34	Sb 41	Nb 48	Bi 55
O 7	S 14	Fe 21	Se 28	Ro & Ru 35	Te 43	Au 49	Th 56

Figure **10.4.2** (Credit: John Alexander Reina Newlands; Source: Commons Wikimedia, Newlands Periodiska System 1866(opens in new window) [commons.wikimedia.org]; License: Public Domain)



This page titled 10.4: Organizing the Elements is shared under a CK-12 license and was authored, remixed, and/or curated by CK-12 Foundation via source content that was edited to the style and standards of the LibreTexts platform.

- **6.1: Early History of the Periodic Table** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.4.1: Mendeleev's Periodic Table

Learning Objectives

- Know the patterns that emerged in the discovery of the periodic table.



Figure **10.4.1.1** (Credit: Laura Guerin; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

When you study for a test, how do you approach the task?

One useful way to study for a test is to use flashcards. To make them, write down vocabulary words, foreign language terms, math formulas, chemistry reactions—anything you want to learn. Then, sort these cards into categories and topics that go together. This organization of information helps you see patterns in the material so that you can tie different ideas together and make better sense of them.

The periodic table was first built using a set of cards. With this strategy, Dmitri Mendeleev, the creator of the periodic table, could organize and rearrange material until patterns emerged. And he did this while trying to find an easy way to help his students learn the material.

Mendeleev's Periodic Table

In 1869, Russian chemist and teacher Dmitri Mendeleev (1836-1907) published a periodic table of the elements. The following year, German chemist Lothar Meyer independently published a very similar table. Mendeleev is generally given more credit than Meyer because his table was published first, and because of several key insights that he made regarding the table.



Figure **10.4.1.2**: Dmitri Mendeleev. (Credit: Unknown; Source: wikimedia; License: Public Domain)

Mendeleev was writing a chemistry textbook for his students and wanted to organize all of the known elements at that time according to their chemical properties. He famously organized the information for each element onto separate note cards that were then easy to rearrange as needed. He discovered that when he placed them in order of increasing atomic mass, certain similarities in chemical behavior repeated at regular intervals. This type of a repeating pattern is called "periodic". A pendulum that swings back and forth in a given time interval is periodic, as is the movement of the moon around the Earth.

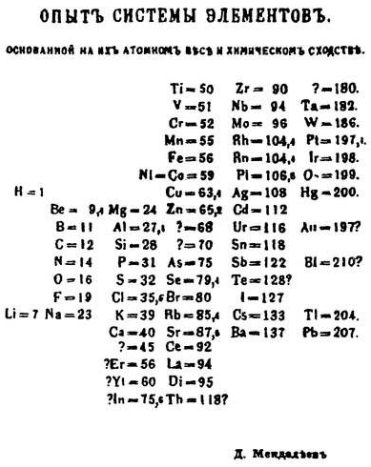


Figure **10.4.1.3**: Mendeleev's 1869 periodic table. (Credit: Dmitri Mendeleev; Source: wikimedia; License: Public Domain)

In the figure above, atomic mass increases from top to bottom of vertical columns, with successive columns going left to right. As a result, elements that are in the same horizontal row are groups of elements that were known to exhibit similar chemical properties. One of Mendeleev's insights is illustrated by the elements tellurium (**Te**) and iodine (**I**). Notice that tellurium is listed before iodine even though its atomic mass is higher. Mendeleev reversed the order because he knew that the properties of iodine were much more similar to those of fluorine (**F**), chlorine (**Cl**), and bromine (**Br**) than they were to oxygen (**O**), sulfur (**S**), and selenium (**Se**). He simply assumed that there was an error in the determination of one or both of the atomic masses. This turned out not to be the case, but Mendeleev was indeed correct to group these two elements as he did.

Notice that there are several places in the table that have no chemical symbol, but are instead labeled with a question mark. Between zinc (***Zn***) and arsenic (**As**) are two such missing elements. Mendeleev believed that elements with atomic masses of 68 and 70 would eventually be discovered and that they would fit chemically into each of those spaces. Listed in the table below are other properties that Mendeleev predicted for the first of these two missing elements, which he called "eka-aluminum", compared with the element gallium.

Mendeleev's predicted properties for Eka-aluminum and gallium.

	Eka-Aluminum (**Ea**)	**Gallium** (**Ga**)
Atomic Mass	**68 amu**	**69.9 amu**
Melting Point	Low	**30.15°C**
Density	**5.9 g/cm^3**	**5.94 g/cm^3**
Formula of Oxide	**Ea_2O_3**	**Ga_2O_3**

The element gallium was discovered four years after the publication of Mendeleev's table, and its properties matched up remarkably well with eka-aluminum, fitting into the table exactly where he had predicted. This was also the case with the element that followed gallium, which was eventually named germanium.

Mendeleev's periodic table gained wide acceptance with the scientific community and earned him credit as the discoverer of the periodic law. Element number 101, synthesized in 1955, is named **mendelevium** after the founder of the periodic table. It was, however, several years after Mendeleev died before the several discrepancies with the atomic masses could be explained, and before the reasons behind the repetition of chemical properties could be fully explained.

The following video contains some of the story of Mendeleev's background and career.



Section Summary

- Mendeleev published his periodic table in 1869.
- His organization of elements was based on atomic mass.
- Mendeleev's periodic table made it possible to predict properties of elements that had not yet been discovered.

This page titled 10.4.1: Mendeleev's Periodic Table is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **6.2: Mendeleev's Periodic Table** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.4.2: Periodic Law

Learning Objectives

- Apply the periodic law when referencing the periodic table of elements.



Figure **10.4.2.1**: Copy and Paste Caption here. (Credit: Christopher Fynn; Source: Flickr, Mysore - Fruit vendor(opens in new window) [www.flickr.com]; License: CC-BY 2.0)

How are these items related to one another?

We have all enjoyed looking around a market for delicious foods to eat later at home. When you get to the market you know you need to get fruits, vegetables and grains to ensure you eat a balanced diet. In the market, these items are all grouped together to make it easier to find the type of fruit, vegetable or grain you are looking for; it would be inconvenient if the different fruits were all scattered in different places. The periodic table is organized in a similar way, ensuring similar elements are found in the same group or period.

The Periodic Law

When Mendeleev put his periodic table together, nobody knew about the existence of the atomic nucleus. Later experiments showed that the periodic relationship shown in the periodic table was not based on the atomic mass, but on the atomic number. (The atomic number is a property of the atomic nucleus which will be discussed in greater detail in the next chapter.) As it turns out, the atomic mass increases as the atomic number increases. For most of the periodic table the increases are on a similar scale, but there were some exceptions which were the cause of the discrepancies in Mendeleev's original periodic table.

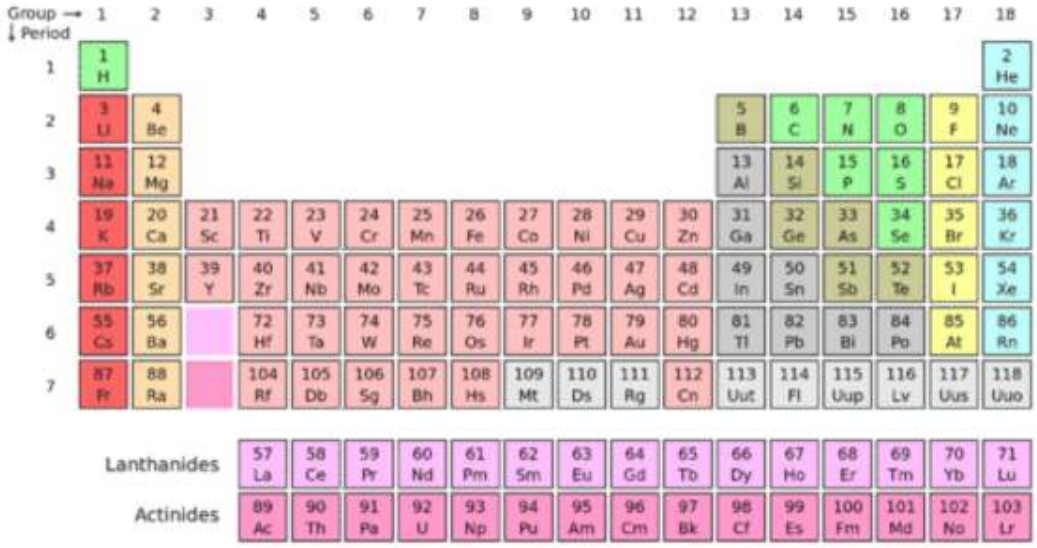


Figure **10.4.2.2** (Credit: User:Cepheus/Wikimedia Commons; Source: Commons Wikimedia, Periodic Table(opens in new window) [commons.wikimedia.org]; License: Public Domain)

When ordered by atomic number, the discrepancies within Mendeleev's table disappeared. Tellurium has an atomic number of 52, while iodine has an atomic number of 53. So even though tellurium does indeed have a greater atomic mass than iodine, it is properly placed before iodine in the periodic table. Mendeleev and Moseley are credited with being most responsible for the modern **periodic law**: When elements are arranged in order of increasing atomic number, there is a periodic repetition of their chemical and physical properties. The result is the periodic table as we know it today. Each new horizontal row of the periodic table corresponds to the beginning of a new **period** because a new principal energy level is being filled with electrons. Elements with similar chemical properties appear at regular intervals, within the vertical columns called **groups**.

PERIODIC LAW

When elements are arranged in order of increasing atomic number, there is a periodic repetition of their chemical and physical properties



Section Summary

- Elements of the periodic table are arranged in order of increasing atomic number.
- The periodic law states: "When elements are arranged in order of increasing atomic number, there is a periodic repetition of their chemical and physical properties."

This page titled 10.4.2: Periodic Law is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **6.3: Periodic Law** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.4.3: Modern Periodic Table- Periods and Groups

Learning Objectives

- Know how groups and periods are used to organize the modern periodic table, and how this relationship expresses the similarities among the elements.

The periodic table has undergone extensive changes in the time since it was originally developed by Mendeleev and Moseley. Many new elements have been discovered, while others have been artificially synthesized. Each fits properly into a **group** of elements with similar properties. The periodic table is an arrangement of the elements in order of their atomic numbers, so that elements with similar properties appear in the same vertical column or group.

The figure below shows the most commonly used form of the periodic table. Each square shows the chemical symbol of the element along with its name. Notice that several of the symbols seem to be unrelated to the name of the element: **Fe** for iron, **Pb** for lead, etc. Most of these are the elements that have been known since ancient times and have symbols based on their Latin names. The atomic number of each element is written above the symbol.

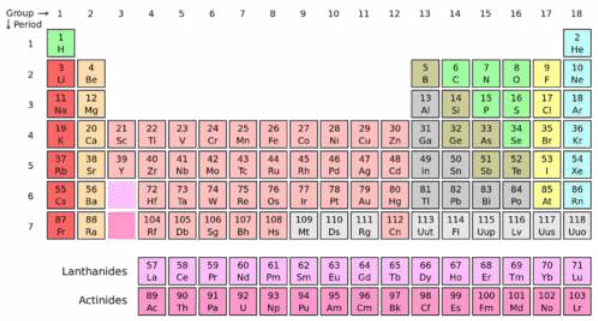


Figure **10.4.3.1** (Credit: User:Cepheus/Wikimedia Commons; Source: wikimedia; License: Public Domain)

A **period** is a horizontal row of the periodic table. There are seven periods in the periodic table, with each one beginning at the far left. A new period begins when a new principal energy level begins filling with electrons. Period 1 has only two elements (hydrogen and helium), while periods 2 and 3 have 8 elements. Periods 4 and 5 have 18 elements. Periods 6 and 7 have 32 elements, because the two bottom rows that are separate from the rest of the table belong to those periods. These two rows are pulled out in order to make the table itself fit more easily onto a single page.

A group is a vertical column of the periodic table, based on the organization of the outer shell electrons. There are a total of 18 groups. There are two different numbering systems that are commonly used to designate groups, and you should be familiar with both. The traditional system used in the United States involves the use of the letters A and B. The first two groups are 1A and 2A, while the last six groups are 3A through 8A. The middle groups use B in their titles. Unfortunately, there was a slightly different system in place in Europe. To eliminate confusion, the International Union of Pure and Applied Chemistry (IUPAC) decided that the official system for numbering groups would be a simple 1 through 18 from left to right. Many periodic tables show both systems simultaneously.

You might notice two rows of elements towards the bottom of the periodic table that seem unconnected to the rest of it. These elements do fit with the rest of the elements, but the width of the periodic table would have to be much greater to fit them in. You can look at the numbers to see where they fit in. These rows of elements are called series, and are named for whichever element is first in the series: lanthanides and actinides.

Most recent depictions of the periodic table show an incomplete seventh period. In fact, many of those seventh period elements were not known before the early twentieth century. Many elements have been synthesized by bombarding known elements with sub-atomic particles such as neutrons or alpha particles. Uranium has been used in this manner to produce elements 93-100. Larger elements (atomic numbers 101 and above) are formed by fusing nuclei of smaller elements together. These synthetic elements tend to be very unstable, often existing for less than a second, so little is known about them.

The video below reviews groups and trends in the periodic table.

Section Summary

- The periodic table is arranged in order of atomic number.
- A period is a horizontal row of the periodic table.
- A group is a vertical row of the periodic table.

Glossary

period
: a horizontal row on the periodic table whose properties are repeated in adjacent rows.

group
: a vertical column of the periodic table, within which are similar properties.

This page titled 10.4.3: Modern Periodic Table- Periods and Groups is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **6.4: Modern Periodic Table- Periods and Groups** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

10.5: Understanding the Periodic Table

As we mentioned in the previous section, the periodic table organizes elements based on their chemical and physical properties. The similarities exist in the groups of the periodic table, and those similarities are repeated along the periods. If you look at the periodic table below, you will notice similar colors in vertical chunks. These correspond to places where there are very similar chemical properties. We will explore these properties in a bit more detail in one of the subsections of this chapter. You might also notice that there are some color schemes that do not exactly match up with vertical chunks. There are a few reasons for this, that will be explored somewhat, but not in a great deal of detail. (The short answer is that there is greater complexity with the internal workings of some elements that gives rise to what might be viewed as exceptions.) Some of these differences are due to trends in physical properties, which have more complex origins than chemical properties. We will not get into the reasoning for that in this chapter, but we will also discuss those differences in physical properties briefly in the first subsection.

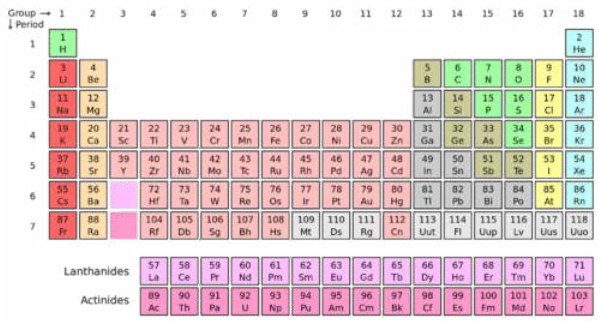


Figure **10.5.1** (Credit: User:Cepheus/Wikimedia Commons; Source: wikimedia; License: Public Domain)

This page titled 10.5: Understanding the Periodic Table is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

10.5.1: Metals, Nonmetals, and Metalloids.

Learning Objectives

- Know where to find metals, nonmetals, and metalloids on the periodic table.
- Know the characteristic properties of metal, nonmetal, and metalloids.

In this subsection we will consider one way of categorizing elements based on a set of properties. We will divide all of the elements into three categories and explore the properties of each of those categories: metals, nonmetal, and metalloids. You can see from Figure **10.5.1.1** that these categories do form arrangements on the periodic table, but not exactly confined to the groups we discussed earlier. The full details of why these do not match exactly is beyond the scope of this text. For now, be aware of the properties of metals, nonmetals, and metalloids and know in which regions of the periodic table each of these can be found.

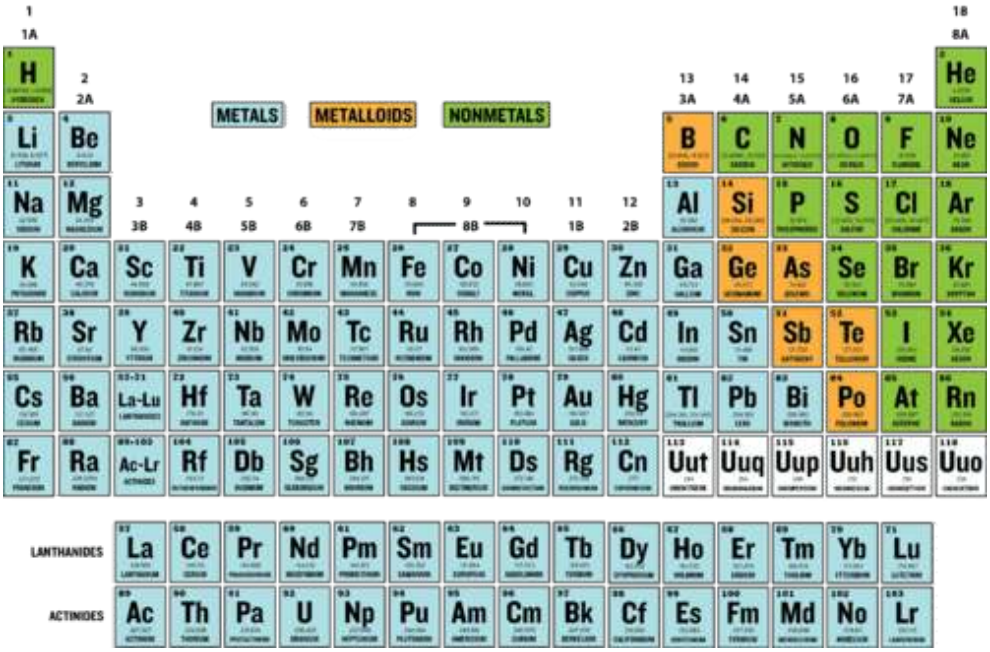


Figure **10.5.1.1** (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))

Metals

A metal is an element that is a good conductor of heat and electricity. Metals are also **malleable**, which means that they can be hammered into very thin sheets without breaking. They are **ductile**, which means that they can be drawn into wires. When a fresh surface of any metal is exposed, it will be very shiny because it reflects light well. This is called **luster**. All metals are solid at room temperature with the exception of mercury (Hg), which is a liquid. Melting points of metals display a very wide variance. The melting point of mercury is -39°C, while the highest melting metal is tungsten (W), with a melting point of 3422°C. The elements in blue in the periodic table below are metals. About 80 percent of the elements are metals.



Figure **10.5.1.2**: Gold jewelry. (Credit: User:Dorieo21/Wikimedia Commons; Source:wikimedia; License: Public Domain)

Nonmetals

In the chemical world, nonmetals might be considered “spare parts", loosely defined as not having the properties of metals. A **nonmetal** is an element that is generally a poor conductor of heat and electricity. Most properties of nonmetals are the opposite of

metals. There is a wider variation in properties among the nonmetals than among the metals. Nonmetals exist in all three states of matter. The majority are gases, such as nitrogen and oxygen. Bromine is a liquid. A few are solids, such as carbon and sulfur. In the solid state, nonmetals are **brittle**, meaning that they will shatter if struck with a hammer. The solids are not lustrous. Melting points are generally much lower than those of metals. The green elements in the periodic table shown in Figure **10.5.1.1** are nonmetals.



Figure **10.5.1.3**: Sulfur. (Credit: Ben Mills (Wikimedia: Benjah-bmm27); Source: wikimedia; License: Public Domain)

Metalloids

Some elements are “none of the above.” They do not fit neatly into the categories of metal or non-metal because of their characteristics. A **metalloid** is an element that has properties that are intermediate between those of metals and nonmetals. Metalloids can also be called semimetals. On the periodic table, the elements colored yellow, which generally border the stair-step line, are considered to be metalloids. Notice that aluminum borders the line, but it is considered to be a metal since all of its properties are like those of metals.

Silicon is a typical metalloid (see figure **10.5.1.4**). It has luster like a metal, but is brittle like a nonmetal. Silicon is used extensively in computer chips and other electronics because its electrical conductivity is in between that of a metal and a nonmetal.



Figure **10.5.1.4**: Silicon. (Credit: User:Enricoros/Wikipedia; Source: wikimedia; License: Public Domain)

Section Summary

- Metals are good conductors of heat and electricity.
- Metals are malleable and ductile
- Most metals are solids at room temperature.
- Nonmetals are generally poor conductors of heat and electricity.
- Properties of nonmetals are usually the opposite of properties of metals
- Nonmetals can be solid, liquid, or gas at room temperature depending upon the element.
- Metalloids are elements with properties intermediate between those of metals and non-metals.

Glossary

brittle
: substance that will will shatter if struck with a large force.

malleable
: substance that is easily shaped.

luster
: property of having a shiny surface.

metal
: element that is shiny, malleable, good conductor of heat and electricity.

metalloid

element that conducts heat and electricity moderately well, and possesses some properties of metals and some properties of nonmetals.

nonmetal

element that appears dull, poor conductor of heat and electricity.

This page titled 10.5.1: Metals, Nonmetals, and Metalloids. is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **6.5: Metals** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.5.2: Families of the Periodic Table

Learning Objectives

- Give the name and location of specific groups on the periodic table, including alkali metals, alkaline earth metals, noble gases, halogens, and transition metals.
- Identify elements that will have the most similar properties to a given element.

Families of the Periodic Table

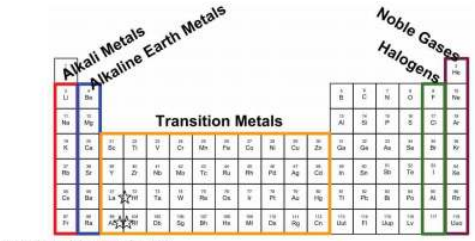
Remember that Mendeleev arranged the periodic table so that elements with the most similar properties were placed in the same group. A **group** is a vertical column of the periodic table. All of the 1A elements have one valence electron. This is what causes these elements to react in the same ways as the other members of the family. The elements in 1A are all very reactive and form compounds in the same ratios with similar properties with other elements. Because of their similarities in their chemical properties, Mendeleev put these elements into the same group. Group 1A is also known as the **alkali metals**. Although most metals tend to be very hard, these metals are actually soft and can be easily cut.

Group 2A is also called the **alkaline earth metals**. Once again, because of their similarities in electron configurations, these elements have similar properties to each other. The same pattern is true of other groups on the periodic table. Remember, Mendeleev arranged the table so that elements with the most similar properties were in the same group on the periodic table.

It is important to recognize a couple of other important groups on the periodic table by their group name. Group 7A (or 17) elements are also called **halogens**. This group contains very reactive nonmetal elements.

The **noble gases** are in group 8A. These elements also have similar properties to each other, the most significant property being that they are extremely unreactive, rarely forming compounds. We will learn the reason for this later, when we discuss how compounds form. The elements in this group are also gases at room temperature.

The **transition metals** are in groups 3-12. These designation of elements contains ten groups. It turns out that for this portion of the periodic table, the variations from one group to the next are not quite as predictable, and grouping them as one large block better represents the chemistry going on in this part of the periodic table.



Families of the periodic table.
CC – Tracy Poulsen

Section Summary

- The vertical columns on the periodic table are called groups or families because of their similar chemical behavior.
- All the members of a family of elements have similar chemical properties.
- Example groups that have very similar properties are alkali metals, alkaline earth metals, noble gases, and halogens.
- Transition metals are a block of several groups that have similar properties across the entire block.

Glossary

alkali metal

element in group 1

alkaline earth metal

element in group 2

group
vertical column of the periodic table

halogen
element in group 17

noble gas
(also, inert gas) element in group 18

period
horizontal row of the periodic table

transition metal
element in columns 3-11

Contributors

-
- Jamie MacArthur

This page titled 10.5.2: Families of the Periodic Table is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

10.6: Inside the Atom

Learning Objectives

- Know the charges and locations within an atom of protons, neutrons, and electrons.

Although atoms are very tiny, they consist of even smaller particles. Mendeleev never actually believed in these tiny particles that younger scientists were beginning to observe towards the end of his life. But as it turns out, these very particles were the very reason why the Periodic Law was even true!

Three main types of particles that make up all atoms are:

- protons, which have a positive electric charge and a large mass.
- electrons, which have a negative electric charge and a small mass.
- neutrons, which are neutral in electric charge and a large mass.

The model in the figure is a representation of how these particles are arranged in an atom. (Note: there are some inaccuracies in this model, particularly related to the position of the electrons. We will learn more about what the electrons are doing later in this chapter.) The particular atom represented by the model is helium, but the particles of all atoms are arranged in a similar way. At the center of the atom is a dense area called the nucleus, where all the protons and neutrons are clustered closely together. The electrons constantly move around the nucleus. Helium has two protons and two neutrons in its nucleus and two electrons moving around the nucleus. Atoms of other elements have different numbers of subatomic particles, but the number of protons always equals the number of electrons. This makes atoms neutral in charge because the positive and negative charges "cancel out." We will explore the atomic nucleus in greater detail later in this text. But it turns out it is the electrons that are of the greatest interest in terms of explaining the periodic law, as we shall begin to investigate in the next section.

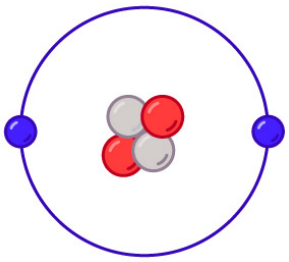


Figure **10.6.1**: Model of a helium atom. (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CK-12 Curriculum Materials license)

Section Summary

- Three main types of particles that make up all atoms are protons, neutrons, and electrons.
- Protons have a positive charge, electrons have a negative charge, and neutrons have no charge.
- Protons and neutrons are located in the nucleus of an atom, and electrons are located outside of the nucleus.

Glossary

electron

negatively charged, subatomic particle of relatively low mass located outside the atomic nucleus

neutron

uncharged, subatomic particle located in the atomic nucleus

nucleus

massive, positively charged center of an atom made up of protons and neutrons

proton

positively charged, subatomic particle located in the atomic nucleus

This page titled 10.6: Inside the Atom is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

10.7: Electrons and Light

Learning Objectives

- Understand how light interacts with electrons to change their energy levels.
- Explain how atomic emission spectrum is different than the spectrum of light from the Sun.

Electrons turn out to be the cause of chemical reactions. This makes a lot of sense if you consider that the electrons are the part of an atom which are on the outside, and therefore most likely to collide with another atom that could lead to a chemical reaction. But electrons behave much differently than the macroscopic matter that we are familiar with. We will develop a bit of an understanding of this behavior throughout the rest of this chapter, and then connect that back to the Periodic Table.

Atomic Emission Spectra

The electrons in an atom tend to be arranged in such a way that the energy of the atom is as low as possible. The **ground state** of an atom is the lowest energy state of the atom. When those atoms are given energy, the electrons absorb the energy and move to a higher energy level. These energy levels of the electrons in atoms are quantized, meaning again that the electron must move from one energy level to another in discrete steps, rather than continuously. An **excited state** of an atom is a state where its potential energy is higher than the ground state. An atom in the excited state is not stable. When it returns back to the ground state, it releases the energy that it had previously gained in the form of electromagnetic radiation.

So how do atoms gain energy in the first place? One way is to pass an electric current through an enclosed sample of a gas at low pressure. Since the electron energy levels are unique for each element, every gas discharge tube will glow with a distinctive color, depending on the identity of the gas (see below).

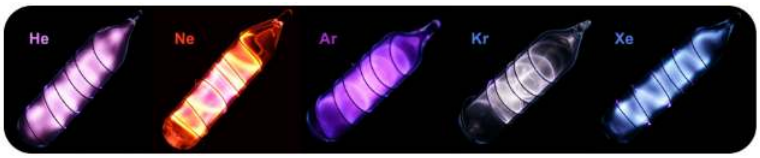


Figure 10.7.1: Gas discharge tubes are enclosed glass tubes filled with a gas at low pressure, through which an electric current is passed. Electrons in the gaseous atoms first become excited, and then fall back to lower energy levels, emitting light of a distinctive color in the process. Shown are gas discharge tubes of helium, neon, argon, krypton, and xenon. (Credit: User:Jurii/Wikimedia Commons, Heinrich Pniok (Wikimedia: Alchemist-hp); Source: Commons Wikimedia, Glowing noble gases(opens in new window) [commons.wikimedia.org]; License: CC by 3.0(opens in new window))

"Neon" signs are familiar examples of gas discharge tubes. However, only signs that glow with the red-orange color seen in the figure are actually filled with neon. Signs of other colors contain different gases or mixtures of gases.

Scientists studied the distinctive pink color of the gas discharge created by hydrogen gas. When a narrow beam of this light was viewed through a prism, the light was separated into four lines of very specific wavelengths (and frequencies since λ and ν are inversely related). An **atomic emission spectrum** is the pattern of lines formed when light passes through a prism to separate it into the different frequencies of light it contains. The figure below shows the atomic emission spectrum of hydrogen. Notice how this behavior differs from the continuous spectrum of visible light that we discussed earlier in this text. This difference was a challenge to solve and resulted in several models of electron behavior, each of which was more bizarre than the one before it. The laws of nature work very different on the scale of electrons than they do in the world we are used to.

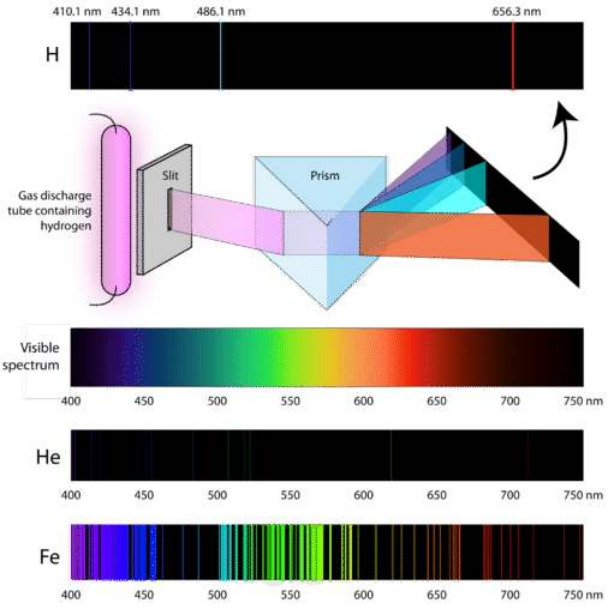


Figure **10.7.2**: When light from a hydrogen gas discharge tube is passed through a prism, the light is split into four visible lines. Each of these spectral lines corresponds to a different electron transition from a higher energy state to a lower energy state. Every element has a unique atomic emission spectra available in the public domain; Source: CK-12 Foundation; H spectrum: Commons Wikimedia, Emission Spectrum- H(opens in new window) [commons.wikimedia.org]; visible spectrum: Commons Wikimedia, Linear Visible Spectrum (opens in new window) [commons.wikimedia.org]; He spectrum: Commons Wikimedia, Helium Emission Spectrum(opens in new window) [commons.wikimedia.org]; Fe spectrum: Commons Wikimedia, Emission Spectrum-Fe(opens in new window) [commons.wikimedia.org]; License: CC BY-NC 3.0(opens in new window)) emission spectrum, as shown by the examples of helium (**He**) and iron (**Fe**). (Credit: Christopher Auyeung, using

Bohr's Atomic Model

Following the discoveries of hydrogen emission spectra and the photoelectric effect, the Danish physicist Niels Bohr (1885-1962) proposed a new model of the atom in 1915. Bohr proposed that electrons do not radiate energy as they orbit the nucleus, but exist in states of constant energy that he called **stationary states**. This means that the electrons orbit at fixed distances from the nucleus (see below). Bohr's work was primarily based on the emission spectra of hydrogen. This is also referred to as the planetary model of the atom. It explained the inner workings of the hydrogen atom. Bohr was awarded the Nobel Prize in physics in 1922 for his work.

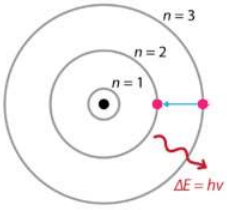


Figure **10.7.3**: Bohr's atomic model hydrogen emission spectra. (Credit: Zachary Wilson; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Bohr explained that electrons can be moved into different orbits with the addition of energy. When the energy is removed, the electrons return back to their ground state, emitting a corresponding amount of energy—a quantum of light, or photon. This was the basis for what later became known as **quantum theory**. This is a theory based on the principle that matter and energy have the properties of both particles and waves. It accounts for a wide range of physical phenomena, including the existence of discrete packets of energy and matter, the uncertainty principle, and the exclusion principle.

According to the Bohr model, often referred to as a **planetary model**, the electrons encircle the nucleus of the atom in specific allowable paths called orbits. When the electron is in one of these orbits, its energy is fixed. The ground state of the hydrogen atom, where its energy is lowest, is when the electron is in the orbit that is closest to the nucleus. The orbits that are further from the nucleus are all of successively greater energy. The electron is not allowed to occupy any of the spaces in between the orbits. An everyday analogy to the Bohr model is the rungs of a ladder. As you move up or down a ladder, you can only occupy specific rungs

and cannot be in the spaces in between rungs. Moving up the ladder increases your potential energy, while moving down the ladder decreases your energy.

Based on the wavelengths of the spectral lines, Bohr was able to calculate the energies that the hydrogen electron would have in each of its allowed energy levels. He then mathematically showed which energy level transitions correspond to the spectral lines in the atomic emission spectrum, as shown in Figure **10.7.4**.

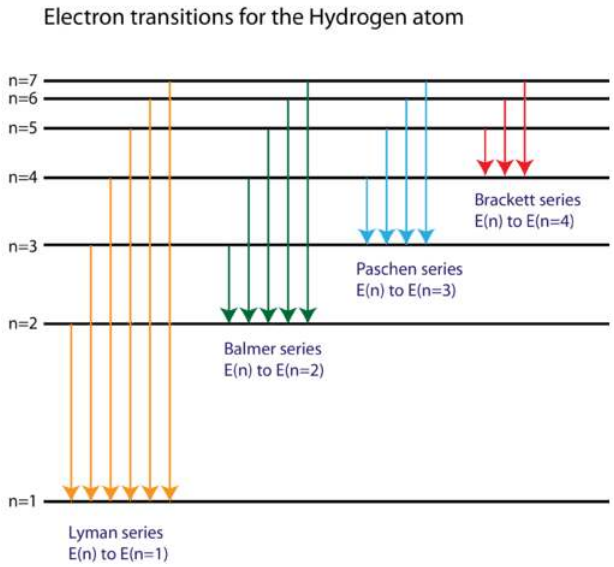


Figure **10.7.4**: The electron energy level diagram for the hydrogen atom. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Bohr's model was a tremendous success in explaining the spectrum of the hydrogen atom. Unfortunately, when the mathematics of the model were applied to atoms with more than one electron, it was not able to correctly predict the frequencies of the spectral lines. While Bohr's model represented a great advancement in the atomic model and the concept of **electron transitions** between energy levels is valid, improvements were needed in order to fully understand all atoms and their chemical behavior. Shortly after Bohr published his planetary model of the atom, several new discoveries were made, which resulted in, yet again, a revised view of the atom.

Summary

- Atomic emission spectra are produced when excited electrons return to the ground state.
- The emitted light of electrons corresponds to energies of the specific electrons.
- The Bohr model postulates that electrons orbit the nucleus at fixed energy levels.
- Orbits further from the nucleus exist at higher energy levels.
- When electrons return to a lower energy level, they emit energy in the form of light.
- Emission lines for hydrogen correspond to energy changes related to electron transitions.
- The Bohr model works only for the hydrogen atom.

Glossary

ground state

lowest energy state of the atom

excited state

state where its potential energy is higher than the ground state

atomic emission spectrum

pattern of lines formed when light emitted from an excited atom passes through a prism to separate it into the different frequencies of light it contains

electron transitions

process by which electrons move between different allowable energy levels within the atom.

This page titled 10.7: Electrons and Light is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **5.5: Atomic Emission Spectra** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **5.6: Bohr's Atomic Model** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **5.7: Spectral Lines of Atomic Hydrogen** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.8: Quantum Mechanics

Learning Objectives

- Explain the dual nature of electrons related to particles and waves.
- Properly explain the possible locations of electrons within an atom based on quantum mechanics.

How do you study something that seemingly makes no sense?

We discuss electrons being in orbits, and it sounds like we can tell where an electron is at any moment. We can draw pictures of electrons in orbit, but the reality is that we just do not know exactly where they are. We will take a quick look at an area of science that even leaves scientists puzzled. When asked about quantum mechanics, Niels Bohr (who proposed the Bohr model of the atom) said, "Anyone who is not shocked by quantum theory has not understood it." Richard Feynman (one of the founders of modern quantum theory) stated, "I think I can safely say that nobody understands quantum theory." So, let's take a short trip into a land that challenges our everyday world...

Quantum Mechanics

Earlier in this text we discussed classical mechanics, the study of the motion of macroscopic objects like people and balls and vehicles. Because of the quantum nature of the electron and other tiny particles moving at high speeds, classical mechanics is inadequate to accurately describe their motion. **Quantum mechanics** is the study of the motion of objects that are atomic or subatomic in size and thus demonstrate wave-particle duality. In classical mechanics, the size and mass of the objects involved effectively obscures any quantum effects, so that such objects appear to gain or lose energies in any amounts. Particles whose motion is described by quantum mechanics gain or lose energy in the small pieces called **quanta**.

One of the fundamental (and hardest to understand) principles of quantum mechanics is that the electron is both a particle and a wave. In the everyday macroscopic world of things we can see, something cannot be both. But this duality can exist in the quantum world of the submicroscopic on the atomic scale. There are even more baffling aspects of quantum mechanics we could explore, but we will focus on developing a model that we can use to make some sense of what is going on with these electrons.

One of the tenants of quantum mechanics is that any moving object also has wave properties. (The specific relationship is called the de Broglie wave equation, but we will not explore it in detail.) The wavelength is inversely proportional to the momentum of the moving object. It turns out that for a large object, this wavelength ends up being so small as to be essentially meaningless. On the other hand, particles with measurable wavelengths are all very small. However, the wave nature of the electron proved to be a key development in a new understanding of the nature of the electron. An electron that is confined to a particular space around the nucleus of an atom can only move around that atom in such a way that its electron wave "fits" the size of the atom correctly. This means that the frequencies of electron waves are **quantized**. Based on the equation

$$E = hf$$

the quantized frequencies mean that electrons can only exist in an atom at specific energies, as Bohr had previously theorized.

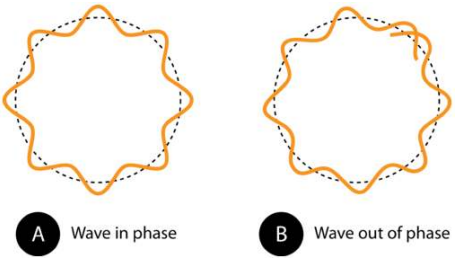


Figure 10.8.1: The circumference of the orbit in (A) allows the electron wave to fit perfectly into the orbit. This is an allowed orbit. In (B), the electron wave does not fit properly into the orbit, so this orbit is not allowed. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Energy Levels and Electron Orbitals

Energy levels (also called electron shells) are fixed distances from the nucleus of an atom where electrons may be found. You can stand on one step or another but not in between the steps. The same goes for electrons. They can occupy one energy level or another but not the space between energy levels. This sounds very much like the Bohr model of the atom we introduced previously, but there are a few key differences that we will note.

The model in Figure **10.8.2** shows the first four energy levels of an atom. Electrons in energy level I (also called energy level K) have the least amount of energy. As you go farther from the nucleus, electrons at higher levels have more energy, and can hold more electrons. The difference between this model and the Bohr model is that we are not specifying the location or direction of movement of the electron, simply the potential energy of the electron based on its distance from the nucleus.

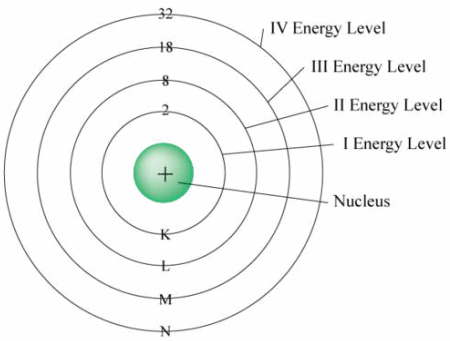


Figure **10.8.2** (Credit: Zachary Wilson; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

You might wonder what the movement of the electron looks like if it is not circling the nucleus like a planet circles the Sun. The mathematical model for the behavior of electrons resulted in what are referred to as **wave functions.** These only indicate the probability of where the electron is most likely to be found (and conversely, were it might not be allowed to exist.)

The location of the electrons in the quantum mechanical model of the atom is often referred to as an **electron cloud**. The electron cloud can be thought of in the following way: Imagine placing a square piece of paper on the floor with a dot in the circle representing the nucleus. Now take a marker and drop it onto the paper repeatedly, making small marks at each point the marker hits. If you drop the marker many, many times, the overall pattern of dots will be roughly circular. If you aim toward the center reasonably well, there will be more dots near the nucleus and progressively fewer dots as you move away from it. Each dot represents a location where the electron could be at any given moment. Because of the uncertainty principle, there is no way to know exactly where the electron is. An electron cloud has variable densities: a high density where the electron is most likely to be and a low density where the electron is least likely to be, as shown in Figure **10.8.3**.

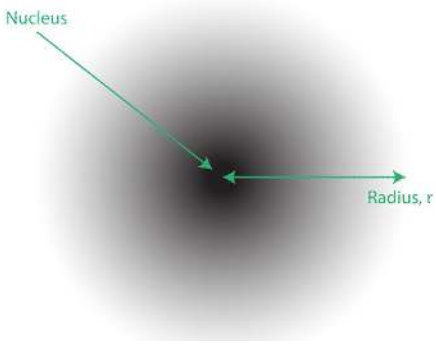


Figure **10.8.3**: An electron cloud: the darker region nearer the nucleus indicates a high probability of finding the electron, while the lighter region further from the nucleus indicates a lower probability of finding the electron. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

In order to specifically define the shape of the cloud, it is customary to refer to the area within which there is a **90** chance of finding the electron. This is called an **orbital**, the three-dimensional region of space that indicates where there is a high probability of finding an electron. Figure **10.8.4** shows some models of electron orbitals. Note that not each of the possible orbitals results in a spherical shape.

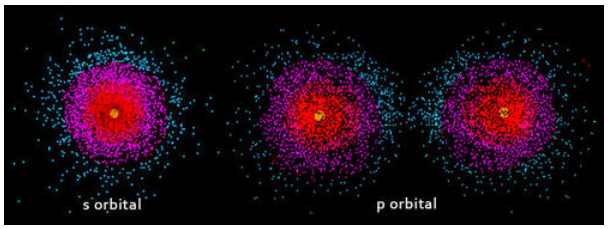


Figure 10.8.4 (Credit: Laura Guerin; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Section Summary

- Quantum mechanics involves the study of material at the atomic level where the particles motion is described by gaining or losing the discrete amounts called quanta.
- In quantum mechanics, electrons exist simultaneously as both a particle and wave.
- This field deals with probabilities since we cannot definitely locate a particle.
- Energy levels (also called electron shells) are fixed distances from the nucleus of an atom where electrons may be found. As you go farther from the nucleus, electrons at higher energy levels have more energy.
- The location of the electron can only be given as a probability that the electron is somewhere in a certain area.

Glossary

quantum mechanics
: study of the motion of objects that are atomic or subatomic in size and thus demonstrate wave-particle duality

quantized
: something that only has certain values allowed

energy levels
: fixed distances from the nucleus of an atom where electrons may be found

wave functions
: mathematical model for the behavior of electrons which indicates the probability of where the electron is most likely to be found

electron cloud
: an indication of the most likely locations of the electron in the quantum mechanical model of the atom

orbital
: three-dimensional region of space that indicates where there is a high probability of finding an electron

---

This page titled 10.8: Quantum Mechanics is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **5.9: Quantum Mechanics** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **5.11: Quantum Mechanical Atomic Model** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

10.9: Electron Shells

Learning Objectives

- Explain the trend of filling energy levels with electrons.
- Explain the connection between valence electrons and the periodic law.

From both the Bohr model and quantum mechanics we now know that electrons are arranged in an orderly manner of energy levels. In this final section, we will see how this particular arrangement of electrons relates back to the periodic law.

Electron Shells

The energy levels for the electrons in an atom are often referred to as **electron shells.** (Note: as we discussed previously, this does not refer to movement according to classical mechanics, but an indication of increasing potential energy as the shell number increases.) These shells are given notations 1n, 2n, etc. to indicate the increasing potential energy of each shell, as indicated in Figure **10.9.1** Each of these energy levels contains one or more **orbitals** that have a limited number of electrons that they may hold. The electrons are generally added to the lowest energy orbital available. When the available orbitals within one energy level have been filled, any additional electrons must be added to the next energy level. In addition to being of a higher potential energy, each successive energy level has a greater number of orbitals it contains, and therefore a greater capacity to contain electrons. As the number of orbitals within an energy level increases, there is a bit more complexity to the order in which the electrons are added. The details of this complexity are beyond the scope of this course, but we will mention some of their implications as it relates to the effects this has on periodic trends.

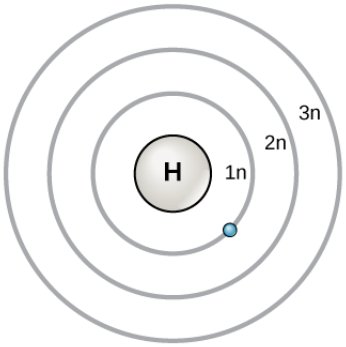


Figure **10.9.1**: The electron shell model is consistent with both the Bohr model and quantum mechanics.

When we begin to add these electrons to an energy level, we start to see a trend emerge that mirrors the trend first observed by Mendeleev. At some point an electron shell will be filled with electrons. Regardless of which electron shell we are talking about, the significance of a filled electron shell has vast implications for the chemical properties of that atom. If we imagine an atom with one additional electron, we end up with an atom that has a single electron in its outer electron shell. Likewise, the significance of an electron shell with a single electron has vast implications on the chemical properties of that atom. Figure **10.9.2** shows models of these electron shells for a portion of the periodic table. Note that each element in group 1 has a single electron in its outer energy shell, and that each element in group 18 has that energy shell filled (though the number of electrons necessary to fill it changes from 2 to 8 after the first shell). Also note that groups 14 and 17 have identical numbers of electrons in their outer energy level. This trend is so significant that we have a special name for the electrons in the outermost electron shell: **valence electrons.** It turns out that most of the chemical properties of greatest importance to the function of life occur because of the behavior of these electrons. We shall explore the behavior of these electrons in more detail later in this text.

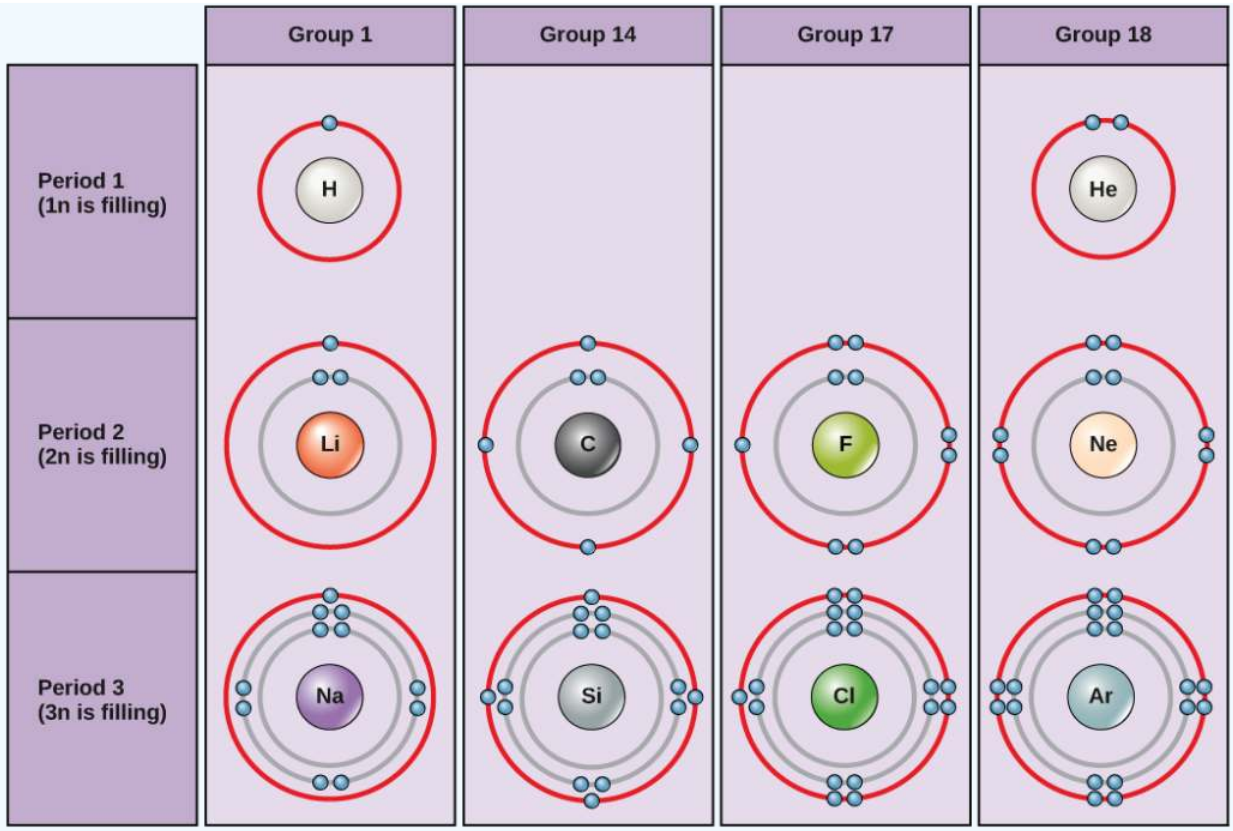


Figure **10.9.2**: electron shell models focusing on the valence electrons for select groups of the period table within the first three periods.

Valence Electrons and the Periodic Law

It turns out that the reason why the periodic law works as a predictor of chemical properties is because of the behavior of the valence electrons. If we look at a periodic table, such as the one shown in Figure **10.9.3**, we can see how some of these trends emerge. The groups which have the greatest similarities in chemical properties were the ones to which we ascribed special names: alkali metals, alkaline earth metals, halogens, and noble gases. Note that these are the positions on the periodic table where the valence electron shell is full, nearly full, or nearly empty. The tendency towards a full valence electron shell is so important that the elements within each of these groups become almost identical in their chemical properties. Elsewhere on the periodic table these trends exist, but are not as pronounced. In fact the areas of the period table we had designated as blocks even though they encompassed several groups (transition metals, lanthanides, actinides) have more complex chemistry because it sometimes involves electrons that are not in the valence shell.

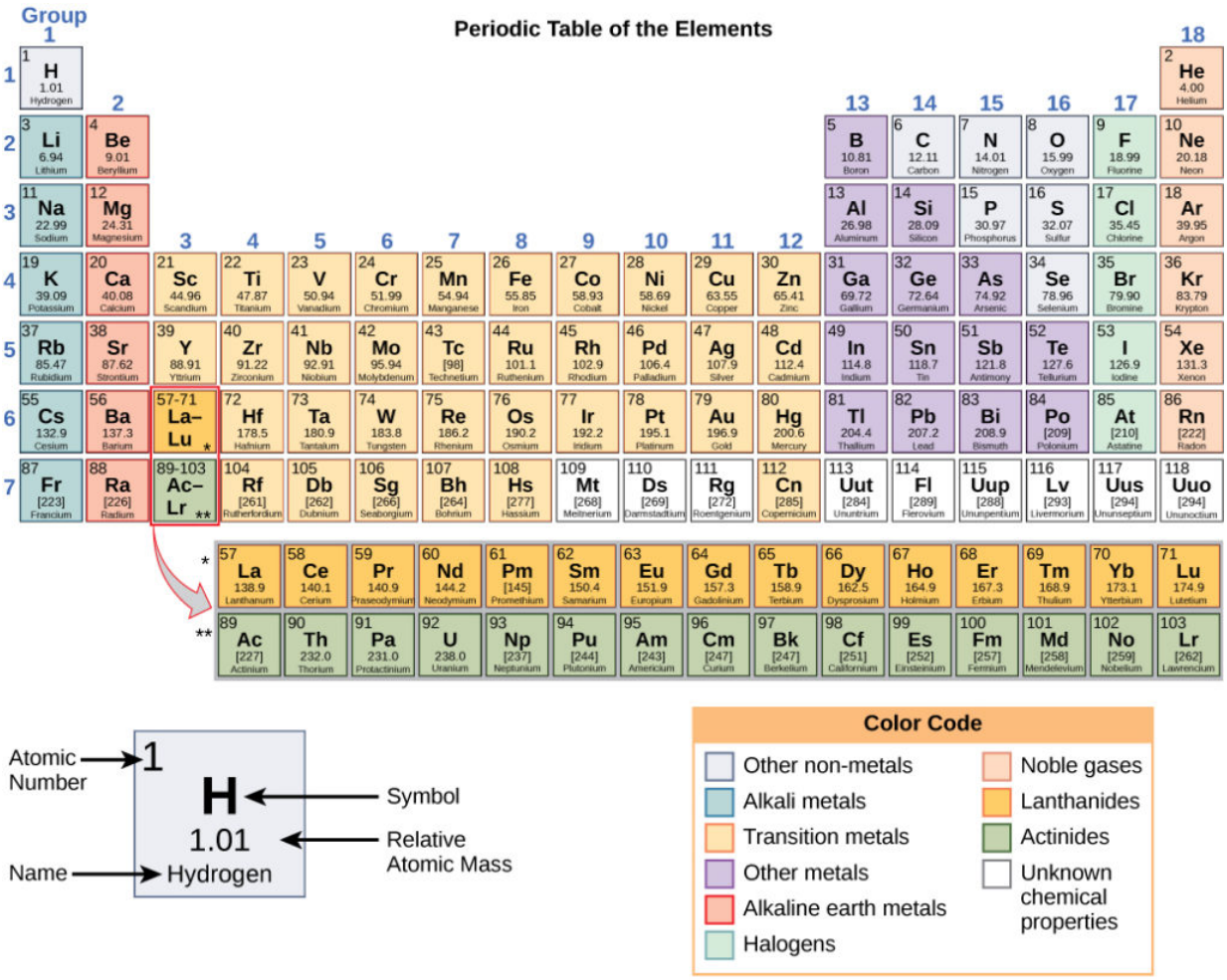


Figure **10.9.3**: The periodic table showing groupings of elements (note: the metalloids have been omitted from this table.)

Ironically, Mendeleev never actually believed in electrons, even though they were the factor that could explain the natural law he had discovered! This is often an unfortunate trend among established scientists who are not able to recognize the contributions of those who come along after them. Nevertheless the progress of science has continued to move along. However, it is an important life lesson to realize that even as teachers there are things we might learn from our students.

Section Summary

- Energy levels are predicted by both the Bohr model and quantum mechanics.
- Electrons typically fill orbitals of the highest vacant energy level.
- Electrons in the outermost energy level are called valence electrons.
- Valence electrons predict chemical properties, and thereby the periodic law.

Glossary

electron shells
: the energy levels of the electrons within an atom.

orbitals
: most precise information typically available about the energy and location of an electron. Capable of containing up to two electrons.

valence electrons
: electrons found in the outermost electron shell.

Contributors and Attributions

CC licensed content, Shared previously

- Jamie MacArthur
- Biology. **Provided by**: OpenStax CNX. **License**: *CC BY: Attribution(opens in new window)*. **License Terms**: Download for free

---

This page titled 10.9: Electron Shells is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

- **4.5: Electrons** has no license indicated.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

11: Chemical Bonds

11.1: Chemical Bond
11.2: Elements, molecules, and compounds
11.3: The Octet Rule
11.4: Ions and Ion Formation
11.5: Ionic Bonding
11.6: Metallic Bonding
11.7: Covalent Bonds and Molecules
11.8: Bond Polarity
11.9: Intermolecular Forces
11.10: Naming Compounds

This page titled 11: Chemical Bonds is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

11.1: Chemical Bond



Figure **11.1.1** (Credit: Flickr: Amelia Wells; Source: flickr; License: CC by 2.0(opens in new window))

Have you ever made cupcakes from scratch?

You mix together flour, sugar, eggs, and other ingredients to make the batter, put the batter into cupcake papers, and then put them into the oven to bake. The cupcakes that come out of the oven after baking are different from any of the individual ingredients that went into the batter. Like the ingredients that join together to make cupcakes, atoms of different elements can join together to form entirely different substances called compounds. In cupcakes, the eggs and other wet ingredients cause the dry ingredients to stick together. What causes elements to stick together in compounds? The answer is chemical bonds.

What Is a Chemical Bond?

A chemical bond is a force of attraction between atoms or ions. Bonds form when atoms share or transfer valence electrons. Valence electrons are the electrons in the outer energy level of an atom that may be involved in chemical interactions. Valence electrons are the basis of all chemical bonds.

Q: Why do you think that chemical bonds form?

A: Chemical bonds form because they give atoms a more stable arrangement of electrons.

Why Bonds Form

To understand why chemical bonds form, consider the common compound known as water, or H_2O. It consists of two hydrogen (H) atoms and one oxygen (O) atom. As you can see in the on the left side of the figure below, each hydrogen atom has just one electron, which is also its sole valence electron. The oxygen atom has six valence electrons. These are the electrons in the outer energy level of the oxygen atom.

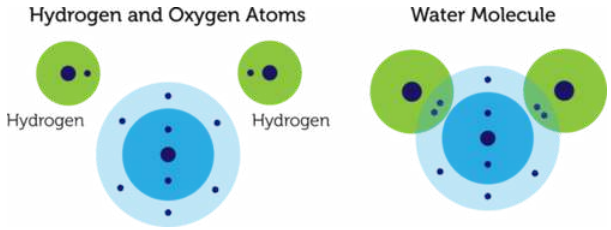


Figure **11.1.2**: Copy and Paste Caption here. (Credit: Christopher AuYeung; Source: CK-12 Foundation; License: CC by-NC 3.0(opens in new window))

In the water molecule on the right in the figure above, each hydrogen atom shares a pair of electrons with the oxygen atom. By sharing electrons, each atom has electrons available to fill its sole or outer energy level. The hydrogen atoms each have a pair of shared electrons, so their first and only energy level is full. The oxygen atom has a total of eight valence electrons, so its outer energy level is full. A full outer energy level is the most stable possible arrangement of electrons. It explains why elements form chemical bonds with each other.

Types of Chemical Bonds

Not all chemical bonds form in the same way as the bonds in water. There are actually three different types of chemical bonds, called covalent, ionic, and metallic bonds. Each type of bond is described in more detail within this chapter.

This page titled 11.1: Chemical Bond is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Elements, molecules, and compounds

Learning Objectives

- Determine if a substance is an element, a compound, or a molecule from its chemical formula.

Substances

Silver and ordinary table salt are two examples of chemical **substances**. A substance is matter that has a uniform and definite composition. All samples of substances, sometimes called pure substances, have identical properties. When chemists run a chemical reaction, they want to use pure materials so they know exactly what they are dealing with. They know that the reaction involves a specific substance, so they expect the same reaction to give the same results each time that it is run.

One characteristic of a substance is that it cannot be separated into components by physical separation techniques (which we will cover in more detail later in this text). The **compound** acetylsalicylic acid (the active ingredient in aspirin) is a substance because it is a chemical compound that cannot be further reduced to other materials by physical means. However, the aspirin that contains the acetylsalicylic acid is a mixture. In addition to the active pain-relieving ingredient, aspirin also may contain waxes, corn starch, and cellulose, among other ingredients. These added materials bind the tablet components into a solid pill and influence how rapidly the material dissolves.



Figure **11.2.1**: Aspirin tablets. (Credit: Heung Soon; Source: pixabay; License: Pixabay License)

Substances may be either **elements** or pure compounds. They may also be solid, liquid, or gas. Some of the elements (such as silver) exist in the solid form. Other elements (mercury) exist as liquids at room temperature. Elements such as helium and radon exist as gases at room temperature. A compound such as water will take the form of a solid, a liquid, or a gas, depending on the temperature.

Compounds

A **compound** is a substance that contains two or more elements chemically combined in a fixed proportion. The elements carbon and hydrogen combine to form many different compounds. One of the simplest is called methane, in which there are always four times as many hydrogen particles as carbon particles. Methane is a pure substance because it always has the same composition. However, it is not an element because it can be broken down into simpler substances—carbon and hydrogen.

Recall that the components of a mixture can be separated from one another by physical means. This is not true for a compound. Table salt is a compound consisting of equal parts of the elements sodium and chlorine. Salt cannot be separated into its two elements by filtering, distillation, or any other physical process. Salt and other compounds can only be decomposed into their elements by a chemical process. A **chemical change** is a change that produces matter with a different composition. Many compounds can be decomposed into their elements by heating. When sugar is heated, it decomposes into carbon and water. Water is still a compound, but one which cannot be broken down into hydrogen and oxygen by heating. Instead, the passage of an electrical current through water will produce hydrogen and oxygen gases.

The properties of compounds are generally very different than the properties of the elements from which the compound is formed. Sodium is an extremely reactive soft metal that cannot be exposed to air or water. Chlorine is a deadly gas. The compound sodium chloride is a white solid which is essential for all living things (see below).

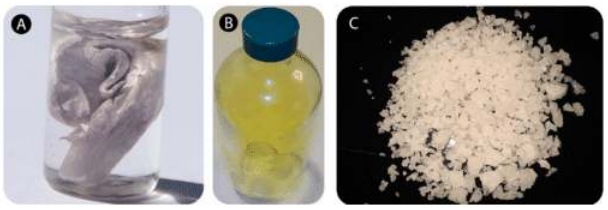


Figure 11.2.1: (A) Sodium is so reactive that it must be stored under oil. (B) Chlorine is a poisonous yellow-green gas. (C) Salt crystals, a compound of sodium and chlorine. (Credit: (A) Hi-Res Images of Chemical Elements; (B) User: Greenhorn1/Wikimedia Commons; (C) User:Miansari66/Wikimedia Commons; Source: (A) images of elements (B) wikimedia (C) wikimedia; License: (A) CC by 3.0(opens in new window) (B) Public Domain; (C) Public Domain)

Chemical Symbols and Formulas

Chemical Formula

In order to illustrate chemical reactions and the elements and **compounds** involved in them, chemists use symbols and formulas. A **chemical symbol** is a one- or two-letter designation of an element. Some examples of chemical symbols are **O** for oxygen, **Zn** for zinc, and **Fe** for iron. The first letter of a symbol is always capitalized. If the symbol contains two letters, the second letter is lower case. The majority of elements have symbols that are based on their English names. However, some of the elements that have been known since ancient times have maintained symbols that are based on their Latin names, as shown below.

Table 1: Symbols and Latin Names for Elements

Chemical Symbol	Name	Latin Name
Na	Sodium	Natrium
K	Potassium	Kalium
Fe	Iron	Ferrum
Cu	Copper	Cuprum
Ag	Silver	Argentum
Sn	Tin	Stannum
Sb	Antimony	Stibium
Au	Gold	Aurum
Pb	Lead	Plumbum

Watch the following video to learn more about the origin of different element's symbols.

Chemical Formulas

Compounds are combinations of two or more elements. A **chemical formula** is an expression that shows the elements in a compound and the relative proportions of those elements. Water is composed of hydrogen and oxygen in a two to one ratio. The chemical formula for water is H_2O. Sulfuric acid is one of the most widely produced chemicals in the United States and is composed of the elements hydrogen, sulfur, and oxygen. The chemical formula for sulfuric acid is H_2SO_4.

Molecules

In the introduction to this chapter, we looked at an example of a water molecule. But what does it mean for something to be a molecule? **Molecules** form between two or more atoms when they share electrons. This is not the only type of chemical bond that can form, so not everything is made of molecules. We will explore these details as we move through the chapter. But what we can say for now is that both elements and compounds can be molecules. The H_2O we looked at previously is a molecule and it is also a compound. We know it is a compound because it contains two different types of elements: H and O. But O_2 is a molecule but not a compound. We can see from the formula that it contains more than one atom, but also that both of those atoms are the same. If a

formula contains more than one atom of the same element it is always a molecule, but also an element. If a formula contains atoms of different elements, it is a compound and might also be a molecule. As we explore the different types of bonding that are possible later in this chapter we will learn when that bonding involves sharing of electrons.

Section Summary

- A substance is a pure material with a uniform and definite composition.
- A substance cannot be separated into components using physical means.
- A substance can be a solid, a liquid, or a gas, dependent on temperature.
- A compound is a substance that contains two or more elements chemically combined in a fixed proportion.
- A chemical change is a change that produces matter with a different composition.
- A chemical symbol is a one- or two-letter designation of an element.
- Compounds are combinations of two or more elements.
- A chemical formula is an expression that shows the elements in a compound and the relative proportions of those elements.
- Some elements have symbols that derive from the Latin name for the element.
- Molecules are atoms bonded together by sharing electrons, and can be either elements or compounds.

Glossary

substance

matter that has a uniform and definite composition.

element

something which can't be broken down into anything simpler by a chemical reaction.

compound

a substance that contains two or more elements chemically combined in a fixed proportion.

chemical change

a change that produces matter with a different composition.

chemical symbol

a one- or two-letter designation of an element.

chemical formula

an expression that shows the elements in a compound and the relative proportions of those elements.a one- or two-letter designation of an element.

molecule

forms between two or more atoms when they share electrons.

This page titled 11.2: Elements, molecules, and compounds is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: Pure Substances** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.12: Compounds** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.15: Chemical Symbols and Formulas** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

11.3: The Octet Rule

Learning Objectives

- Draw an electron configuration for any of the representative elements.
- Explain how the Octet Rule predicts trends in chemical bonding.



Figure **11.3.1** (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CK-12 Curriculum Materials license)

How are electrons organized in atoms?

Graduations are exciting events in one's life. The processional, the ceremonies, even the speeches can be handled because this marks the end of a stage of life and the beginning of another one. The ceremony itself is challenging for those organizing it. There are just enough seats on the stage for the graduates. How do you avoid getting too many people in a row and not enough in the next row? Someone is stationed at the end of the row to count the students as they enter. Only so many are allowed to go into a row, and then you begin to fill the next row. Electrons in atoms behave the same way. There are rules that determine where electrons go in compounds.

Electron Dot Diagrams

Recall that the valence electrons of an atom are the electrons located in the highest occupied principal energy level. Valence electrons are primarily responsible for the chemical properties of elements. The number of valence electrons can be easily determined from the group of the periodic table that an element is located in. Several examples from the second period elements are shown in the table below. In each case, valence electrons are those in the second principal energy level. As one proceeds left to right across a period, the number of valence electrons increases by one.

Table **11.3.1**:

Table showing elements, group numbers, and number of valence electrons

Element	Group Number	Number of Valence Electrons
lithium	1	1
beryllium	2	2
nitrogen	5	5
neon	8	8

When examining chemical bonding, it is necessary to keep track of the valence electrons of each atom. **Electron dot diagrams** are diagrams in which the valence electrons of an atom are shown as dots distributed around the element's symbol. A beryllium atom, with two valence electrons, has the electron dot diagram below:



Figure **11.3.2** (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))

Since electrons repel each other, the dots for a given atom are distributed evenly around the symbol before they are paired. The table below shows the electron dot diagrams for the entire second period.

Table 11.3.2: Electron Dot Diagrams for the Second Period Elements

Electron Dot Diagrams for the Second Period Elements

Group Number	Electron Dot Diagram
1	 Figure 11.3.3 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))
2	 Figure 11.3.4 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))
13	 Figure 11.3.5 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))
14	 Figure 11.3.6 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))
15	 Figure 11.3.7 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))
16	 Figure 11.3.8 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))
17	 Figure 11.3.9 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))
18	 Figure 11.3.10 (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))

Electron dot diagrams would be the same for each element in the same groups as those listed here. The Octet Rule only applies to these representative elements, and we will not be using it for transition metals. It is important to understand these dot diagrams for the representative elements before we are able to understand chemical bonding. The following video provides an additional resource to practice this concept.



Octet Rule

The noble gases are unreactive because of their electron configurations. The noble gas neon has a full outer shell and cannot incorporate any more electrons into the valence shell. The other noble gases have the same outer shell electron configuration, even though they have different numbers of inner-shell electrons.



Figure **11.3.2**: Electron configuration of neon atom. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC-SA 3.0(opens in new window))

American chemist Gilbert Lewis (1875-1946) used this observation to explain the types of ions and molecules that are formed by other elements. He called his explanation the **octet rule**. The octet rule states that atoms tend to form compounds in ways that give them eight valence electrons, and thus the electron configuration of a noble gas. An exception to an octet of electrons is in the case of the first noble gas, helium, which only has two valence electrons. This primarily affects the element hydrogen, which forms stable compounds by achieving two valence electrons. Lithium, an alkali metal with three valence electrons, is also an exception to the octet rule. Lithium tends to lose one electron to take on the electron configuration of the nearest noble gas, helium, leaving it with two valence electrons.

Octet Rule

Atoms tend to form compounds in ways that give them eight valence electrons, and thus the electron configuration of a noble gas.

There are two ways in which atoms can satisfy the octet rule. One way is by sharing their valence electrons with other atoms. The second way is by transferring valence electrons from one atom to another. Atoms of metals tend to lose all of their valence electrons, which leaves them with an octet from the next lowest principal energy level. Atoms of nonmetals tend to gain electrons in order to fill their outermost principal energy level with an octet. We will explore these bonding concepts as we continue to move through the chapter.

Summary

- Electron dot diagrams show the valence electrons for an atom.
- The dot diagrams are the same for each element in the representative element groups.
- Atoms form compounds in ways that give them eight valence electrons.

- Metals tend to lose electrons to achieve this configuration.
- Nonmetals tend to gain electrons to achieve this configuration.

Glossary

electron dot diagrams

diagrams in which the valence electrons of an atom are shown as dots distributed around the element's symbol.

octet rule

Atoms tend to form compounds in ways that give them eight valence electrons, and thus the electron configuration of a noble gas.

This page titled 11.3: The Octet Rule is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **8.2: Octet Rule** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **8.1: Electron Dot Diagrams** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

11.4: Ions and Ion Formation

Learning Objectives

- Use the Octet Rule to predict the charge that ions of a given element will form.
- Correctly refer to ions as anions or cations depending on their charge.

Atoms to Ions

Atoms can gain electrons. They can also lose electrons. In either case, they become **ions**. Ions are atoms that have a positive or negative charge because they have unequal numbers of protons and electrons. If atoms lose electrons, they become positive ions, or **cations**. If atoms gain electrons, they become negative ions, or **anions**. Consider the example of fluorine (see Figure **11.4.1**). A fluorine atom has nine protons and nine electrons, so it is electrically neutral. If a fluorine atom gains an electron, it becomes a fluoride ion with an electric charge of -1.

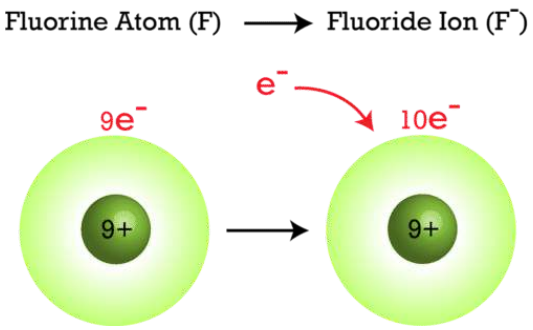


Figure **11.4.1** (Credit: Christopher AuYeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Names and Symbols

Like fluoride, other negative ions usually have names ending in *–ide*. Positive ions, on the other hand, are just given the element name followed by the word *ion*. For example, when a sodium atom loses an electron, it becomes a positive sodium ion. The charge of an ion is indicated by a plus (+) or minus sign (-), which is written to the right of and just above the ion's chemical symbol. For example, the fluoride ion is represented by the symbol F^-, and the sodium ion is represented by the symbol Na^+. If the charge is greater than one, a number is used to indicate it. For example, magnesium (Mg) may lose two electrons to form an ion with a charge of plus two. This ion would be represented by the symbol Mg^{2+}. This and some other common ions are listed with their symbols in the table below.

Table ***PageIndex1***: Some Common Ions

Cations		Anions	
Name of Ion	Chemical Symbol	Name of Ion	Chemical Symbol
Calcium ion	Ca^{2+}	Chloride	Cl^-
Hydrogen ion	H^+	Fluoride	F^-
Magnesium ion	Mg^{2+}	Bromide	Br^-
Aluminum ion	Al^{3+}	Oxide	O^{2-} The process in which an atom becomes an ion is called ionization. It may occur when atoms are exposed to high levels of radiation. The radiation may give their outer electrons enough energy to escape from the attraction of the positive nucleus. However, most ions form when atoms transfer electrons to or from other atoms or

How Ions Form molecules.

For example, sodium atoms may transfer electrons to chlorine atoms. This forms positive sodium ions (Na^+) and negative chloride ions (Cl^-).



? Exercise 11.4.1

Why do you think atoms lose electrons to, or gain electrons from, other atoms?

Answer

Atoms form ions by losing or gaining electrons because it makes them more stable and this state takes less energy to maintain. The most stable state for an atom is to have its outermost energy level filled with the maximum possible number of electrons. In the case of metals such as lithium, with just one electron in the outermost energy level, a more stable state

> can be achieved by losing that one outer electron. In the case of nonmetals such as fluorine, which has seven electrons in the outermost energy level, a more stable state can be achieved by gaining one electron and filling up the outer energy level. These are two applications of the Octet Rule which we introduced earlier in this chapter.

Based on the Octet Rule, we can make some predictions for the charges that will form for elements based on their position on the periodic table. These are listed in Figure **11.4.2**. You might notice some variability towards the bottom of the periodic table for some of these elements. These are situations where the Octet Rule is not followed and we will ignore them for this class. Look at the third or fourth period and make sure you feel comfortable predicting the charge of these elements based on their dot structure. Elements following the Octet Rule will gain or lose electrons to become isoelectronic with a noble gas. This means that they have the same number of electrons.

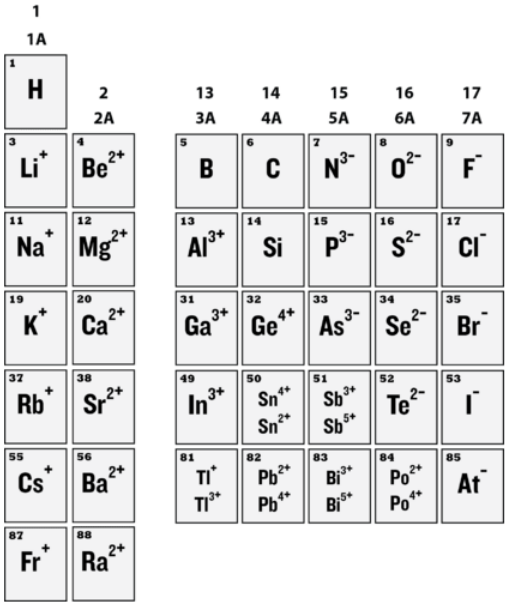


Figure **11.4.2**: Ion charges. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Properties of Ions

Ions are highly reactive, especially as gases. They usually react with ions of opposite charge to form neutral compounds. For example, positive sodium ions and negative chloride ions react to form the neutral compound sodium chloride, commonly known as table salt. This occurs because oppositely charged ions attract each other. Ions with the same charge, on the other hand, repel each other. Ions are also deflected by a magnetic field, as you saw in the opening image of the northern lights.

Section Summary

- Atoms have equal numbers of positive protons and negative electrons, so they are neutral in electric charge.
- Atoms can gain or lose electrons and become ions, which are atoms that have a positive or negative charge because they have unequal numbers of protons and electrons.
- Anions are negative ions formed by accepting electrons.
- Cations form when an atom loses one or more electrons.
- The outermost principal energy level of an anion is usually an octet.
- The process in which an atom becomes an ion is called ionization. It may occur when atoms are exposed to high levels of radiation or when atoms transfer electrons to or from other atoms.
- Ions are reactive, attracted or repulsed by other charged particles, and deflected by a magnetic field.

Glossary

anions

negative ions that form when atoms gain electrons.

cations

positive ions that form when atoms lose electrons.

ions

atoms that have a positive or negative charge because they have unequal numbers of protons and electrons.

This page titled 11.4: Ions and Ion Formation is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **1.3: Vectors, Scalars, and Coordinate Systems** by OpenStax is licensed CC BY 4.0.
- **8.4: Anion Formation** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

11.5: Ionic Bonding

Learning Objectives

- Predict the formulas of the ionic compounds that form according to the Octet Rule.
- Describe the structure of ionic compounds.



Figure **11.5.1** (Credit: Leon Brooks; Source: Commons Wikimedia, Waneroo Beach(opens in new window) [commons.wikimedia.org]; License: Public Domain)

Does the sea really have salt in it?

We can get common table salt from several sources. It can be mined in the solid form in salt mines, or found as a solid in deposits. We can also get salt from the ocean, but it really does not exist as a salt when in solution. The sodium ions and chloride ions are dissolved, but not combined into a structure until all the water is removed.

Most of the rocks and minerals that make up the Earth's crust are composed of positive and negative ions held together by ionic bonding. An ionic compound is an electrically neutral compound consisting of positive and negative ions. You are very familiar with some ionic compounds, such as sodium chloride (NaCl). A sodium chloride crystal consists of equal numbers of positive sodium ions ($\mathrm{Na^+}$) and negative chloride ions ($\mathrm{Cl^-}$).

Ionic Bonds

Oppositely charged particles attract each other. This attractive force is often referred to as an **electrostatic force**. An **ionic bond** is the electrostatic force that holds ions together in an **ionic compound**. The strength of the ionic bond is directly dependent upon the quantity of the charges and inversely dependent on the distance between the charged particles. A cation with a $2+$ charge will make a stronger ionic bond than a cation with a $1+$ charge. A larger ion makes a weaker ionic bond because of the greater distance between its electrons and the nucleus of the oppositely charged ion.

Electron Dot Diagrams

We will use sodium chloride as an example to demonstrate the nature of the ionic bond and how it forms. As you know, sodium is a metal and loses its one valence electron to become a cation. Chlorine is a nonmetal and gains one electron in becoming an anion. Both achieve a noble-gas electron configuration. However, electrons cannot be simply "lost" to nowhere in particular. A more accurate way to describe what is happening is that a single electron is transferred from the sodium atom to the chlorine atom, as shown below.

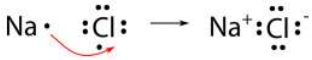


Figure **11.5.2** (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

The ionic bond is the attraction of the $\mathrm{Na^+}$ ion for the $\mathrm{Cl^-}$ ion. It is conventional to show the cation without dots around the symbol to emphasize that the original energy level that contained the valence electron is now empty. The anion is now shown with a complete octet of electrons.

For a compound such as magnesium chloride, it is not quite as simple. Because magnesium has two valence electrons, it needs to lose both to achieve the noble-gas configuration. Therefore, two chlorine atoms will be needed.

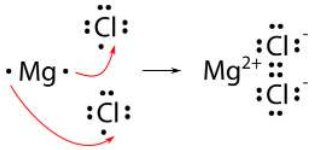


Figure **11.5.3** (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

The final formula for magnesium chloride is $MgCl_2$.

Ionic Crystal Structure

Electron dot diagrams show the nature of the electron transfer that takes place between metal and nonmetal atoms. However, ionic compounds do not exist as discrete molecules, as the dot diagrams may suggest. In order to minimize the potential energy of the system, ionic compounds take on the form of an extended three-dimensional array of alternating cations and anions. This maximizes the attractive forces between the oppositely charged ions. The figure below shows two different ways of representing the ionic crystal lattice. A ball and stick model makes it easier to see how individual ions are oriented with respect to one another. A space filling diagram is a more accurate representation of how the ions pack together in the **crystal**.

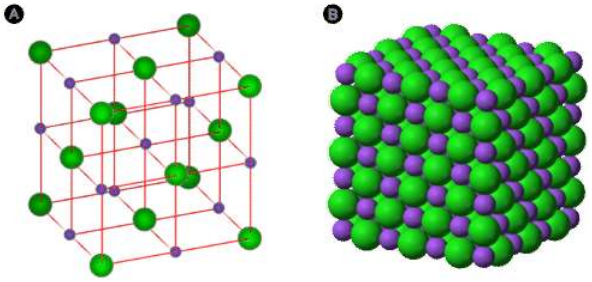


Figure **11.5.2**: Two models of a sodium chloride crystal are shown. The purple spheres represent the Na^+ ions, while the green spheres represent the Cl^- ions. (A) In an expanded view, the distances between ions are exaggerated, more easily showing the coordination numbers of each ion. (B) In a space filling model, the electron clouds of the ions are in contact with each other. (Credit: (A) Eloy; (B) Ben Mills (Wikimedia: Benjah-bmm27); Source: (A) wikimedia; (B) wikimedia; License: Public Domain)

Naturally occurring sodium chloride **(halite)** does not look at first glance like the neat diagrams shown above. It is only when we use modern techniques to analyze the crystal structure at the atomic level that we can see the true regularity of the organized ions.



Figure **11.5.3**: Halite crystals. (Credit: Ingo Wölbern (Wikimedia: Iwoelbern); Source: wikimedia [commons.wikimedia.org]; License: Public Domain)

Section Summary

- An ionic compound contains positive and negative ions.
- An ionic bond is electrostatic in nature.
- Electron dot diagrams can be used to illustrate electron movements and ion formation.
- Ionic compounds take on the form of extended three-dimensional arrays of cations and anions.
- The arrangement maximizes the attractive force between oppositely-charged ions.

Glossary

ionic bond

the electrostatic force that holds ions together in an ionic compound.

ionic compound

a compound formed by the attraction between ions.

crystal

a structure of matter that is an extended three-dimensional array of repeating patterns of particles that make up that form of matter.

This page titled 11.5: Ionic Bonding is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **8.6: Ionic Bonding** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **8.7: Ionic Crystal Structure** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

11.6: Metallic Bonding

Learning Objectives

- Explain why metallic bonding occurs.
- Describe the properties of metals.
- Explain what an alloy is composed of.



Figure 11.6.1: A nineteenth century copper plate. (Credit: John Pearson; Source: wikimedia; License: Public Domain)

Why do metals behave the way they do?

The image below is of a copper plate that was made in 1893. The utensil has a great deal of elaborate decoration, and the item is very useful. What would have happened if this plate was made of copper (I) chloride instead? Copper (I) chloride does contain copper, after all. However, the CuCl would end up as a powder when a metalworker pounded on it to shape it. Metals behave in unique ways. The bonding that occurs in a metal is responsible for its distinctive properties: luster, malleability, ductility, and excellent conductivity.

The Metallic Bond

Pure metals are crystalline solids, but unlike ionic compounds, every point in the crystal lattice is occupied by an identical atom. The electrons in the outer energy levels of a metal are mobile and capable of drifting from one metal atom to another. This means that the metal is more properly viewed as an array of positive ions surrounded by a sea of mobile valence electrons. Electrons which are capable of moving freely throughout the empty orbitals of the metallic crystal are called **delocalized electrons** (see below). A **metallic bond** is the attraction of the stationary metal cations to the surrounding mobile electrons.

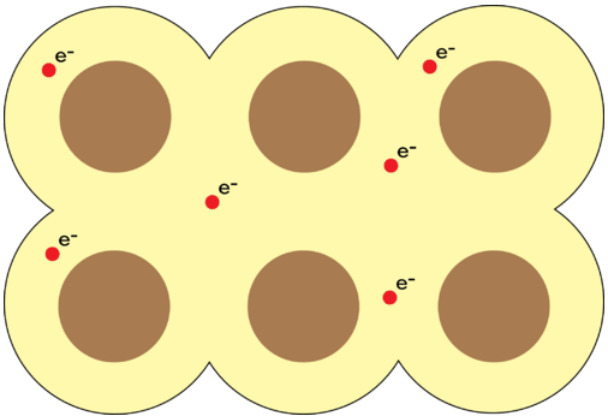


Figure 11.6.2: In a metal, the stationary metal cations are surrounded by a sea of mobile valence electrons that are not associated with any one cation. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Properties of Metals

The metallic bonding model explains the physical properties of metals. Metals conduct electricity and heat very well because of their free-flowing electrons. As electrons enter one end of a piece of metal, an equal number of electrons flow outward from the other end. When light is shone onto the surface of a metal, its electrons absorb small amounts of energy and become excited into one of its many empty orbitals. The electrons immediately fall back down to lower energy levels and emit light. This process is

responsible for the high **luster** of metals. Recall that ionic compounds are very brittle. Application of a force results in like-charged ions in the crystal coming too close to one another, causing the crystal to shatter. When a force is applied to a metal, the free-flowing electrons can slip in between the stationary cations and prevent them from coming in contact. Imagine ball bearings that have been coated with oil sliding past one another. As a result, metals are very **malleable** and **ductile**. They can be hammered into shapes, rolled into thin sheets, or pulled into thin wires.

Alloys

An **alloy** is a mixture composed of two or more elements, at least one of which is a metal. You are probably familiar with some alloys such as brass and bronze. **Brass** is an alloy of copper and zinc. **Bronze** is an alloy of copper and tin. Alloys are commonly used in manufactured items because the properties of these metal mixtures are often superior to a pure metal. Bronze is harder than copper and more easily cast. Brass is very malleable and its acoustic properties make it useful for musical instruments.



Figure **11.6.3**: Bronze, an alloy of copper and tin, has been in use since ancient times. The Bronze Age saw the increased use of metals rather than stone for weapons, tools, and decorative objects. Brass, an alloy of copper and zinc, is widely used in musical instruments like the trumpet and trombone. (Credit: (A) Davide Ferro; (B) Guillaume Piolle; Source: (A) wikimedia (B) wikimedia; License: CC by 2.0(opens in new window))

Steels are a very important class of alloys. The many types of steels are primarily composed of iron, with various amounts of the elements carbon, chromium, manganese, nickel, molybdenum, and boron. Steels are widely used in building construction because of their strength, hardness, and resistance to corrosion. Most large modern structures like skyscrapers and stadiums are supported by a steel skeleton.

Section Summary

- The metallic bond is responsible for the properties of metals.
- Metals conduct electricity and heat well.
- Metals are ductile and malleable.
- Metals have luster.
- Alloys are mixtures of materials, at least one of which is a metal.

Glossary

malleable
substance that is easily shaped.

luster
property of having a shiny surface.

delocalized electrons
electrons which are capable of moving freely throughout the empty orbitals of a chemically bonded substance.

metallic bond
attraction of the stationary metal cations to the surrounding mobile electrons.

alloy

a mixture composed of two or more elements, at least one of which is a metal.

This page titled 11.6: Metallic Bonding is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **8.10: Metallic Bonding** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **8.12: Alloys** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

11.7: Covalent Bonds and Molecules

Learning Objectives

- Explain how covalent bonds form by sharing electrons.
- Describe the characteristic properties of molecules.

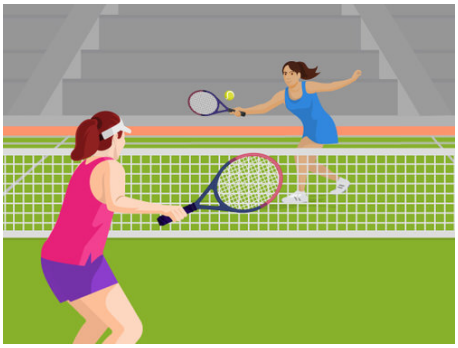


Figure **11.7.1** (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CK-12 Curriculum Materials license)

In a tennis match, two players keep hitting the ball back and forth. The ball bounces from one player to the other, over and over again. The ball keeps the players moving together on the court. What if the two players represented the nuclei of two atoms and the ball represented valence electrons? What would the back and forth movement of the ball represent? The answer is a covalent bond.

Covalent Bonds: Sharing Electrons

A covalent bond is the force of attraction that holds together two atoms that share a pair of valence electrons. The shared electrons are attracted to the nuclei of both atoms. This forms a molecule consisting of two or more atoms. Covalent bonds form only between atoms of nonmetals.

The two atoms that are held together by a covalent bond may be atoms of the same element or different elements. When atoms of different elements form covalent bonds, a new substance, called a covalent compound, results. All covalent compounds are also molecules. Water is an example of a covalent compound. A water molecule is modeled in the figure below. A molecule is the smallest particle of a covalent compound that still has the properties of the compound.

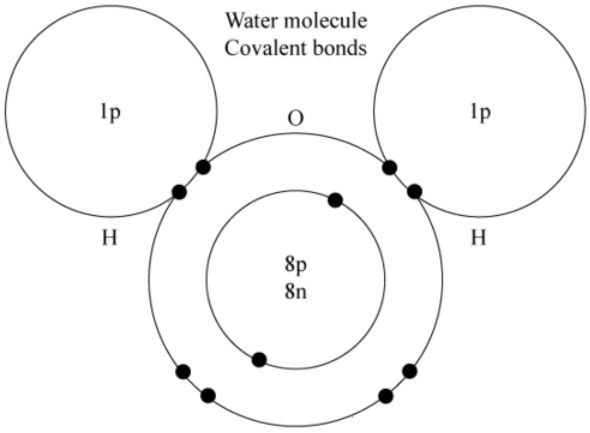


Figure **11.7.2** (Credit: Zachary Wilson; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Exercise 11.7.1

How many valence electrons does the oxygen atom (O) share with each hydrogen atom (H)? How many covalent bonds hold the water molecule together? (Refer to Figure **11.7.2**)

Answer

The oxygen atom shares one pair of valence electrons with each hydrogen atom. Each pair of shared electrons represents one covalent bond, so two covalent bonds hold the water molecule together.

The diagram in the figure below shows an example of covalent bonds between two atoms of the same element, in this case two atoms of oxygen. The diagram represents an oxygen molecule, so it's not a new compound. Oxygen normally occurs in diatomic ("two-atom") molecules. Several other elements also occur as diatomic molecules: hydrogen, nitrogen, and all but one of the halogens (fluorine, chlorine, bromine, and iodine).

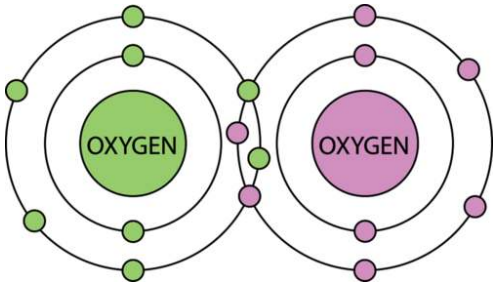


Figure **11.7.3** (Credit: Christopher AuYeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

? Exercise 11.7.2

How many electrons do these two oxygen atoms share? How many covalent bonds hold the oxygen molecule together?(Refer to Figure **11.7.3**)

Answer

The two oxygen atoms share two pairs of electrons, so two covalent bonds hold the oxygen molecule together.

Why Covalent Bonds Form

Covalent bonds form because they give atoms a more stable arrangement of electrons. Nonmetals in molecules follow the Octet Rule by sharing electrons in the formation of a covalent bond in such a way that each of the atoms involved in the bond can attain a noble-gas electron configuration. The shared electrons are "counted" for each of the atoms involved in the sharing. For hydrogen (H2)(H2), the shared pair of electrons means that each of the atoms is able to attain the electron configuration of helium, the noble gas with two electrons. For atoms other than hydrogen, the sharing of electrons will usually provide each of the atoms with eight valence electrons

Multiple Covalent Bonds

Some molecules are not able to satisfy the octet rule by making only single covalent bonds between the atoms. Consider the compound ethene, which has a molecular formula of $\mathbf{C_2H_4}$. The carbon atoms are bonded together, with each carbon also bonded to two hydrogen atoms.

$$\text{two } \mathbf{C} \text{ atoms } = 2 \times 4 = 8 \text{ valence electrons}$$

$$\text{four } \mathbf{H} \text{ atoms } = 4 \times 1 = 4 \text{ valence electrons}$$

$$= \text{total of 12 valence electrons in the molecule}$$

If the Lewis electron dot structure was drawn with a single bond between the carbon atoms and with the octet rule followed, it would look like this:

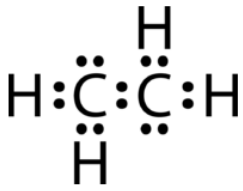


Figure **11.7.2**: Incorrect dot structure of ethene. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

This Lewis structure is incorrect because it contains a total of 14 electrons. However, the Lewis structure can be changed by eliminating the lone pairs on the carbon atoms and having the carbon atoms share two pairs, instead of only one pair, of electrons.

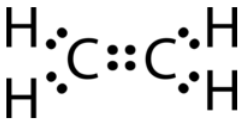


Figure **11.7.3**: Correct dot structure for ethene. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

A **double covalent bond** is a covalent bond formed by atoms that share two pairs of electrons. The double covalent bond that occurs between the two carbon atoms in ethane can also be represented by a structural formula and with a molecular model, as shown in the figure below.

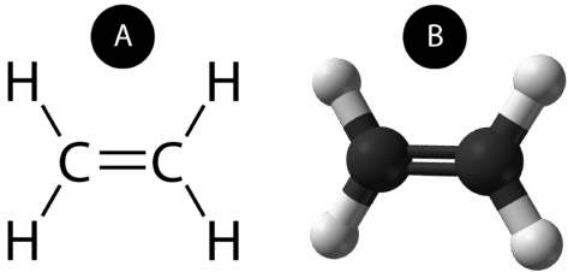


Figure **11.7.4**: (A) The structural model for C_2H_4 consists of a double covalent bond between the two carbon atoms and single bonds to the hydrogen atoms. (B) Molecular model of C_2H_4. (Credit: Ben Mills (Wikimedia: Benjah-bmm27); Source: (A) wikimedia; (B) wikimedia; License: Public Domain)

A **triple covalent bond** is a covalent bond formed by atoms that share three pairs of electrons. The element nitrogen is a gas that composes the majority of Earth's atmosphere. A nitrogen atom has five valence electrons, which can be shown as one pair and three single electrons. When combining with another nitrogen atom to form a diatomic molecule, the three single electrons on each atom combine to form three shared pairs of electrons.

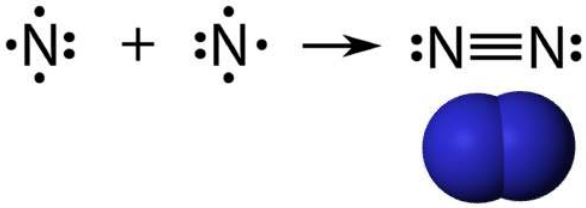


Figure **11.7.5**: Triple bond in N_2. (Credit: CK-12 Foundation - Joy Sheng, using 3D molecular structure by Ben Mills (Wikimedia: Benjah-bmm27); Source: Molecular structure: wikimedia ; License: CC BY-NC-SA 3.0(opens in new window); Molecule Public Domain)

Each nitrogen atom follows the octet rule with one lone pair of electrons, and six electrons that are shared between the atoms.

Properties of Molecules

The covalent bonds of molecules are responsible for many of their properties, regardless of whether they are elements or compounds. Because valence electrons are shared in molecules, rather than transferred between atoms as they are in ionic compounds, molecules have very different properties than ionic compounds.

- Many covalent compounds, especially those containing carbon and hydrogen, burn easily. In contrast, many ionic compounds do not burn.
- Many covalent compounds do not dissolve in water, whereas most ionic compounds dissolve well in water.
- Unlike ionic compounds, covalent compounds do not have freely moving electrons, so they cannot conduct electricity.
- The individual molecules of covalent compounds are more easily separated than the ions in a crystal, so most covalent compounds have relatively low boiling points. This explains why many of them are liquids or gases at room temperature.

Section Summary

- A covalent bond is the force of attraction that holds together two atoms that share a pair of valence electrons. Covalent bonds form only between atoms of nonmetals.
- The two atoms that are held together in a covalent bond may be atoms of the same element or different elements. When atoms of different elements bond together, it forms a covalent compound.
- Covalent bonds form because the shared electrons fill each atom's outer energy level and this is the most stable arrangement of electrons.
- Lewis electron-dot structures show the bonding in covalent molecules.
- Covalent compounds contain two or more nonmetallic elements held together by covalent bonds, in which atoms share pairs of valence electrons. A molecule is the smallest particle of a covalent compound that still has the properties of the compound.
- Covalent bonds are responsible for many of the properties of covalent compounds. Covalent compounds have relatively low boiling points, cannot conduct electricity, and may not dissolve in water.
- Covalent bonds between atoms can be indicated either with dots (:) or a dash (−).
- Lewis structures can be drawn for molecules that share multiple pairs of electrons.
- A double covalent bond is a covalent bond formed by atoms that share two pairs of electrons.
- A triple covalent bond is a covalent bond formed by atoms that share three pairs of electrons.

This page titled 11.7: Covalent Bonds and Molecules is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.
- **9.7: Multiple Covalent Bonds** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

11.8: Bond Polarity

Learning Objectives

- Explain the difference between nonpolar covalent and polar covalent bonds.
- Explain how bond polarity can give rise to molecular polarity.



Figure **11.8.1** (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CK-12 Curriculum Materials license)

What makes people share?

Have you ever spent time with someone you really didn't like? You had nothing in common with them and did not want to have anything to do with them. On the other hand, there are people that you enjoy being with. You have a lot in common and like to share with them. Atoms work the same way. If there are strong differences in their attraction of electrons, one atom gets the electrons and the other atom loses them. If they are similar, they share electrons to form a covalent bond.

Bond Polarity

Electronegativity is defined as the ability of an atom to attract electrons when the atoms are in a compound. Electronegativities of elements are shown in the periodic table below.

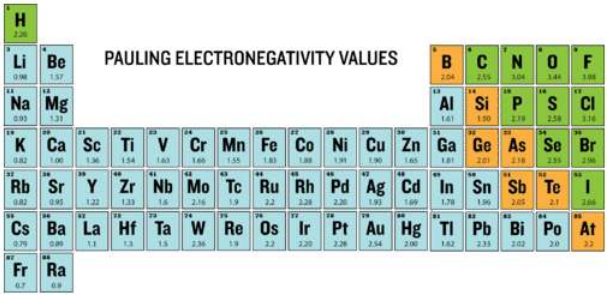


Figure **11.8.2**: Electronegativities of elements. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

The degree to which a given bond is ionic or covalent is determined by calculating the difference in electronegativity between the two atoms involved in the bond. We won't perform those calculations in this course, but we can discuss the meaning of them.

Nonpolar Covalent Bonds

A bond in which the electronegativity difference is small is considered to be mostly covalent in character. However, at this point we need to distinguish between two general types of covalent bonds. A **nonpolar covalent bond** is a covalent bond in which the bonding electrons are shared equally between the two atoms. In a nonpolar covalent bond, the distribution of electrical charge is balanced between the two atoms.

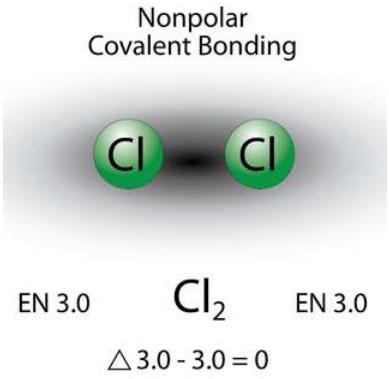


Figure **11.8.3**: A nonpolar covalent bond is one in which the distribution of electron density between the two atoms is equal. (Credit: Jodi So; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

The two chlorine atoms share the pair of electrons in the single covalent bond equally, and the electron density surrounding the Cl_2 molecule is symmetrical.

Polar Covalent Bonds

A bond in which the electronegativity difference between the atoms is between 0.4 and 1.7 is called a polar covalent bond. A **polar covalent bond** is a covalent bond in which the atoms have an unequal attraction for electrons, and so the sharing is unequal. In a polar covalent bond, sometimes simply called a polar bond, the distribution of electrons around the molecule is no longer symmetrical.

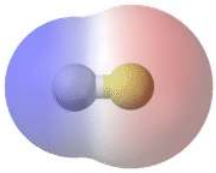


Figure **11.8.4**: In the polar covalent bond of **HF**, the electron density is unevenly distributed. There is a higher density (red) near the fluorine atom, and a lower density (blue) near the hydrogen atom. (Credit: Ben Mills (Wikimedia: Benjah-bmm27); Source: Commons Wikimedia, Hydrogen fluoride elpot transparent 3D balls (opens in new window) [commons.wikimedia.org]; License: Public Domain)

An easy way to illustrate the uneven electron distribution in a polar covalent bond is to use the Greek letter delta (δ).

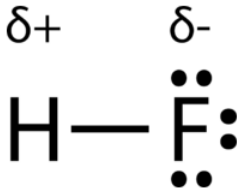


Figure **11.8.5**: Use of δ to indicate partial charge. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

The atom with the greater electronegativity acquires a partial negative charge, while the atom with the lesser electronegativity acquires a partial positive charge. The delta symbol is used to indicate that the quantity of charge is less than one. A crossed arrow can also be used to indicate the direction of greater electron density.

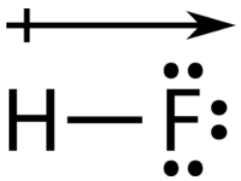


Figure **11.8.6**: Use of crossed arrow to indicate polarity. (Credit: Joy Sheng
Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Polar Molecules

A **polar molecule** is a molecule in which one end of the molecule is slightly positive, while the other end is slightly negative. A diatomic molecule that consists of a polar covalent bond, such as **HF**, is a polar molecule. The two electrically charged regions on either end of the molecule are called poles, similar to a magnet having a north and a south pole. A molecule with two poles is called a **dipole** (see figure below). Hydrogen fluoride is a dipole.



Figure 11.8.2: A dipole is any molecule with a positive end and a negative end, resulting from unequal distribution of electron density throughout the molecule. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Polar molecules orient themselves in the presence of an electric field with the positive ends of the molecule being attracted to the negative plate, while the negative ends of the molecules are attracted to the positive plate (see figure below).

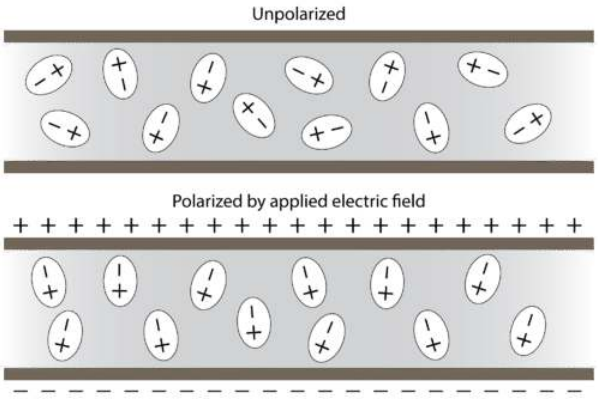


Figure 11.8.3: Polar molecules are randomly oriented in the absence of an applied electric field (top). In an electric field, the molecules orient themselves because of the attraction of opposite charges (bottom). (Credit: Zachary Wilson; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

For molecules with more than two atoms, the molecular geometry must also be taken into account when determining if the molecule is polar or nonpolar. The figure below shows a comparison between carbon dioxide and water. Carbon dioxide (CO_2) is a linear molecule. The oxygen atoms are more electronegative than the carbon atom, so there are two individual dipoles pointing outward from the C atom to each O atom. However, since the dipoles are of equal strength and are oriented this way, they cancel out, and the overall molecular polarity of CO_2 is zero.

Water is a bent molecule because of the two lone pairs on the central oxygen atom. The individual dipoles point from the H atoms toward the O atom. Because of the shape, the dipoles do not cancel each other out, and the water molecule is polar. In the figure below, the net dipole is shown in blue and points upward.

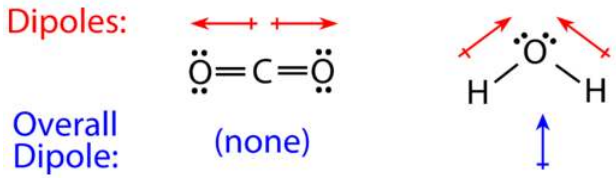


Figure 11.8.4: The molecular geometry of a molecule affects its polarity. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Some other molecules are shown in the figure below. Notice that a tetrahedral molecule such as CH_4 is nonpolar. However, if one of the peripheral H atoms is replaced with another atom that has a different electronegativity, the molecule becomes polar. A trigonal planar molecule (BF_3) may be nonpolar if all three peripheral atoms are the same, but a trigonal pyramidal molecule (NH_3) is polar.

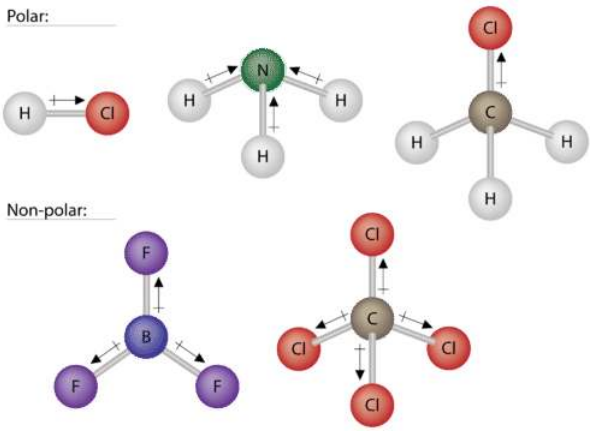


Figure 11.8.5: Some examples of polar and nonpolar molecules based on molecular geometry. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Section Summary

- The electronegativity of an atom determines how strongly it attracts electrons to itself.
- The polarity of a bond is affected by the electronegativity values of the two atoms involved in that bond.
- Polar molecules result from differences in the electronegativity of the atoms in the molecule.
- Dipoles that are directly opposite one another cancel each other out.

Glossary

electronegativity

The ability of an atom to attract electrons when the atoms are in a compound.

nonpolar covalent bond

A covalent bond in which the bonding electrons are shared equally between the two atoms.

polar covalent bond

A covalent bond in which the atoms have an unequal attraction for electrons, and so the sharing is unequal.

polar molecule

A molecule in which one end of the molecule is slightly positive, while the other end is slightly negative.

dipole

A molecule with two poles.

This page titled 11.8: Bond Polarity is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **9.16: Bond Polarity** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **9.17: Polar Molecules** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

11.9: Intermolecular Forces

Learning Objectives

- Explain the difference between dipole-dipole forces, dispersion forces, and hydrogen bonds

Molecules are attracted to each other, but not as strongly as a chemical bond such as an ionic bond. However these attractions are important as we will see. In this section we will discuss some different forms of these attractions, which are called **intermolecular forces.**

Dipole-Dipole Forces

Dipole-dipole forces are the attractive forces that occur between polar molecules. A molecule of hydrogen chloride has a partially positive hydrogen atom and a partially negative chlorine atom. In a collection of many hydrogen chloride molecules, the molecules will align themselves so that the oppositely charged regions of neighboring molecules are near each other.

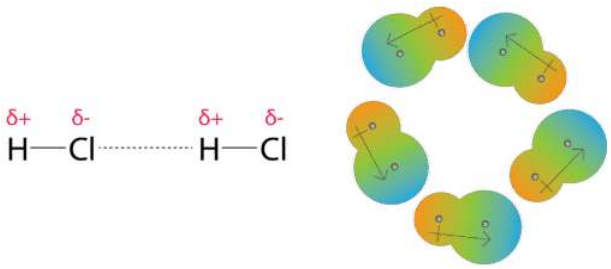


Figure **11.9.1**: Dipole-dipole forces are a result of the attraction of the positive end of one dipole to the negative end of a neighboring dipole. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Dipole-dipole forces are similar in nature to ionic bonds, but much weaker.

London Dispersion Forces

Dispersion forces are also considered a type of van der Waals force and are the weakest of all intermolecular forces. They are often called London dispersion forces after Fritz London (1900-1954), who first proposed their existence in 1930. **London dispersion forces** are the intermolecular forces that occur between atoms, and between nonpolar molecules as a result of the motion of electrons.

The electron cloud of a helium atom contains two electrons, which can normally be expected to be equally distributed spatially around the nucleus. However, at any given moment the electron distribution may be uneven, resulting in an instantaneous dipole. This weak and temporary dipole subsequently influences neighboring helium atoms through electrostatic attraction and repulsion. It induces a dipole on nearby helium atoms (see figure below).

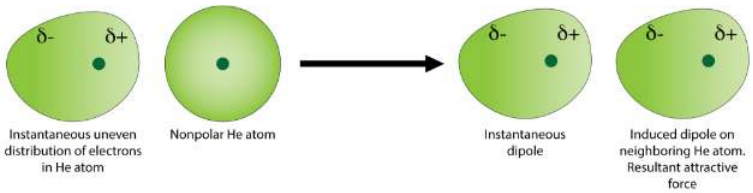


Figure **11.9.3**: A short-lived or instantaneous dipole in a helium atom. (Credit: Zachary Wilson; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

The instantaneous and induced dipoles are weakly attracted to one another. The strength of dispersion forces increases as the number of electrons in the atoms or nonpolar molecules increases.

The halogen group consists of four elements that all take the form of nonpolar diatomic molecules. The table below shows a comparison of the melting and boiling points for each.

Melting and Boiling Points of Halogens

Table 11.9.1: Melting and Boiling Points of Halogens

Molecule	Total Number of Electrons	Melting Point (°C)	Boiling Point (°C)	Physical State at Room Temperature
F_2	18	-220	-188	gas
Cl_2	34	-102	-34	gas
Br_2	70	-7	59	liquid
I_2	106	114	184	solid

The dispersion forces are strongest for iodine molecules because they have the greatest number of electrons. The relatively stronger forces result in melting and boiling points that are the highest of the halogen group. These forces are strong enough to hold iodine molecules close together in the solid state at room temperature. The dispersion forces are progressively weaker for bromine, chloride, and fluorine; this is illustrated in their steadily lower melting and boiling points. Bromine is a liquid at room temperature, while chlorine and fluorine are gases whose molecules are much further apart from one another. Intermolecular forces are nearly nonexistent in the gas state, and so the dispersion forces in chlorine and fluorine only become measurable as the temperature decreases and they condense into the liquid state.

Hydrogen Bonding

The attractive force between water molecules is a dipole interaction. The hydrogen atoms are bound to the highly electronegative oxygen atom (which also possesses two lone pair sets of electrons, making for a very polar bond). The partially positive hydrogen atom of one molecule is then attracted to the oxygen atom of a nearby water molecule (see figure below).

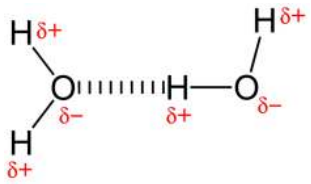


Figure **11.9.3**: A hydrogen bond in water occurs between the hydrogen atom of one water molecule and the lone pair of electrons on an oxygen atom of a neighboring water molecule. (Credit: Ben Mills (Wikimedia: Benjah-bmm27); Source: wikimedia; License: Public Domain)

A **hydrogen bond** is an intermolecular attractive force in which a hydrogen atom that is covalently bonded to a small, highly electronegative atom is attracted to a lone pair of electrons on an atom in a neighboring molecule. Hydrogen bonds are very strong compared to other dipole interactions. The strength of a typical hydrogen bond is about **5%** of that of a covalent bond.

Hydrogen bonding occurs only in molecules where hydrogen is covalently bonded to one of three elements: fluorine, oxygen, or nitrogen. These three elements are so electronegative that they withdraw the majority of the electron density in the covalent bond with hydrogen, leaving the **H** atom very electron-deficient. The **H** atom nearly acts as a bare proton, leaving it very attracted to lone pair electrons on a nearby atom.

The hydrogen bonding that occurs in water leads to some unusual, but very important, properties. Most molecular compounds that have a mass similar to water are gases at room temperature. Because of the strong hydrogen bonds, water molecules are able to stay condensed in the liquid state. The figure below shows how the bent shape, and two hydrogen atoms per molecule, allows each water molecule to be able to hydrogen bond to two other molecules.

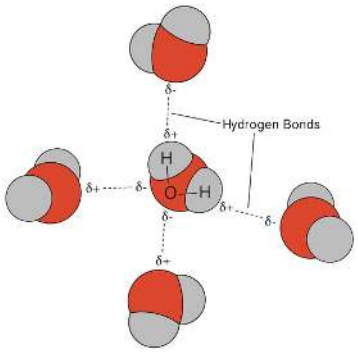


Figure 11.9.4: Multiple hydrogen bonds occur simultaneously in water because of its bent shape and the presence of two hydrogen atoms per molecule. (Credit: Laura Guerin; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

In the liquid state, the hydrogen bonds of water can break and reform as the molecules flow from one place to another. When water is cooled, the molecules begin to slow down. Eventually, when water is frozen to ice, the hydrogen bonds become permanent and form a very specific network (see figure below).

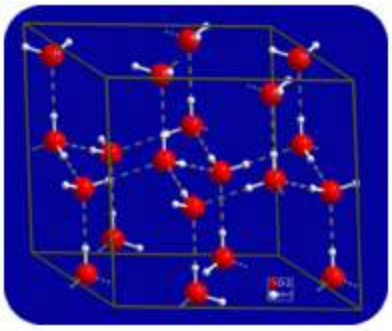


Figure 11.9.5: When water freezes to ice, the hydrogen bonding network becomes permanent. Each oxygen atom has an approximately tetrahedral geometry—two covalent bonds and two hydrogen bonds. (Credit: User:Materialscientist/Wikimedia Commons; Source: wikimedia; License: Public Domain)

The bent shape of the molecules leads to gaps in the hydrogen bonding network of ice. Ice has a very unusual property—its solid state is less dense than its liquid state. Ice floats in water. Virtually all other substances are denser in the solid state than in the liquid state. Hydrogen bonds play a very important biological role in the physical structures of proteins and nucleic acids.

Section Summary

- Van der Waals forces are weak interactions between molecules that involve dipoles.
- Polar molecules have permanent dipole-dipole interactions.
- Nonpolar molecules can interact by way of London dispersion forces.
- Hydrogen bonds form when a **H** attached to a **N**, **O**, or **F** atom interacts with another **N**, **O**, or **F** atom.

Glossary

dipole-dipole forces

The attractive forces that occur between polar molecules.

dispersion forces

The intermolecular forces that occur between atoms, and between nonpolar molecules as a result of the motion of electrons.

hydrogen bond

an intermolecular attractive force in which a hydrogen atom that is covalently bonded to a small, highly electronegative atom is attracted to a lone pair of electrons on an atom in a neighboring molecule.

intermolecular forces

Attractions between molecules which are not as strong as a chemical bond.

This page titled 11.9: Intermolecular Forces is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **9.18: Van der Waals Forces** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **9.19: Hydrogen Bonding** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.

11.10: Naming Compounds

Learning Objectives

- Correctly name a covalent compound.

We have discussed many kinds of compounds and the bonds that hold them together. The rules for naming ionic compounds are a bit involved, so we won't cover them in this text. Instead we will focus only on the methods to name covalent compounds.

To name simple covalent compounds, follow these rules:

- Start with the name of the element closer to the left side of the periodic table.
- Follow this with the name of element closer to the right of the periodic table. Give this second name the suffix *–ide*.
- Use prefixes to represent the numbers of the different atoms in each molecule of the compound. The most commonly used prefixes are shown in the Table below.

Table **11.10.1**

Naming and Writing Formulas for Covalent Compounds

Number	Prefix
1	*mono-* (or none)
2	*di-*
3	*tri-*
4	*tetra-*
5	*penta-*
6	*hexa-*

Exercise 11.10.1

What is the name of the compound that contains three oxygen atoms and two nitrogen atoms?

Answer

The compound is named dinitrogen trioxide. Nitrogen is named first because it is farther to the left in the periodic table than oxygen. Oxygen is given the *-ide* suffix because it is the second element named in the compound. The prefix *di-* is added to nitrogen to show that there are two atoms of nitrogen in each molecule of the compound. The prefix *tri-* is added to oxygen to show that there are three atoms of oxygen in each molecule.

In the chemical formula for a covalent compound, the numbers of the different atoms in a molecule are represented by subscripts. For example, the formula for the compound named carbon dioxide is CO_2.

Exercise 11.10.2

What is the chemical formula for dinitrogen trioxide?

Answer

The chemical formula is N_2O_3.

Section Summary

- A few rules can be applied to name simple covalent compounds: name the left-most element in the periodic table first, add *-ide* to the second named element, and use prefixes for the numbers of atoms.

This page titled 11.10: Naming Compounds is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **1.3: Vectors, Scalars, and Coordinate Systems** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

12: Physical Properties

12.1: Introduction to Physical Properties
12.2: Physical and Chemical
12.3: Classifying Matter
12.4: The Physical States of Matter
12.5: Transitions between Phases
12.6: Energy and Phase Transitions
12.7: Moles- Counting Atoms
12.7.1: Moles and Avogrado's Number
12.7.2: Moles and Mass
12.8: Solutions and Solubility
12.9: Concentrations of Solutions
12.10: Soap- an Application of Physical Properties

This page titled 12: Physical Properties is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

12.1: Introduction to Physical Properties



Figure **12.1.1** (Credit: JillWellington; Source: pixabay; License: Pixabay License)

Ahhhh! A tall glass of ice-cold lemonade is really refreshing on a hot day. Lemonade is a combination of lemon juice, water, and sugar. Do you know what kind of matter lemonade is? It is obviously not an element because it consists of more than one substance. Is it a compound? Not all combined substances are compounds. Think about everything we have learned regarding chemical bonding. There were specific amounts of each element necessary in each chemical bond in order for the octet rule to work. But we know that is not how it works for something like lemonade. When you ask someone how they would like their lemonade, they will often tell you how much sugar they want along with it. There is something different going on with lemonade, and in fact with most of the matter we interact with on a daily basis. Lemonade is what is called a mixture.

In this chapter we will learn about properties of matter that are unrelated to chemical reactions. We can come up with some categories for matter. We will discuss keeping track of how much matter we have in both mixtures and pure substances. We will talk about phase changes, some types of mixtures and their properties, and how many of these properties show up in things we do on a day to day basis.

This page titled 12.1: Introduction to Physical Properties is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Physical and Chemical

Learning Objectives

- Recognize examples of physical properties of a substance.
- Recognize examples of physical changes a substance might undergo.

In this section we will more thoroughly distinguish between physical and chemical in terms of both properties and changes. We will explore those physical properties and changes in more detail in this chapter, and then we will look at the chemical properties and changes in the next chapter.

Physical Properties

A **physical property** is a characteristic of a substance that can be observed or measured without changing the identity of the substance. Silver is a shiny metal that conducts electricity very well. It can be molded into thin sheets—a property called malleability. Salt is dull, brittle, and conducts electricity when it has been dissolved into water (which it does quite easily). Physical properties of matter include color, hardness, malleability, solubility, electrical conductivity, density, **melting point**, and **boiling point**.

For the elements, color does not vary much from one element to the next. The vast majority of elements are colorless, silver, or gray. Some elements do have distinctive colors: sulfur and chlorine are yellow, copper is (of course) copper-colored, and elemental bromine is red.

Density can be a very useful parameter for identifying an element. Of the materials that exist as solids at room temperature, iodine has a very low density compared to zinc, chromium, and tin. Gold has a very high density, as does platinum.

Hardness is a measure of a material's ability to resist plastic deformation (i.e.- scratching or denting the material). Hardness helps determine how an element (especially a metal) might be used. Many elements are fairly soft (silver and gold, for example) while others (such as titanium, tungsten, and chromium) are much harder. Carbon is an interesting example of hardness. In graphite (the "lead" found in pencils) the carbon is very soft, while the carbon in a diamond is roughly seven times as hard.



Figure **12.2.1**: Pencil (Credit: Public Domain; Source: wikimedia; License: Public Domain)



Figure **12.2.2**: Diamond ring (Credit: User:Ikkyu2/Wikimedia Commons; Source:wikimedia; License: Public Domain)

Melting point and boiling point are somewhat unique identifiers, especially in the case of compounds. In addition to giving some idea as to the identity of the compound, important information can be obtained about the purity of the material.

Physical Change

As an ice cube melts, its shape changes as it acquires the ability to flow. However, its composition does not change. **Melting** is an example of a **physical change**. A physical change is a change to a sample of matter in which some properties of the material change, but the identity of the matter does not. Physical changes can further be classified as reversible or irreversible. The melted ice cube may be refrozen, so melting is a reversible physical change. Physical changes that involve a change of state are all

reversible. Other changes of state include **vaporization** (liquid to gas), **freezing** (liquid to solid), and **condensation** (gas to liquid). Dissolving is also a reversible physical change. When salt is dissolved into water, the salt is said to have entered the aqueous state. The salt may be regained by boiling off the water, leaving the salt behind.



Figure **12.2.2**: Melting ice in the Beaufort Sea. (Credit: Courtesy of Rear Admiral Harley D. Nygren, NOAA; Source: wikimedia; License: Public Domain)

When a piece of wood is ground into sawdust, that change is irreversible since the sawdust cannot be reconstituted into the same piece of wood that it was before. Cutting grass or pulverizing a rock are examples of irreversible physical changes. Chopping wood for a fire also represents an irreversible physical change, since the pieces cannot be put back together to form the tree.



Figure **12.2.3**: Chopping wood is an irreversible physical change. (Credit: stevepb; Source: pixabay; License: Pixabay License)

Physical Changes vs. Chemical Changes

What is the difference between physical and chemical changes? Watch the video to find out.



Section Summary

- A physical property is a characteristic of a substance that can be observed or measured without changing the identity of the substance.
- Physical properties include color, density, hardness, melting point, and boiling point.
- A physical change is a change to a sample of matter in which some properties of the material change, but the identity of the matter does not.

- A physical change can be reversible where the original form of the matter can be restored, or irreversible where the original form cannot be restored.
- Melting ice and grinding wood into sawdust are examples of physical changes.
- A chemical change produces different materials than the ones we started with.

Glossary

hardness

A measure of a material's ability to resist plastic deformation (i.e.- scratching or denting the material).

physical property

A characteristic of a substance that can be observed or measured without changing the identity of the substance.

physical change

A change to a sample of matter in which some properties of the material change, but the identity of the matter does not.

This page titled 12.2: Physical and Chemical is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.3: Physical Properties** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.6: Physical Change** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.14: Chemical Change** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **10.2: Magnets** by OpenStax is licensed CC BY 4.0.

12.3: Classifying Matter

Learning Objectives

- Identify the characteristics of substances, homogeneous mixtures, heterogeneous mixtures, solutions, colloids, and suspensions.
- Identify techniques to separate mixtures.

Earlier in this text we discussed **substances** such as elements and compounds. But at the start of this chapter we looked at an example illustrating that most of the matter we interact with on a daily basis is not in the form of a substance, but a mixture. In this section we will look at what differentiates a mixture from a substance, as well as some categories of mixtures.

What Is a Mixture?

A **mixture** is a combination of two or more substances in any proportion. This is different from a compound, which consists of substances in fixed proportions. The substances in a mixture also do not combine chemically to form a new substance, as they do in a compound. Instead, they just intermingle and keep their original properties. The lemonade pictured above is a mixture because it does not have fixed proportions of ingredients. It could have more or less lemon juice, for example,

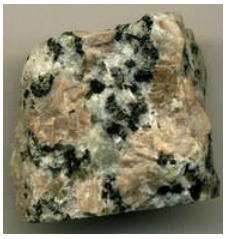


Figure **12.3.1**: This rock is a mixture of smaller rocks and minerals. (Credit: James St. John (Flickr: jsj1771); Source:flickr; License: CC by 2.0(opens in new window))

Homogenous Mixtures

A **homogeneous** mixture is a mixture in which the composition is uniform throughout the mixture. The lemonade in the introductory section is an example of a homogeneous mixture (unless it has chunks of lemon in it, then it would be another type of mixture we will discuss shortly). A homogeneous mixture has the same composition throughout. Another example of a homogeneous mixture is salt water. Often it is easy to confuse a homogeneous mixture with a pure substance because they are both uniform. The difference is that the composition of the **substance** is always the same. The amount of salt in the salt water can vary from one sample to another, so it is not a substance but a mixture. However, because the dissolved material is present in the same amount throughout the solution it is a homogeneous mixture. Substances will have the same composition in any sample. Homogeneous mixtures will have the same composition throughout a given sample.

Heterogeneous Mixtures

A **heterogeneous mixture** is a mixture in which the composition is not uniform throughout the mixture. Vegetable soup is a heterogeneous mixture. Any given spoonful of soup will contain varying amounts of the different vegetables and other components of the soup. The rock in Figure **12.3.1** is an example of a heterogeneous mixture. A heterogeneous mixture varies in its composition. The black nuggets, for example, are not distributed evenly throughout the rock.

A **phase** is any part of a sample that has a uniform composition and properties. By definition, a pure substance or a homogeneous mixture consists of a single phase. A heterogeneous mixture consists of two or more phases. When oil and water are combined, they do not mix evenly, but instead form two separate layers. Each of the layers is called a phase.



Figure 12.3.2: Oil and water do not mix, instead forming two distinct layers called phases. The oil phase is less dense than the water phase, and so the oil floats on top of the water. (Credit: Flickr: Yortw; Source: flickr

http://www.flickr.com/photos/yortw/5470226807/(opens in new window); License: CC by 2.0(opens in new window))

In the case of vegetable soup, one phase would be the liquid soup itself. This phase has vitamins, minerals, and other components dissolved in the water. This phase would be homogeneous. The carrots, peas, corn, or other vegetables represent other phases of the soup.The various vegetables are not mixed evenly in the soup, but are spread around at random.

There are a large number of heterogeneous mixtures around us. Soil is composed of a variety of substances, and is often of different composition depending on the sample taken. One shovelful may come up with dirt and grass, while the next shovelful could contain an earthworm.

Smog is another example of a heterogeneous mixture. This murky collection of pollutants can be a mixture of water and contaminants from burning gasoline or plastics, mixed with nitric oxide derivatives and ozone. You can see that the smog distribution in the air illustrated below is not evenly spread out, but varies from one part of the atmosphere to another.

? Exercise 12.3.1

Which of the following are mixtures: salt water, salad dressing, air is a mixture of gases, the rock pictured in Figure 12.3.1. Identify each of these which is a mixture as either homogeneous or heterogeneous.

Answer

All of these are mixtures.

The rock and the salad dressing are visibly different throughout, which identifies them as a heterogeneous mixtures. (Although depending on the type of salad dressing, it might be homogeneous. Any list of ingredients in a food that contains more than one is always going to be a mixture, and some of those that have only one might also be mixtures.)

Salt water and air are both homogeneous mixtures: the mixing occurs on the level of molecules. The air is a mixture of mostly oxygen and nitrogen, though there are other trace components.



Figure 12.3.2: Smog in New York City. (Credit: Courtesy of Dr. Edwin P. Ewing, Jr./CDC; Source: wikimedia; License: Public Domain)

Types of Mixtures

Mixtures have different properties depending on the size of their particles. As we will see, some of these are homogeneous mixtures and some of them are heterogeneous mixtures. Three types of mixtures based on particle size are solutions, suspensions, and

colloids, all of which are described in the table below.

Table 12.3.1: Solutions, Suspensions, and Colloids

Type of Mixture	Description
Solutions  Figure 12.3.3 (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CK-12 Curriculum Materials license)	A **solution** is a homogeneous mixture with tiny particles. The particles are too small to see and also too small to settle or be filtered out of the mixture. When the salt is thoroughly mixed into the water in this glass, it will form a solution. The salt will no longer be visible in the water, and it won't settle to the bottom of the glass.
Colloids  Figure 12.3.4 (Credit: Ray Bouknight; Source: flickr; License: CC by 2.0(opens in new window))	A **colloid** is a mixture with medium-sized particles. The particles are large enough to disperse light, but not large enough to settle or be filtered out of the mixture. Most scientists would classify these as heterogeneous mixtures, but some texts classify them as homogeneous mixtures. The gelatin in this dish is a colloid. It looks red because you can see the red gelatin particles in the mixture. However, the particles are too small to settle to the bottom of the dish.
Suspensions  Figure 12.3.5 (Credit: jlarbig; Source: pixabay); License: Pixabay License)	A **suspension** is a heterogeneous mixture with large particles. The particles are large enough to see and also to settle or be filtered out of the mixture. The salad dressing in this bottle is a suspension. It contains oil, vinegar, herbs, and spices. If the bottle sits undisturbed for very long, the mixture will separate into its component parts. That's why you should shake it before you use it.

? Exercise 12.3.2

If you buy a can of paint at a paint store, a store employee may put the can on a shaker machine to mix up the paint in the can. What type of mixture is the paint?

Answer

The paint is a suspension. Some of the components of the paint settle out of the mixture when it sits undisturbed for a long time. This explains why you need to shake (or stir) the paint before you use it.

? Exercise 12.3.3

The milk you buy in the supermarket has gone through a process called homogenization. This process breaks up the cream in the milk into smaller particles. As a result, the cream doesn't separate out of the milk no matter how long it sits on the shelf. Which type of mixture is homogenized milk?

Answer

Homogenized milk is a colloid. The particles in the milk are large enough to see—that's why milk is white instead of clear like water, which is the main component of milk. However, the particles are not large enough to settle out of the mixture.

Separating Mixtures

The components of a mixture keep their own identity when they combine, so they retain their physical properties. Examples of physical properties include boiling point, ability to dissolve, and particle size. When components of mixtures vary in physical properties such as these, processes such as boiling, dissolving, or filtering can be used to separate them.

Look at the figure below of the Great Salt Lake in Utah. The water in the lake is a solution of salt and water. Do you see the white salt deposits near the shore? How did the salt separate from the salt water? Water has a lower boiling point than salt, and it evaporates in the heat of the sun. With its higher boiling point, the salt doesn't get hot enough to evaporate, so it is left behind.



Figure **12.3.6**: "Great Salt Lake" (Credit: John Morgan; Source: flickr; License: CC by 2.0(opens in new window))

Q: Suppose you have a mixture of salt and pepper. What properties of the salt and pepper might allow you to separate them?

A: Salt dissolves in water but pepper does not. If you mix salt and pepper with water, only the salt will dissolve, leaving the pepper floating in the water. You can separate the pepper from the water by pouring the mixture through a filter, such as a coffee filter.

Q: After you separate the pepper from the salt water, how could you separate the salt from the water?

A: You could heat the water until it boils and evaporates. The salt would be left behind.

Section Summary

- A substance is a pure material with a uniform and definite composition.
- A substance cannot be separated into components using physical means.
- A mixture is a combination of two or more substances in any proportions. The substances in a mixture do not combine chemically, so they retain their physical properties.
- A homogeneous mixture has the same composition throughout. A heterogeneous mixture varies in its composition.
- A homogeneous mixture is a mixture in which the composition is uniform throughout the mixture.
- All solutions would be considered homogeneous.
- A heterogeneous mixture is a mixture in which the composition is not uniform throughout the mixture.
- A phase is a separate layer in a heterogeneous mixture.
- Mixtures can be classified on the basis of particle size into three different types: solutions, suspensions, and colloids.
- The components of a mixture retain their own physical properties. These properties can be used to separate the components by filtering, boiling, or other physical processes.

Glossary

mixture

a combination of two or more substances in any proportion.

homogeneous mixture

a mixture in which the composition is uniform throughout the mixture.

heterogeneous mixture

a mixture in which the composition is not uniform throughout the mixture.

phase

any part of a sample that has a uniform composition and properties.

solution

a homogeneous mixture with tiny particles.

colloid

a mixture with particles large enough to disperse light, but not large enough to settle or be filtered out of the mixture.

suspension

a heterogeneous mixture with particles large enough to see and also to settle or be filtered out of the mixture.

This page titled 12.3: Classifying Matter is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: Pure Substances** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.8: Homogeneous Mixture** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.9: Heterogeneous Mixtures** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.4: Time, Velocity, and Speed** by OpenStax is licensed CC BY 4.0.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

12.4: The Physical States of Matter

Learning Objectives

- Describe the characteristic properties of solids, liquids, and gases



Figure 12.4.1 (Credit: Iceberg: Courtesy of Rear Admiral Harley D. Nygren, NOAA; Beach: User:Wicki/Wikimedia Commons; Volcano: Courtesy of E. Klett, US Fish and Wildlife Service; Source: Iceberg: wikimedia; Beach: wikimedia; Volcano: wikimedia; License: Public Domain)

Why is the state of water different in each picture?

Water can take many forms. At low temperatures (below $0°C$), water is a solid. When at "normal" temperatures (between $0°C$ and $100°C$), it is a liquid. At temperatures above $100°C$, water is a gas (steam).

The state of water depends on the temperature. Each state (solid, liquid, and gas) has its own unique set of physical properties.

Matter and its States

Matter typically exists in one of three states: **solid**, **liquid**, or **gas**. There is a fourth state of matter called **plasma**, which rarely exists on earth, but we will omit this from our current discussion. The state a given substance exhibits is also a physical property. Some substances exist as gases at room temperature (oxygen and carbon dioxide), while others, like water and mercury metal, exist as liquids. Most metals exist as solids at room temperature. All substances can exist in any of these three states.

Liquid

Liquids have the following characteristics:

- No definite shape (takes the shape of its container).
- Has definite volume.
- Particles are free to move over each other, but are still attracted to each other.

A familiar liquid is mercury metal. Mercury is an anomaly. It is the only metal we know of that is liquid at room temperature. Mercury also has an ability to stick to itself (surface tension), which is a property that all liquids exhibit. Mercury has a relatively high surface tension, and this makes it very unique. Here you can see mercury in its common liquid form.

If we heat liquid mercury to its boiling point of $357°C$, and contain it under the right pressure conditions, we would notice all particles in the liquid state go into the gas state.

Gas

Gases have the following characteristics:

- No definite shape (takes the shape of its container).
- No definite volume.
- Particles move in random motion with little or no attraction to each other.
- Highly compressible.

Solid

Solids are defined by the following characteristics:

- Definite shape (rigid).
- Definite volume.
- Particles vibrate around fixed axes.

If we were to cool a sample of liquid mercury to its freezing point of $-39°C$, and had it contained under the right pressure conditions, we would notice all of the liquid particles would go into the solid state.



Plasma

Plasma is a state of matter that resembles a gas but has certain properties that gases do not have. Like a gas, plasma consists of particles of matter that can pull apart and spread out, so it lacks a fixed volume and a fixed shape. Unlike a gas, plasma can conduct electricity and respond to a magnetic field. That's because plasma consists of electrically charged particles called ions, instead of uncharged particles such as atoms or molecules.

Plasma are defined by the following characteristics:

- particles are charged ions and free electrons
- no definite shape
- no definite volume
- conducts electricity
- responds to magnetic field

Section Summary

- Three states of matter exist: solid, liquid, and gas.
- Solids have a definite shape and volume.
- Liquids have a definite volume, but take the shape of their container.
- Gases have no definite shape or volume.

This page titled 12.4: The Physical States of Matter is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.5: States of Matter** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

12.5: Transitions between Phases

Learning Objectives

- Identify the names of the phase changes between solid, liquid, and gas phases.

We have introduced the different physical states of matter and mentioned that transitions between these physical states are a type of physical change. We will explore these changes in more detail in this section.

Evaporation

A puddle of water left undisturbed eventually disappears. The liquid molecules escape into the gas phase, becoming water vapor. **Vaporization** is the process in which a liquid is converted to a gas. **Evaporation** is the conversion of a liquid to its vapor below the boiling temperature of the liquid. If the water is instead kept in a closed container, the water vapor molecules do not have a chance to escape into the surroundings, and so the water level does not change. As some water molecules become vapor, an equal number of water vapor molecules condense back into the liquid state. **Condensation** is the change of state from a gas to a liquid.

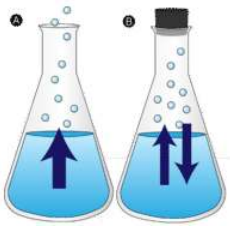


Figure 12.5.1: Evaporation (A) and condensation (B). (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

In some parts of the world, people are able to cool off using a device known as a "swamp cooler". This piece of equipment traces its origin back to the ancient Egyptians, who hung wet blankets across the doors of their homes. As the warm air passed through the blankets, water would evaporate and cool the air. The royalty went one step further, and had servants fan wet cloths over jugs of water to get more evaporation and cooling. The origin of the term "swamp cooler" is not known; they certainly do not work in a swamp. The best conditions for cooling include a high temperature (over 80°F) and a low humidity (preferably less than 30%). These coolers work well in desert areas, but do not provide any cooling in the humid areas of the country.

Boiling Point

As a liquid is heated, the average kinetic energy of its particles increases. The rate of evaporation increases as more and more molecules are able to escape the liquid's surface into the vapor phase. Eventually a point is reached when the molecules all throughout the liquid have enough kinetic energy to **vaporize**. At this point, the liquid begins to boil. The **boiling point** is the temperature at which the vapor pressure of a liquid is equal to the external pressure. Figure 12.5.2 illustrates the boiling of liquid.

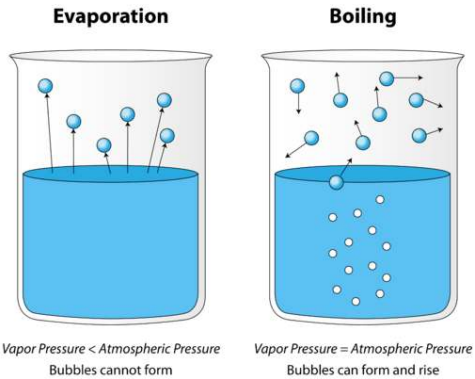


Figure 12.5.2: Comparison between evaporation and boiling. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

In the picture on the left, the liquid is below its boiling point, yet some of the liquid evaporates. On the right, the temperature has been increased until bubbles begin to form in the body of the liquid. When the vapor pressure inside the bubbles is equal to the external atmospheric pressure, the bubbles rise to the surface of the liquid and burst. The temperature at which this process occurs is the boiling point of the liquid.

The normal boiling point is the temperature at which the vapor pressure of the liquid is equal to standard pressure. Because atmospheric pressure can change based on location, the boiling point of a liquid changes with the external pressure. The normal boiling point is a constant because it is defined relative to the standard atmospheric pressure of $760\ \text{mm Hg}$ (or $1\ \text{atm}$ or $101.3\ \text{kPa}$).

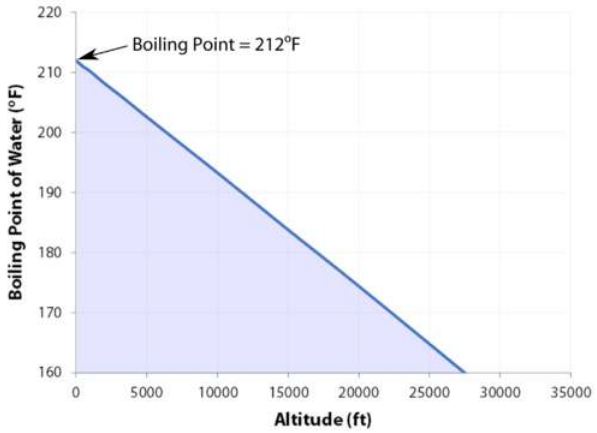


Figure 12.5.3: Influence of altitude on the boiling point of water. (Credit: Joy Sheng; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Melting Point

Solids are similar to liquids in that both are condensed states, with particles that are far closer together than those of a gas. However, while liquids are fluid, solids are not. The particles of most solids are packed tightly together in an orderly arrangement. The motion of individual atoms, ions, or molecules in a solid is restricted to **vibrational motion** about a fixed point. Solids are almost completely **incompressible** and are the densest of the three states of matter.

As a solid is heated, its particles vibrate more rapidly as it absorbs kinetic energy. Eventually, the organization of the particles within the solid structure begins to break down, and the solid starts to melt. The **melting point** is the temperature at which a solid changes into a liquid. At its melting point, the disruptive vibrations of the particles of the solid overcome the attractive forces operating within the solid. As with boiling points, the melting point of a solid is dependent on the strength of those attractive forces. Sodium chloride (NaCl) is an ionic compound that consists of a multitude of strong ionic bonds. Sodium chloride melts at 801°C. Ice (solid H_2O) is a molecular compound whose molecules are held together by hydrogen bonds. Though hydrogen bonds are the strongest of the intermolecular forces, the strength of hydrogen bonds is much less than that of ionic bonds. The melting point of ice is 0 °C.

The melting point of a solid is the same as the freezing point of the liquid. At that temperature, the solid and liquid states of the substance are in equilibrium. For water, this equilibrium occurs at 0°C.

$$\text{H}_2\text{O}\,(s) \rightleftharpoons \text{H}_2\text{O}\,(l)$$

We tend to think of solids as those materials that are solid at room temperature. However, all materials have melting points of some sort. Gases become solids at extremely low temperatures, and liquids will also become solid if the temperature is low enough.

The table below gives the melting points of some common materials.

Table 12.5.1: Melting Points of Common Materials

Materials	Melting Point (°C)
Hydrogen	-259
Oxygen	-219
Diethyl ether	-116
Ethanol	-114
Water	0
Pure silver	961
Pure gold	1063
Iron	1538

Sublimation

Solids also have a vapor pressure, though it is generally much less than that of a liquid. A snow bank will gradually disappear even if the temperature stays below 0°C. The snow does not melt, but instead passes directly from the solid state to the vapor state. **Sublimation** is the change of state from a solid to a gas, without passing through the liquid state.

"Dry ice", or solid carbon dioxide, is a substance that sublimes at atmospheric pressures. Dry ice is very cold (−78°C) and so is used as a coolant for goods, such as ice cream, that must remain frozen during shipment. Because the dry ice sublimes rather than melts, there is no liquid mess associated with its change of state as it warms. As you may have seen in demonstrations, dry ice merely transforms from a solid to a cold, steamy looking gas. (The steam you see is actual water that has condensed around the very cold carbon dioxide gas that is forming as the dry ice sublimes.)

Summary of State Changes

All of the changes of state that occur between solid, liquid, and gas are summarized in the diagram in the figure below. **Freezing** is the opposite of melting, and both represent the equilibrium between the solid and liquid states. **Evaporation** occurs when a liquid turns to a gas. **Condensation** is the opposite of vaporization, and both represent the equilibrium between the liquid and gas states. **Deposition** is the opposite of sublimation, and both represent the equilibrium between the solid and gas states.

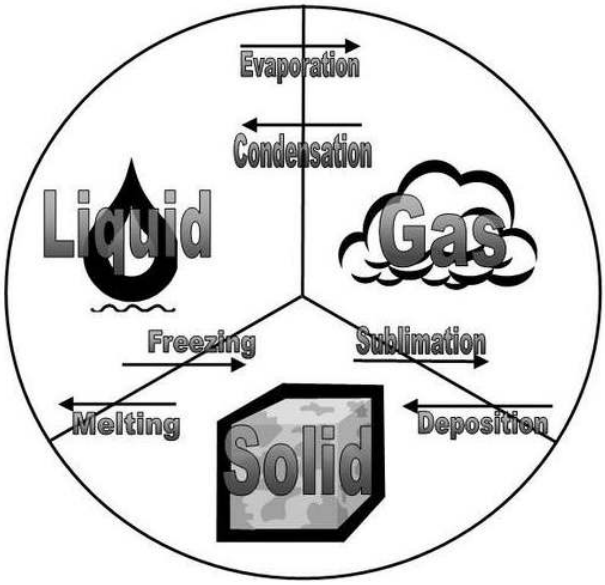


Figure 12.5.4: Solid, liquid, and gas states with the terms for each change of state that occurs between them.

Section Summary

- The boiling point is the temperature at which the vapor pressure of a liquid is equal to the external pressure.
- As the altitude increases, the boiling point decreases.
- Evaporation is the conversion of a liquid to its vapor below the boiling temperature of the liquid.
- Condensation is the change of state from a gas to a liquid.
- As the temperature increases, the rate of evaporation increases.
- The melting point is the temperature at which a solid changes into a liquid.
- Intermolecular forces have a strong influence on melting point.
- Sublimation is the change of state from a solid to a gas, without passing through the liquid state.
- Freezing is the opposite of melting.
- Deposition is the opposite of sublimation.

Glossary

vaporization

The process in which a liquid is converted to a gas.

evaporation

The conversion of a liquid to its vapor below the boiling temperature of the liquid.

condensation

The change of state from a gas to a liquid.

boiling point

The temperature at which the vapor pressure of a liquid is equal to the external pressure.

melting point

The temperature at which a solid changes into a liquid.

sublimation

The change of state from a solid to a gas, without passing through the liquid state.

freezing

the change of state from a liquid to a solid.

deposition

the change of state from a gas to a solid, without passing through the liquid state.

This page titled 12.5: Transitions between Phases is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **13.7: Evaporation** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **13.9: Boiling** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **13.11: Melting** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **13.12: Sublimation** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **13.18: Heating and Cooling Curves** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

12.6: Energy and Phase Transitions

Learning Objectives

- Predict the direction of energy flow associated with each phase change.
- Interpret a heating curve.

In the previous subsection we discussed transitions between different physical states. But why do these transitions occur? You might associate something that freezes with a low temperature and something that melts with a high temperature. We have seen frozen water and melted wax that are consistent with this perception. And based on what we already know from our study of energy earlier in this text, we might assume that heat transfers can lead to phase changes. This is correct, and we will explore this concept in this section.

Heat and Evaporation

In order for a liquid molecule to escape into the gas state, the molecule must have enough kinetic energy to overcome the intermolecular attractive forces in the liquid. Recall that a given liquid sample will have molecules with a wide range of kinetic energies. Liquid molecules that have this certain threshold kinetic energy escape the surface and become vapor. As a result, the liquid molecules that remain now have lower kinetic energy. As evaporation occurs, the temperature of the remaining liquid decreases. You have observed the effects of evaporative cooling. On a hot day, the water molecules in your perspiration absorb body heat and evaporate from the surface of your skin. The evaporation process leaves the remaining perspiration cooler, which in turn absorbs more heat from your body.

A given liquid will evaporate more quickly when it is heated. This is because the heating process results in a greater fraction of the liquid's molecules having the necessary kinetic energy to escape the surface of the liquid. The figure below shows the kinetic energy distribution of liquid molecules at two temperatures. The numbers of molecules that have the required kinetic energy to evaporate are shown in the shaded area under the curve at the right. The higher temperature liquid (T_2) has more molecules that are capable of escaping into the vapor phase than the lower temperature liquid (T_1).

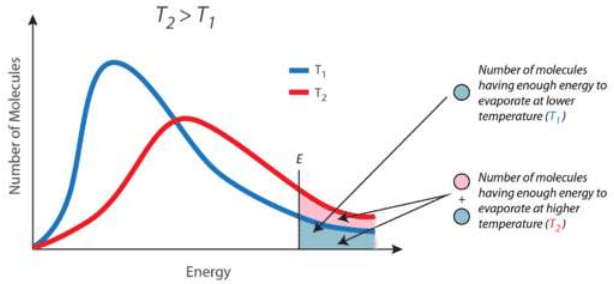


Figure **12.6.2**: Kinetic energy distribution curves for a liquid at two temperatures T_1 and T_2. The shaded area represents the molecules with enough kinetic energy to escape the liquid and become vapor. (Credit: Christopher Auyeung; Source: CK-12 Foundation; License: CC BY-NC 3.0(opens in new window))

Heating Curves

Imagine that you have a block of ice that is at a temperature of **−30°C**, well below its **melting** point. The ice is in a closed container. As heat is steadily added to the ice block, the water molecules will begin to vibrate faster and faster as they absorb kinetic energy. Eventually, when the ice has warmed to **0°C**, the added energy will start to break apart the hydrogen bonding that keeps the water molecules in place when it is in the **solid** form. As the ice melts, its temperature does not rise. All of the energy that is being put into the ice goes into the melting process and not into any increase in temperature. During the melting process, the two states—solid and **liquid**—are in equilibrium with one another. If the system was isolated at that point and no energy was allowed to enter or leave, the ice-water mixture at **0°C** would remain. Temperature is always constant during a change of state.

Continued heating of the water after the ice has completely melted will now increase the kinetic energy of the liquid molecules and the temperature will rise. Assuming that the atmospheric pressure is standard, the temperature will rise steadily until it reaches **100°C**. At this point, the added energy from the heat will cause the liquid to begin to vaporize. As with the previous state change,

the temperature will remain at 100°C while the water molecules are going from the liquid to the **gas** or vapor state. Once all the liquid has completely boiled away, continued heating of the steam (since the container is closed) will increase its temperature above 100°C.

The experiment described above can be summarized in a graph called a heating curve (figure below).

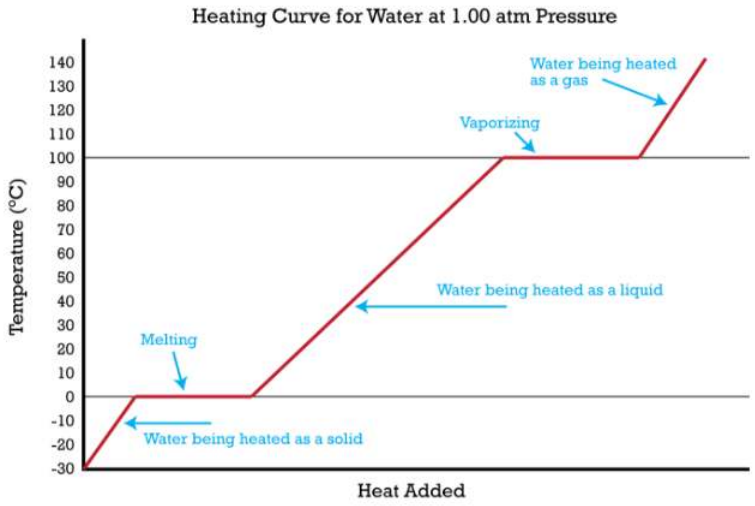


Figure 12.6.1: In the heating curve of water, the temperature is shown as heat is continually added. Changes of state occur during plateaus, because the temperature is constant.

The change of state behavior of all substances can be represented with a heating curve of this type. The melting and boiling points of the substance can be determined by the horizontal lines or plateaus on the curve. Other substances have melting and boiling points that are different from those of water. An exception to this blueprint heating curve is for a substance such as carbon dioxide, which sublimes rather than melts at standard pressure. The heating curve for carbon dioxide would have only one plateau, at the **sublimation** temperature of CO_2.

The entire experiment could be run in reverse. Steam above 100°C could be steadily cooled down to 100°C, at which point it would condense to liquid water. The water could then be cooled to 0°C, at which point continued cooling would freeze the water to ice. The ice could then be cooled to a point below 0°C. This could be diagrammed in a cooling curve that would be the reverse of the heating curve.

Summary of Energy and Phase Changes

Some physical changes of state require energy to be added for them to occur, while others require that energy be removed in order for them to occur. Changes of state that require energy to be added are evaporation, boiling, melting, and sublimation. Changes of state that require energy to be removed are condensation, freezing, and deposition. Whether energy is being added or removed, those changes of state will occur at a constant temperature. When a substance is not at the temperature of a phase change, any energy added to that substance will cause the temperature to increase. Likewise, any energy removed from it will cause the temperature to decrease.

Section Summary

- A change of state can be brought about by putting heat into a system or removing heat from the system.
- The temperature of a system will not change as long as the substance is undergoing a change from solid to liquid to gas, as well as the reverse.
- Freezing is the opposite of melting.
- Evaporation occurs when a liquid turns to a gas.
- Condensation is the opposite of vaporization.
- Deposition is the opposite of sublimation.

This page titled 12.6: Energy and Phase Transitions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **13.18: Heating and Cooling Curves** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **13.7: Evaporation** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

12.7: Moles- Counting Atoms



Figure **12.7.1** (Credit: User:OSX/Wikimedia Commons; Source: wikimedia; License: Public Domain)

Is there an easier way to load this truck?

When the weather is nice, many people begin to work on their yards and homes. For many projects, sand is needed as a foundation for a walk or to add to other materials. You could order up twenty million grains of sand and have people really stare at you. You could order by the pound, but that takes a lot of time weighing out. The best bet is to order by the yard, meaning a cubic yard. The loader can easily scoop up what you need and put it directly in your truck.

It turns out there are also convenient ways of counting how much we have in chemistry. We will introduce some of those in this chapter. The first section will relate to counting atoms, and the second section will relate counting atoms to mass.

12.7: Moles- Counting Atoms is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.7.1: Moles and Avogrado's Number

Learning Objectives

- Use Avogrado's number to convert between moles and atoms.

It certainly is easy to count objects such as bananas, or something as large as elephants (as long as you stay out of their way). However, counting grains of sugar from a sugar canister would take a long, long time. Atoms and molecules are extremely small—far, far smaller than grains of sugar. Counting atoms or molecules is not only unwise, it is absolutely impossible. One drop of water contains about 10^{22} molecules of water. If you counted 10 molecules every second for 50 years, without stopping, you would have counted only 1.6×10^{10} molecules. Put another way, at that counting rate, it would take you over 30 trillion years to count the water molecules in one tiny drop.

Chemists of the past needed a name that could stand for a very large number of items. Amadeo Avogadro (1776-1856), an Italian scientist, provided such a number. He is responsible for the counting unit of measure called the mole. A **mole** (mol) is the amount of a substance that contains 6.02×10^{23} representative particles of that substance. The mole is the SI unit for amount of a substance. Just like the dozen and the gross, it is a name that stands for a number. There are therefore 6.02×10^{23} water molecules in a mole of water molecules. There also would be 6.02×10^{23} bananas in a mole of bananas, if such a huge number of bananas ever existed.



Figure 12.7.1.2: Italian scientist Amadeo Avogadro, whose work led to the concept of the mole as a counting unit in chemistry. (Credit: C. Sentier; Source: wikimedia; License: Public Domain)

The number 6.02×10^{23} is called **Avogadro's number**, the number of representative particles in a mole. It is an experimentally determined number. A **representative particle** is the smallest unit in which a substance naturally exists. For the majority of elements, the representative particle is the atom. Iron, carbon, and helium consist of iron atoms, carbon atoms, and helium atoms, respectively. Seven elements exist in nature as diatomic molecules and they are H_2, N_2, O_2, F_2, Cl_2, Br_2, and I_2. The representative particle for these elements is the molecule. Likewise, all molecular compounds such as H_2O and CO_2 exist as molecules and so the molecule is their representative particle. For ionic compounds such as NaCl and $Ca(NO_3)_2$, the representative particle is the formula unit. A mole of any substance contains Avogadro's number (6.02×10^{23}) of representative particles.



Conversions Between Moles and Atoms

Using our unit conversion techniques, we can use the mole label to convert back and forth between the number of particles and moles.

Example 12.7.1.1: Converting Number of Particles to Moles

The element carbon exists in two primary forms: graphite and diamond. How many moles of carbon atoms is 4.72×10^{24} atoms of carbon?

Solution

Step 1: List the known quantities and plan the problem.

Known

- number of C atoms $= 4.72 \times 10^{24}$
- 1 mole $= 6.02 \times 10^{23}$ atoms

Unknown

- 4.72 x 10^{24} = ? mol C

One conversion factor will allow us to convert from the number of C atoms to moles of C atoms.

Step 2: Calculate.

$$4.72 \times 10^{24}\ \text{atoms C} \times \frac{1\ \text{mol C}}{6.02 \times 10^{23}\ \text{atoms C}} = 7.84\ \text{mol C}$$

Step 3: Think about your result.

The given number of carbon atoms was greater than Avogadro's number, so the number of moles of C atoms is greater than 1 mole. Since Avogadro's number is a measured quantity with three significant figures, the result of the calculation is rounded to three significant figures.



Section Summary

- A mole of any substance contains Avogadro's number (6.02×10^{23}) of representative particles.
- Methods are described for conversions between moles, atoms, and molecules.

This page titled 12.7.1: Moles and Avogrado's Number is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **10.1: Avogadro's Number** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **10.2: Conversions Between Moles and Atoms** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

- **10.3: Molar Mass** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

12.7.2: Moles and Mass

Learning Objectives

- Use molar mass to convert between moles and mass.

Molar mass is defined as the mass of one mole of representative particles of a substance. By looking at a periodic table, we can conclude that the molar mass of lithium is $6.94\ \text{g}$, the molar mass of zinc is $65.38\ \text{g}$, and the molar mass of gold is $196.97\ \text{g}$. Each of these quantities contains 6.02×10^{23} atoms of that particular element. The units for molar mass are grams per mole, or g/mol. The molar mass is listed on the periodic table for each element, typically underneath the symbol of the element.

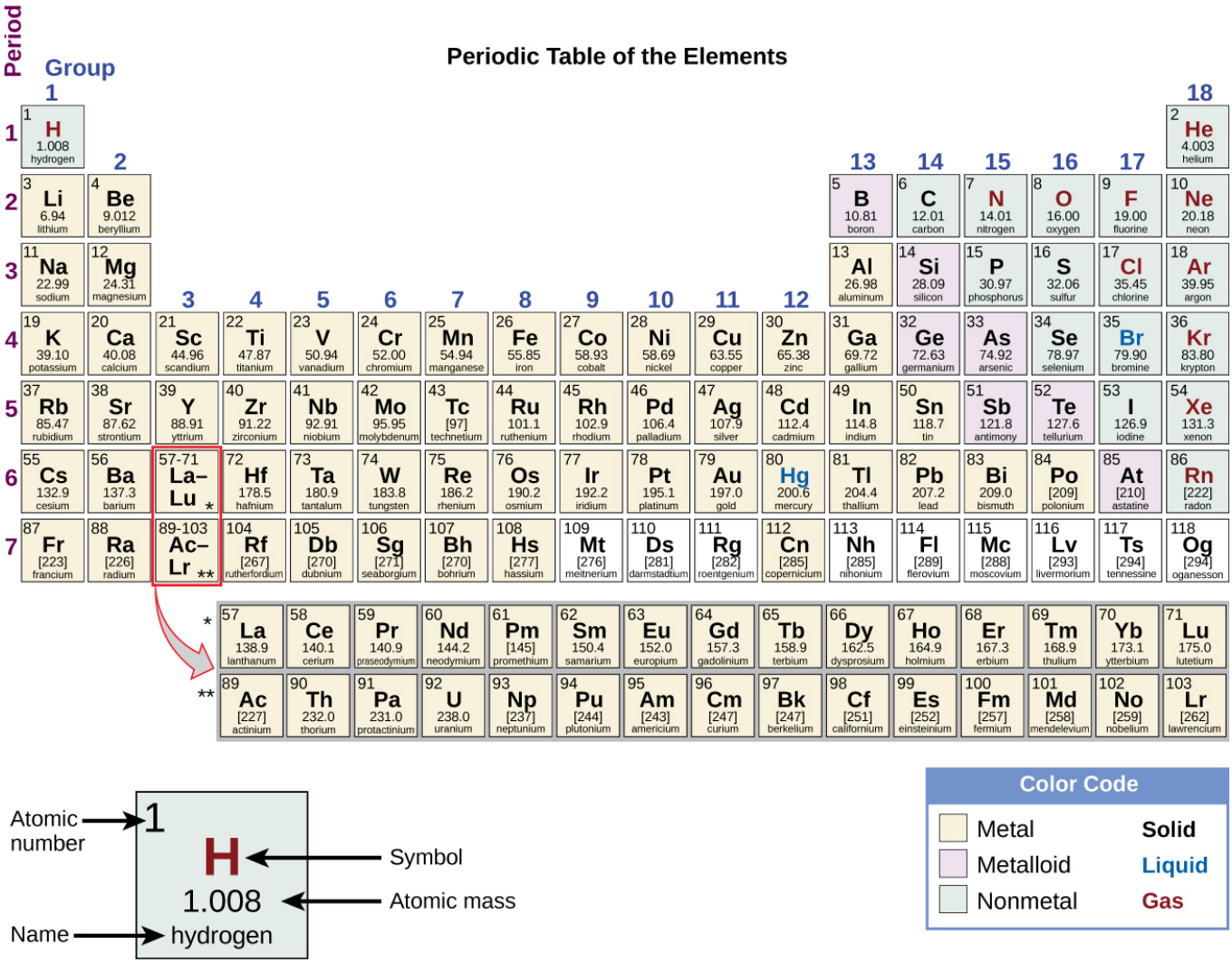


Figure 12.7.2.3: periodic table including molar mass (this same number is also referred to as the atomic mass) Source: Open Stax, as compiled by Libretexts.

Molar Masses of Compounds

The molecular formula of the compound carbon dioxide is CO_2. One molecule of carbon dioxide consists of 1 atom of carbon and 2 atoms of oxygen. We can calculate the mass of one molecule of carbon dioxide by adding together the masses of 1 atom of carbon and 2 atoms of oxygen:

$$12.01\ \text{amu} + 2\,(16.00\ \text{amu}) = 44.01\ \text{amu}$$

The **molecular mass** of a compound is the mass of one molecule of that compound. The molecular mass of carbon dioxide is $44.01\ \text{amu}$.

The molar mass of any compound is the mass in grams of one mole of that compound. One mole of carbon dioxide molecules has a mass of $44.01\ \text{g}$, while one mole of sodium sulfide formula units has a mass of $78.04\ \text{g}$. The molar masses are $44.01\ \text{g/mol}$ and $78.04\ \text{g/mol}$ respectively. In both cases, that is the mass of 6.02×10^{23} representative particles. The representative particle of CO_2 is the molecule, while for Na_2S it is the formula unit.

Example 12.7.2.1: Molar Mass of a Compound

Calcium nitrate, $Ca(NO_3)_2$, is used as a component in fertilizer. Determine the molar mass of calcium nitrate.

Solution

Step 1: List the known and unknown quantities and plan the problem.

Known

- Formula $= Ca(NO_3)_2$
- Molar mass $Ca = 40.08\ \text{g/mol}$
- Molar mass $N = 14.01\ \text{g/mol}$
- Molar mass $O = 16.00\ \text{g/mol}$

Unknown

- molar mass $Ca(NO_3)_2$

First we need to analyze the formula. Since the Ca lacks a subscript, there is one Ca atom per formula unit. The 2 outside the parentheses means that there are two nitrate ions per formula unit and each nitrate ion consists of one nitrogen atom and three oxygen atoms per formula unit. Thus, $1\ \text{mol}$ of calcium nitrate contains $1\ \text{mol}$ of Ca atoms, $2\ \text{mol}$ of N atoms, and $6\ \text{mol}$ of O atoms.

Step 2: Calculate

Use the molar masses of each atom together with the number of atoms in the formula and add together.

$$1\ \text{mol Ca} \times \frac{40.08\ \text{g Ca}}{1\ \text{mol Ca}} = 40.08\ \text{g Ca}$$

$$2\ \text{mol N} \times \frac{14.01\ \text{g N}}{1\ \text{mol N}} = 28.02\ \text{g N}$$

$$6\ \text{mol O} \times \frac{16.00\ \text{g O}}{1\ \text{mol O}} = 96.00\ \text{g O}$$

$$\text{Molar mass of } Ca(NO_3)_2 = 40.08\ \text{g} + 28.02\ \text{g} + 96.00\ \text{g} = 164.10\ \text{g/mol}$$



Conversions Between Moles and Mass

The molar mass of any substance is the mass in grams of one mole of representative particles of that substance. The representative particles can be atoms, molecules, or formula units of ionic compounds. This relationship is frequently used in the laboratory. Suppose that for a certain experiment, you need 3.00 moles of calcium chloride ($\mathrm{CaCl_2}$). Since calcium chloride is a solid, it would be convenient to use a balance to measure the mass that is needed. The molar mass of $\mathrm{CaCl_2}$ is $110.98\ \mathrm{g/mol}$. The conversion factor that can be used is then based on the equality that $1\ \mathrm{mol} = 110.98\ \mathrm{g\ CaCl_2}$. Dimensional analysis will allow you to calculate the mass of $\mathrm{CaCl_2}$ that you should measure.

$$3.00\ \mathrm{mol\ CaCl_2} \times \frac{110.98\ \mathrm{g\ CaCl_2}}{1\ \mathrm{mol\ CaCl_2}} = 333\ \mathrm{g\ CaCl_2}$$

When you measure the mass of $333\ \mathrm{g}$ of $\mathrm{CaCl_2}$, you are measuring 3.00 moles of $\mathrm{CaCl_2}$.



Figure 12.7.2.4: Calcium chloride is used as a drying agent and as a road deicer. (Credit: Martin Walker (User:Walkerma/Wikimedia Commons); Source: wikimedia; License: Public Domain)

Example 12.7.2.1: Converting Moles to Mass

Chromium metal is used for decorative electroplating of car bumpers and other surfaces. Find the mass of 0.560 moles of chromium.

Solution

Step 1: List the known quantities and plan the problem.

Known

- Molar mass of $\mathrm{Cr} = 52.00\ \mathrm{g/mol}$
- $0.560\ \mathrm{mol\ Cr}$

Unknown

- $0.560\ \mathrm{mol\ Cr}$ = ? g

One conversion factor will allow us to convert from the moles of Cr to mass.

Step 2: Calculate.

$$0.560\ \mathrm{mol\ Cr} \times \frac{52.00\ \mathrm{g\ Cr}}{1\ \mathrm{mol\ Cr}} = 29.1\ \mathrm{g\ Cr}$$

Step 3: Think about your result.

Since the desired amount was slightly more than one half of a mole, the mass should be slightly more than one half of the molar mass. The answer has three significant figures because of the $0.560\ \mathrm{mol}$.

A similar conversion factor utilizing molar mass can be used to convert from the mass of a substance to moles. In a laboratory situation, you may perform a reaction and produce a certain amount of a product which can be massed. It will often then be necessary to determine the number of moles of the product that was formed. The next problem illustrates this situation.

Section Summary

- Calculations are described for the determination of molar mass of an atom or a compound.
- Calculations involving conversions between moles of a material and the mass of that material are described.

12.7.2: Moles and Mass is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

- **10.1: Avogadro's Number** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **10.3: Molar Mass** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **10.4: Conversions Between Moles and Mass** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

12.8: Solutions and Solubility

Learning Objectives

- Explain the effects of pressure and temperature on the solubility in various types of solution.

You might recall that we had introduced the idea of a solution as a type of homogeneous mixture in which the particles that make it up are very tiny. In this section we will discuss some of the properties of solutions.

Solute-Solvent Combinations

Although much of the time we will be focusing on water based solutions, it turns out that you do not need water in order to have a solution. In fact, you do not even necessarily need to have a liquid in order to have a solution. There are many examples of solutions that do not involve water at all, or that involve solutes that are not solids. The table below summarizes the possible combinations of solute-solvent states, along with examples of each.

Table **12.8.1**: Solute-Solvent Combinations

Solute State	Solvent State	Example
liquid	gas	water in air
gas	gas	oxygen in nitrogen (gas mixture)
solid	liquid	salt in water
liquid	liquid	alcohol in water
gas	liquid	carbon dioxide in water
solid	solid	zinc in copper (brass alloy)
liquid	solid	mercury in silver and tin (dental amalgam)

Gas-Gas Solutions

Our air is a homogenous mixture of many different gases and therefore qualifies as a solution. Approximately 78% of the atmosphere is nitrogen, making it the solvent for this solution. The next major constituent is oxygen (about 21%), followed by the inert gas argon (0.9%), carbon dioxide (0.03%), and trace amounts of neon, methane, helium, and other gases.

Solid-Solid Solutions

Solid-solid solutions such as brass, bronze, and sterling silver are called alloys. Bronze (composed mainly of copper with added tin) was widely used in making weapons in times past, dating back to at least 2400 B.C. This metal alloy was hard and tough, but was eventually replaced by iron. Although it is possible for these solutions to exist in the gas phase, they must be formed in the liquid phase.

Liquid-Liquid Solutions

Perhaps the most familiar liquid-solid solution is dental amalgam, used to fill teeth when there is a cavity. Approximately 50% of the amalgam material is liquid mercury to which a powdered alloy of silver, tin, and copper is added. Mercury is used because it binds well with the solid metal alloy. However, the use of mercury-based dental amalgam has gone under question in recent years, because of concerns regarding the toxicity of mercury.

How Temperature Influences Solubility

The **solubility** of a substance is the amount of that substance that is required to form a saturated solution in a given amount of solvent at a specified temperature. Solubility is often measured as the grams of solute per $100\ g$ of solvent. The solubility of sodium chloride in water is $36.0\ g$ per $100\ g$ water at $20°C$. The temperature must be specified because solubility varies with temperature.

For gases, the pressure must also be specified. Solubility is specific for a particular solvent. In this examples considered here, we will consider solubility of material in water as solvent.

The solubility of the majority of solid substances increases as the temperature increases. However, the effect is difficult to predict and varies widely from one solute to another. The temperature dependence of solubility can be visualized with the help of a **solubility curve**, a graph of the solubility vs. temperature (Figure 12.8.1).

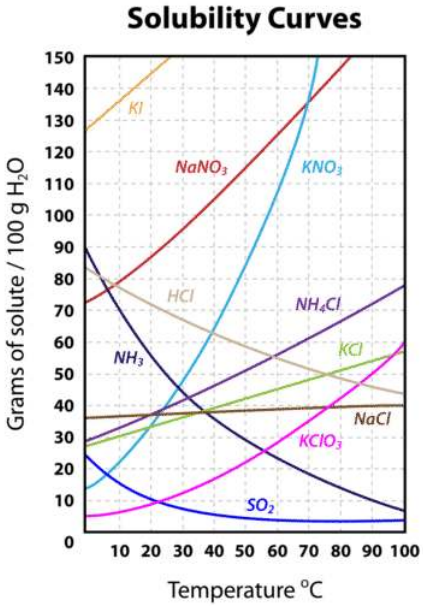


Figure 12.8.1: Solubility curves for several compounds.

Notice how the temperature dependence of $NaCl$ is fairly flat, meaning that an increase in temperature has relatively little effect on the solubility of $NaCl$. The curve for KNO_3, on the other hand, is very steep, and so an increase in temperature dramatically increases the solubility of KNO_3.

Several substances—HCl, NH_3, and SO_2—have solubility that decreases as temperature increases. They are all gases at standard pressure. When a solvent with a gas dissolved in it is heated, the kinetic energy of both the solvent and solute increase. As the kinetic energy of the gaseous solute increases, its molecules have a greater tendency to escape the attraction of the solvent molecules and return to the gas phase. Therefore, the solubility of a gas decreases as the temperature increases.

Henry's Law

Pressure has very little effect on the solubility of solids or liquids, but has a significant effect on the solubility of gases. Gas solubility increases as the partial pressure of a gas above the liquid increases. Suppose a certain volume of water is in a closed container with the space above it occupied by carbon dioxide gas at standard pressure. Some of the CO_2 molecules come into contact with the surface of the water and dissolve into the liquid. Now suppose that more CO_2 is added to the space above the container, causing a pressure increase. In this case, more CO_2 molecules are in contact with the water and so more of them dissolve. Thus, the solubility increases as the pressure increases. As with a solid, the CO_2 that is undissolved reaches an equilibrium with the dissolved CO_2, represented by the equation:

$$CO_2\,(g) \rightleftharpoons CO_2\,(aq)$$

At equilibrium, the rate of gaseous CO_2 dissolution is equal to the rate of dissolved CO_2 coming out of the solution.

When carbonated beverages are packaged, they are done so under high CO_2 pressure so that a large amount of carbon dioxide dissolves in the liquid. When the bottle is open, the equilibrium is disrupted because the CO_2 pressure above the liquid decreases. Immediately, bubbles of CO_2 rapidly exit the solution and escape out of the top of the open bottle. The amount of dissolved CO_2 decreases. If the bottle is left open for an extended period of time, the beverage becomes "flat" as more and more CO_2 comes out of the liquid.

The relationship of gas solubility to pressure is described by Henry's law, named after English chemist William Henry (1774-1836). **Henry's Law** states that the solubility of a gas in a liquid is directly proportional to the partial pressure of the gas above the liquid. Henry's law can be written as follows:

$$\frac{S_1}{P_1} = \frac{S_2}{P_2}$$

Section Summary

- The solubility of a substance is the amount of that substance that is required to form a saturated solution in a given amount of solvent at a specified temperature.
- A solubility curve is a graph of the solubility vs. temperature.
- The solubility of a solid in water increases with an increase in temperature.
- Gas solubility decreases as the temperature increases.
- Gas solubility increases with increasing pressure.

Glossary

solubility

The amount of a substance that is required to form a saturated solution in a given amount of solvent at a specified temperature.

solubility curve

A graph of the solubility vs. temperature.

This page titled 12.8: Solutions and Solubility is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **16.1: Solute-Solvent Combinations** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **16.4: How Temperature Influences Solubility** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **16.6: Henry's Law** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

12.9: Concentrations of Solutions

Learning Objectives

- Calculate the concentration of a solution in either percent concentration or molarity.

Mixtures, including solutions, have variable compositions from one sample to another. Because of this variability, it becomes important in some cases to specify the amount of each substance present in a particular sample of a solution. We will explore some methods of keeping track of the amount of a substance present in a solution in this section.

Saturated and Unsaturated Solutions

Table salt (NaCl) readily dissolves in water. Suppose that you have a beaker of water to which you add some salt, stirring until it dissolves. You add more and that dissolves. You keep adding more and more salt, eventually reaching a point at which no more of the salt will dissolve—no matter how long or how vigorously you stir it. Why? On the molecular level, we know that action of the water causes the individual ions to break apart from the salt crystal and enter the solution, where they remain hydrated by water molecules. What also happens is that some of the dissolved ions collide back again with the crystal and remain there. **Recrystallization** is the process of dissolved solute returning to the solid state. At some point, the rate at which the solid salt is dissolving becomes equal to the rate at which the dissolved solute is recrystallizing. When that point is reached, the total amount of dissolved salt remains unchanged. **Solution equilibrium** is the physical state described by the opposing processes of dissolution and recrystallization occurring at the same rate. The solution equilibrium for the dissolution of sodium chloride can be represented by one of two equations:

$$\mathrm{NaCl}\,(s) \rightleftharpoons \mathrm{NaCl}\,(aq)$$

While the above equation shows the change of state back and forth between solid and aqueous solution, the preferred equation also shows the dissociation that occurs as an ionic solid dissolves:

$$\mathrm{NaCl}\,(s) \rightleftharpoons \mathrm{Na^+}\,(aq) + \mathrm{Cl^-}\,(aq)$$

When the solution equilibrium point is reached and no more solute will dissolve, the solution is said to be saturated. A **saturated solution** is a solution that contains the maximum amount of solute that is capable of dissolving. At $20°\mathrm{C}$, the maximum amount of NaCl that will dissolve in $100.\ \mathrm{g}$ of water is $36.0\ \mathrm{g}$. If any more NaCl is added past this point, it will not dissolve, because the solution is saturated. What if more water is added to the solution instead? In that case, more NaCl would be capable of dissolving in the additional solvent. An **unsaturated solution** is a solution that contains less than the maximum amount of solute that is capable of being dissolved. The figure below illustrates the above process and shows the distinction between unsaturated and saturated.

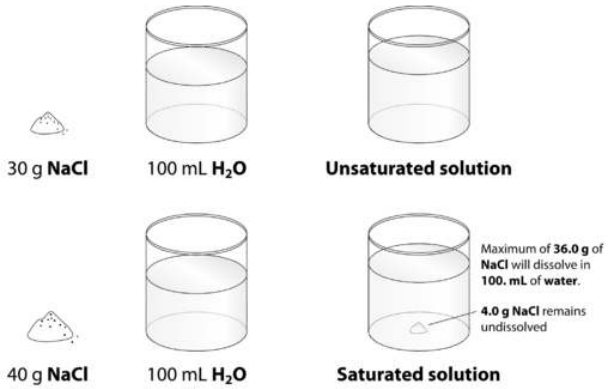


Figure 12.9.1: When $30.0\ \mathrm{g}$ of NaCl is added to $100\ \mathrm{mL}$, it all dissolves, forming an unsaturated solution. When $40.0\ \mathrm{g}$ is added, $36.0\ \mathrm{g}$ dissolves and $4.0\ \mathrm{g}$ remains undissolved, forming a saturated solution.

How can you tell if a solution is saturated or unsaturated? If more solute is added and it does not dissolve, then the original solution was saturated. If the added solute dissolves, then the original solution was unsaturated. A solution that has been allowed to reach equilibrium, but which has extra undissolved solute at the bottom of the container, must be saturated.

For solutions that are saturated, the amount of a substance present in it can be obtained from tabulated data for that substance at the given pressure and temperature. But most of the time we are not dealing with a saturated solution, so we want to be able to express this more generally.

Percent Solutions

One way to describe the concentration of a solution is by the percent of a solute in the solvent. The percent can further be determined in one of two ways: (1) the ratio of the mass of the solute divided by the mass of the solution or (2) the ratio of the volume of the solute divided by the volume of the solution.

Mass Percent

When the solute in a solution is a solid, a convenient way to express the concentration is by mass percent $\left(\frac{\text{mass}}{\text{mass}}\right)$, which is the grams of solute per $100\ \text{g}$ of solution.

$$\text{Percent by mass} = \frac{\text{mass of solute}}{\text{mass of solution}} \times 100\%$$

Suppose that a solution was prepared by dissolving $25.0\ \text{g}$ of sugar into $100\ \text{g}$ of water. The percent by mass would be calculated by:

$$\text{Percent by mass} = \frac{25\ \text{g sugar}}{125\ \text{g solution}} \times 100\% = 20\%\ \text{sugar}$$

Sometimes you may want to make up a particular mass of solution of a given percent by mass, and need to calculate what mass of the solvent to use. For example, you need to make $3000\ \text{g}$ of a 5% solution of sodium chloride. You can rearrange and solve for the mass of solute:

$$\text{mass of solute} = \frac{\text{percent by mass}}{100\%} \times \text{mass of solution} = \frac{5\%}{100\%} \times 3000\ \text{g} = 150\ \text{g NaCl}$$

You would need to weigh out $150\ \text{g}$ of NaCl and add it to $2850\ \text{g}$ of water. Notice that it was necessary to subtract the mass of the NaCl $(150\ \text{g})$ from the mass of solution $(3000\ \text{g})$ to calculate the mass of the water that would need to be added.

Volume Percent

The percentage of solute in a solution can more easily be determined by volume when the solute and solvent are both liquids. The volume of the solute divided by the volume of the solution, expressed as a percent, yields the percent by volume $\left(\frac{\text{volume}}{\text{volume}}\right)$ of the solution. If a solution is made by adding $40\ \text{mL}$ of ethanol to $200\ \text{mL}$ of water, the percent by volume is:

$$\begin{aligned}\text{Percent by volume} &= \frac{\text{volume of solute}}{\text{volume of solution}} \times 100\% \\ &= \frac{40\ \text{mL ethanol}}{240\ \text{mL solution}} \times 100\% \\ &= 16.7\%\ \text{ethanol}\end{aligned}$$

Frequently, ingredient labels on food products and medicines have amounts listed as percentages (see figure below).

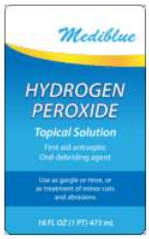


Figure 12.9.1: Hydrogen peroxide is commonly labeled as a 3% by volume solution for use as a disinfectant.

Molarity

Chemists primarily need the concentration of solutions to be expressed in a way that accounts for the number of particles that react according to a particular chemical equation. Since percentage measurements are based on either mass or volume, they are generally

not useful for chemical reactions. A concentration unit based on moles is preferable. The **molarity** (M) of a solution is the number of moles of solute dissolved in one liter of solution. To calculate the molarity of a solution, divide the moles of solute by the volume of the solution expressed in liters:

$$\text{Molarity (M)} = \frac{\text{moles of solute}}{\text{liters of solution}} = \frac{\text{mol}}{\text{L}}$$

Note that the volume is in liters of solution and not liters of solvent. When a molarity is reported, the unit is the symbol M and is read as "molar". For example a solution labeled as $1.5\ \text{M}\ NH_3$ is read as "1.5 molar ammonia solution".

Example 12.9.1

A solution is prepared by dissolving $42.23\ \text{g}$ of NH_4Cl into enough water to make $500.0\ \text{mL}$ of solution. Calculate its molarity.

Solution

Step 1: List the known quantities and plan the problem.

Known

- Mass $= 42.23\ \text{g}\ NH_4Cl$
- Molar mass $NH_4Cl = 53.50\ \text{g/mol}$
- Volume solution $= 500.0\ \text{mL} = 0.5000\ \text{L}$

Unknown

The mass of the ammonium chloride is first converted to moles. Then the molarity is calculated by dividing by liters. Note that the given volume has been converted to liters.

Step 2: Solve.

$$42.23\ \text{g}\ NH_4Cl \times \frac{1\ \text{mol}\ NH_4Cl}{53.50\ \text{g}\ NH_4Cl} = 0.7893\ \text{mol}\ NH_4Cl$$

$$\frac{0.7893\ \text{mol}\ NH_4Cl}{0.5000\ \text{L}} = 1.579\ \text{M}$$

Step 3: Think about your result.

The molarity is $1.579\ \text{M}$, meaning that a liter of the solution would contain $1.579\ \text{mol}\ NH_4Cl$.

In a laboratory situation, a chemist must frequently prepare a given volume of solutions of a known molarity. The task is to calculate the mass of the solute that is necessary. The molarity equation can be rearranged to solve for moles, which can then be converted to grams.

Example 12.9.2

A chemist needs to prepare $3.00\ \text{L}$ of a $0.250\ \text{M}$ solution of potassium permanganate ($KMnO_4$). What mass of $KMnO_4$ does she need to make the solution?

Solution

Step 1: List the known quantities and plan the problem.

Known

- Molarity $= 0.250\ \text{M}$
- Volume $= 3.00\ \text{L}$
- Molar mass $KMnO_4 = 158.04\ \text{g/mol}$

Unknown

Moles of solute is calculated by multiplying molarity by liters. Then, moles is converted to grams.

Step 2: Solve.

$$\text{mol KMnO}_4 = 0.250\ \text{M KMnO}_4 \times 3.00\ \text{L} = 0.750\ \text{mol KMnO}_4$$

$$0.750\ \text{mol KMnO}_4 \times \frac{158.04\ \text{g KMnO}_4}{1\ \text{mol KMnO}_4} = 119\ \text{g KMnO}_4$$

Step 3: Think about your result.

When $119\ \text{g}$ of potassium permanganate is dissolved into water to make $3.00\ \text{L}$ of solution, the molarity is $0.250\ \text{M}$.

Section Summary

- Techniques for calculation of percent mass and percent volume solution concentrations are described.
- The molarity (M) of a solution is the number of moles of solute dissolved in one liter of solution.
- Calculations using the concept of molarity are described.

Glossary

recrystallization

The process of dissolved solute returning to the solid state.

solution equilibrium

The physical state described by the opposing processes of dissolution and recrystallization occurring at the same rate.

saturated solution

A solution that contains the maximum amount of solute that is capable of dissolving.

unsaturated solution

A solution that contains less than the maximum amount of solute that is capable of being dissolved.

concentration

Measure of the amount of solute that has been dissolved in a given amount of solvent or solution.

This page titled 12.9: Concentrations of Solutions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **16.3: Saturated and Unsaturated Solutions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **16.7: Percent Solutions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **16.8: Molarity** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

12.10: Soap- an Application of Physical Properties

Learning Objectives

- Describe the mechanism by which soaps exert their cleansing action.

Physical properties are all around us. Most of the matter that we encounter is a mixture of some sort. Because we know that the matter exists as mixtures, we are able to purify it if necessary using physical means. Evaporation of sweat from our bodies cools us off as our body heat provides the energy for the phase change. Henry's Law explains the carbonation of carbonated beverages, as well as why there are more fish to catch in shady water. We will focus on a few of the ways physical properties are important in our lives.

Personal Cleanliness

Soap is a salt of a fatty acid used in a variety of cleansing and lubricating products. In a domestic setting, soaps are surfactants usually used for washing, bathing, and other types of housekeeping. In industrial settings, soaps are used as thickeners, components of some lubricants, and precursors to catalysts.

When used for cleaning, soap solubilizes particles and grime, which can then be separated from the article being cleaned. In hand washing, as a surfactant, when lathered with a little water, soap kills microorganisms by disorganizing their membrane lipid bilayer and denaturing their proteins. It also emulsifies oils, enabling them to be carried away by running water.[2]

Industrially manufactured **bar soaps** became available in the late 18th century, as advertising campaigns in Europe and America promoted popular awareness of the relationship between cleanliness and health.[35] In modern times, the use of soap has become commonplace in industrialized nations due to a better understanding of the role of hygiene in reducing the population size of pathogenic microorganisms (Figure **12.10.1**). The details of how soap is created are beyond the scope of this text, but we can discuss how it works in a bit of detail.



Figure **12.10.1** Modern Soap Shop in Tubeingen , Germany. Source: Wikipedia

The cleansing action of soap is determined by its polar and non-polar structures in conjunction with an application of solubility principles. The long hydrocarbon chain is of course non-polar and **hydrophobic** (repelled by water). The "salt" end of the soap molecule is ionic and **hydrophilic** (water soluble). Examples of soap and detergent molecules, are shown in Figure **12.10.4**. The use of such compounds as cleaning agents is facilitated by their surfactant character, which lowers the surface tension of water, allowing it to penetrate and wet a variety of materials.

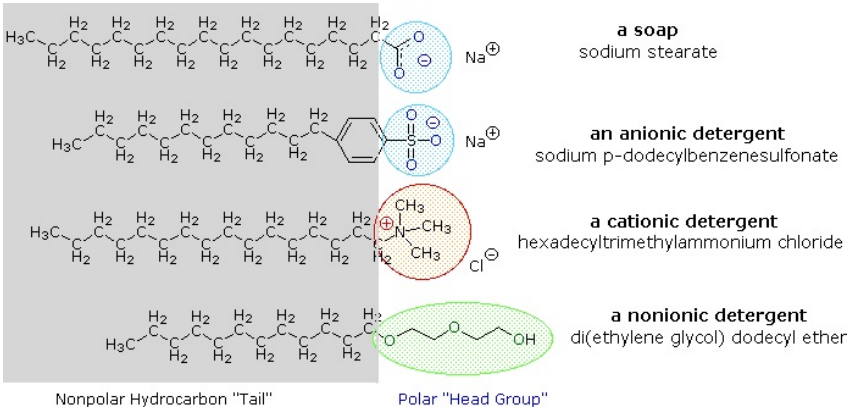


Figure 12.10.4 Soap and detergent molecules with different polar structures.

When soap is added to water, the ionic-salt end of the molecule is attracted to water and dissolved in it. The non-polar hydrocarbon end of the soap molecule is repelled by water. A drop or two of soap in water forms a monolayer (Figure 12.10.5) on the water surface as shown in the graphics on the left. The soap molecules "stand up" on the surface as the polar carboxyl salt end is attracted to the polar water. The non-polar hydrocarbon tails are repelled by the water, which makes them appear to stand up.

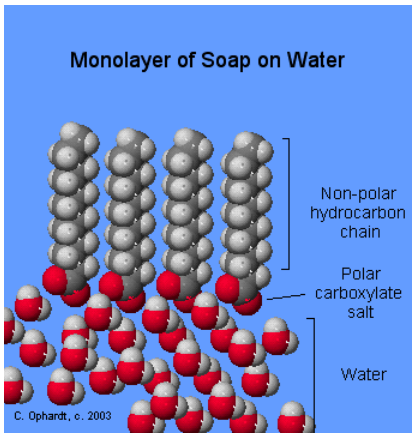


Figure 12.10.5 Monolayer of soap on water.

Water alone is not able to penetrate grease or oil because they are of opposite polarity. When grease or oil (non-polar hydrocarbons) are mixed with a soap- water solution, the soap molecules work as a "bridge" between polar water molecules and non-polar oil molecules. Soap molecules have both properties of non-polar and polar at opposite ends of the molecule.

The oil is a pure hydrocarbon so it is non-polar. The non-polar hydrocarbon tail of the soap dissolves into the oil. That leaves the polar carboxylate ion of the soap molecules are sticking out of the oil droplets, the surface of each oil droplet is negatively charged. As a result, the oil droplets repel each other and remain suspended in solution (this is called an emulsion) to be washed away by a stream of water. The outside of the droplet is also coated with a layer of water molecules.

The graphic on Figure 12.10.6 although not strictly a representation of the above description is a micelle that works in much the same fashion. The oil would be a the center of the micelle.

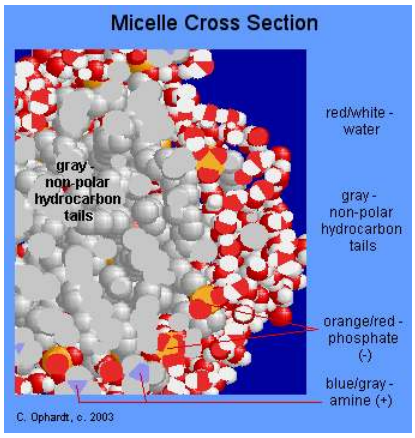


Figure **12.10.6** A micelle with oil at the center.

The followign video about the power of soap to kill the coronavirus was recorded in the spring of 2020. Soaps are quite effective at killing viruses.

Can Soap Really Kill the Coronavirus ?

***Video* 12.10.1** *Soap and coronavirus.*

Section Summary

- Soap, as a cleaning agent, solubilizes particles and grime, which can then be easily removed from the article being cleaned.
- Soap works by having some parts of the soap that dissolve in water and others that dissolve in oils.

Glossary

soap
: A salt of a fatty acid used in a variety of cleansing and lubricating products.

hydrophobic
: Repelled by water.

hydrophilic
: Water soluble.

Contributors

- Remixed by Jamie MacArthur
- Wikipedia
- Libretext: Organic Chemistry (McMurry)

- William Reusch, Professor Emeritus (Michigan State U.), Virtual Textbook of Organic Chemistry
- Charles Ophardt, Professor Emeritus, Elmhurst College; Virtual Chembook(opens in new window)
- Andrea Kubisch, Courtney Korff (UCD)
- Marisa Alviar-Agnew (Sacramento City College)

This page titled 12.10: Soap- an Application of Physical Properties is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **21.1: Cleaning with Soap** is licensed CC BY-NC-SA 4.0.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

13: Chemical Reactions

13.1: Introduction to Chemical Changes
13.2: Aspects of a Chemical Change
13.3: Chemical Equations- Representing Chemical Changes
13.4: Balancing Equations
13.5: Energy and Chemical Reactions
13.6: Entropy and Chemical Reactions
13.7: Classifying Reactions
13.8: Acid Base Reactions
13.9: Redox Reactions- The Transfer of Electrons
13.10: Polymerization Reactions

This page titled 13: Chemical Reactions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

13.1: Introduction to Chemical Changes



Figure **13.1.1** (Credit: Jon Sullivan; Source; License: Public Domain)

Do you like to cook?

Cooking is a valuable skill that can benefit everyone. Whether it is fixing a simple grilled cheese sandwich or preparing an elaborate meal, cooking demonstrates some basic concepts of chemistry. When you bake bread, you mix flour, sugar, yeast, and water together. After baking, this mixture has been changed to form bread, another substance that has different characteristics and qualities from the original materials. The process of baking has produced chemical changes in the ingredients that result in bread being made.

Chemical Change

Most of the elements we know about do not exist freely in nature. Sodium cannot be found by itself (unless we prepare it in the laboratory) because it interacts too easily with other materials. On the other hand, the element helium does not interact with other elements to any extent. We can isolate helium from natural gas during the process of drilling for oil.

A chemical change produces different materials than the ones we started with. One aspect of the science of chemistry is the study of the changes that matter undergoes. If chemistry was simply a study of elements that did nothing, life would be very boring (in fact, life would not exist since the elements are what make up our bodies and sustain us). But the processes of change that take place when different chemicals are combined produce all the materials that we use daily.

One type of chemical change (already mentioned) is when two elements combine to form a compound. Another type involves the breakdown of a compound to produce the elements that make it up. If we pass an electric current through bauxite (aluminum oxide, the raw material for aluminum metal), we get metallic aluminum as a product.



Figure **13.1.1**: Electrolytic production of aluminum. (Credit: Ronny Jacques; Source; License: Public Domain)

However, the vast majority of chemical changes involve one compound being transformed into another compound. There are literally millions of possibilities when we take this approach to chemical change. New compounds can be made to produce better fabrics that are easier to clean and maintain; they can help preserve food so it doesn't spoil as quickly; we can make new medicines to treat diseases; even products, such as airbags (as seen in the simulation below), use a chemical change to improve the world around us. In this chapter we will learn more about these chemical changes: how to keep track of them with equations, how to balance those equations, how to categorize them, and how to think about the energy involved with them. We will end with some important examples of reaction types.

This page titled 13.1: Introduction to Chemical Changes is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.14: Chemical Change** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

13.2: Aspects of a Chemical Change

Learning Objectives

- Identify chemical properties and chemical changes.
- Determine if a chemical change has occurred based on observations.

Have you ever left your bicycle out in the rain?

It is not a good idea because the rain can cause the bicycle to rust. You start to get a reddish-orange build-up of a rough deposit on the metal. It may start with the chain, but can spread to other parts of the bicycle, especially if there are scratches that create a bare metal surface. The formation of rust is a chemical process that takes place when iron is exposed to water and oxygen. It is estimated that damage due to rust costs U.S. businesses, military, and government over 276 billion dollars a year—a very expensive chemical process.

Chemical Properties

A **chemical property** describes the ability of a substance to undergo a specific chemical change. A chemical property of iron is that it is capable of combining with oxygen to form iron oxide, the chemical name of rust. The more general term for rusting and other similar processes is corrosion. Other terms that are commonly used to describe chemical changes are: burn, rot, explode, decompose, and ferment. Chemical properties are very useful in identifying substances. However, unlike physical properties, chemical properties can only be observed as the substance is in the process of being changed into a different substance.

Chemical Change

A **chemical change** is also called a chemical reaction. A chemical reaction is a process that occurs when one or more substances are changed into one or more new substances. Zinc (Zn) is a silver gray element that can be ground into a powder. If zinc is mixed at room temperature with powdered sulfur (S), a bright yellow element, the result will simply be a mixture of zinc and sulfur. No chemical reaction occurs. However, if energy is provided to the mixture in the form of heat, the zinc will chemically react with the sulfur to form the compound zinc sulfide (ZnS). The figure below shows the substances involved in this reaction.

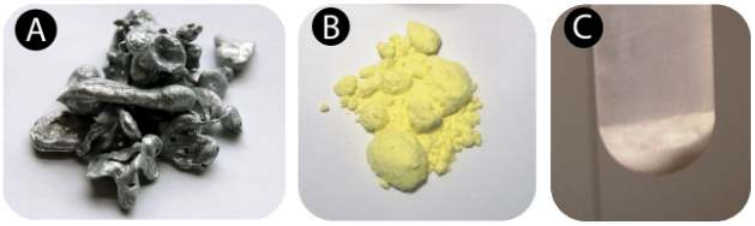


Figure 13.2.1: Zinc and sulfur are two elements that undergo a chemical reaction when heated to form the compound zinc sulfide. (Credit: (A) Ben Mills (Benjah-bmm27); (B) User:Vineyard/Wikimedia Commons; (C) Courtesy of A. Eisen/NOAA; Source: (A); (B); (C); License: (A) Public Domain; (B) CC by 3.0(opens in new window); (C) Public Domain

Recognizing Chemical Reactions

How can a scientist tell if a **chemical reaction** is taking place? There are four visual clues that indicate that a chemical reaction is likely occurring:

1. A change in color occurs during the reaction.
2. A **gas** is produced during the reaction.
3. A solid product called a **precipitate** is produced in the reaction.
4. A transfer of energy occurs as a result of the reaction.

Mercury (II) oxide is a red solid. When it is heated to a temperature above $500°\mathrm{C}$, it easily decomposes into mercury and oxygen gas. The red color of the mercury oxide reactant becomes the silver color of mercury. The color change is a sign that the reaction is occurring.



Video **13.2.1**: Mercuric oxide decomposing into mercury and oxygen (Source)

When zinc reacts with hydrochloric acid, the reaction bubbles vigorously as hydrogen gas is produced. The production of a gas is also an indication that a chemical reaction is occurring.



Video **13.2.2**: Zinc reacting with hydrochloric acid produces bubbles of hydrogen gas. (Source)

When a colorless solution of lead (II) nitrate is added to a colorless solution of potassium iodide, a yellow solid called a precipitate is instantly produced. A precipitate is a solid product that forms from a reaction and settles out of a liquid solution. The formation of a precipitate is an indication of a chemical reaction.



Video**13.2.3**: A yellow precipitate of solid lead (II) iodide forms immediately when solutions of lead (II) nitrate and potassium iodide are mixed. (Source)

All chemical changes involve a transfer of energy. When zinc reacts with hydrochloric acid, the test tube becomes very warm as energy is released during the reaction. Some other reactions absorb energy. While energy changes are a potential sign of a chemical reaction, care must be taken to ensure that a chemical reaction is indeed taking place. Physical changes also involve a transfer of energy. The melting of a solid absorbs energy, while the condensation of a gas releases energy. The only way to be certain that a

chemical reaction has taken place is to test the composition of the substances after the change has taken place, to see if they are different from the starting substances.

The following videos show some additional examples of chemical changes occurring.



Section Summary

- A chemical property describes the ability of a substance to undergo a specific chemical change.
- A chemical reaction is a process that occurs when one or more substances are changed into one or more new substances.
- There are four visual clues that indicate that a chemical reaction is likely occurring:
  - A change of color occurs during the reaction.
  - A gas is produced during the reaction.
  - A solid product called a precipitate is produced in the reaction.
  - A transfer of energy occurs as a result of the reaction.

Glossary

chemical property

Describes the ability of a substance to undergo a specific chemical change.

chemical change (chemical reaction)

A process that occurs when one or more substances are changed into one or more new substances.

precipitate

A solid product that occurs in some reactions.

This page titled 13.2: Aspects of a Chemical Change is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.16: Chemical Properties and Chemical Reactions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.18: Recognizing Chemical Reactions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

13.3: Chemical Equations- Representing Chemical Changes

Learning Objectives

- Represent chemical changes using both words and symbols.

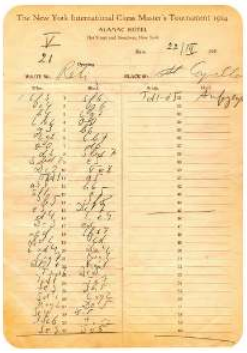


Figure **13.3.1** (Author Unknown; Source: Commons Wikimedia, Capablanca(opens in new window) [commons.wikimedia.org]; License: Public Domain)

How do chess players monitor their moves in a game?

Suppose you were walking along and noticed a piece of paper on the ground with markings on it. You pick it up and see the paper in the picture above.

To most people, these notes are meaningless (maybe they're a secret spy code). But to a chess player, these symbols tell the story of a chess game. Each abbreviation describes a chess piece or a move during the game. The use of special symbols allows chess players to "see" the game without having to read a wordy and possibly incomplete description of what happened. We will learn similar tactics for analyzing the results of chemical equations.

Chemical Symbols and Formulas

In order to illustrate chemical reactions and the elements and **compounds** involved in them, chemists use symbols and formulas. A **chemical symbol** is a one- or two-letter designation of an element. Compounds are combinations of two or more elements. A **chemical formula** is an expression that shows the elements in a compound and the relative proportions of those elements. Please review this topic before moving forward, as it will be important for understanding what comes next.

Chemical Equation

The reaction between zinc and sulfur can be shown in what is called a **chemical equation**. In words, we could write the reaction as:

$$\text{zinc} + \text{sulfur} \rightarrow \text{zinc sulfide}$$

The more convenient way to express a chemical reaction is to use the symbols and formulas of the substances involved:

$$\text{Zn} + \text{S} \rightarrow \text{ZnS}$$

The substance(s) to the left of the arrow in a chemical equation are called reactants. A **reactant** is a substance that is present at the start of a chemical reaction. The substance(s) to the right of the arrow are called **products**. A product is a substance that is present at the end of a chemical reaction. In the equation above, the zinc and sulfur are the reactants that chemically combine to form the zinc sulfide product.

There is a standard way of writing chemical equations. The reactants are all written on the left-hand side of the equation, with the products on the right-hand side. Unlike a math equation, a chemical equation does not use an equal sign. Instead, the arrow is called a yield sign and so the equation is described as "reactants yield products".

$$\text{reactants} \rightarrow \text{products}$$

The presence of the arrow also indicates that the reaction goes in one direction under the conditions indicated. There are reactions which can be easily reversed, but we will not take those up right now.

There are a wide variety of reactions possible: elements may form compounds (as seen in the reaction above), compounds may form elements (water will break down in the presence of an electric current to form hydrogen gas and oxygen gas) or compounds may combine, break apart, or rearrange to form new materials.

It is often important to know the physical states of the reactants and products taking part in a reaction. To do this, put the appropriate symbol in parentheses after each formula: (s) for solid, (l) for liquid, (g) for gas, and (aq) for an aqueous (water-based) solution. The previous reaction becomes:

$$\mathrm{CH_4}\,(g) + \mathrm{O_2}\,(g) \rightarrow \mathrm{CO_2}\,(g) + \mathrm{H_2O}\,(l)$$

The table below shows a listing of symbols used in chemical equations. Some, such as the double arrow which represents equilibrium, and the use of a catalyst in a reaction, will be detailed in other concepts.

Table **13.3.1**: Symbols Used in Chemical Equations

Symbols	Description
$+$	Used to separate multiple reactants or products.
$\rightarrow$	Yield sign; separates reactants from products.
$\rightleftharpoons$	Replaces the yield sign for reversible reactions that reach equilibrium.
(s)	Reactant or product in the solid state.
(l)	Reactant or product in the liquid state.
(g)	Reactant or product in the gaseous state.
(aq)	Reactant or product in an aqueous solution (dissolved in water).
$\xrightarrow{\mathrm{Pt}}$	Formula written above the arrow is used as a catalyst in the reaction.
$\xrightarrow{\Delta}$	Triangle indicates that the reaction is being heated.

Word Equations

You can describe a chemical reaction by writing a **word equation**. When silver metal is exposed to sulfur, it reacts to form silver sulfide. Silver sulfide is commonly known as tarnish and turns the surface of silver objects dark and streaky black (see figure below). The sulfur that contributes to tarnish can come from traces of sulfur in the air, or from foods such as eggs. The word equation for the process is:

$$\text{Silver} + \text{sulfur} \rightarrow \text{Silver sulfide}$$

The silver and the sulfur are the reactants in the equation, while the silver sulfide is the product.



Figure **13.3.2:** You can see dark spots of tarnished (silver sulfide) forming on this ring as it reacts with sulfur compounds in the air. (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CK-12 Curriculum Materials license)

Another common chemical reaction is the burning of methane gas. Methane is the major component of natural gas and is commonly burned on a gas stove or in a Bunsen burner (figure below). Burning is a chemical reaction in which some type of fuel is

reacted with oxygen gas. The products of the reaction in the burning of methane, as well as other fuels, are carbon dioxide and water. The word equation for this reaction is:

$$\text{Methane} + \text{oxygen} \rightarrow \text{carbon dioxide} + \text{water}$$

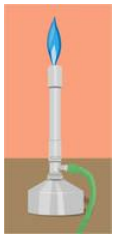


Figure 13.3.3: A Bunsen burner is commonly used to heat substances in a chemistry lab. Methane is reacted with oxygen to form carbon dioxide and water. (Credit: CK-12 Foundation; Source: CK-12 Foundation; License: CK-12 Curriculum Materials license)

Word equations can be very useful, but do have one major drawback—they cannot be used for any quantitative work. A word equation does not tell how many moles of each material are needed, or how many moles of product are formed. This is something we will consider in more detail in the next subsection.

The following video shows how cooking is a form of chemical reaction, as we stated at the introduction to this chapter.



Section Summary

- A chemical symbol is a one- or two-letter designation of an element.
- A chemical formula is an expression that shows the elements in a compound and the relative proportions of those elements.
- A chemical equation describes a chemical reaction.
- Reactants are starting materials and are written on the left-hand side of the equation.
- Products are the end result of the reaction and are written on the right-hand side of the equation.
- A chemical equation is representation a chemical reaction using chemical formulas reactants and products, as well as symbols.
- Symbols used in chemical equations are described and explained.
- Word equations are used to describe the conversion of reactants to products.

Glossary

product

A substance that is present at the end of a chemical reaction.

reactant

A substance that is present at the start of a chemical reaction.

chemical equation

The use of symbols and formulas to express a chemical change.

This page titled 13.3: Chemical Equations- Representing Chemical Changes is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **2.15: Chemical Symbols and Formulas** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **2.17: Reactants and Products** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **11.2: Chemical Equations** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **11.1: Word Equations** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

13.4: Balancing Equations

Learning Objectives

- Balance a chemical equation.

Any leftovers?

When you cook a meal, quite often there are leftovers because you prepared more than people would eat at one sitting. Sometimes when you repair a piece of equipment, you end up with what are called "pocket parts"—small pieces you put in your pocket because you're not sure where they belong. Chemistry tries to avoid leftovers and pocket parts. In normal chemical processes, we cannot create or destroy matter (law of conservation of mass). If we start out with ten carbon atoms, we need to end up with ten carbon atoms. John Dalton's atomic theory said that chemical reactions basically involve the rearrangement of atoms. Chemical equations need to follow these principles in order to be correct.

Balancing Chemical Equations

A **balanced equation** is a chemical equation in which mass is conserved and there are equal numbers of atoms of each element on both sides of the equation. We can write a chemical equation for the reaction of carbon with hydrogen gas to form methane ($\mathbf{CH_4}$):

$$\underset{\text{2 C atoms}}{\mathbf{C}\,(s)} + \underset{\text{2 H atoms}}{\mathbf{H_2}\,(g)} \rightarrow \underset{\text{1 C atom, 4 H atoms}}{\mathbf{CH_4}\,(g)}$$

In order to write a correct equation, you must first write the correct skeleton equation with the correct chemical formulas. Recall that hydrogen is a diatomic molecule and so is written as $\mathbf{H_2}$.

When we count the number of atoms of both elements, shown under the equation, we see that the equation is not balanced. There are only 2 atoms of hydrogen on the reactant side of the equation, while there are 4 atoms of hydrogen on the product side. We can balance the above equation by adding a coefficient of 2 in front of the formula for hydrogen.

$$\mathbf{C}\,(s) + \mathbf{2H_2}\,(g) \rightarrow \mathbf{CH_4}\,(g)$$

A **coefficient** is a small in magnitude whole number placed in front of a formula in an equation to signify the quantity needed to balance it. The 2 in front of the $\mathbf{H_2}$ means that there are a total of $\mathbf{2 \times 2 = 4}$ atoms of hydrogen as reactants. Visually, the reaction looks like the figure below.

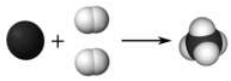


Figure **13.4.2**: Reaction between carbon and hydrogen to form methane. (Credit: Ben Mills (Wikimedia: Benjah-bmm27; Source, and Source); License: Public Domain)

In the balanced equation, there is one atom of carbon and four atoms of hydrogen on both sides of the arrow. Below are guidelines for writing and balancing chemical equations.

1. Determine the correct chemical formulas for each reactant and product.
2. Write the skeleton equation.
3. Count the number of atoms of each element that appears as a reactant and as a product. If a polyatomic ion is unchanged on both sides of the equation, count it as a unit.
4. Balance each element one at a time by placing coefficients in front of the formulas.
    a. It is best to begin by balancing elements that only appear in one chemical formula on each side of the equation.
    b. It is best to end with an element by itself (if possible).
    c. No coefficient is written for a 1.
    d. NEVER change the subscripts in a chemical formula—you can only balance equations by using coefficients.
5. Check each atom or polyatomic ion to be sure that they are equal on both sides of the equation.
6. Make sure that all coefficients are in the lowest possible ratio. If necessary, reduce to the lowest ratio.

Example 13.4.1: Balancing Chemical Equations

Aqueous solutions of lead (II) nitrate and sodium chloride are mixed. The products of the reaction are an aqueous solution of sodium nitrate and a solid precipitate of lead (II) chloride. Write the balanced chemical equation for this reaction.

Solution

Step 1: Plan the problem.

Follow the steps for writing and balancing a chemical equation listed in the text.

Step 2: Solve.

Write the skeleton equation with the correct formulas.

$$\mathrm{Pb(NO_3)_2}\,(aq) + \mathrm{NaCl}\,(aq) \rightarrow \mathrm{NaNO_3}\,(aq) + \mathrm{PbCl_2}\,(s)$$

Count the number of each atom or polyatomic ion on both sides of the equation.

Reactants	Products
1 Pb atom	1 Pb atom
2 NO_3^- ions	1 NO_3^- ions
1 Na atom	1 Na atom
1 Cl atom	2 Cl atoms

The nitrate ions and the chlorine atoms are unbalanced. Start by placing a 2 in front of the NaCl. This increases the reactant counts to 2 Na atoms and 2 Cl atoms. Then place a 2 in front of the $\mathrm{NaNO_3}$. The result is:

$$\mathrm{Pb(NO_3)_2}\,(aq) + 2\mathrm{NaCl}\,(aq) \rightarrow 2\mathrm{NaNO_3}\,(aq) + \mathrm{PbCl_2}\,(s)$$

The new count for each atom and polyatomic ion becomes:

Reactants	Products
1 Pb atom	1 Pb atom
2 NO_3^- ions	2 NO_3^- ions
2 Na atom	2 Na atom
2 Cl atom	2 Cl atoms

Step 3: Think about your result.

The equation is now balanced since there are equal numbers of atoms of each element on both sides of the equation.

MISCONCEPTION ALERT: YOU CAN CHANGE ANY NUMBER YOU WANT TO BALANCE AN EQUATION

The substances in a balanced equation are described by their chemical formula. The chemical formula will contain subscripts to indicate the number of atoms of a given element if that substance has more than one atom of that element in its formula. If we change subscripts in an attempt to balance an equation, we are actually changing the identity of the substances in the chemical reaction that the balanced equation is about. Imagine if you wanted a drink of water and when you put it in your mouth realized that it was actually hydrogen peroxide! The difference between water (H_2O) and hydrogen peroxide (H_2O_2) is significant, but the difference in the chemical formula is merely a subscript. Any time you change the subscript in a chemical equation you are making a similar change to the substances involved. This isn't balancing the equation, it is saying something else instead of answering the question.



Section Summary

- The process of balancing chemical equations is described.

This page titled 13.4: Balancing Equations is shared under a CK-12 license and was authored, remixed, and/or curated by CK-12 Foundation via source content that was edited to the style and standards of the LibreTexts platform.

- **11.3: Balancing Equations** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **10.2: Magnets** by OpenStax is licensed CC BY 4.0.

13.5: Energy and Chemical Reactions

Learning Objectives

- Describe how energy might interact with a chemical reaction, both graphically and in words.

Gunpowder was originally developed by the Chinese in the ninth century CE, primarily for rockets. This material is composed of charcoal, sulfur, and saltpeter (potassium nitrate). The explosive reaction that occurs involves the conversion of the charcoal to carbon dioxide, with the potassium nitrate providing the extra oxygen needed for a rapid reaction. Although sulfur was included to stabilize the product, gunpowder is still highly explosive.

Chemical Potential Energy

As was discussed earlier in this text, energy is the capacity for doing work or supplying heat. When you fill your car with gasoline, you are providing it with potential energy. **Chemical potential energy** is the energy stored in the chemical bonds of a substance. The various chemicals that make up gasoline contain a large amount of chemical potential energy that is released when the gasoline is burned in a controlled way in the engine of the car. The release of that energy does two things: some of the potential energy is transformed into work, which is used to move the car; at the same time, some of the potential energy is converted to heat and makes the car's engine very hot. The energy changes of a system occur as either heat or work, or some combination of both.



Figure **13.5.1**: A dragster is able to accelerate because of the chemical potential energy of its fuel. The burning of the fuel also produces large amounts of heat. (CC BY-NC; CK-12)

Dynamite is another example of chemical potential energy. The major component of dynamite is nitroglycerin, a very unstable material. By mixing it with diatomaceous earth, the stability is increased and it is less likely to explode if it receives a physical shock. When ignited, the nitroglycerin explodes rapidly, releasing large amounts of nitrogen and other gases along with a massive amount of heat.



Figure **13.5.2**: Dynamite explosion. (CC BY-NC; CK-12)

Exothermic and Endothermic Processes

When physical or chemical changes occur, they are generally accompanied by a transfer of energy. The **law of conservation of energy** states that in any physical or chemical process, energy is neither created nor destroyed. In other words, the entire energy in the universe is conserved. In order to better understand the energy changes taking place during a reaction, we need to define two parts of the universe: *the system* and *the surroundings*. The **system** is the specific portion of matter in a given space that is being studied during an experiment or an observation. The **surroundings** is everything in the universe that is not part of the system. In practical terms for a laboratory chemist, the system is the particular chemicals being reacted, while the surroundings is the immediate vicinity within the room. During most processes, energy is exchanged between the system and the surroundings. If the system loses a certain amount of energy, that same amount of energy is gained by the surroundings. If the system gains a certain amount of energy, that energy is supplied by the surroundings.

A chemical reaction or physical change is **endothermic** if heat is absorbed by the system from the surroundings. In the course of an endothermic process, the system gains heat from the surroundings, and so the temperature of the surroundings decreases. The quantity of heat for a process is represented by the letter q. The sign of q for an endothermic process is positive because the system is gaining heat. A chemical reaction or physical change is **exothermic** if heat is released by the system into the surroundings. Because the surroundings are gaining heat from the system, the temperature of the surroundings increases. The sign of q for an exothermic process is negative because the system is losing heat.

Scientists have define a quantity called enthalpy to make it more useful to keep track of the flow of heat in a reaction from a thermodynamics perspective. The definition of enthalpy is beyond the scope of this text, but it is defined in such a way that the heat flow at constant pressure is equal to the change in enthalpy (represented as (ΔH). Most reactions of interest to chemists occur at constant pressure, so this creates a convenient way of taking a non-conservative property such a heat flow and equating it with a state function such as enthalpy, which is more convenient mathematically. Throughout the rest of this text we will refer to change in enthalpy, (ΔH), and heat flow somewhat interchangeably.

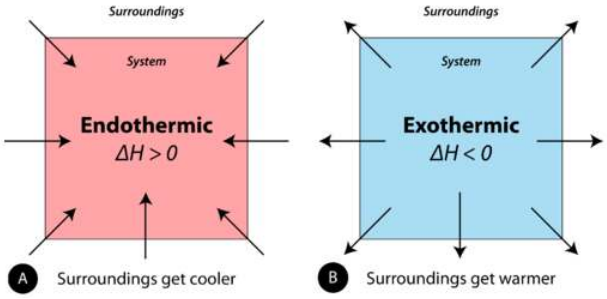


Figure 13.5.3: (A) Endothermic reaction. (B) Exothermic reaction. (CC BY-NC; CK-12)

Activation Energy



Figure 13.5.4: 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.

Potential Energy Diagrams

The energy changes that occur during a chemical reaction can be shown in a diagram called a potential energy diagram, or sometimes called a reaction progress curve. A **potential energy diagram** shows the change in potential energy of a system as reactants are converted into products. The figure below shows basic potential energy diagrams for an endothermic (A) and an exothermic (B) reaction. Recall that the enthalpy change (ΔH) is positive for an endothermic reaction and negative for an exothermic reaction. This can be seen in the potential energy diagrams. The total potential energy of the system increases for the endothermic reaction as the system absorbs energy from the surroundings. The total potential energy of the system decreases for the exothermic reaction as the system releases energy to the surroundings.

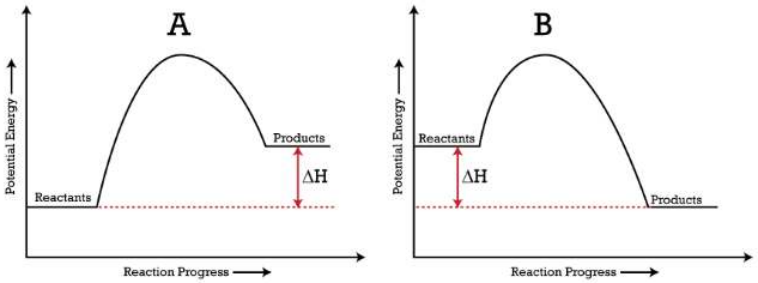


Figure **13.5.5**: A potential energy diagram shows the total potential energy of a reacting system as the reaction proceeds. (A) In an endothermic reaction, the energy of the products is greater than the energy of the reactants and ΔH is positive. (B) In an exothermic reaction, the energy of the products is lower than the energy of the reactants and ΔH is negative. (CC BY-NC; CK-12)

The activation energy for a reaction is illustrated in the potential energy diagram by the height of the hill between the reactants and the products. For this reason, the activation energy of a reaction is sometimes referred to as the activation energy barrier. Reacting particles must have enough energy so that when they collide, they can overcome that barrier (see figure below).

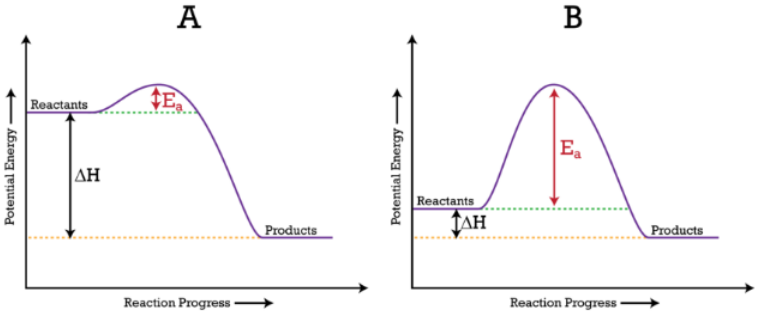


Figure **13.5.6**: The activation energy (E_a) of a reaction is the barrier that must be overcome for the reactants to be able to become products. (A) The activation energy is low, meaning that the reaction is likely to be fast. (B) The activation energy is high, meaning that the reaction is likely to be slow. (CC BY-NC; CK-12)

Section summary

- Chemical potential energy is energy available in the chemical bonds of a compound.
- A potential energy diagram shows the change in potential energy of a system as reactants are converted into products.
- Potential energy diagrams for endothermic and exothermic reactions are described.
- Diagrams of activation energy and reaction progress are given.

Glossary

chemical potential energy

The energy stored in the chemical bonds of a substance.

system

The specific portion of matter in a given space that is being studied during an experiment or an observation.

surroundings

Everything in the universe that is not part of the system.

endothermic

A chemical or physical change in which heat is absorbed by the system from the surroundings.

exothermic

A chemical or physical change in which heat is released by the system into the surroundings.

activation energy

The minimum energy that colliding particles must have in order to undergo a reaction.

This page titled 13.5: Energy and Chemical Reactions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **17.1: Chemical Potential Energy** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **17.3: Exothermic and Endothermic Processes** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **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/.
- **18.4: Potential Energy Diagrams** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

13.6: Entropy and Chemical Reactions

Learning Objectives

- Describe the entropy changes associated with various chemical and physical processes.
- Describe the relationship between enthalpy, entropy, and spontaneity.

When the pieces of a jigsaw puzzle are dumped from the box, the pieces naturally hit the table in a very random pattern. In order to put the puzzle together, a great deal of work must be done to overcome the natural disorder of the pieces. The pieces need to be turned right-side up, then sorted by color or edge (some people like to put the border together first). Finally comes the challenge of finding the exact spot of each piece of the puzzle, in order to obtain the finished picture.

Entropy

As was discussed earlier in this text, natural processes tend towards increasing entropy.

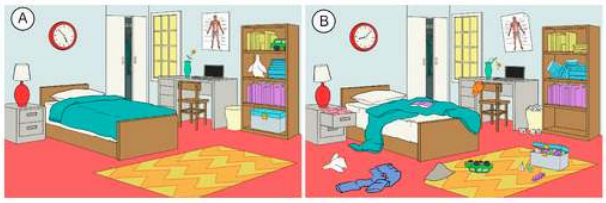


Figure **13.6.1**: The messy room on the right has more entropy than the highly ordered room on the left.

Chemical reactions also tend to proceed in such a way as to increase the total entropy of the system. How can you tell if a certain reaction shows an increase or a decrease in entropy? The molecular state of the reactants and products provide certain clues. The general cases below illustrate entropy at the molecular level.

1. For a given substance, the entropy of the liquid state is greater than the entropy of the solid state. Likewise, the entropy of the gas is greater than the entropy of the liquid. Therefore, entropy increases in processes in which solid or liquid reactants form gaseous products. Entropy also increases when solid reactants form liquid products.
2. Entropy increases when a substance is broken up into multiple parts. The process of dissolution increases entropy because the solute particles become separated from one another when a solution is formed.
3. Entropy increases as temperature increases. An increase in temperature means that the particles of the substance have greater kinetic energy. The faster-moving particles have more disorder than particles that are moving slowly at a lower temperature.
4. Entropy generally increases in reactions in which the total number of product molecules is greater than the total number of reactant molecules. An exception to this rule is when a gas is produced from nongaseous reactants.

These examples serve to illustrate how the entropy change in a reaction can be predicted:

$\mathrm{Cl_2}\,(g) \rightarrow \mathrm{Cl_2}\,(l)$

The entropy is decreasing because a gas is becoming a liquid.

$\mathrm{CaCO_3}\,(s) \rightarrow \mathrm{CaO}\,(s) + \mathrm{CO_2}\,(g)$

The entropy is increasing because a gas is being produced and the number of molecules is increasing.

$\mathrm{N_2}\,(g) + 3\mathrm{H_2}\,(g) \rightarrow 2\mathrm{NH_3}\,(g)$

The entropy is decreasing because four total reactant molecules are forming two total product molecules. All are gases.

$\mathrm{AgNO_3}\,(aq) + \mathrm{NaCl}\,(aq) \rightarrow \mathrm{NaNO_3}\,(aq) + \mathrm{AgCl}\,(s)$

The entropy is decreasing because a solid is formed from aqueous reactants.

$\mathrm{H_2}\,(g) + \mathrm{Cl_2}\,(g) \rightarrow 2\mathrm{HCl}\,(g)$

The entropy change is unknown (but likely not zero), because there are equal numbers of molecules on both sides of the equation, and all are gases.

Standard Entropy

All molecular motion ceases at absolute zero ($0\ \text{K}$). Therefore, the entropy of a pure crystalline substance at absolute zero is defined to be equal to zero. As the temperature of the substance increases, its entropy increases because of an increase in molecular motion. The absolute or **standard entropy** of substances can be measured. The symbol for entropy is S and the standard entropy of a substance is given by the symbol $S°$, indicating that the standard entropy is determined under standard conditions. The units for entropy are $\text{J/K} \cdot \text{mol}$. Standard entropies for a few substances are shown in the table below.

Standard Entropy Values at $25°\text{C}$

Table 13.6.1: Standard Entropy Values at 25°C

Substance	$S°\ (\text{J/K} \cdot \text{mol})$
$\text{H}_2\ (g)$	131.0
$\text{O}_2\ (g)$	205.0
$\text{H}_2\text{O}\ (l)$	69.9
$\text{H}_2\text{O}\ (g)$	188.7
C (graphite)	5.69
C (diamond)	2.4

We will not perform calculations using entropy in this class, but we will see how having a numerical value for it will be important for other decisions about the energy associated with chemical reactions.

Spontaneous Reactions

Reactions are favorable when they result in a decrease in enthalpy and an increase in entropy of the system. When both of these conditions are met, the reaction occurs naturally. A **spontaneous reaction** is a reaction that favors the formation of products at the conditions under which the reaction is occurring. A roaring bonfire is an example of a spontaneous reaction, since it is exothermic (there is a decrease in the energy of the system as energy is released to the surroundings as heat). The products of a fire are composed partly of gases such as carbon dioxide and water vapor. The entropy of the system increases during a combustion reaction. The combination of energy decrease and entropy increase dictates that combustion reactions are spontaneous reactions.



Figure **13.6.1**: Bonfire.

A **nonspontaneous reaction** is a reaction that does not favor the formation of products at the given set of conditions. In order for a reaction to be nonspontaneous, it must be endothermic, accompanied by a decrease in entropy, or both. Our atmosphere is composed primarily of a mixture of nitrogen and oxygen gases. One could write an equation showing these gases undergoing a chemical reaction to form nitrogen monoxide:

$$N_2(g) + O_2(g) \rightarrow 2NO(g)$$

Fortunately, this reaction is nonspontaneous at normal temperatures and pressures. It is a highly endothermic reaction with a slightly positive entropy change (ΔS). Nitrogen monoxide is capable of being produced at very high temperatures and has been observed to form as a result of lightning strikes.

One must be careful not to confuse the term spontaneous with the notion that a reaction occurs rapidly. A spontaneous reaction is one in which product formation is favored, even if the reaction is extremely slow. A piece of paper will not suddenly burst into flames, although its combustion is a spontaneous reaction. What is missing is the required activation energy to get the reaction started. If the paper were to be heated to a high enough temperature, it would begin to burn, at which point the reaction would proceed spontaneously until completion.

In a reversible reaction, one reaction direction may be favored over the other. Carbonic acid is present in carbonated beverages. It decomposes spontaneously to carbon dioxide and water, according to the following reaction.

$$H_2CO_3(aq) \rightleftharpoons CO_2(g) + H_2O(l)$$

If you were to start with pure carbonic acid in water and allow the system to come to equilibrium, more than 99% of the carbonic acid would be converted into carbon dioxide and water. The forward reaction is spontaneous because the products of the forward reaction are favored at equilibrium. In the reverse reaction, carbon dioxide and water are the reactants, and carbonic acid is the product. When carbon dioxide is bubbled into water, less than 1% is converted to carbonic acid when the reaction reaches equilibrium. The reverse reaction, as written above, is not spontaneous.

Free Energy

Many chemical reactions and physical processes release energy that can be used to do other things. When the fuel in a car is burned, some of the released energy is used to power the vehicle. **Free energy** is energy that is available to do work. Spontaneous reactions release free energy as they proceed. Recall that the determining factors for spontaneity of a reaction are the enthalpy and entropy changes that occur for the system. The free energy change of a reaction is a mathematical combination of the enthalpy change and the entropy change.

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

The symbol for free energy is G, in honor of American scientist Josiah Gibbs (1839-1903), who made many contributions to thermodynamics. The change in Gibbs free energy is equal to the change in enthalpy minus the mathematical product of the change in entropy, multiplied by the Kelvin temperature. Each thermodynamic quantity in the equation is for substances in their standard states. The usual units for ΔH are $\mathrm{kJ/mol}$, while ΔS is often reported in $\mathrm{J/K \cdot mol}$. It is necessary to change the units for ΔS to $\mathrm{kJ/K \cdot mol}$, so that the calculation of ΔG is in $\mathrm{kJ/mol}$.

A spontaneous reaction is one that releases free energy, and so the sign of ΔG must be negative. Since ΔH and ΔS can be either positive or negative, depending on the characteristics of the particular reaction, there are four different general outcomes for ΔG, and these are outlined in the table below.

Enthalpy, Entropy, and Free Energy Changes

Table 13.6.1: Enthalpy, Entropy, and Free Energy Changes

ΔH	ΔS	ΔG
- value (exothermic)	+ value (disordering)	always negative
+ value (endothermic)	+ value (disordering)	negative at higher temperatures
- value (exothermic)	- value (ordering)	negative at lower temperatures
+ value (endothermic)	- value (ordering)	never negative

Keep in mind that the temperature in the Gibbs free energy equation is the Kelvin temperature and so can only be positive. When ΔH is negative and ΔS is positive, the sign of ΔG will always be negative, and the reaction will be spontaneous at all temperatures.

This corresponds to both driving forces being in favor of product formation. When ΔH is positive and ΔS is negative, the sign of ΔG will always be positive, and the reaction can never be spontaneous. This corresponds to both driving forces working against product formation.

When one driving force favors the reaction, but the other does not, it is the temperature that determines the sign of ΔG. Consider first an endothermic reaction (positive ΔH) that also displays an increase in entropy (positive ΔS). It is the entropy term that favors the reaction. Therefore, as the temperature increases, the $T\Delta S$ term in the Gibbs free energy equation will begin to predominate and ΔG will become negative. A common example of a process that falls into this category is the melting of ice. At a relatively low temperature (below $273\ \text{K}$), the melting is not spontaneous because the positive ΔH term "outweighs" the $T\Delta S$ term. When the temperature rises above $273\ \text{K}$, the process becomes spontaneous because the larger T value has tipped the sign of ΔG over to being negative.

When the reaction is exothermic (negative ΔH) but undergoes a decrease in entropy (negative ΔS), it is the enthalpy term that favors the reaction. In this case, a spontaneous reaction is dependent upon the $T\Delta S$ term being small relative to the ΔH term, so that ΔG is negative. The freezing of water is an example of this type of process. It is spontaneous only at a relatively low temperature. Above $273\ \text{K}$, the larger $T\Delta S$ value causes the sign of ΔG to be positive, and freezing does not occur.

Section Summary

- Situations involving entropy changes are described.
- Entropy can be measured and tabulated.
- Relationships between enthalpy, entropy, and free energy are described.

Glossary

spontaneous reaction

A reaction that favors the formation of products at the conditions under which the reaction is occurring.

nonspontaneous reaction

A reaction that does not favor the formation of products at the given set of conditions.

free energy

Energy that is available to do work.

This page titled 13.6: Entropy and Chemical Reactions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **20.1: Entropy** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **20.2: Standard Entropy** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **20.3: Spontaneous and Nonspontaneous Reactions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **20.4: Free Energy** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

13.7: Classifying Reactions

Learning Objectives

- Categorize a reaction as combination, decomposition, single replacement, or double replacement.

There are millions of chemical reactions that occur in the world. Just like we use the periodic table to help us remember properties of the elements, we can group chemical reactions into categories based on similar aspects. We will not look at all the possible reaction types in this section, but we will introduce some important categories for simple reactions. In the following sections we will discuss some more specific reaction types based on how they influence our lives.

Combination Reactions

A **combination reaction** is a reaction in which two or more substances combine to form a single new substance. Combination reactions can also be called synthesis reactions. The general form of a combination reaction is:

$$\mathrm{A + B \rightarrow AB}$$

One combination reaction is two elements combining to form a compound. Solid sodium metal reacts with chlorine gas to produce solid sodium chloride.

$$\mathrm{2Na\,(\mathit{s}) + Cl_2\,(\mathit{g}) \rightarrow 2NaCl\,(\mathit{s})}$$



Video 13.7.1: Sodium reacting with chlorine to make table salt. (Source)

In order to write and balance the equation correctly, it is important to remember the seven elements that exist in nature as diatomic molecules: $\mathrm{H_2}$, $\mathrm{N_2}$, $\mathrm{O_2}$, $\mathrm{F_2}$, $\mathrm{Cl_2}$, $\mathrm{Br_2}$, and $\mathrm{I_2}$.

One sort of combination reaction that occurs frequently is the reaction of an element with oxygen to form an oxide. Metals and nonmetals both react readily with oxygen under most conditions. Magnesium reacts rapidly and dramatically when ignited, combining with oxygen from the air to produce a fine powder of magnesium oxide.

$$\mathrm{2Mg\,(\mathit{s}) + O_2\,(\mathit{g}) \rightarrow 2MgO\,(\mathit{s})}$$

Decomposition Reactions

A **decomposition reaction** is a reaction in which a compound breaks down into two or more simpler substances. The general form of a decomposition reaction is:

$$\mathrm{AB \rightarrow A + B}$$

Most decomposition reactions require an input of energy in the form of heat, light, or electricity.

Binary compounds are compounds composed of just two elements. The simplest kind of decomposition reaction is when a binary compound decomposes into its elements. Mercury (II) oxide, a red solid, decomposes when heated to produce mercury and oxygen gas.

$$\mathrm{2HgO\,(\mathit{s}) \rightarrow 2Hg\,(\mathit{l}) + O_2\,(\mathit{g})}$$



Video **13.7.2**: Mercuric oxide decomposing into mercury and oxygen
(Source)

A reaction is also considered to be a decomposition reaction even when one or more of the products are still compounds. A metal carbonate decomposes into a metal oxide and carbon dioxide gas. For example, calcium carbonate decomposes into calcium oxide and carbon dioxide:

$$\mathrm{CaCO_3}\,(s) \rightarrow \mathrm{CaO}\,(s) + \mathrm{CO_2}\,(g)$$

Metal hydroxides decompose on heating to yield metal oxides and water. Sodium hydroxide decomposes to produce sodium oxide and water:

$$\mathrm{2NaOH}\,(s) \rightarrow \mathrm{Na_2O}\,(s) + \mathrm{H_2O}\,(g)$$

Some unstable acids decompose to produce nonmetal oxides and water. Carbonic acid decomposes easily at room temperature into carbon dioxide and water:

$$\mathrm{H_2CO_3}\,(aq) \rightarrow \mathrm{CO_2}\,(g) + \mathrm{H_2O}\,(l)$$

Single-Replacement Reactions

A **single-replacement reaction** is a reaction in which one element replaces a similar element in a compound. The general form of a single-replacement (also called single-displacement) reaction is:

$$\mathrm{A + BC \rightarrow AC + B}$$

In this general reaction, element A is a metal and replaces element B (also a metal) in the compound. When the element that is doing the replacing is a nonmetal, it must replace another nonmetal in a compound, and the general equation becomes:

$$\mathrm{Y + XZ \rightarrow XY + Z}$$

(Where Y is a nonmetal and replaces the nonmetal Z in the compound with X.)

Metal Replacement

Magnesium is a more reactive metal than copper. When a strip of magnesium metal is placed in an aqueous solution of copper (II) nitrate, it replaces the copper. The products of the reaction are aqueous magnesium nitrate and solid copper metal.

$$\mathrm{Mg}\,(s) + \mathrm{Cu(NO_3)_2}\,(aq) \rightarrow \mathrm{Mg(NO_3)_2}\,(aq) + \mathrm{Cu}\,(s)$$

This subcategory of single-replacement reactions is called a metal replacement reaction because it is a metal that is being replaced (copper).

Hydrogen Replacement

Many metals react easily with acids and when they do so, one of the products of the reaction is hydrogen gas. Zinc reacts with hydrochloric acid to produce aqueous zinc chloride and hydrogen (figure below).

$$\mathrm{Zn}\,(s) + \mathrm{2HCl}\,(aq) \rightarrow \mathrm{ZnCl_2}\,(aq) + \mathrm{H_2}\,(g)$$

In a hydrogen replacement reaction, the hydrogen in the acid is replaced by an active metal.



Video 13.7.3: Zinc reacting with hydrochloric acid produces bubbles of hydrogen gas. (Source)

Some metals are so reactive that they are capable of replacing the hydrogen in water. The products of such a reaction are the metal hydroxide and hydrogen gas. All group 1 metals undergo this type of reaction. Sodium reacts vigorously with water to produce aqueous sodium hydroxide and hydrogen (see figure below).

$$2\text{Na}\,(s) + 2\text{H}_2\text{O}\,(l) \rightarrow 2\text{NaOH}\,(aq) + \text{H}_2\,(g)$$



Figure 13.7.1: Sodium metal reacts vigorously with water, giving off hydrogen gas. A large piece of sodium will often generate so much heat that the hydrogen will ignite. (Credit: User:Ajhalls/Wikimedia Commons; Source; License: Public Domain)

Double-Replacement Reactions

A **double-replacement reaction** is a reaction in which the positive and negative ions of two ionic compounds exchange places to form two new compounds. The general form of a double-replacement (also called double-displacement) reaction is:

$$\text{AB} + \text{CD} \rightarrow \text{AD} + \text{CB}$$

In this reaction, A and C are positively-charged cations, while B and D are negatively-charged anions. Double-replacement reactions generally occur between substances in aqueous solution. In order for a reaction to occur, one of the products is usually a solid precipitate, a gas, or a molecular compound such as water.

Formation of a Precipitate

A precipitate forms in a double-replacement reaction when the cations from one of the reactants combine with the anions from the other reactant to form an insoluble ionic compound. When aqueous solutions of potassium iodide and lead (II) nitrate are mixed, the following reaction occurs:

$$2\text{KI}\,(aq) + \text{Pb}(\text{NO}_3)_2\,(aq) \rightarrow 2\text{KNO}_3\,(aq) + \text{PbI}_2\,(s)$$

There are very strong attractive forces that occur between Pb^{2+} and I^- ions and the result is a brilliant yellow precipitate (see figure below). The other product of the reaction, potassium nitrate, remains soluble.



Video**13.7.4**: A yellow precipitate of solid lead (II) iodide forms immediately when solutions of lead (II) nitrate and potassium iodide are mixed. (Source)

Section Summary

- Combination reactions occur when two or more substances combine to form a new substance.
- A definition of decomposition reaction and example reactions are given.
- The activity series describes the relative reactivities of metals and halogens.
- The double-replacement reaction generally takes the form of AB + CD → AD + CB where A and C are positively-charged cations, while B and D are negatively-charged anions.
- In a double replacement reactions, typically one of the products is a precipitate, a gas, or a molecular compound.

Glossary

combination reaction

A reaction in which two or more substances combine to form a single new substance.

decomposition reaction

A reaction in which a compound breaks down into two or more simpler substances.

single-replacement reaction

A reaction in which one element replaces a similar element in a compound.

double-replacement reaction

A reaction in which the positive and negative ions of two ionic compounds exchange places to form two new compounds.

This page titled 13.7: Classifying Reactions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **11.4: Combination Reactions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **11.5: Decomposition Reactions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **11.7: Single Replacement Reactions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **11.9: Double Replacement Reactions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

13.8: Acid Base Reactions

Learning Objectives

- Identify acids and bases both in terms of their characteristic properties and in terms of their chemical reactivity.
- Identify acidic or basic solutions using the pH scale.

You are probably familiar with the terms acid, base, and pH. It turns out that acids and bases show up in our daily lives all the time: the foods we eat, the cleaning supplies we use, and even breathing involve acid base chemistry. In the section we will discuss these concepts in more detail as we explore acid base reactions.

Acids

Acids are very common in some of the foods that we eat. Citrus fruits such as oranges and lemons contain citric acid and ascorbic acid, which is better known as vitamin C. Carbonated sodas contain phosphoric acid. Vinegar contains acetic acid. Your own stomach utilizes hydrochloric acid to digest food.

Acids are a distinct class of compounds because of the properties of their aqueous solutions. These properties are:

1. Aqueous solutions of acids are electrolytes, meaning that they conduct electrical current. Some acids are strong electrolytes because they ionize completely in water, yielding a great many ions. Other acids are weak electrolytes that exist primarily in a non-ionized form when dissolved in water.
2. Acids have a sour taste. Lemons, vinegar, and sour candies all contain acids.
3. Acids change the color of certain acid-base indicates. Two common indicators are litmus and phenolphthalein. Blue litmus turns red in the presence of an acid, while phenolphthalein turns colorless.
4. Acids react with active metals to yield hydrogen gas. Recall that an activity series is a list of metals in descending order of reactivity. Metals that are above hydrogen in the activity series will replace the hydrogen from an acid in a single-replacement reaction, as shown below:

$$\mathrm{Zn}\,(s) + \mathrm{H_2SO_4}\,(aq) \rightarrow \mathrm{ZnSO_4}\,(aq) + \mathrm{H_2}\,(g)$$

5. Acids react with bases to produce a salt compound and water. When equal moles of an acid and a base are combined, the acid is neutralized by the base. The products of this reaction are an ionic compound, which is labeled as a salt, and water.

Bases

Bases have properties that mostly contrast with those of acids.

1. Aqueous solutions of bases are also electrolytes. Bases can be either strong or weak, just as acids can.
2. Bases often have a bitter taste and are found in foods less frequently than acids. Many bases, like soaps, are slippery to the touch.
3. Bases also change the color of indicators. Litmus turns blue in the presence of a base while phenolphthalein turns pink.
4. Bases do not react with metals in the way that acids do.
5. Bases react with acids to produce a salt and water.



Figure* 13.8.1*: *Phenolphthalein indicator in presence of base.*

Please note that tasting chemicals and touching them are NOT good lab practices and should be avoided—in other words, do not do this at home.

Bases are less common as foods, but they are nonetheless present in many household products. Many cleaners contain ammonia, a base. Sodium hydroxide is found in drain cleaner. **Antacids**, which combat excess stomach acid, are comprised of bases such as magnesium hydroxide or sodium hydrogen carbonate.

Arrhenius Acids and Bases

Swedish chemist Svante Arrhenius (1859-1927) was the first to propose a theory to explain the observed behavior of acids and bases. Because of their ability to conduct a current, he knew that both acids and bases contained ions in solution. An **Arrhenius acid** is a compound that ionizes to yield hydrogen ions ($\mathbf{H^+}$) in aqueous solution. An **Arrhenius base** is a compound that ionizes to yield hydroxide ions ($\mathbf{OH^-}$) in aqueous solution.

Acids are molecular compounds with ionizable hydrogen atoms. Only hydrogen atoms that are part of a highly polar covalent bond are ionizable. Hydrogen chloride ($\mathbf{HCl}$) is a gas at room temperature and under normal pressure. The $\mathbf{H{-}Cl}$ bond in hydrogen chloride is a polar bond. The hydrogen atom is electron deficient because of the higher electronegativity of the chlorine atom. Consequently, the hydrogen atom is attracted to the lone pair of electrons in a water molecule when $\mathbf{HCl}$ is dissolved in water. The result is that the $\mathbf{H{-}Cl}$ bond breaks, with both bonding electrons remaining with the $\mathbf{Cl}$, forming a chloride ion. The $\mathbf{H^+}$ ion attaches to the water molecule, forming a polyatomic ion called the hydronium ion. The **hydronium ion** ($\mathbf{H_3O^+}$) can be thought of as a water molecule with an extra attached hydrogen ion.

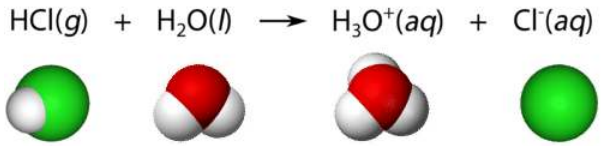


Figure **13.8.1**: Formation of a hydronium ion. (CC BY-NC; CK-12)

Equations showing the ionization of an acid in water are frequently simplified by omitting the water molecule:

$$\mathbf{HCl}\,(g) \rightarrow \mathbf{H^+}\,(aq) + \mathbf{Cl^-}\,(aq)$$

This is merely a simplification of the previous equation, but it is commonly used. Any hydrogen ions in an aqueous solution will be attached to water molecules as hydronium ions.

The pH Scale

Expressing the acidity of a solution by using the molarity of the hydrogen ion is cumbersome because the quantities are generally very small. Danish scientist Søren Sørensen (1868-1939) proposed an easier system for indicating the concentration of $\mathbf{H^+}$ called the pH scale. The letters pH stand for the power of the hydrogen ion. The **pH** of a solution is the negative logarithm of the hydrogen-ion concentration:

$$\mathbf{pH} = -\log\,[\mathbf{H^+}]$$

In pure water or a neutral solution, the $[\mathbf{H^+}] = \mathbf{1.0 \times 10^{-7}\ M}$. Substituting into the pH expression:

$$\mathbf{pH} = -\log\left[1.0 \times 10^{-7}\right] = -\,(-7.00) = 7.00$$

The pH of pure water or any neutral solution is thus 7.00. For recording purposes, the numbers to the right of the decimal point in the pH value are the significant figures. Since $\mathbf{1.0 \times 10^{-7}}$ has two significant figures, the pH can be reported as 7.00.

A logarithmic scale condenses the range of acidity to numbers that are easy to use. Consider a solution with $[\mathbf{H^+}] = \mathbf{1.0 \times 10^{-4}\ M}$. That is a hydrogen-ion concentration that is 1000 times higher than the concentration in pure water. The pH of such a solution is 4.00, a difference of just 3 pH units. Notice that when the $[\mathbf{H^+}]$ is written in scientific notation and the coefficient is 1, the pH is simply the exponent with the sign changed. The pH of a solution with $[\mathbf{H^+}] = \mathbf{1 \times 10^{-2}\ M}$ is 2 and the pH of a solution with $[\mathbf{H^+}] = \mathbf{1 \times 10^{-10}\ M}$ is 10.

As we saw earlier, a solution with $[\mathbf{H^+}]$ higher than $\mathbf{1.0 \times 10^{-7}}$ is acidic, while a solution with $[\mathbf{H^+}]$ lower than $\mathbf{1.0 \times 10^{-7}}$ is basic. Consequently, solutions with a pH of less than 7 are acidic, while those with a pH higher than 7 are basic. Figure **13.8.1** illustrates this relationship, along with some examples of various solutions.

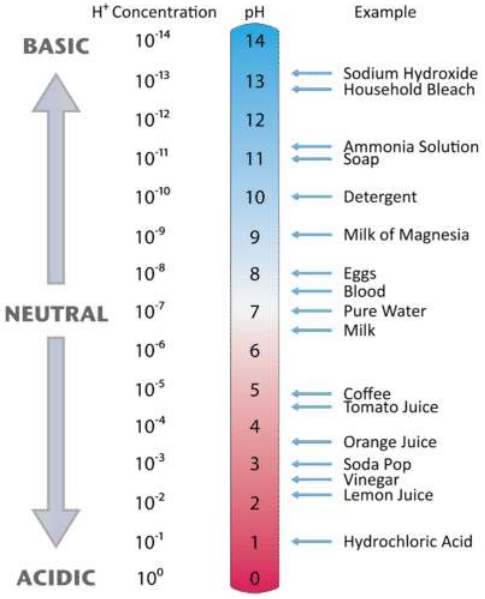


Figure **13.8.1**: The pH values for several common materials. (CC BY-NC; CK-12)

Section Summary

- Characteristic properties of acids and bases are provided.
- An Arrhenius acid is a compound that ionizes to yield hydrogen ions (H^+) in aqueous solution.
- An Arrhenius base is a compound that ionizes to yield hydroxide ions (OH−)(OH−) in aqueous solution.
- The pH of a solution is the negative logarithm of the hydrogen-ion concentration.
- Solutions with a pH of less than 7 are acidic, while those with a pH higher than 7 are basic.
- pH values for several common materials are listed.

Glossary

Arrhenius acid

A compound that ionizes to yield hydrogen ions (H^+) in aqueous solution.

Arrhenius base

A compound that ionizes to yield hydroxide ions (OH^-) in aqueous solution.

hydronium ion (H_3O^+)

A water molecule with an extra hydrogen ion attached to it.

pH

The negative logarithm of the hydrogen-ion concentration.

This page titled 13.8: Acid Base Reactions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **21.1: Properties of Acids** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **21.2: Properties of Bases** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **21.3: Arrhenius Acids** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **21.9: The pH Scale** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

13.9: Redox Reactions- The Transfer of Electrons

Learning Objectives

- Identify substances which are oxidized or reduced in a chemical reaction.

Earlier in this text we discussed the function of batteries in supplying a source of electrical energy. All batteries function on the basis of electron transfer reactions, commonly known as redox reactions. There are several other examples of important redox reactions: corrosion of a metal, combustion of a flammable substance, and digestion of our food all involve redox reactions. Some of these are reactions we want to have occur, while others we might want to prevent. The more we know about redox reactions, the better we will be able to design systems to make use of the ones we want and avoid the ones that we don't want.

Historically, the terms oxidation and reduction had to do with how oxygen behaved in a chemical reaction. A chemical was oxidized if it reacted with oxygen, and reduced if it reacted in a way that lost oxygen. Eventually this process was generalized to include other types of reactions in which electrons are being transferred. An **oxidation-reduction reaction** is a reaction that involves the full or partial transfer of electrons from one reactant to another. **Oxidation** is the full or partial loss of electrons (gaining of oxygen is only one way to do this). **Reduction** is the full or partial gain of electrons (loss of oxygen is only one way to do this). A **redox reaction** is another term for an oxidation-reduction reaction.

In the course of a chemical reaction between a metal and a nonmetal, electrons are transferred from the metal atoms to the nonmetal atoms. For example, when zinc metal is mixed with sulfur and heated, the compound zinc sulfide is produced. Two valence electrons from each zinc atom are transferred to each sulfur atom.

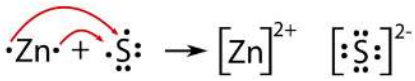


Figure 13.9.1: Reaction between zinc and sulfur.

Since the zinc is losing electrons in the reaction, it is being oxidized. The sulfur is gaining electrons and is thus being reduced. An **oxidation-reduction reaction** is a reaction that involves the full or partial transfer of electrons from one reactant to another. **Oxidation** is the full or partial loss of electrons or the gain of oxygen. **Reduction** is the full or partial gain of electrons or the loss of oxygen. A **redox reaction** is another term for an oxidation-reduction reaction.

Each of these processes can be shown in a separate equation called a half-reaction. A **half-reaction** is an equation that shows either the oxidation or the reduction reaction that occurs during a redox reaction.

$$\text{Oxidation: } \mathrm{Zn} \rightarrow \mathrm{Zn^{2+}} + 2\mathrm{e^-}$$
$$\text{Reduction: } \mathrm{S} + 2\mathrm{e^-} \rightarrow \mathrm{S^{2-}}$$

It is important to remember that the two half-reactions occur simultaneously. The resulting ions that are formed are then attracted to one another in an ionic bond.

In the reaction above, zinc is being oxidized by losing electrons. However, there must be another substance present that gains those electrons and in this case that is the sulfur. In other words, the sulfur is causing the zinc to be oxidized. Sulfur is called the oxidizing agent. The zinc causes the sulfur to gain electrons and become reduced and so the zinc is called the reducing agent. The **oxidizing agent** is a substance that causes oxidation by accepting electrons. The **reducing agent** is a substance that causes reduction by losing electrons. The simplest way to think of this is that the oxidizing agent is the substance that is reduced, while the reducing agent is the substance that is oxidized. The example below shows how to analyze a redox reaction.

Example 13.9.1

When chlorine gas is bubbled into a solution of sodium bromide, a reaction occurs which produces aqueous sodium chloride and bromine. Determine what is being oxidized and what is being reduced. Identify the oxidizing and reducing agents.

$$\mathrm{Cl_2}\,(g) + 2\mathrm{NaBr}\,(aq) \rightarrow 2\mathrm{NaCl}\,(aq) + \mathrm{Br_2}\,(l)$$

Solution:

Step 1: Plan the problem.

Break the reaction down into a net ionic equation and then into half-reactions. The substance that loses electrons is being oxidized and is the reducing agent. The substance that gains electrons is being reduced and is the oxidizing agent.

Step 2: Solve.

$$Cl_2\,(g) + \cancel{2Na^+\,(aq)} + 2Br^-\,(aq) \rightarrow \cancel{2Na^+\,(aq)} + 2Cl^-\,(aq) + Br_2\,(l)$$

$$Cl_2\,(g) + 2Br^-\,(aq) \rightarrow 2Cl^-\,(aq) + Br_2\,(l) \quad \text{(net ionic equation)}$$

$$\text{Oxidation: } Cl_2\,(g) + 2e^- \rightarrow 2Cl^-\,(aq)$$

$$\text{Reduction: } 2Br^-\,(aq) \rightarrow Br_2\,(l) + 2e^-$$

The Cl_2 is being reduced and is the oxidizing agent. The Br^- is being oxidized and is the reducing agent.

Section Summary

- An oxidation-reduction reaction, or redox reaction, is a reaction that involves the full or partial transfer of electrons from one reactant to another.
- Oxidation is the full or partial loss of electrons or the gain of oxygen.
- Reduction is the full or partial gain of electrons or the loss of oxygen.
- Oxidation and reduction reactions must occur together; neither can happen alone in a reaction.
- A half-reaction is an equation that shows either the oxidation or the reduction reaction that occurs during a redox reaction.
- Oxidation and reduction reactions must occur together; neither can happen alone in a reaction.
- The oxidizing agent is a substance that causes oxidation by accepting electrons.
- The reducing agent is a substance that causes reduction by losing electrons.

Glossary

oxidation-reduction reaction (redox reaction)

A reaction that involves the transfer of electrons from one reactant to another.

oxidation

The loss of electrons.

reduction

The gain of electrons.

half reaction

An equation that shows either the oxidation or the reduction reaction that occurs during a redox reaction.

oxidizing agent

A substance that causes oxidation by accepting electrons.

reducing agent

A substance that causes reduction by losing electrons.

This page titled 13.9: Redox Reactions- The Transfer of Electrons is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **22.2: Redox Reactions and Ionic Compounds** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **22.3: Oxidizing and Reducing Agents** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

13.10: Polymerization Reactions

Learning Objectives

- Explain the difference between an addition polymer and a condensation polymer.

We reap the benefits of using Styrofoam containers, but do not often consider where they end up. Styrofoam materials do not break down quickly under exposure to the elements. When buried in a landfill, styrofoam will remain intact for a long time. The good news is that there is not a lot of this pollutant found in landfills (maybe about 0.5% by weight of the total mass of garbage). There is no good way to recycle Styrofoam at present, but in the future, a creative scientist may change that. Styrofoam is an example of a polymer.

Polymers are very different from the other kinds of organic molecules that you have seen so far. Whereas other compounds are of relatively low molar mass, polymers are giant molecules of very high molar mass. Polymers are the primary components of all sorts of plastics and related compounds. A **polymer** is a large molecule formed of many smaller molecules covalently bonded in a repeating pattern. The small molecules which make up the polymer are called **monomers**. Polymers generally form either from an addition reaction or a condensation reaction. The "small molecules" that polymers are comprised of are often much more complicated than the compounds we have been exploring earlier in this chapter. We will not get into the details of these more advanced structures in this text, but will provide a few examples of types of polymer reactions.

Addition Polymers

An **addition polymer** is a polymer formed by chain addition reactions between monomers that contain a double bond. Molecules of ethene can polymerize with each other under the right conditions to form the polymer called polyethylene.

$$n\mathrm{CH_2{=}CH_2} \rightarrow -(\mathrm{CH_2CH_2})_n-$$

The letter n stands for the number of monomers that are joined in repeated fashion to make the polymer, and can have a value in the hundreds or even thousands.

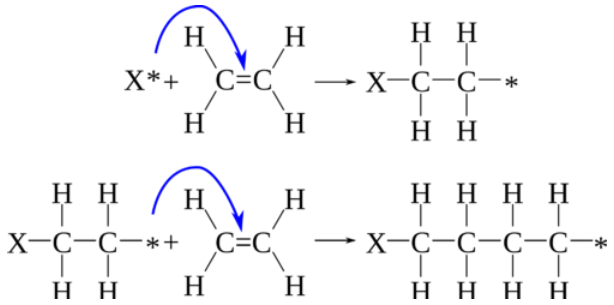


Figure 13.10.1: Polyethylene synthesis.

Polyethylene can have different properties depending on the length of the polymer chains, and on how efficiently they pack together. Some common products made from different forms of polyethylene include plastic bottles, plastic bags, and harder plastic objects such as milk crates.

Several other kinds of unsaturated monomers can be polymerized, and are components in common household products. Polypropylene is stiffer than polyethylene, and is in plastic utensils and some other types of containers.

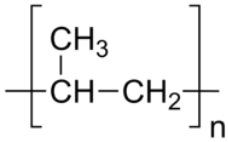


Figure 13.10.2: Polypropylene structure.

Polystyrene is used in insulation and in molded items such as coffee cups.

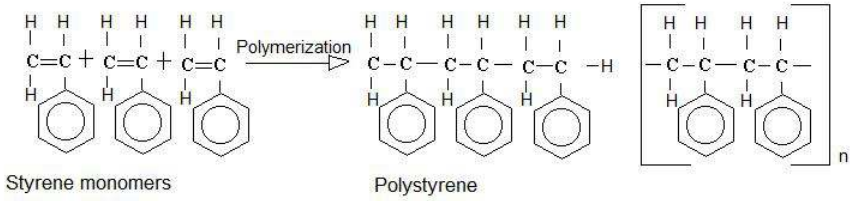


Figure **13.10.3**: Polystyrene synthesis and structure.

Polyvinyl chloride (PVC) is extensively used for plumbing pipes.



Figure **13.10.4**: Polyvinyl chloride.

Polyisoprene is a polymer of isoprene and is better known as rubber. It is produced naturally by rubber trees, but several variants have been developed which demonstrate improvements on the properties of natural rubber.

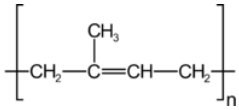


Figure **13.10.5**: Polyisoprene.

Condensation Polymers

A **condensation polymer** is a polymer formed by condensation reactions. Monomers of condensation polymers must contain two functional groups so that each monomer can link up with two other monomers. One type of condensation polymer is called a polyamide. An amide is characterized by the functional group shown below wherein the carbon of a carbonyl group is bonded to the nitrogen of an amine.

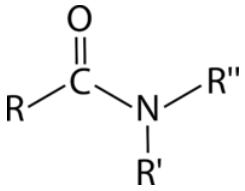


Figure **13.10.1**: Amide.

One pair of monomers that can form a polyamide is that of adipic acid and hexanediamine. Adipic acid is a carboxylic acid with two carboxyl groups on either end of the molecule. Hexanediamine has amino groups on either end of a six-carbon chain. When these molecules react with each other, a molecule of water is eliminated, classifying it as a condensation reaction (see figure below).

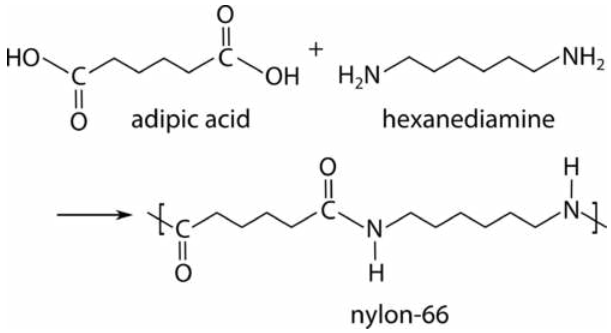


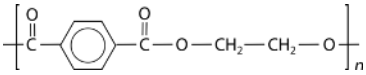
Figure **13.10.2**: Nylon synthesis.

The polymer that results from the repetition of the condensation reaction is a polyamide called nylon-66. Nylon-66 was first invented in 1935 and has been used in all sorts of products. Polyamides, including Nylon-66, are commonly found in fibers and clothing, cooking utensils, fishing line, and carpeting—among many other applications.



Figure **13.10.3**: Nylon spatula.

Polyester is another common type of condensation polymer. Recall that esters are formed from the reaction of an alcohol with a carboxylic acid. When both the acid and alcohol have two functional groups, the ester is capable of being polymerized. One such polyester is called polyethylene terephthalate (PET) and is formed from the reaction of ethylene glycol with terephthalic acid. The structure of PET is shown below.



PET is used in tires, photographic film, food packaging, and clothing. Polyester fabric is used in permanent-press clothing. Its resistance to wrinkling comes from the cross-linking of the polymer strands.

Section Summary

- A polymer is a large molecule formed of many smaller molecules covalently bonded in a repeating pattern; they are the primary components of all sorts of plastics and related compounds.
- The small molecules which make up polymers are called monomers.
- Polymers generally form either from an addition reaction or a condensation reaction.
- An addition polymer is a polymer formed by chain addition reactions between monomers that contain a double bond.
- A condensation polymer is a polymer formed by condensation reactions which release a small molecule such as water.

Glossary

polymer

A large molecule formed of many smaller molecules covalently bonded in a repeating pattern.

monomers

The small molecules which make up a polymer.

addition polymer

A polymer formed by chain addition reactions between monomers that contain a double bond.

condensation polymer

a polymer formed by condensation reactions, such as a reaction that forms water molecules as a side product.

This page titled 13.10: Polymerization Reactions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **25.19: Polymerization - Addition Polymers** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **25.20: Polymerization - Condensation Polymers** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

CHAPTER OVERVIEW

14: Nuclear Radiation

14.1: Introduction to Radioactivity
14.2: The Atomic Nucleus
14.3: Discovery of Radioactivity
14.4: Nuclear Decay Processes
14.5: Nuclear Equations
14.6: Half-Life
14.7: What Holds the Nucleus Together
14.8: The Energy Within the Nucleus and all Matter
14.9: Nuclear Binding Energy
14.9.1: Nuclear Fission
14.9.2: Nuclear Fusion
14.10: Nuclear Radiation in Our Lives
14.11: Process of Science- Lise Meitner and Otto Hahn

This page titled 14: Nuclear Radiation is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur.

14.1: Introduction to Radioactivity

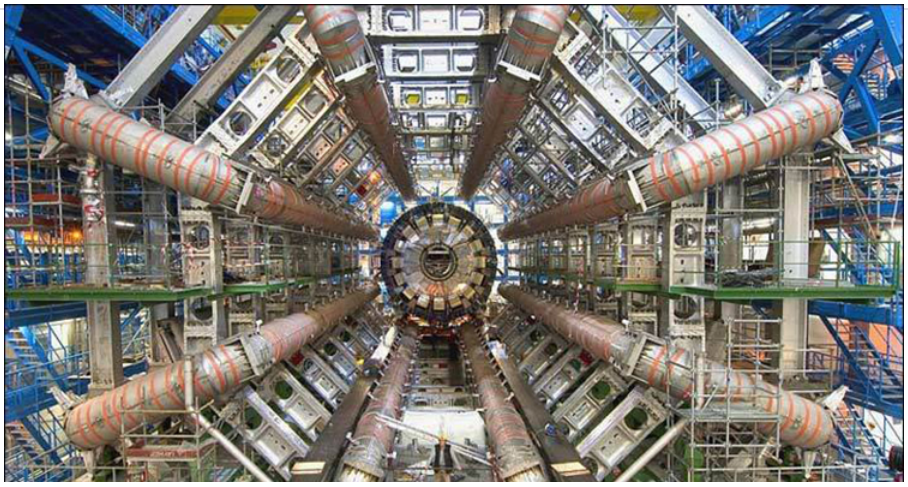


Figure **14.1.1**: Part of the Large Hadron Collider at CERN, on the border of Switzerland and France. The LHC is a particle accelerator, designed to study fundamental particles. (credit: Image Editor, Flickr)

There is an ongoing quest to find substructures of matter. At one time, it was thought that atoms would be the ultimate substructure, but just when the first direct evidence of atoms was obtained, it became clear that they have a substructure and a *nucleus*. The nucleus itself has spectacular characteristics. For example, certain nuclei are unstable, and their decay emits radiations with energies millions of times greater than atomic energies. Some of the mysteries of nature, such as why the core of the earth remains molten and how the sun produces its energy, are explained by nuclear phenomena. The exploration of *radioactivity* and the nucleus revealed fundamental and previously unknown particles, forces, and conservation laws. That exploration has evolved into a search for further underlying structures, such as quarks. In this chapter, we will start with the fundamentals of nuclear radioactivity and the nucleus, moving on to some of the important applications of nuclear physics. Quarks and leptons are beyond the scope of this text.

Figure **14.1.2** shows the structures and length scales involved in study of nucleus and elementary particles. In atomic and molecular physics, we were already dealing with structures at nanometer scale, about 500 times shorter than the wavelength of green light. The nuclear scale goes 100,000 times smaller; the size of largest stable nuclei is about 10 femtometers, or 10^{-14} m. All other objects we are studying in nuclear and particle physics are smaller than this, with the most elementary particles, quarks and leptons, having no known size and being considered as "point masses."

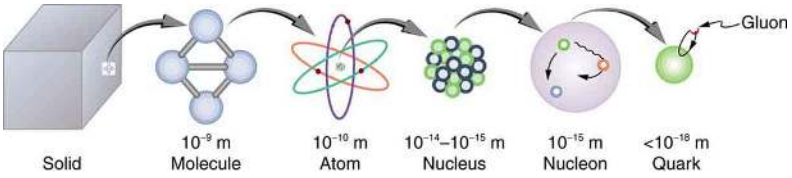


Figure **14.1.2**: The properties of matter are based on substructures called molecules and atoms. Atoms have the substructure of a nucleus with orbiting electrons, the interactions of which explain atomic properties. Protons and neutrons, the interactions of which explain the stability and abundance of elements, form the substructure of nuclei. Protons and neutrons are not fundamental—they are composed of quarks. Like electrons and a few other particles, quarks may be the fundamental building blocks of all there is, lacking any further substructure. But the story is not complete, because quarks and electrons may have substructure smaller than details that are presently observable.

This page titled 14.1: Introduction to Radioactivity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **14.1: Introduction to Nuclear and Particle Physics** by OpenStax is licensed CC BY 4.0.

14.2: The Atomic Nucleus

Learning Objectives

- Explain the atomic number, mass number, and isotopes as it relates to the particles within the nucleus of an atom.
- Calculate the average atomic mass.

We introduced the idea of an atomic nucleus in a previous chapter. Now we will discuss this part of the atom in more detail. Although the electrons are responsible for the chemical reactions of elements. The particles inside the nucleus are important to understand for other changes, such as nuclear radiation. These particles are collectively called **nucleons**, and consist of protons and neutrons. In this section we will discuss a few ways we keep track of and discuss the particles within the nucleus.

Atomic Number

The atomic number (Z) of an element is the number of protons in the nucleus of each atom of that element. This means that the number of protons is the characteristic which makes each element unique compared to all other elements. Elements are different because of their atomic number. The **periodic table** displays all of the known elements and is arranged in order of increasing atomic number. In this table, an element's atomic number is indicated above the elemental symbol. Hydrogen, at the upper left of the table, has an atomic number of 1. Every hydrogen atom has one proton in its nucleus. Following on the table is helium, whose atoms have two protons in the nucleus. Lithium atoms have three protons, and so forth.

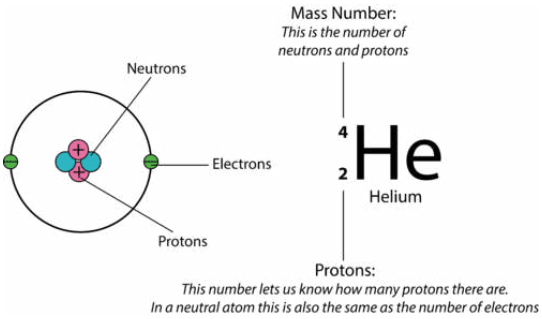


Figure 14.2.1: How can you determine the atomic number of an element. (Credit: User:Materialscientist/Wikimedia Commons; (Source) (opens in new window); License: Public Domain)

Mass Number

Rutherford showed that the vast majority of the mass of an atom is concentrated in its nucleus, which is composed of protons and neutrons. The **mass number** (A) is defined as the total number of protons and neutrons in an atom. It can be calculated by adding the number of neutrons and the number of protons (atomic number) together.

Mass number = atomic number + number of neutrons

Consider Table 14.2.1 below that shows data from the first six elements of the periodic table.

Table 14.2.1: Atoms of the First Six Elements

Name	Symbol	Protons	Neutrons	Electrons	Atomic Number	Mass Number
Hydrogen	H	1	0	1	1	1
Helium	He	2	2	2	2	4
Lithium	Li	3	4	3	3	7
Beryllium	Be	4	5	4	4	9
Boron	B	5	6	5	5	11
Carbon	C	6	6	6	6	12

Consider the element helium. Its atomic number is 2, so it has two protons in its nucleus. Its nucleus also contains two neutrons. Since $2 + 2 = 4$, we know that the mass number of the helium atom is 4. Finally, the helium atom also contains two electrons since the number of electrons must equal the number of protons. This example may lead you to believe that atoms have the same number of protons and neutrons, but further examination of the table above will show that this is not the case. Lithium, for example, has three protons and four neutrons, leaving it with a mass number of 7.

Knowing the mass number and the atomic number of an atom allows you to determine the number of neutrons present in that atom by subtraction.

$$\text{Number of neutrons} = \text{mass number} - \text{atomic number}$$

Example 14.2.1

Atoms of the element chromium (Cr) have an atomic number of 24 and a mass number of 52. How many neutrons are in the nucleus of a chromium atom?

Solution

To determine this, you subtract the atomic number from the mass number, as shown above:

$$52 - 24 = 28 \text{ neutrons in a chromium atom}$$

The composition of any atom can be illustrated with a shorthand notation using the atomic number and the mass number. Both are written before the chemical symbol, with the mass number written as a superscript and the atomic number written as a subscript. The chromium atom discussed above would be written as:

$$^{52}_{24}\text{Cr}$$

Another way to refer to a specific atom is to write the mass number of the atom after the name, separated by a hyphen. The above atom would be written as chromium-52.

Isotopes

Although John Dalton stated in his atomic theory of 1804 that all atoms of an element are identical, the discovery of the neutron began to show that this assumption was not correct. The study of radioactive materials (elements that spontaneously give off particles to form new elements) by Frederick Soddy (1877-1956) gave important clues about the internal structure of atoms. His work showed that some substances with different radioactive properties and different atomic masses were in fact the same element. He coined the term **isotope** from the Greek roots isos (ίσος "equal") and topos (τόπος "place"). He described isotopes as, "Put colloquially, their atoms have identical outsides but different insides." Soddy won the Nobel Prize in Chemistry in 1921 for his work.

As stated earlier, not all atoms of a given element are identical. Specifically, the number of neutrons can be variable for many elements. As an example, naturally occurring carbon exists in three forms. Each carbon atom has the same number of protons (6), which is its atomic number. Each carbon atom also contains six electrons in order to maintain electrical neutrality. However the number of neutrons varies as six, seven, or eight. Isotopes are atoms that have the same number atomic number, but different mass numbers due to a change in the number of neutrons.

When referring to a single type of nucleus, we often use the term **nuclide** and identify it by the notation $^{Z}_{A}\text{X}$, where X is the symbol for the element, A is the mass number, and Z is the atomic number (for example, $^{14}_{6}\text{C}$). Often a nuclide is referenced by the name of the element followed by a hyphen and the mass number. For example, $^{14}_{6}\text{C}$ is called "carbon-14."

The other two isotopes of carbon can be referred to as carbon-12 ($^{12}_{6}C$), and carbon-13 ($^{13}_{6}C$). A carbon atom is one of these three different nuclides. Most elements naturally consist of combinations of isotopes. Carbon has three natural isotopes, while some heavier elements can have many more. Tin has ten stable isotopes, the most of any element.

While the presence of isotopes affects the mass of an atom, it does not affect its chemical reactivity. Chemical behavior is governed by the number of electrons and the number of protons. Carbon-13 behaves chemically in exactly the same way as the more plentiful carbon-12.

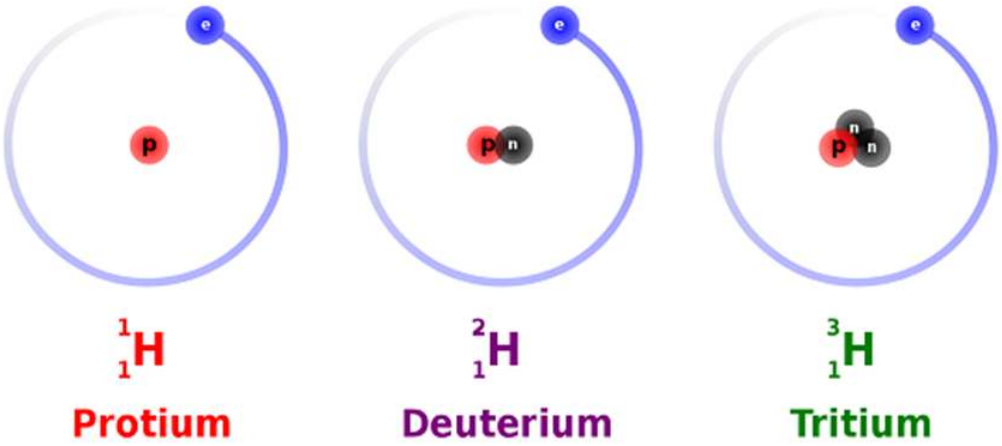


Figure **14.2.2**:Thee isotopes of hydrogen. Each has one proton, but they have zero, one, or two neutrons. Although these isotopes have specific names, most isotopes do not. (Credit: User:Dirk Hünniger/Wikimedia Commons; (Source) (opens in new window); License: CC BY-SA 3.0)
)

Mass Number

Average Atomic Mass

Most elements occur naturally as a mixture of two or more isotopes. The table below shows the natural isotopes of several elements, along with the **percent natural abundance** of each.

Table **14.2.1**: Atomic Masses and Percents of Abundance of Some Natural Isotopes

Element	Isotope (Symbol)	Percent Natural Abundance	Atomic Mass (amu)	Average Atomic Mass (amu)
Hydrogen	$^{1}_{1}H$	99.985	1.0078	1.0080
	$^{2}_{1}H$	0.015	2.0141	
	$^{3}_{1}H$	negligible	3.0160	
Carbon	$^{12}_{6}C$	98.89	12.000	12.011
	$^{13}_{6}C$	1.11	13.003	
	$^{14}_{6}C$	trace	14.003	
Oxygen	$^{16}_{8}O$	99.759	15.995	15.999
	$^{17}_{8}O$	0.037	16.995	
	$^{18}_{8}O$	0.204	17.999	
Chlorine	$^{35}_{17}Cl$	75.77	34.969	35.453
	$^{37}_{17}Cl$	24.23	36.966	
Copper	$^{63}_{29}Cu$	69.17	62.930	63.546
	$^{65}_{29}Cu$	30.83	64.928	

For some elements, one particular isotope predominates greatly over the other isotopes. Naturally occurring hydrogen is nearly all hydrogen-1 and naturally occurring oxygen is nearly all oxygen-16. For many other elements, however, more than one isotope may exist in more substantial quantities. Chlorine (atomic number 17) is a yellowish-green toxic gas. About three quarters of all chlorine atoms have 18 neutrons, giving those atoms a mass number of 35. About one quarter of all chlorine atoms have 20 neutrons, giving

those atoms a mass number of 37. Were you to simply calculate the arithmetic average of the precise **atomic masses**, you would get 36.

$$\frac{(34.969 + 36.966)}{2} = 35.968 \text{ amu}$$

Clearly the actual average atomic mass from the last column of the table is significantly lower. Why? We need to take into account the percent natural abundance of each isotope, in order to calculate the weighted average. The atomic mass of an element is the weighted average of the atomic masses of the naturally occurring isotopes of that element. The sample problem below demonstrates how to calculate the atomic mass of chlorine.

Example 14.2.1

Use the atomic masses of each of the two isotopes of chlorine along with their respective percent abundances to calculate the average atomic mass of chlorine.

Solution

Step 1: List the known and unknown quantities and plan the problem.

Known

- Chlorine-35: atomic mass $= 34.969$ amu and percent abundance $= 75.77\%$
- Chlorine-37: atomic mass $= 36.966$ amu and percent abundance $= 24.23\%$

Unknown

- Average atomic mass of chlorine

Change each percent abundance into decimal form by dividing by 100. Multiply this value by the atomic mass of that isotope. Add together for each isotope to get the average atomic mass.

Step 2: Calculate.

chlorine-35	$0.7577 \times 34.969 = 26.50$ amu
chlorine-37	$0.2423 \times 36.966 = 8.957$ amu
average atomic mass	$26.50 + 8.957 = 35.46$ amu

Note: Applying significant figure rules results in the 35.45 amu result without excessive rounding error. In one step:

$$(0.7577 \times 34.969) + (0.2423 \times 36.966) = 35.46 \text{ amu}$$

Step 3: Think about your result.

The calculated average atomic mass is closer to 35 than to 37 because a greater percentage of naturally occurring chlorine atoms have the mass number of 35. It agrees with the value from the table above.

Section Summary

- Rutherford proposed a model of the atomic nucleus which had a solid core.
- The atomic number (Z) of an element is the number of protons in the nucleus of each atom of that element.
- The number of electrons is equal to the number of protons in an atom of an element.
- The mass number is defined as the total number of protons and neutrons in an atom.
- The mass number = number of neutrons + atomic number.
- Isotopes are atoms that have the same atomic number, but different mass numbers due to a change in the number of neutrons.
- The term nuclide refers to the nucleus of a given isotope of an element.
- The atomic mass of an atom equals the sum of the protons and the neutrons.
- The atomic mass of an element is the weighted average of the atomic masses of the naturally occurring isotopes of that element.
- Calculations of atomic mass use the percent abundance of each isotope.

Glossary

nucleons

The particles inside the nucleus of an atom (protons and neutrons).

mass number (A)

The total number of protons and neutrons in an atom.

atomic number (Z)

The number of protons in the nucleus of each atom of that element.

isotopes

Atoms which have the same number of protons (and hence atomic number), but different number of neutrons (and hence different mass numbers).

nuclide

A single type of atomic nucleus.

atomic mass

The weighted average of all of the naturally occurring isotopes of a given element.

This page titled 14.2: The Atomic Nucleus is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **4.17: Mass Number** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **4.18: Isotopes** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **4.20: Calculating Average Atomic Mass** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **4.15: Atomic Nucleus** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

14.3: Discovery of Radioactivity

Learning Objectives

- Describe the discovery of nuclear radiation.

In 1895, Wilhelm Röntgen produced great interest around what caused the phenomenon of x-rays. One researcher of this phenomenon was Henri Becquerel. Becquerel studied the fluorescent properties of uranium salts, believing they had something to do with x-rays. He soon learned that uranium could expose a photographic plate without an external input of energy (thought to be needed to produce uranium fluorescence). Becquerel also showed that the uranium salt emissions could be affected by a magnetic field, which was not true for x-rays.



Figure **14.3.1**: Becquerel's experiment: uranium exposure to photographic plate.

Pierre and Marie Curie studied the properties of uranium salts with the express purpose of identifying the details of these emissions. They were the first to coin the term "**radioactivity**", meaning the spontaneous emission of radiation in the form of particles or high energy photons, resulting from a nuclear reaction. The major contributions to the work came from Marie, who showed that the amount of radioactivity present was due to the amount of a specific element, and not due to some chemical reaction. She discovered the element polonium and named it after her native Poland. Madame Curie shared the 1903 Nobel Prize in Physics with her husband Pierre and Henri Becquerel. She won the Nobel Prize in Chemistry in 1911, the sole recipient. Marie Curie was the first person to win multiple Nobel Prizes. There have only been four people to ever win multiple Nobel Prizes, and only two to win Prizes in different categories.

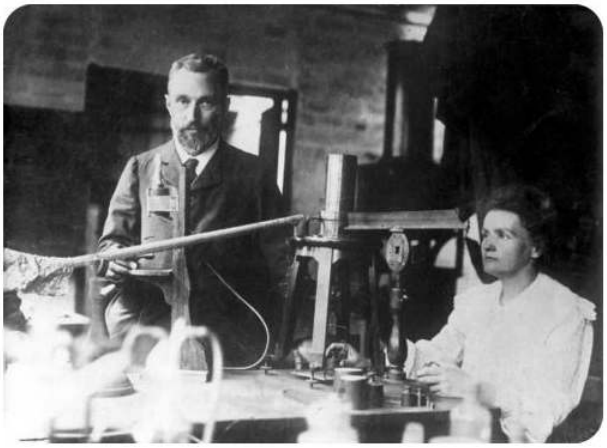


Figure **14.3.2**: Pierre and Marie Curie in their lab.

Ernest Rutherford, a later researcher, was able to demonstrate three different types of radioactive emissions. These emission types differed in terms of mass, charge, and their ability to penetrate materials. He designated them simply as alpha (α) emissions, beta (β) emissions, and gamma (γ) emissions. These emissions will be explored in greater detail later in this chapter, along with other nuclear processes that have since been discovered.

Radioactivity involves the spontaneous emission of material and/or energy from the nucleus of an atom. The most common radioactive atoms have high atomic numbers and contain a large excess of neutrons. Some typical radioactive elements are technetium (atomic number 43), promethium (atomic number 61), and all elements atomic number 84 (polonium) and higher. There are four primary types of emission, either involving the release of a particle from the nucleus or the release of energy. In many instances, both energy and a particle are produced by the radioactive event.

It should be noted that some elements considered to be stable do have radioactive isotopes. Carbon-14 is radioactive, but this isotope is only a small fraction of the total amount of carbon in existence (about one part per trillion carbon atoms). Hydrogen also has a radioactive isotope known as hydrogen-3 or tritium, again a very small fraction of the total elemental hydrogen present.

Section Summary

- Henri Becquerel showed that uranium salt emissions could be affected by a magnetic field, which was not true for x-rays.
- Marie and Pierre Curie were the first to coin the term *radioactivity*, meaning the spontaneous emission of radiation in the form of particles or high energy photons, resulting from a nuclear reaction.
- Ernest Rutherford, a later researcher, was able to demonstrate three different types of radioactive emissions: alpha (α) emissions, beta (β) emissions, and gamma (γ) emissions.

This page titled 14.3: Discovery of Radioactivity is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **24.1: Discovery of Radioactivity** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

14.4: Nuclear Decay Processes

Learning Objectives

- Describe the composition and properties of the different types of radiation.

Radioactive decay involves the emission of a particle and/or energy as one atom changes into another. In most instances, the atom changes its identity to become a new element. There are four different types of emissions that occur. We will explore each of them in this section, as well as some of their properties.

Alpha Emission

Alpha (α) decay involves the release of helium ions from the nucleus of an atom. This ion consists of two protons and two neutrons and has a $2+$ charge. Release of an α-particle produces a new atom that has an atomic number two less than the original atom and an atomic weight that is four less. A typical alpha decay reaction is the conversion of uranium-238 to thorium:

$$^{238}_{92}\text{U} \rightarrow {}^{234}_{90}\text{Th} + {}^{4}_{2}\alpha$$

We see a decrease of two in the atomic number (uranium to thorium) and a decrease of four in the atomic weight (238 to 234). Usually the emission is not written with atomic number and weight indicated since it is a common particle whose properties should be memorized. Quite often the alpha emission is accompanied by gamma (γ) radiation, a form of energy release. Many of the largest elements in the periodic table are alpha-emitters.

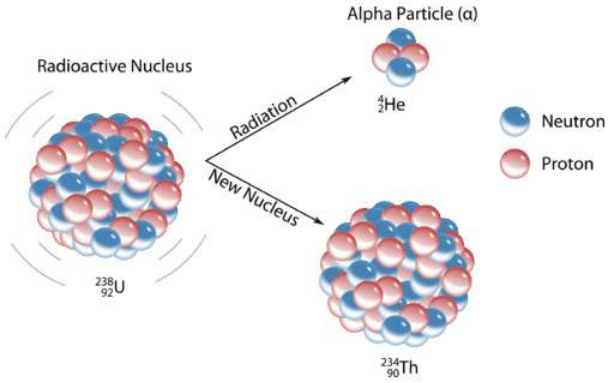


Figure 14.4.1: Emission of an alpha particle from the nucleus.

Beta Emission

Beta (β) decay is a more complicated process. Unlike the α-emission, which simply expels a particle, the β-emission involves the transformation of a neutron in the nucleus to a proton and an electron. The electron is then ejected from the nucleus. In the process, the atomic number increases by one while the atomic weight stays the same. As is the case with α-emissions, β-emissions are often accompanied by γ-radiation.

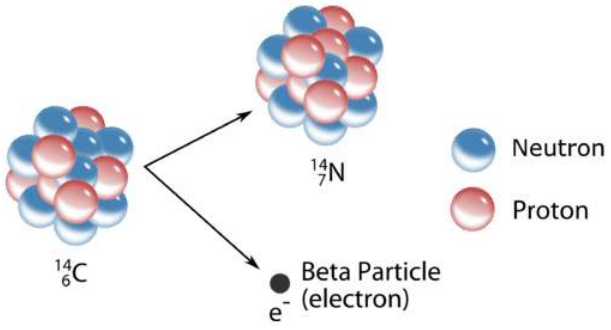


Figure 14.4.2: Beta emission.

A typical beta decay process involves carbon-14, often used in radioactive dating techniques. The reaction forms nitrogen-14 and an electron:

$$^{14}_{6}\mathrm{C} \rightarrow {}^{14}_{7}\mathrm{N} + {}^{0}_{-1}\mathrm{e}$$

Again, the beta emission is usually simply indicated by the Greek letter β; memorization of the process is necessary in order to follow nuclear calculations in which the Greek letter β appears without further notation.

Positron Emission

A **positron** is a positive electron (a form of antimatter). This rare type of emission occurs when a proton is converted to a neutron and a positron in the nucleus, with ejection of the positron. The atomic number will decrease by one, while the atomic weight does not change. A positron is often designated by β^{+}.

Carbon-11 emits a positron to become boron-11:

$$^{11}_{6}\mathrm{C} \rightarrow {}^{11}_{5}\mathrm{B} + {}^{0}_{+1}\beta$$

Electron Capture

An alternate way for a nuclide to increase its neutron to proton ratio is by a phenomenon called electron capture. In electron capture, an electron from an inner orbital is captured by the nucleus of the atom, and combined with a proton to form a neutron. For example, silver-106 undergoes electron capture to become palladium-106.

$$^{106}_{47}\mathrm{Ag} + {}^{0}_{-1}\mathrm{e} \rightarrow {}^{106}_{46}\mathrm{Pd}$$

Note that the overall result of electron capture is identical to positron emission. The atomic number decreases by one while the mass number remains the same.

Gamma Emission

Gamma (γ) radiation is high energy electromagnetic radiation, as introduced earlier in this text. It may be released by itself or, more commonly, in association with other radiation events. There is no change of atomic number or atomic weight in a simple γ-emission. Often, an isotope may produce γ-radiation as a result of a transition in a metastable isotope. This type of isotope may just "settle", with a shifting of particles in the nucleus. The composition of the atom is not altered, but the nucleus could be considered more "comfortable" after the shift. This shift increases the stability of the isotope from the energetically unstable (or "metastable") isotope to a more stable form of the nucleus.

Penetrating Ability of Emissions

The various emissions differ considerably in their ability to go through matter, known as their **penetrating ability**. The α-particle has the least penetrating power since it is the largest and slowest emission. It can be blocked by a sheet of paper or a human hand. Beta particles are more penetrating than alpha particles, but can be stopped by a thin sheet of aluminum. Of the three basic types of emissions, gamma particles are the most penetrating. A thick lead shield is required to stop gamma emissions. Positrons represent a special case in that they annihilate when they come in contact with electrons. The collision of a positron and an electron results in the formation of two gamma emissions that go 180 degrees away from each other.

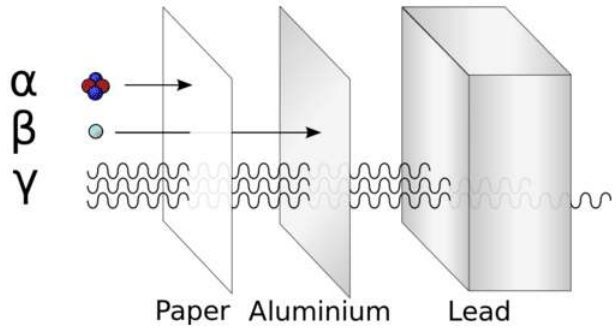


Figure **14.4.3**: Penetrating ability of radioactive emissions.

Blocking of alpha particles can easily be accomplished by as little as $10\ \text{mm}$ plastic or paper. Beta emissions represent a somewhat different situation. The negative charge on a beta particle has the potential for activating the element being used to block the radiation. Lead and tungsten are large atoms with many protons and neutrons in their nuclei. While the beta electron may be blocked, the target material could become irradiated in the process.

High-density materials are much more effective protection against gamma emissions than low-density ones. Gamma rays are usually blocked effectively by lead shielding. The thickness of the shielding will determine the effectiveness of the protection offered by the lead.

Section Summary

- Radioactive decay involves the emission of a particle and/or energy as one atom changes into another.
- Alpha (α) decay involves the release of helium ions from the nucleus of an atom.
- Beta (β) decay involves the transformation of a neutron in the nucleus to a proton and an electron.
- A positron is a positive electron (a form of antimatter). This rare type of emission occurs when a proton is converted to a neutron and a positron in the nucleus, with ejection of the positron.
- Gamma (γ) radiation is simply energy. It may be released by itself or, more commonly, in association with other radiation events.
- The penetrating ability of a radioactive emission is its ability to go through matter.
- The relative penetrating abilities of radioactive emissions are described.

Glossary

Alpha (α) decay
The release of helium ions from the nucleus of an atom.

Beta (β) decay
The release of an electron from the nucleus, as a neutron transforms into a proton.

positron emission
The release of a positive electron from the nucleus, as a proton transforms into a neutron.

Gamma (γ)
High energy electromagnetic radiation.

Penetrating ability
Ability of nuclear radiation to go through matter.

Electron capture (EC)
An electron from an inner orbital is captured by the nucleus of the atom, and combined with a proton to form a neutron.

This page titled 14.4: Nuclear Decay Processes is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **24.2: Nuclear Decay Processes** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **24.9: Penetrating Ability of Emissions** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

14.5: Nuclear Equations

Learning Objectives

- Write and balance nuclear equations.

A balanced chemical reaction equation reflects the fact that during a chemical reaction, bonds break and form, and atoms are rearranged, but the total numbers of atoms of each element are conserved and do not change. A balanced nuclear reaction equation indicates that there is a rearrangement during a nuclear reaction, but of subatomic particles rather than atoms. Nuclear reactions also follow conservation laws, and they are balanced in two ways:

1. The sum of the mass numbers of the reactants equals the sum of the mass numbers of the products.
2. The sum of the charges of the reactants equals the sum of the charges of the products.

If the atomic number and the mass number of all but one of the particles in a nuclear reaction are known, we can identify the particle by balancing the reaction. For instance, we could determine that $^{17}_{8}\text{O}$ is a product of the nuclear reaction of $^{14}_{7}\text{N}$ and $^{4}_{2}\text{He}$ if we knew that a proton, $^{1}_{1}\text{H}$, was one of the two products. Example **14.5.1** shows how we can identify a nuclide by balancing the nuclear reaction.

Example 14.5.1: Balancing Equations for Nuclear Reactions

The reaction of an α particle with magnesium-25 ($^{25}_{12}\text{Mg}$) produces a proton and a nuclide of another element. Identify the new nuclide produced.

Solution

The nuclear reaction can be written as:

$$^{25}_{12}\text{Mg} + {}^{4}_{2}\text{He} \rightarrow {}^{1}_{1}\text{H} + {}^{A}_{Z}\text{X}$$

where

- **A** is the mass number and
- **Z** is the atomic number of the new nuclide, **X**.

Because the sum of the mass numbers of the reactants must equal the sum of the mass numbers of the products:

$$25 + 4 = A + 1$$

so

$$A = 28$$

Similarly, the charges must balance, so:

$$12 + 2 = Z + 1$$

so

$$Z = 13$$

Check the periodic table: The element with nuclear charge = +13 is aluminum. Thus, the product is $^{28}_{13}\text{Al}$.

Exercise 14.5.1

The nuclide $^{125}_{53}\text{I}$ combines with an electron and produces a new nucleus and no other massive particles. What is the equation for this reaction?

Answer

$$^{125}_{53}\text{I} + {}^{0}_{-1}\text{e} \rightarrow {}^{125}_{52}\text{Te}$$

The two general kinds of nuclear reactions are nuclear decay reactions and nuclear transmutation reactions**.** In a **nuclear decay reaction**, also called radioactive decay, an unstable nucleus emits radiation and is transformed into the nucleus of one or more other elements. The resulting daughter nuclei have a lower mass and are lower in energy (more stable) than the parent nucleus that decayed. In contrast, in a **nuclear transmutation reaction**, a nucleus reacts with a subatomic particle or another nucleus to form a product nucleus that is *more massive* than the starting material. As we shall see, nuclear decay reactions occur spontaneously under all conditions, but nuclear transmutation reactions occur only under very special conditions, such as the collision of a beam of highly energetic particles with a target nucleus or in the interior of stars. We begin this section by considering the different classes of radioactive nuclei, along with their characteristic nuclear decay reactions and the radiation they emit.

> *Nuclear decay reactions occur spontaneously under all conditions, whereas nuclear transmutation reactions are induced.*

Nuclear Decay Reactions

Just as we use the number and type of atoms present to balance a chemical equation, we can use the number and type of nucleons present to write a balanced nuclear equation for a nuclear decay reaction. This procedure also allows us to predict the identity of either the parent or the daughter nucleus if the identity of only one is known. Regardless of the mode of decay, the total number of nucleons is conserved in all nuclear reactions.

To describe nuclear decay reactions, chemists have extended the $^{A}_{Z}\text{X}$ notation for nuclides to include radioactive emissions. Table **14.5.1** lists the name and symbol for each type of emitted radiation. The most notable addition is the **positron**, a particle that has the same mass as an electron but a positive charge rather than a negative charge.

Table **14.5.1**: Nuclear Decay Emissions and Their Symbols

Identity	Symbol	Charge	Mass (amu)
helium nucleus	$^{4}_{2}\alpha$	+2	4.001506
electron	$^{0}_{-1}\beta$ or β^{-}	−1	0.000549
photon	$^{0}_{0}\gamma$	—	—
neutron	$^{1}_{0}\text{n}$	0	1.008665
proton	$^{1}_{1}\text{p}$	+1	1.007276
positron	$^{0}_{+1}\beta$ or β^{+}	+1	0.000549

Like the notation used to indicate isotopes, the upper left superscript in the symbol for a particle gives the mass number, which is the total number of protons and neutrons. For a proton or a neutron, $A = 1$. Because neither an electron nor a positron contains protons or neutrons, its mass number is 0. The numbers should not be taken literally, however, as meaning that these particles have zero mass; ejection of a beta particle (an electron) simply has a negligible effect on the mass of a nucleus.

Similarly, the lower left subscript gives the charge of the particle. Because protons carry a positive charge, $Z = +1$ for a proton. In contrast, a neutron contains no protons and is electrically neutral, so $Z = 0$. In the case of an electron, $Z = −1$, and for a positron, $Z = +1$. Because γ rays are high-energy photons, both A and Z are 0. In some cases, two different symbols are used for particles that are identical but produced in different ways. For example, the symbol $^{0}_{-1}\text{e}$, which is usually simplified to e^{-}, represents a free electron or an electron associated with an atom, whereas the symbol $^{0}_{-1}\beta$, which is often simplified to β^{-}, denotes an electron that originates from within the nucleus, which is a β particle. Similarly, $^{4}_{2}\text{He}^{2+}$ refers to the nucleus of a helium atom, and $^{4}_{2}\alpha$ denotes an identical particle that has been ejected from a heavier nucleus.

There are six fundamentally different kinds of nuclear decay reactions, and each releases a different kind of particle or energy. The essential features of each reaction are shown in Figure **14.5.1**. The most common are alpha and beta decay and gamma emission, but the others are essential to an understanding of nuclear decay reactions.

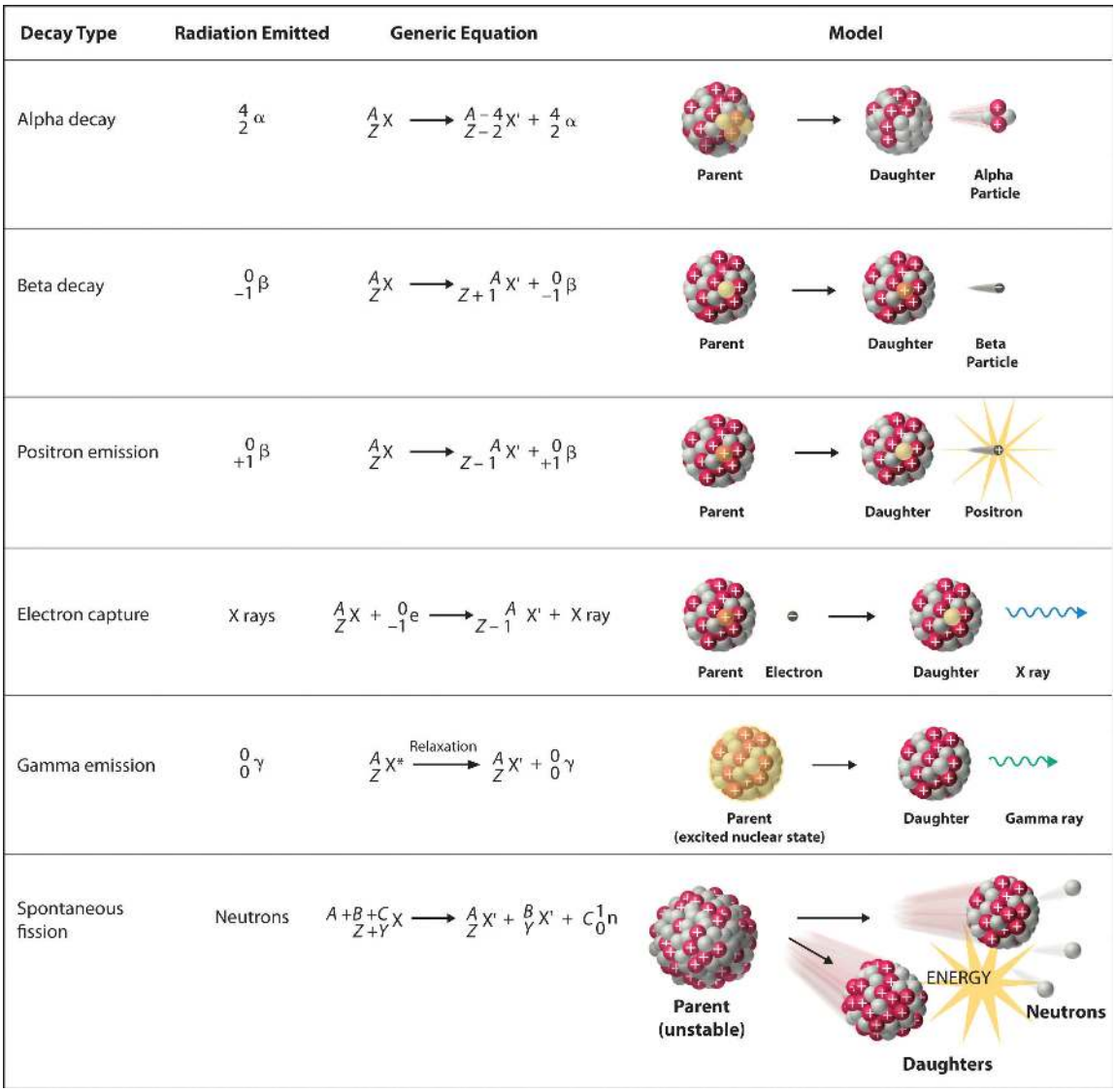


Figure **14.5.1**: Common Modes of Nuclear Decay

The different types of decay are alpha, beta, positron emission, electron capture, gamma emission, and spontaneous fission.

Alpha α Decay

Many nuclei with mass numbers greater than 200 undergo **alpha (α) decay**, which results in the emission of a helium-4 nucleus as an **alpha (α) particle**, ${}^{4}_{2}\alpha$. The general reaction is as follows:

$$\underset{\text{parent}}{{}^{A}_{Z}\mathrm{X}} \rightarrow \underset{\text{daughter}}{{}^{A-4}_{Z-2}\mathrm{X}'} + \underset{\text{alpha particle}}{{}^{4}_{2}\alpha} \tag{14.5.1}$$

The daughter nuclide contains two fewer protons and two fewer neutrons than the parent. Thus, α-particle emission produces a daughter nucleus with a mass number $A - 4$ and a nuclear charge $Z - 2$ compared to the parent nucleus. Radium-226, for example, undergoes alpha decay to form radon-222:

$${}^{226}_{88}\mathrm{Ra} \rightarrow {}^{222}_{86}\mathrm{Rn} + {}^{4}_{2}\alpha \tag{14.5.2}$$

Because nucleons are conserved in this and all other nuclear reactions, the sum of the mass numbers of the products, 222 + 4 = 226, equals the mass number of the parent. Similarly, the sum of the atomic numbers of the products, 86 + 2 = 88, equals the atomic number of the parent. Thus, the nuclear equation is balanced.

> *Just as the total number of atoms is conserved in a chemical reaction, the total number of nucleons is conserved in a nuclear reaction.*

Beta β^- Decay

Nuclei that contain too many neutrons often undergo **beta (β) decay**, in which a neutron is converted to a proton and a high-energy electron that is ejected from the nucleus as a β particle:

$$\underset{\substack{\text{unstable}\\\text{neutron in}\\\text{nucleus}}}{{}^{1}_{0}\text{n}} \rightarrow \underset{\substack{\text{proton}\\\text{retained}\\\text{by nucleus}}}{{}^{1}_{1}\text{p}} + \underset{\substack{\text{beta particle}\\\text{emitted by}\\\text{nucleus}}}{{}^{0}_{-1}\beta} \tag{14.5.3}$$

The general reaction for beta decay is therefore

$$\underset{\text{parent}}{{}^{A}_{Z}\text{X}} \rightarrow \underset{\text{daughter}}{{}^{A}_{Z+1}\text{X}'} + \underset{\text{beta particle}}{{}^{0}_{-1}\beta} \tag{14.5.4}$$

Although beta decay does not change the mass number of the nucleus, it does result in an increase of +1 in the atomic number because of the addition of a proton in the daughter nucleus. Thus, beta decay decreases the neutron-to-proton ratio, moving the nucleus toward the band of stable nuclei. For example, carbon-14 undergoes beta decay to form nitrogen-14:

$${}^{14}_{6}\text{C} \rightarrow {}^{14}_{7}\text{N} + {}^{0}_{-1}\beta$$

Once again, the number of nucleons is conserved, and the charges are balanced. The parent and the daughter nuclei have the same mass number, 14, and the sum of the atomic numbers of the products is 6, which is the same as the atomic number of the carbon-14 parent.

Positron β^+ Emission

Because a positron has the same mass as an electron but opposite charge, **positron emission** is the opposite of beta decay. Thus positron emission is characteristic of neutron-poor nuclei, which decay by transforming a proton to a neutron and emitting a high-energy positron:

$${}^{1}_{1}\text{p}^{+} \rightarrow {}^{1}_{0}\text{n} + {}^{0}_{+1}\beta^{+} \tag{14.5.5}$$

The general reaction for positron emission is therefore

$$\underset{\text{parent}}{{}^{A}_{Z}\text{X}} \rightarrow \underset{\text{daughter}}{{}^{A}_{Z-1}\text{X}'} + \underset{\text{positron}}{{}^{0}_{+1}\beta^{+}}$$

Like beta decay, positron emission does not change the mass number of the nucleus. In this case, however, the atomic number of the daughter nucleus is lower by 1 than that of the parent. Thus, the neutron-to-proton ratio has increased, again moving the nucleus closer to the band of stable nuclei. For example, carbon-11 undergoes positron emission to form boron-11:

$${}^{11}_{6}\text{C} \rightarrow {}^{11}_{5}\text{B} + {}^{0}_{+1}\beta^{+}$$

Nucleons are conserved, and the charges balance. The mass number, 11, does not change, and the sum of the atomic numbers of the products is 6, the same as the atomic number of the parent carbon-11 nuclide.

Electron Capture

A neutron-poor nucleus can decay by either positron emission or **electron capture (EC)**, in which an electron in an inner shell reacts with a proton to produce a neutron:

$${}^{1}_{1}\text{p} + {}^{0}_{-1}\text{e} \rightarrow {}^{1}_{0}\text{n} \tag{14.5.6}$$

When a second electron moves from an outer shell to take the place of the lower-energy electron that was absorbed by the nucleus, an x-ray is emitted. The overall reaction for electron capture is thus

$$\underset{\text{parent}}{{}^{A}_{Z}\text{X}} + \underset{\text{electron}}{{}^{0}_{-1}\text{e}} \rightarrow \underset{\text{daughter}}{{}^{A}_{Z-1}\text{X}'} + \text{x-ray}$$

Electron capture does not change the mass number of the nucleus because both the proton that is lost and the neutron that is formed have a mass number of 1. As with positron emission, however, the atomic number of the daughter nucleus is lower by 1 than that of the parent. Once again, the neutron-to-proton ratio has increased, moving the nucleus toward the band of stable nuclei. For example, iron-55 decays by electron capture to form manganese-55, which is often written as follows:

$$^{55}_{26}\text{Fe} \xrightarrow{\text{EC}} {}^{55}_{25}\text{Mn} + \text{x-ray}$$

The atomic numbers of the parent and daughter nuclides differ in Equation 20.2.11, although the mass numbers are the same. To write a balanced nuclear equation for this reaction, we must explicitly include the captured electron in the equation:

$$^{55}_{26}\text{Fe} + {}^{0}_{-1}\text{e} \rightarrow {}^{55}_{25}\text{Mn} + \text{x-ray}$$

Both positron emission and electron capture are usually observed for nuclides with low neutron-to-proton ratios, but the decay rates for the two processes can be very different.

Gamma γ Emission

Many nuclear decay reactions produce daughter nuclei that are in a nuclear excited state, which is similar to an atom in which an electron has been excited to a higher-energy orbital to give an electronic excited state. Just as an electron in an electronic excited state emits energy in the form of a photon when it returns to the ground state, a nucleus in an excited state releases energy in the form of a photon when it returns to the ground state. These high-energy photons are γ rays. **Gamma (γ) emission** can occur virtually instantaneously, as it does in the alpha decay of uranium-238 to thorium-234, where the asterisk denotes an excited state:

$$^{238}_{92}\text{U} \rightarrow \underset{\substack{\text{excited}\\ \text{nuclear}\\ \text{state}}}{^{234}_{90}\text{Th}^*} + {}^{4}_{2}\alpha \xrightarrow{\text{relaxation}} {}^{234}_{90}\text{Th} + {}^{0}_{0}\gamma$$

If we disregard the decay event that created the excited nucleus, then

$$^{234}_{88}\text{Th}^* \rightarrow {}^{234}_{88}\text{Th} + {}^{0}_{0}\gamma$$

or more generally,

$$^{A}_{Z}\text{X}^* \rightarrow {}^{A}_{Z}\text{X} + {}^{0}_{0}\gamma$$

Gamma emission can also occur after a significant delay. For example, technetium-99*m* has a half-life of about 6 hours before emitting a γ ray to form technetium-99 (the *m* is for metastable). Because γ rays are energy, their emission does not affect either the mass number or the atomic number of the daughter nuclide. Gamma-ray emission is therefore the only kind of radiation that does not necessarily involve the conversion of one element to another, although it is almost always observed in conjunction with some other nuclear decay reaction.

✔ Example 14.5.2

Write a balanced nuclear equation to describe each reaction.

a. the beta decay of $^{35}_{16}\text{S}$
b. the decay of $^{201}_{80}\text{Hg}$ by electron capture
c. the decay of $^{30}_{15}\text{P}$ by positron emission

Given: radioactive nuclide and mode of decay

Asked for: balanced nuclear equation

Strategy:

A Identify the reactants and the products from the information given.

B Use the values of *A* and *Z* to identify any missing components needed to balance the equation.

Solution

a.

A We know the identities of the reactant and one of the products (a β particle). We can therefore begin by writing an equation that shows the reactant and one of the products and indicates the unknown product as ${}^{A}_{Z}\mathrm{X}$:

$$ {}^{35}_{16}\mathrm{S} \rightarrow {}^{A}_{Z}\mathrm{X} + {}^{0}_{-1}\beta $$

B Because both protons and neutrons must be conserved in a nuclear reaction, the unknown product must have a mass number of $A = 35 - 0 = 35$ and an atomic number of $Z = 16 - (-1) = 17$. The element with $Z = 17$ is chlorine, so the balanced nuclear equation is as follows:

$$ {}^{35}_{16}\mathrm{S} \rightarrow {}^{35}_{17}\mathrm{Cl} + {}^{0}_{-1}\beta $$

b.

A We know the identities of both reactants: ${}^{201}_{80}\mathrm{Hg}$ and an inner electron, ${}^{0}_{-1}\mathrm{e}$. The reaction is as follows:

$$ {}^{201}_{80}\mathrm{Hg} + {}^{0}_{-1}\mathrm{e} \rightarrow {}^{A}_{Z}\mathrm{X} $$

B Both protons and neutrons are conserved, so the mass number of the product must be $A = 201 + 0 = 201$, and the atomic number of the product must be $Z = 80 + (-1) = 79$, which corresponds to the element gold. The balanced nuclear equation is thus

$$ {}^{201}_{80}\mathrm{Hg} + {}^{0}_{-1}\mathrm{e} \rightarrow {}^{201}_{79}\mathrm{Au} $$

c.

A As in part (a), we are given the identities of the reactant and one of the products—in this case, a positron. The unbalanced nuclear equation is therefore

$$ {}^{30}_{15}\mathrm{P} \rightarrow {}^{A}_{Z}\mathrm{X} + {}^{0}_{+1}\beta $$

B The mass number of the second product is $A = 30 - 0 = 30$, and its atomic number is $Z = 15 - 1 = 14$, which corresponds to silicon. The balanced nuclear equation for the reaction is as follows:

$$ {}^{30}_{15}\mathrm{P} \rightarrow {}^{30}_{14}\mathrm{Si} + {}^{0}_{+1}\beta $$

? Exercise 14.5.2

Write a balanced nuclear equation to describe each reaction.

a. ${}^{11}_{6}\mathrm{C}$ by positron emission
b. the beta decay of molybdenum-99
c. the emission of an α particle followed by gamma emission from ${}^{185}_{74}\mathrm{W}$

Answer a

$$ {}^{11}_{6}\mathrm{C} \rightarrow {}^{11}_{5}\mathrm{B} + {}^{0}_{+1}\beta $$

Answer d

$$ {}^{99}_{42}\mathrm{Mo} \rightarrow {}^{99m}_{43}\mathrm{Tc} + {}^{0}_{-1}\beta $$

Answer c

$$ {}^{185}_{74}\mathrm{W} \rightarrow {}^{181}_{72}\mathrm{Hf} + {}^{4}_{2}\alpha + {}^{0}_{0}\gamma $$

Radioactive Decay Series

The nuclei of all elements with atomic numbers greater than 83 are unstable. Thus, all isotopes of all elements beyond bismuth in the periodic table are radioactive. Because alpha decay decreases Z by only 2, and positron emission or electron capture decreases Z by only 1, it is impossible for any nuclide with $Z > 85$ to decay to a stable daughter nuclide in a single step, except via nuclear fission. Consequently, radioactive isotopes with $Z > 85$ usually decay to a daughter nucleus that is radiaoctive, which in turn decays to a second radioactive daughter nucleus, and so forth, until a stable nucleus finally results. This series of sequential alpha- and

beta-decay reactions is called a **radioactive decay series**. We will discuss some of the implications of the naturally occuring decay series later in this chapter as we address half-life.

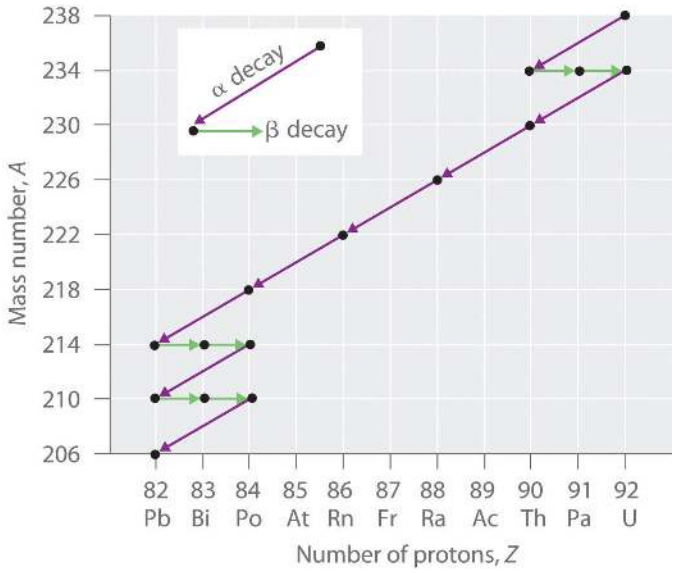


Figure **14.5.2** A Radioactive Decay Series. Three naturally occurring radioactive decay series are known to occur currently: the uranium-238 decay series, the decay of uranium-235 to lead-207, and the decay of thorium-232 to lead-208.

Graph of mass number against number of protons. The purple lines are alpha decay which are linear while the green lines are beta decay and are horizontal and parallel to the x axis.

Nuclear Transmutation Reactions

The same process of balancing the masses in nuclear decay reactions can also be effective in balancing more complicated nuclear reactions. These include induced nuclear transmutations as well as some naturally occurring processes. Induced nuclear transmutations are necessary for the synthesis of elements with more protons than uranium (92), and are referred to as transuranium elements. We will explore two additional types of nuclear transmutation later in this chapter: fission, and fusion. Both of these can be either induced or naturally occurring.

Section Summary

- Nuclear equations are balanced by making sure the number of protons and neutrons remain the same on both sides of the reaction.
- Nuclear reactions might be either naturally occurring or induced.

Glossary

nuclear decay reaction (radioactive decay)

An unstable nucleus spontaneously emits radiation and is transformed into the nucleus of one or more other elements.

nuclear transmutation reaction

a nucleus reacts with a subatomic particle or another nucleus to form a product nucleus that is *more massive* than the starting material.

radioactive decay series

series of sequential alpha- and beta-decay reactions that eventually results in a non-radioactive element.

14.5: Nuclear Equations is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

- **21.1: Radioactivity** is licensed CC BY-NC-SA 3.0.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

14.6: Half-Life

Learning Objectives

- Perform calculations based on the half-life concept

Uranium isotopes produce plutonium-239 as a decay product. The plutonium can be used in nuclear weapons and is a power source for nuclear reactors, which generate electricity. This isotope has a half-life of 24,100 years, causing concern in regions where radioactive plutonium has accumulated and is stored. At some storage sites, the waste is slowly leaking into the groundwater and contaminating nearby rivers. The 24,100 year half-life means that it will be present in the earth for a very long time.

Half-Life

Radioactive materials lose some of their activity each time a decay event occurs. This loss of activity can be estimated by determining the **half-life** of an isotope. The half-life is defined as the period of time needed for one half of a given quantity of a substance to undergo a change. For a radioisotope, every time a decay event occurs, a count is detected on the Geiger counter, or other measuring device. A specific isotope might have a total count of $30,000$ cpm. In one hour, the count could be $15,000$ cpm (half of the original count). So, the half-life of that isotope is one hour. Some isotopes have long half-lives—the half-life of U-234 is 245,000 years. Other isotopes have shorter half-lives. I-131, used in thyroid scans, has a half-life of 8.02 days.

Half-Life Calculations

Information on the half-life of an isotope can be used to calculate how much radioactivity of that isotope will be present after a certain period of time. There is a formula that allows calculation at any time after the initial count, but we are just going to look at loss of activity after different half-lives. The isotope I-125 is used in certain laboratory procedures and has a half-life of 59.4 days. If the initial activity of a sample of I-125 is $32,000$ cpm, how much activity will be present in 178.2 days? Begin by determining how many half-lives are represented by 178.2 days:

$$\frac{178.2 \text{ days}}{59.4 \text{ days/half-life}} = 3 \text{ half-lives}$$

Then, simply count activity:

$$\begin{aligned} \text{initial activity } (t_0) &= 32,000 \text{ cpm} \\ \text{after one half-life} &= 16,000 \text{ cpm} \\ \text{after two half-lives} &= 8,000 \text{ cpm} \\ \text{after three half-lives} &= 4,000 \text{ cpm} \end{aligned}$$

Be sure to keep in mind that the initial count is at time zero (t_0), and to subtract from that count at the first half-life. The second half-life has an activity of half the previous count (not the initial count).

For the more mathematically inclined, the following formula can be used to calculate the amount of radioactivity remaining after a given time:

$$N_t = N_0 \times (0.5)^{\text{number of half-lives}}$$

where $N_t =$ activity at time t and $N_0 =$ initial activity at time $t = 0$.

If we have an initial activity of $42,000$ cpm, what will the activity be after four half-lives?

$$\begin{aligned} N_t &= N_0 \, (0.5)^4 \\ &= (42,000) \, (0.5) \, (0.5) \, (0.5) \, (0.5) \\ &= 2625 \text{ cpm} \end{aligned}$$

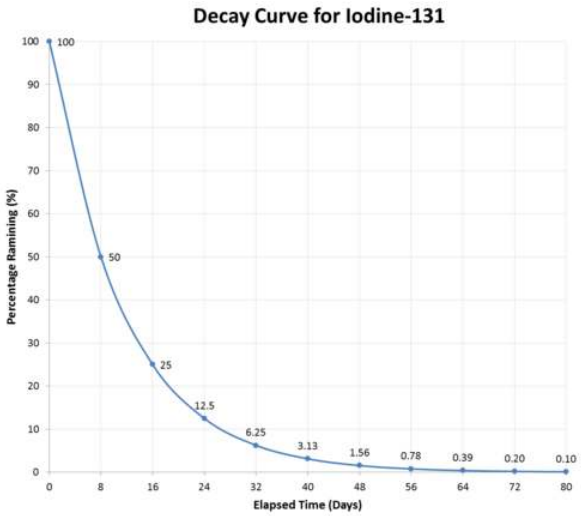


Figure **14.6.1**: Typical radioactive decay curve.

The graph above illustrates a typical decay curve for a radioactive material. The activity decreases by one-half during each succeeding half-life. Half-lives of different elements vary considerably, as shown in the table below:

Table **14.6.1**: Isotope Half-Lives

Isotope	Decay Mode	Half-Life
Cobalt-60	beta	5.3 years
Neptunium-237	alpha	2.1 million years
Polonium-214	alpha	0.00016 seconds
Radium-224	alpha	3.7 days
Tritium (^{3}H-3)	beta	12 years

Radioactive Decay Series

Earlier in this text, we discussed **radioactive decay series**. Now that we have introduced the idea of half-life, we can discuss these series in more detail. Specifically we will look at when we see a decay series, when we do not see such a decay series, and the evidence we might find that such a decay is occurring or has previously occurred.

The most common is the uranium-238 decay series, which produces lead-206 in a series of 14 sequential alpha- and beta-decay reactions (Figure **14.6.2**). Although a radioactive decay series can be written for almost any isotope with $Z > 85$, only two others occur naturally: the decay of uranium-235 to lead-207 (in 11 steps) and thorium-232 to lead-208 (in 10 steps). A fourth series, the decay of neptunium-237 to bismuth-209 in 11 steps, is known to have occurred on the primitive Earth. With a half-life of “only” 2.14 million years, all the neptunium-237 present when Earth was formed decayed long ago, and today all the neptunium on Earth is synthetic.

Due to these radioactive decay series, small amounts of very unstable isotopes are found in ores that contain uranium or thorium. These rare, unstable isotopes should have decayed long ago to stable nuclei with a lower atomic number, and they would no longer be found on Earth. Because they are generated continuously by the decay of uranium or thorium, however, their amounts have reached a steady state, in which their rate of formation is equal to their rate of decay. In some cases, the abundance of the daughter isotopes can be used to date a material or identify its origin.

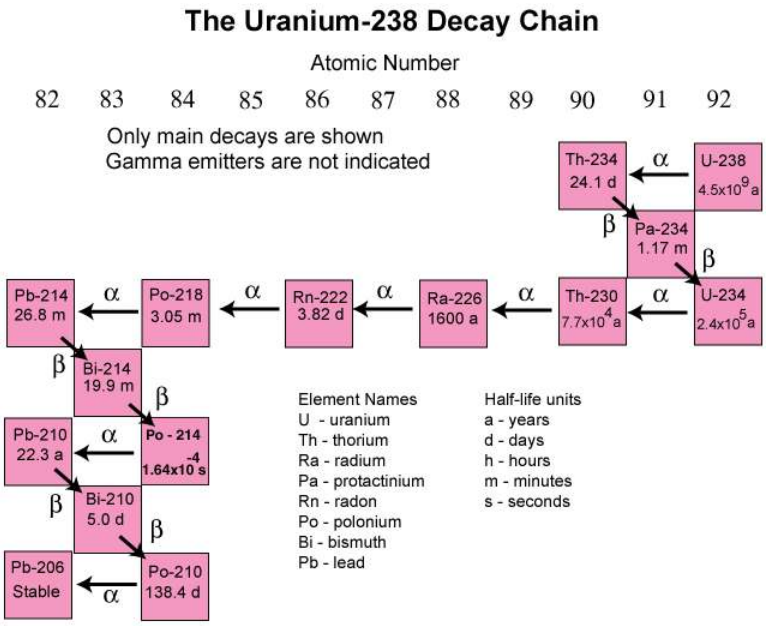


Figure **14.6.2**: Decay of uranium to stable end-product.

Section Summary

- Radioactive materials lose some of their activity each time a decay event occurs.
- Radioactive loss of activity can be estimated by determining the half-life of an isotope.
- Half-life is defined as the period of time needed for one half of a given quantity of a substance to undergo a change.
- Half-lives of different elements vary considerably.

This page titled 14.6: Half-Life is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **24.4: Half-Life** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **21.1: Radioactivity** is licensed CC BY-NC-SA 3.0.

14.7: What Holds the Nucleus Together

Learning Objectives

- Explain the nuclear forces.

A variety of experiments indicate that a nucleus behaves something like a tightly packed ball of nucleons, as illustrated in Figure **14.7.2**. These nucleons have large kinetic energies and, thus, move rapidly in very close contact. What forces hold a nucleus together? The nucleus is very small and its protons, being positive, exert tremendous repulsive forces on one another. (The Coulomb force increases as charges get closer, since it is proportional to $1/r^2$, even at the tiny distances found in nuclei.) The answer is that two previously unknown forces hold the nucleus together and make it into a tightly packed ball of nucleons.



Figure **14.7.2**: A model of the nucleus.

These forces are called the *weak and strong nuclear forces*. Nuclear forces are so short ranged that they fall to zero strength when nucleons are separated by only a few fm. However, like glue, they are strongly attracted when the nucleons get close to one another. The strong nuclear force is about 100 times more attractive than the repulsive EM force, easily holding the nucleons together. Nuclear forces become extremely repulsive if the nucleons get too close, making nucleons strongly resist being pushed inside one another, something like ball bearings.

The fact that nuclear forces are very strong is responsible for the very large energies emitted in nuclear decay. During decay, the forces do work, and since work is force times the distance, a large force can result in a large emitted energy. In fact, we know that there are *two* distinct nuclear forces because of the different types of nuclear decay—the strong nuclear force is responsible for α decay, while the weak nuclear force is responsible for β decay.

The many stable and unstable nuclei we have explored, and the hundreds we have not discussed, can be arranged in a table called the **chart of the nuclides**, a simplified version of which is shown in Figure **14.7.3**. Nuclides are located on a plot of N versus Z. Examination of a detailed chart of the nuclides reveals patterns in the characteristics of nuclei, such as stability, abundance, and types of decay, analogous to but more complex than the systematics in the periodic table of the elements.

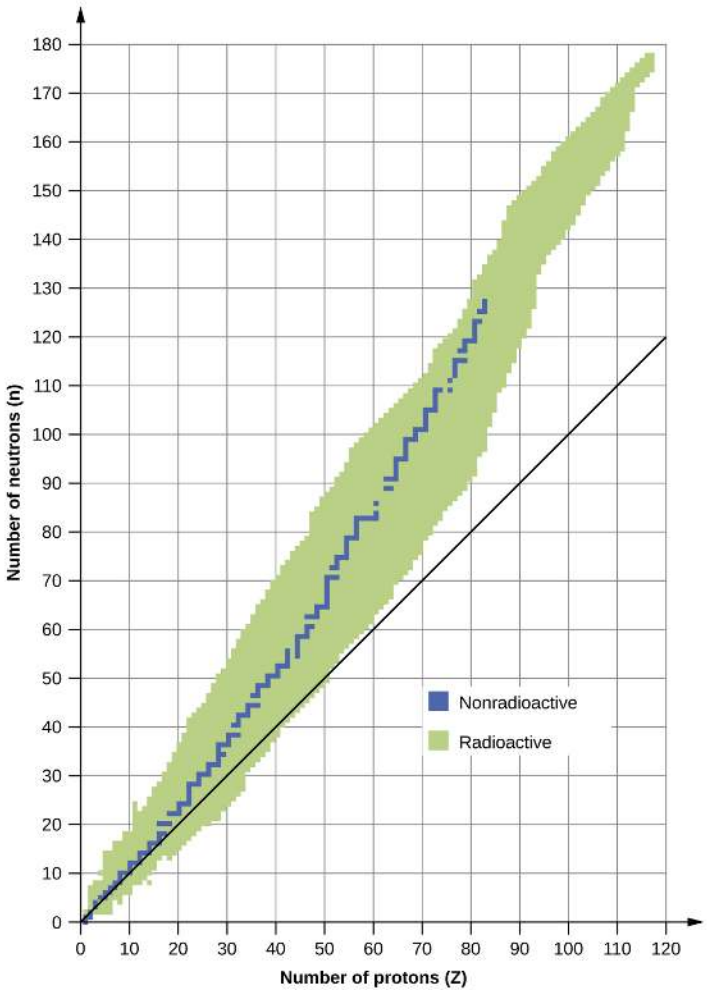


Figure 14.7.3: Simplified chart of the nuclides, a graph of N versus Z for known nuclides. The patterns of stable and unstable nuclides reveal characteristics of the nuclear forces. The dashed line is for $N = Z$. Numbers along diagonals are mass numbers A. (Source.)

In principle, a nucleus can have any combination of protons and neutrons, but Figure **14.7.3** shows a definite pattern for those that are stable. For low-mass nuclei, there is a strong tendency for N and Z to be nearly equal. This means that the nuclear force is more attractive when $N = Z$. More detailed examination reveals greater stability when N and Z are even numbers—nuclear forces are more attractive when neutrons and protons are in pairs. For increasingly higher masses, there are progressively more neutrons than protons in stable nuclei. This is due to the ever-growing repulsion between protons. Since nuclear forces are short ranged, and the Coulomb force is long ranged, an excess of neutrons keeps the protons a little farther apart, reducing Coulomb repulsion. Decay modes of nuclides out of the region of stability consistently produce nuclides closer to the region of stability. There are more stable nuclei having certain numbers of protons and neutrons, called **magic numbers**. Magic numbers indicate a shell structure for the nucleus in which closed shells are more stable. Nuclear shell theory has been very successful in explaining nuclear energy levels, nuclear decay, and the greater stability of nuclei with closed shells. We have been producing ever-heavier transuranic elements since the early 1940s, and we have now produced the element with $Z = 118$. There are theoretical predictions of an island of relative stability for nuclei with such high Z s.



Figure **14.7.4**: The German-born American physicist Maria Goeppert Mayer (1906–1972) shared the 1963 Nobel Prize in physics with J. Jensen for the creation of the nuclear shell model. This successful nuclear model has nucleons filling shells analogous to electron shells in atoms. It was inspired by patterns observed in nuclear properties. (credit: Nobel Foundation via Wikimedia Commons)

Section Summary

- There are two nuclear forces, the weak and the strong.
- Systematics in nuclear stability seen on the chart of the nuclides indicate that there are shell closures in nuclei for values of Z and N equal to the magic numbers, which correspond to highly stable nuclei.

Glossary

chart of the nuclides
a table comprising stable and unstable nuclei

magic numbers
a number that indicates a shell structure for the nucleus in which closed shells are more stable

nucleons
the particles found inside nuclei

nucleus
a region consisting of protons and neutrons at the center of an atom

nuclide
a type of atom whose nucleus has specific numbers of protons and neutrons

protons
the positively charged nucleons found in a nucleus

This page titled 14.7: What Holds the Nucleus Together is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **14.4: Substructure of the Nucleus** by OpenStax is licensed CC BY 4.0.

14.8: The Energy Within the Nucleus and all Matter

Learning Objectives

Compute total energy of a relativistic object.

Describe rest energy, and explain how it can be converted to other forms.

Explain why massive particles cannot reach c.

A tokamak is a form of experimental fusion reactor, which can change mass to energy. Accomplishing this requires an understanding of relativistic energy. Nuclear reactors are proof of the conservation of relativistic energy.

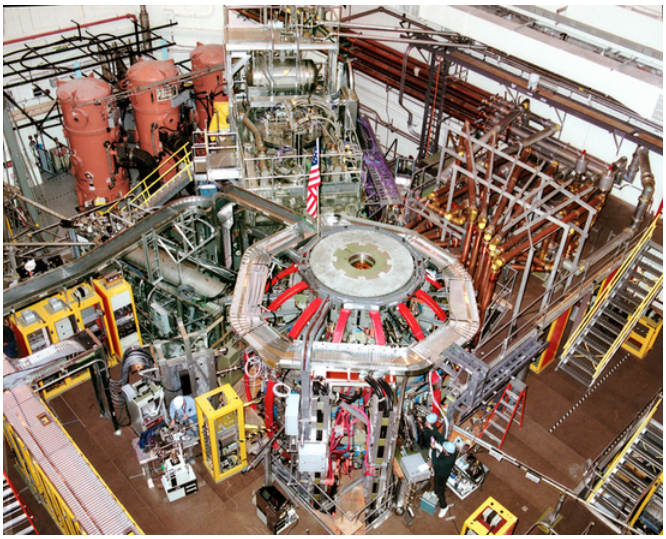


Figure **14.8.1**: The National Spherical Torus Experiment (NSTX) has a fusion reactor in which hydrogen isotopes undergo fusion to produce helium. In this process, a relatively small mass of fuel is converted into a large amount of energy. (credit: Princeton Plasma Physics Laboratory)

Conservation of energy is one of the most important laws in physics. Not only does energy have many important forms, but each form can be converted to any other. We know that classically the total amount of energy in a system remains constant. Relativistically, energy is still conserved, provided its definition is altered to include the possibility of mass changing to energy, as in the reactions that occur within a nuclear reactor. Relativistic energy is intentionally defined so that it will be conserved in all inertial frames, just as is the case for relativistic momentum. As a consequence, we learn that several fundamental quantities are related in ways not known in classical physics. All of these relationships are verified by experiment and have fundamental consequences. The altered definition of energy contains some of the most fundamental and spectacular new insights into nature found in recent history.

Total Energy and Rest Energy

The first postulate of relativity states that the laws of physics are the same in all inertial frames. Einstein showed that the law of conservation of energy is valid relativistically, if we define energy to include a relativistic factor.

Definition: Total Energy

Total energy E is defined to be

$$E = \gamma mc^2, \tag{14.8.1}$$

where m is mass, c is the speed of light, $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$, and v is the velocity of the mass relative to an observer.

There are many aspects of the total energy E that we will discuss—among them are how kinetic and potential energies are included in E, and how E is related to relativistic momentum. But first, note that at rest, total energy is not zero. Rather, when $v = 0$, we have $\gamma = 1$, and an object has rest energy.

Definition: Rest Energy

Rest energy is

$$E_0 = mc^2. \tag{14.8.2}$$

This is the correct form of Einstein's most famous equation, which for the first time showed that energy is related to the mass of an object at rest. For example, if energy is stored in the object, its rest mass increases. This also implies that mass can be destroyed to release energy. The implications of these first two equations regarding relativistic energy are so broad that they were not completely recognized for some years after Einstein published them in 1907, nor was the experimental proof that they are correct widely recognized at first. Einstein, it should be noted, did understand and describe the meanings and implications of his theory.

Example 14.8.1: Calculating Rest Energy: Rest Energy is Very Large

Calculate the rest energy of a 1.00-g mass.

Strategy

One gram is a small mass—less than half the mass of a penny. We can multiply this mass, in SI units, by the speed of light squared to find the equivalent rest energy.

Solution

1. Identify the knowns: $m = 1.00 \times 10^{-3}\ kg$; $c = 3.00 \times 10^{8}\ m/s$
2. Identify the unknown: E_0
3. Choose the appropriate equation: $E_0 = mc^2$
4. Plug the knowns into the equation:

$$\begin{aligned} E_0 &= mc^2 \\ &= (1.00 \times 10^{-3}\ kg)(3.00 \times 10^{8}\ m/s)^2 \\ &= 9.00 \times 10^{13}\ kg \cdot m^2/s^2 \end{aligned}$$

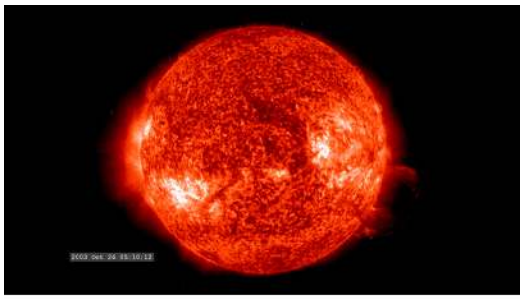
5. Convert units, we see the rest mass energy is:

$$E_0 = 9.00 \times 10^{13}\ J. \tag{14.8.3}$$

Discussion

This is an enormous amount of energy for a 1.00-g mass. We do not notice this energy, because it is generally not available. Rest energy is large because the speed of light c^2 is a very large number, so that mc^2 is huge for any macroscopic mass. The $9.00 \times 10^{13}\ J$ rest mass energy for 1.00 g is about twice the energy released by the Hiroshima atomic bomb and about 10,000 times the kinetic energy of a large aircraft carrier. If a way can be found to convert rest mass energy into some other form (and all forms of energy can be converted into one another), then huge amounts of energy can be obtained from the destruction of mass.

Today, the practical applications of *the conversion of mass into another form of energy*, such as in nuclear weapons and nuclear power plants, are well known. But examples also existed when Einstein first proposed the correct form of relativistic energy, and he did describe some of them. Nuclear radiation had been discovered in the previous decade, and it had been a mystery as to where its energy originated. The explanation was that, in certain nuclear processes, a small amount of mass is destroyed and energy is released and carried by nuclear radiation. But the amount of mass destroyed is so small that it is difficult to detect that any is missing. Although Einstein proposed this as the source of energy in the radioactive salts then being studied, it was many years before there was broad recognition that mass could be and, in fact, commonly is converted to energy (Figure 14.8.1).



(a)



(b)

Figure 14.8.2: The Sun (a) and the Susquehanna Steam Electric Station (b) both convert mass into energy—the Sun via nuclear fusion, the electric station via nuclear fission. (credits: (a) NASA/Goddard Space Flight Center, Scientific Visualization Studio; (b) U.S. government)

Because of the relationship of rest energy to mass, we now consider mass to be a form of energy rather than something separate. There had not even been a hint of this prior to Einstein's work. Such conversion is now known to be the source of the Sun's energy, the energy of nuclear decay, and even the source of energy keeping Earth's interior hot.

Stored Energy and Potential Energy

What happens to energy stored in an object at rest, such as the energy put into a battery by charging it, or the energy stored in a toy gun's compressed spring? The energy input becomes part of the total energy of the object and, thus, increases its rest mass. All stored and potential energy becomes mass in a system. Why is it we don't ordinarily notice this? In fact, conservation of mass (meaning total mass is constant) was one of the great laws verified by 19th-century science. Why was it not noticed to be incorrect? The following example helps answer these questions.

Example 14.8.2: Calculating Rest Mass: A Small Mass Increase due to Energy Input

A car battery is rated to be able to move 600 ampere-hours ($\cdot h$) of charge at 12.0 V.

a. Calculate the increase in rest mass of such a battery when it is taken from being fully depleted to being fully charged.
b. What percent increase is this, given the battery's mass is 20.0 kg?

Strategy

In part (a), we first must find the energy stored in the battery, which equals what the battery can supply in the form of electrical potential energy. Since $PE_{elec} = qV$, we have to calculate the charge q in $600\ A \cdot h$, which is the product of the current I and the time t. We then multiply the result by 12.0 V. We can then calculate the battery's increase in mass using $\Delta E = PE_{elec} = (\Delta m)c^2$.

Part (b) is a simple ratio converted to a percentage.

Solution for (a)

1. Identify the knowns: $I \cdot t = 600\ A \cdot h$; $V = 12.0\ V$; $c = 3.00 \times 10^8\ m/s$

2. Identify the unknown: δm
3. Choose the appropriate equation: $PE_{elec} = (\Delta m)c^2$
4. Rearrange the equation to solve for the unknown: $\Delta m = \frac{PE_{elec}}{c^2}$
5. Plug the knowns into the equation:

$$\Delta m = \frac{PE_{elec}}{c^2} = \frac{qV}{c^2} = \frac{(It)V}{c^2} = \frac{(600\,A \cdot h)(12.0\,V)}{(3.00 \times 10^8)^2}. \tag{14.8.4}$$

Write amperes A as coulombs per second (C/s), and convert hours to seconds.

$$\Delta m = \frac{(600\,C/s \cdot h(\frac{3600\,s}{1\,h})(12.0\,J/C)}{3.00 \times 10^8\,m/s)^2} \tag{14.8.5}$$

$$= \frac{(2.16 \times 10^6\,C)(12.0\,J/C)}{(3.00 \times 10^8\,m/s)^2} \tag{14.8.6}$$

Using the conversion $1\,kg \cdot m^2/c^2 = 1\,J$, we can write the mass as $\delta m = 2.88 \times 10^{-10}\,kg$.

Solution for (b)

1. Identify the knowns: $\Delta m = 2.88 \times 10^{-10}\,kg$; $m = 20.0\,kg$
2. Identify the unknown: % change
3. Choose the appropriate equation: $\%\,increase = \frac{\Delta m}{m} \times 100\%$
4. Plug the knowns into the equation:

$$\%\,increase = \frac{\delta m}{m} \times 100\% = \frac{2.88 \times 10^{-10}\,kg}{20.0\,kg} \times 100\% = 1.44 \times 10^{-9}\% \tag{14.8.7}$$

Discussion

Both the actual increase in mass and the percent increase are very small, since energy is divided by c^2, a very large number. We would have to be able to measure the mass of the battery to a precision of a billionth of a percent, or 1 part in 10^{11}, to notice this increase. It is no wonder that the mass variation is not readily observed. In fact, this change in mass is so small that we may question how you could verify it is real. The answer is found in nuclear processes in which the percentage of mass destroyed is large enough to be measured. The mass of the fuel of a nuclear reactor, for example, is measurably smaller when its energy has been used. In that case, stored energy has been released (converted mostly to heat and electricity) and the rest mass has decreased. This is also the case when you use the energy stored in a battery, except that the stored energy is much greater in nuclear processes, making the change in mass measurable in practice as well as in theory.

Kinetic Energy and the Ultimate Speed Limit

Kinetic energy is energy of motion. Classically, kinetic energy has the familiar expression $\frac{1}{2}mv^2$. The relativistic expression for kinetic energy is obtained from the work-energy theorem. This theorem states that the net work on a system goes into kinetic energy. If our system starts from rest, then the work-energy theorem is

$$W_{net} = KE. \tag{14.8.8}$$

Relativistically, at rest we have rest energy $E_0 = mc^2$. The work increases this to the total energy $E = \gamma mc^2$. Thus,

$$W_{net} = E - E_0 = \gamma mc^2 - mc^2 = (\gamma - 1)mc^2. \tag{14.8.9}$$

Relativistically, we have $W_{net} = KE_{rel}$.

Definition: Relativistic Kinetic Energy

Relativistic kinetic energy is

$$KE_{rel} = (\gamma - 1)mc^2. \tag{14.8.10}$$

When motionless, we have $v = 0$ and

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = 1, \tag{14.8.11}$$

so that $KE_{rel} = 0$ at rest, as expected. But the expression for relativistic kinetic energy (such as total energy and rest energy) does not look much like the classical $\frac{1}{2}mv^2$ To show that the classical expression for kinetic energy is obtained at low velocities, we note that the binomial expansion for γ at low velocities gives

$$\gamma = 1 + \frac{1}{2}\frac{v^2}{c^2}. \tag{14.8.12}$$

Entering this into the expression for relativistic kinetic energy gives

$$KE_{rel} = \left[\frac{1}{2}\frac{v^2}{c^2}\right] mc^2 = \frac{1}{2}mv^2 = KE_{class}. \tag{14.8.13}$$

So, in fact, relativistic kinetic energy does become the same as classical kinetic energy when $v << c$.

It is even more interesting to investigate what happens to kinetic energy when the velocity of an object approaches the speed of light. We know that γ becomes infinite as v approaches c, so that KE_{rel} also becomes infinite as the velocity approaches the speed of light (Figure 14.8.1). An infinite amount of work (and, hence, an infinite amount of energy input) is required to accelerate a mass to the speed of light.

Definition: Speed of Light

No object with mass can attain the speed of light.

So the speed of light is the ultimate speed limit for any particle having mass. All of this is consistent with the fact that velocities less than c always add to less than c. Both the relativistic form for kinetic energy and the ultimate speed limit being c have been confirmed in detail in numerous experiments. No matter how much energy is put into accelerating a mass, its velocity can only approach—not reach—the speed of light.

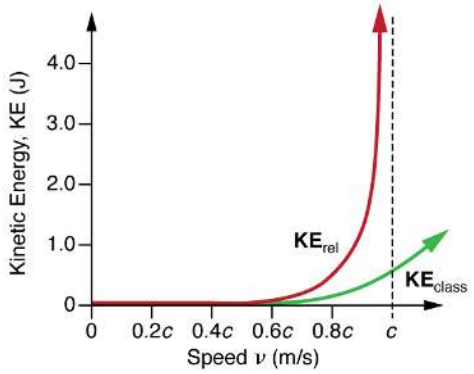


Figure 14.8.3: This graph of KE_{rel} versus velocity shows how kinetic energy approaches infinity as velocity approaches the speed of light. It is thus not possible for an object having mass to reach the speed of light. Also shown is KE_{class}, the classical kinetic energy, which is similar to relativistic kinetic energy at low velocities. Note that much more energy is required to reach high velocities than predicted classically.

Example 14.8.3: Comparing Kinetic Energy: Relativistic Energy Versus Classical Kinetic Energy

An electron has a velocity $v = 0.990c$.

a. Calculate the kinetic energy in MeV of the electron.
b. Compare this with the classical value for kinetic energy at this velocity. (The mass of an electron is $9.11 \times 10^{-31}\ kg$.)

Strategy

The expression for relativistic kinetic energy is always correct, but for (a) it must be used since the velocity is highly relativistic (close to c). First, we will calculate the relativistic factor γ, and then use it to determine the relativistic kinetic energy. For (b), we will calculate the classical kinetic energy (which would be close to the relativistic value if v were less than a few percent of c) and see that it is not the same.

Solution for (a)

1. Identify the knowns: $v = 0.990c$; $m = 9.11 \times 10^{-31}\ kg$
2. Identify the unknown: KE_{rel}
3. Choose the appropriate equation $KE_{rel} = (\gamma - 1)mc^2$
4. Plug the knowns into the equation:

First calculate γ. We will carry extra digits because this is an intermediate calculation.

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - \frac{(0.990c)^2}{c^2}}} = \frac{1}{\sqrt{1 - (0.990)^2}} = 7.0888 \tag{14.8.14}$$

Next, we use this value to calculate the kinetic energy.

$$\begin{aligned} KE_{rel} &= (\gamma - 1)mc^2 \\ &= (7.0888 - 1)(9.11 \times 10^{-31}\ kg)(3.00 \times 10^8\ m/s)^2 = 4.99 \times 10^{-13}\ J \end{aligned} \tag{14.8.15}$$

5. Convert units:

$$KE_{rel} = (4.99 \times 10^{-13}\ J)\left(\frac{1\ MeV}{1.60 \times 10^{-13}\ J}\right) = 3.12\ MeV \tag{14.8.16}$$

Solution for (b)

1. List the knowns: $v = 0.990c$; $m = 9.11 \times 10^{-31}\ kg$
2. List the unknown: KE_{class}
3. Choose the appropriate equation: $KE_{class} = \frac{1}{2}mv^2$
4. Plug the knowns into the equation:

$$KE_{class} = \frac{1}{2}mv^2 \tag{14.8.17}$$

$$= \frac{1}{2}(9.11 \times 10^{-31}\ kg)(0.990)^2(3.00 \times 10^8\ m/s)^2 \tag{14.8.18}$$

$$= 4.02 \times 10^{-14}\ J \tag{14.8.19}$$

5. Convert units:

$$KE_{class} = 4.02 \times 10^{-14}\left(\frac{1\ MeV}{1.60 \times 10^{-13}\ J}\right) = 0.251\ MeV \tag{14.8.20}$$

Discussion

As might be expected, since the velocity is 99.0% of the speed of light, the classical kinetic energy is significantly off from the correct relativistic value. Note also that the classical value is much smaller than the relativistic value. In fact, $KE_{rel}/KE_{class} = 12.4$ here. This is some indication of how difficult it is to get a mass moving close to the speed of light. Much more energy is required than predicted classically. Some people interpret this extra energy as going into increasing the mass of the system, but, as discussed in Relativistic Momentum, this cannot be verified unambiguously. What is certain is that ever-increasing amounts of energy are needed to get the velocity of a mass a little closer to that of light. An energy of 3 MeV is a very small amount for an electron, and it can be achieved with present-day particle accelerators. SLAC, for example, can accelerate electrons to over $50 \times 10^9\ eV = 50,000 MeV$.

Is there any point in getting v a little closer to c than 99.0% or 99.9%? The answer is yes. We learn a great deal by doing this. The energy that goes into a high-velocity mass can be converted to any other form, including into entirely new masses. (See Figure.) Most of what we know about the substructure of matter and the collection of exotic short-lived particles in nature has been learned this way. Particles are accelerated to extremely relativistic energies and made to collide with other particles, producing totally new species of particles. Patterns in the characteristics of these previously unknown particles hint at a basic substructure for all matter.



Figure 14.8.4: The Fermi National Accelerator Laboratory, near Batavia, Illinois, was a subatomic particle collider that accelerated protons and antiprotons to attain energies up to 1 Tev (a trillion electronvolts). The circular ponds near the rings were built to dissipate waste heat. This accelerator was shut down in September 2011. (credit: Fermilab, Reidar Hahn)

Section Summary

- Relativistic energy is conserved as long as we define it to include the possibility of mass changing to energy.
- Total Energy is defined as: $E = \gamma mc^2$, where $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$.
- Rest energy is $E_0 = mc^2$, meaning that mass is a form of energy. If energy is stored in an object, its mass increases. Mass can be destroyed to release energy.
- We do not ordinarily notice the increase or decrease in mass of an object because the change in mass is so small for a large increase in energy.
- The relativistic work-energy theorem is $W_{net} = E - E_0 = \gamma mc^2 = (\gamma - 1)mc^2$.
- Relativistically, $W_{net} = KE_{rel}$, where KE_{rel} is the relativistic kinetic energy.
- Relativistic kinetic energy is $KE_{rel} = (\gamma - 1)mc^2$, where $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$. At low velocities, relativistic kinetic energy reduces to classical kinetic energy.
- **No object with mass can attain the speed of light** because an infinite amount of work and an infinite amount of energy input is required to accelerate a mass to the speed of light.

Glossary

total energy

defined as $E = \gamma mc^2$, where $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$

rest energy

the energy stored in an object at rest: $E_0 = mc^2$

relativistic kinetic energy

the kinetic energy of an object moving at relativistic speeds: $KE_{rel} = (\gamma - 1)mc^2$, where $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$

This page titled 14.8: The Energy Within the Nucleus and all Matter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jamie MacArthur.

- **28.6: Relativistic Energy** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/college-physics.

14.9: Nuclear Binding Energy

Learning Objectives

Explain trends in the relative stability of nuclei.

As a simple example of the energy associated with the strong nuclear force, consider the helium atom composed of two protons, two neutrons, and two electrons. The total mass of these six subatomic particles may be calculated as:

$$\underbrace{(2\times1.0073\,\text{amu})}_{\text{protons}}+\underbrace{(2\times1.0087\,\text{amu})}_{\text{neutrons}}+\underbrace{(2\times0.00055\,\text{amu})}_{\text{electrons}}=4.0331\,\text{amu}$$

However, mass spectrometric measurements reveal that the mass of an atom is 4.0026 amu, less than the combined masses of its six constituent subatomic particles. This difference between the calculated and experimentally measured masses is known as the mass defect of the atom. In the case of helium, the **mass defect** indicates a "loss" in mass of 4.0331 amu – 4.0026 amu = 0.0305 amu. The loss in mass accompanying the formation of an atom from protons, neutrons, and electrons is due to the conversion of that mass into energy that is evolved as the atom forms. The **nuclear binding energy** is the energy produced when the atoms' nucleons are bound together; this is also the energy needed to break a nucleus into its constituent protons and neutrons. In comparison to chemical bond energies, nuclear binding energies are *vastly* greater, as we will learn in this section. Consequently, the energy changes associated with nuclear reactions are vastly greater than are those for chemical reactions.

The conversion between mass and energy is most identifiably represented by the **mass-energy equivalence equation** covered earlier, which results in massive amounts of energy for even tiny amounts of mass. Because the energy changes for breaking and forming bonds are so small compared to the energy changes for breaking or forming nuclei, the changes in mass during all ordinary chemical reactions are virtually undetectable. The most energetic chemical reactions exhibit enthalpies on the order of *thousands* of kJ/mol, which is equivalent to mass differences in the nanogram range (10^{-9} g). On the other hand, nuclear binding energies are typically on the order of *billions* of kJ/mol, corresponding to mass differences in the milligram range (10^{-3} g).

The relative stability of a nucleus is correlated with its **binding energy per nucleon**, the total binding energy for the nucleus divided by the number or nucleons in the nucleus. As we shall see, nuclear reactions will occur that lead towards maximizing the nuclear binding energy per nucleon. Whether this results in increasing the number of nucleons or decreasing the number of nucleons depends upon where the nucelus is on the curve shown in Figure 14.9.1. In the subsections which follow, we will explore each of these possibilities in more detail.

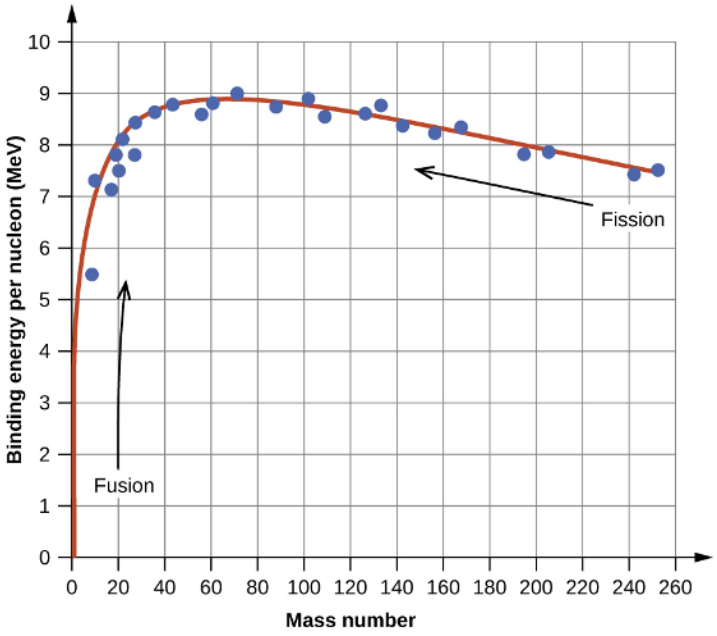


Figure 14.9.2: The binding energy per nucleon is largest for nuclides with mass number of approximately 56.

Glossary

mass defect

The difference between the calculated and experimentally measured masses.

nuclear binding energy

The energy needed to break a nucleus into its constituent protons and neutrons.

binding energy per nucleon

The total binding energy for the nucleus divided by the number or nucleons in the nucleus.

This page titled 14.9: Nuclear Binding Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **21.1: Nuclear Structure and Stability** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.
- **1.5: Acceleration** by OpenStax is licensed CC BY 4.0.

14.9.1: Nuclear Fission

Learning Objectives

- Explain the process of nuclear fission.

Nuclear fission was first discovered by Lise Meitner and Otto Hahn, in the 1930s. They began their work by bombarding uranium with neutrons, hoping to create larger elements. Instead, they were very surprised to find **Ba**-141, a much smaller element. It was many years before Lise's contribution to this discovery was fully understood, in part because she had to complete much of her work as a refugee who had to flee Germany during World War II. Hahn was later a sole recipient of a Nobel Prize for this work. Meitner was critical of Hahn for not giving her credit for her work, not speaking out against the Nazi regime, and critical of her colleagues in nuclear physics for the use of this research to create the atomic bomb.

Nuclear Fission

Radioactive decay by the release of alpha or beta particles is not the only way new isotopes are formed. When a neutron collides with a nucleus, the nucleus splits into two isotopes, each of which is roughly half the mass of the original atom. A small amount of mass is "left over" and released as energy, as predicted by Einstein's famous equation $E = mc^2$, that relates mass and energy. This process is known as **nuclear fission**. The neutron must be a "slow" neutron, traveling at a speed that is approximately that of the molecules of a gas at the same temperature in the system producing the neutrons. High-speed ("fast") neutrons will not result in nuclear fission.

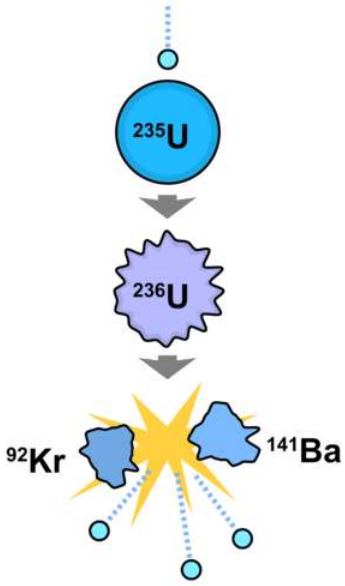


Figure **14.9.1.1**: Fission of a uranium nucleus produced by collision with a neutron.

The example above illustrates the basic nuclear fission process. A neutron (generally produced by some controlled process, not usually a natural event) collides with an atom of **U**-235. Momentarily, a **U**-236 atom forms, which then splits into two smaller atoms (**Kr**-93 and **Ba**-141) in the diagram. This process results in the release of three new neutrons, which can then initiate fission reactions with more atoms. We will see later how this propagation of neutrons can be employed in a reactor for the generation of electricity.

An extended version of this process can be seen in the figure below. Not every collision of a neutron with **U**-235 results in a fission reaction. A neutron from the initial fission process may strike an atom of **U**-238, which does not continue the process. Another neutron may not collide with a nucleus, and be lost in the environment. However, a third neutron produced from the initial collision can collide with more **U**-235, and continue the chain reaction to produce more neutrons.

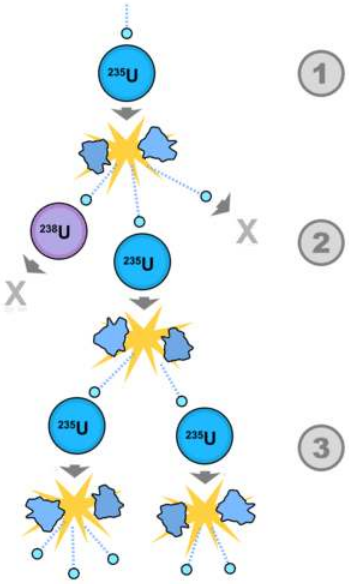


Figure 14.9.1.2: Fission reaction with U-235.

Typical nuclear fission reactions balance in terms of mass. The total mass of the reactants is equal to the total mass of the products:

$$^{235}_{92}\text{U} + ^{1}_{0}\text{n} \rightarrow ^{92}_{36}\text{Kr} + ^{142}_{56}\text{Ba} + 2^{1}_{0}\text{n} + \text{energy}$$

There are a total of 236 mass units on the left of the equation and 236 mass units on the right. In the same manner, we see 92 protons on the left and 92 on the right. The energy that is released is the binding energy that holds the nucleus together.

Another set of fission products from U-235 can be seen in the following reaction:

$$^{235}_{92}\text{U} + ^{1}_{0}\text{n} \rightarrow ^{95}_{42}\text{Mo} + ^{139}_{57}\text{La} + 2^{1}_{0}\text{n} + \text{energy}$$

Again, we see that the total number of mass units and of protons is equal on both sides of the equation.

Nuclear Power Generation

The generation of electricity is critical for operation of businesses, health care delivery, schools, homes, and other areas requiring the use of electrical power. According to 2011 statistics, coal is used for 42% of the total power generated, with natural gas being employed for another 25%. Nuclear power plants are employed in about 19% of the cases, with renewable energy sources supplying the last 13%. All of these fuels are used to heat water to generate steam. The steam then turns a turbine to generate electricity.

The diagram below shows the layout of a typical nuclear power plant. The radioactive rods are in the red container along with water, which is heated to steam. The energy for this heat comes from fission reactions of uranium. The steam passes through the turbine and causes the turbine to spin, generating electricity. As the steam condenses, it is run through a cooling tower to lower its temperature. The water then recirculates through the reactor core to be used again.

The control rods play an important role in the modulation of the nuclear chain reaction (usually a collision of a neutron with uranium). Each collision produces more neutrons than were present initially. If left unsupervised, the reaction would soon get out of control. Rods are commonly made of boron or a number of metals and metal alloys. The purpose of the control rods is to absorb neutrons to regulate the rate of the chain reaction, so that the water does not overheat and destroy the reactor.

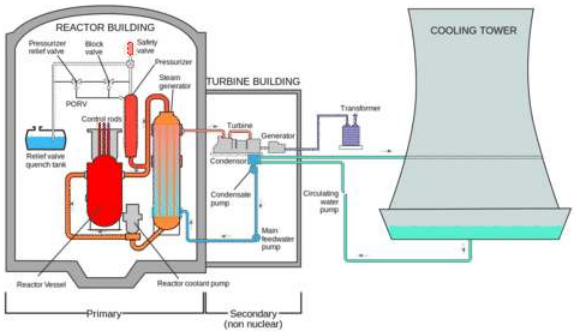


Figure 14.9.1.1: Schematic for a nuclear power plant.

Nuclear power is also used to propel ships. The turbine can be connected to a propeller system. The rotating turbine shaft will turn the propeller to move the ship.



Figure 14.9.1.2: Nuclear submarine.

Section Summary

- Nuclear fission is the process of a neutron colliding with a nucleus. The nucleus splits into two isotopes, each of which is roughly half the mass of the original atom. A small amount of mass is "left over" and released as energy.
- Examples of nuclear fission processes are illustrated.
- The importance of nuclear power in generating electricity is described.
- The operation of a nuclear power plant is described.

This page titled 14.9.1: Nuclear Fission is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **24.6: Nuclear Fission Processes** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **24.7: Nuclear Power Generation** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

14.9.2: Nuclear Fusion

Learning Objectives

- Explain the process of nuclear fusion.

A number of reactions take place in the sun that cannot be duplicated on Earth. Some of these reactions involve the formation of large elements from smaller ones. So far, we have only been able to observe formation of very small elements here on Earth. The reaction sequence observed appears to be the following: hydrogen-1 atoms collide to form the larger hydrogen isotopes, hydrogen-2 (deuterium) and hydrogen-3 (tritium). In the process, positrons and gamma rays are formed. The positrons will collide with any available electrons and annihilate, producing more gamma rays. In the process, tremendous amounts of energy are produced to keep us warm and continue supplying reactions.

In contrast to nuclear fission, which results in smaller isotopes being formed from larger ones, the goal of **nuclear fusion** is to produce larger materials from the collision of smaller atoms. The forcing of the smaller atoms together results in tighter packing, and the release of energy. As seen in the figure below, energy is released in the formation of the larger atom, helium (**He**) from the fusion of hydrogen-2 and hydrogen-3, as well as from the expulsion of a neutron.

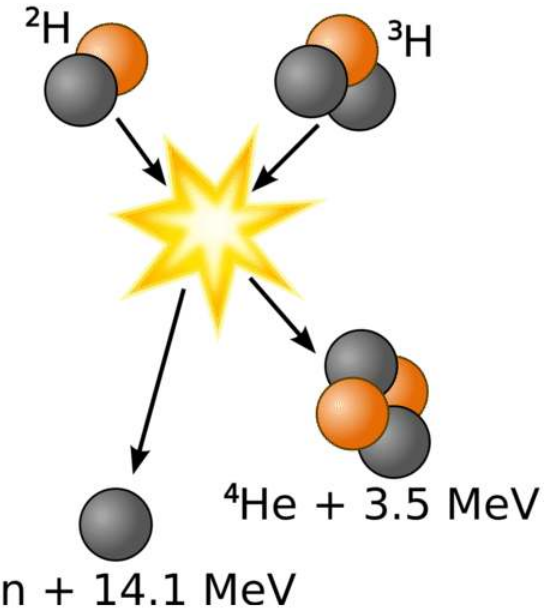


Figure **14.9.2.1**: Nuclear fusion reaction between deuterium and tritium.

This release of energy is what drives research on fusion reactors today. If such a reaction could be accomplished efficiently on Earth, it could provide a clean source of nuclear energy. Unlike fission reactions, nuclear fusion does not produce radioactive products that represent hazards to living systems.

Nuclear fusion reactions in the laboratory have been extraordinarily difficult to achieve. Extremely high temperatures (i.e., millions of degrees) are required. Methods must be developed to force the atoms together and hold them together long enough to react. The neutrons released during the fusion reactions can interact with atoms in the reactor, and convert them to radioactive materials. There has been some success in the field of nuclear fusion reactions, but the journey to feasible fusion power is still a long and uncertain one.

Section Summary

- Nuclear fusion produces larger materials from the collision of smaller atoms. The forcing of the smaller atoms together results in tighter packing, and the release of energy.
- Examples of nuclear fusion reactions are given.

This page titled 14.9.2: Nuclear Fusion is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **24.8: Nuclear Fusion** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

14.10: Nuclear Radiation in Our Lives

Learning Objectives

- Explain some of the effects radiation has on our health.

Nuclear radiation is ever present in our lives. In this section we will look at a few areas related specifically to health effects of radiation. Radiation is largely dangerous, but there are also places in medical treatment where it can be useful.

Background Radiation

We are all exposed to a small amount of radiation in our daily lives. This **background radiation** comes from naturally occurring sources and from human-produced radiation. Exposure to x-rays and nuclear medicine isotopes, ground sources, and cosmic radiation account for almost half of the background exposure of the average American. Radon gas, formed from the decay of uranium and thorium isotopes, is responsible for a little over half the total amount of background radiation. See the table below for background sources.

Sources of Background Radiation

Table 14.10.1: Sources of Background Radiation

radon	54%
consumer products	3%
nuclear medicine	4%
cosmic radiation	8%
terrestrial	8%
internal	11%
x-rays	11%
other	1%

The Problem of Radon

Small amounts of uranium and thorium are found in the soil of a large number of areas in the U.S. When radioactive isotopes of these elements decay, radon is one of the products formed. Radon is a colorless, odorless gas and is chemically inert, as it is one of the noble gases.

Radon is also radioactive and can easily be inhaled into the lungs. Over time, this internal radon exposure can lead to the development of lung cancer. The incidence of lung cancer in smokers exposed to radon is much higher than in non-smokers exposed to radon, since smoking has already produced some lung damage, and the radon simply makes the damage worse. Radon exposure is highest in homes lacking good air circulation to move the gas out of the residence. There are a number of inexpensive approaches to decreasing your exposure to radon. A good start is to test your living area for radon with a radon test kit.



Figure 14.10.1: Radon test kit.

Effects of Radiation

In order to better understand how cellular radiation damage occurs, we need to take a quick review of how the cell functions. **DNA** in the nucleus is responsible for protein synthesis and for regulation of many cellular functions. In the process of protein synthesis, DNA partially unfolds to produce **messenger RNA** (mRNA). The mRNA leaves the nucleus and interacts with ribosomes, transfer RNA, amino acids, and other cellular constituents in the cytoplasm. Through a complex series of reactions, proteins are produced to carry out a number of specialized processes within the organism. Anything that disturbs this flow of reactions can produce damage to the cell.

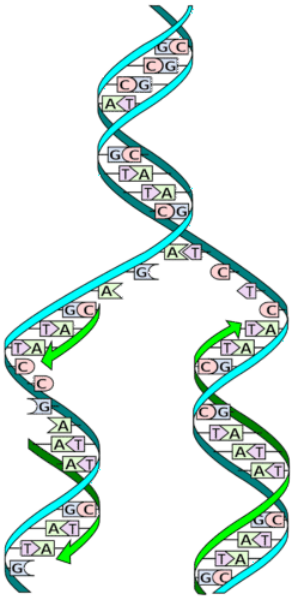


Figure 14.10.1: DNA replication.

The major effect of ionizing radiation on the cell is the disruption of the DNA strand. With the DNA structure damaged, the cell cannot reproduce in its normal fashion. Protein synthesis is affected, as are a number of processes necessary for proper cell function. One common effect is the generation of cancer cells. These cells have an abnormal structure due to the damaged DNA. In addition, they usually grow rapidly since the normal control processes regulating cell growth have been changed by the altered composition of the DNA. Tissue damage is also common in people with severe exposure to radiation.

Effects of Radiation on Humans

We can see two general types of effects when humans are exposed to radiation. Low-level exposure can lead to development of cancer. The regulatory processes regulating cell growth are disrupted, leading to uncontrolled growth of abnormal cells. Acute exposure can produce nausea, weakness, skin burns, and internal tissue damage. Cancer patients receiving radiation therapy

experience these symptoms; the radiation is targeted to a specific site in the body so that the damage is primarily to the cancer cells, and the patient is able to recover from the exposure.

Radioisotopes in Medical Diagnosis and Treatment

Radioisotopes are widely used to diagnose disease and as effective treatment tools. For diagnosis, the isotope is administered, and then located in the body using a scanner of some sort. The decay product (often gamma emission) can be located, and the intensity measured. The amount of isotope taken up by the body can then give information as to the extent of the medical problem.

An isotope of iodine (I-131) is used in both the diagnosis and treatment of thyroid cancer. The thyroid will normally absorb iodine to produce the iodine-containing thyroid hormones. An overactive thyroid gland will absorb the radioactive material, which can then destroy excess thyroid tissue or any cancer of the thyroid. The material is sometimes used to image cancers in other parts of the body.

Technetium-99m is perhaps the most widely used radioisotope in diagnosis and treatment (the "m" stands for *metastable*). This isotope decays to Tc-99 and a gamma emission of low intensity, making the radiation damage fairly negligible. The half-life is about six hours, so it will remain in the body for some time. Tc-99m can be used to look at cardiac damage. The isotope flows in the bloodstream; if there is less blood flow in the heart, there will be less isotope concentrated in the heart muscle. Similar information can be obtained for blood flow in the brain.

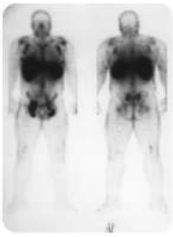


Figure **14.10.2**: Radioisotope scan to detect tumors.

Isotopes can be very useful in scans to locate cancer cells. The patient in the above image has multiple tumors that have spread (metastasized) from the main tumor. A radioisotope has been attached to **antibodies** that bind to specific cancer cells. The very dark spots in the armpits, neck, and groin represent areas where tumor cells exist.

Many other examples could be presented. There are presently over 25 different isotopes in use for diagnosis and treatment. A very partial list can be seen in the table below.

Radioisotopes Employed in Diagnosis and/or Treatment

Table 14.10.1: Radioisotopes Employed in Diagnosis and/or Treatment

Isotope	Half-Life	Application
Cr-51	28 days	Label red blood cells.
Fe-59	446 days	Study iron metabolism in spleen.
Xe-133	5 days	Study lung function.
Ho-166	26 hours	Cancer treatment.

Section Summary

- Background radiation is defined.
- Sources of background radiation are listed.
- Anything that affects DNA replication and protein synthesis can damage a cell.
- Effects of ionizing radiation on protein synthesis are listed.
- The impact of ionizing radiation on human health is discussed.

This page titled 14.10: Nuclear Radiation in Our Lives is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Jamie MacArthur via source content that was edited to the style and standards of the LibreTexts platform.

- **24.5: Background Radiation** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **24.10: Effects of Radiation** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.
- **24.11: Radioisotopes in Medical Diagnosis and Treatment** by CK-12 Foundation is licensed CK-12. Original source: https://flexbooks.ck12.org/cbook/ck-12-chemistry-flexbook-2.0/.

14.11: Process of Science- Lise Meitner and Otto Hahn

There are a few rightfully famous scientists from this chapter who could be highlighted: Albert Einstein and Marie Curie. These two are usually listed at the very top of any list of scientists of the 20th century, and among the top five to ten of any list of scientists of all time. Because of their relative recency, they are more understandable to us than someone like Galileo or Newton who lived in times much different than our own. Not only were their scientific accomplishments extremely significant, but their personal stories were quite compelling. Marie Curie was the first women to win a Nobel Prize, and one of only a few scientists to win more than one. Einstein was a Jewish refugee who won a Nobel Prize (but not for his most famous and groundbreaking work), and was also a scientific celebrity in many regards. Both of them have been fixtures of popular culture for decades, inspiring many books and even movies.

Instead of highlighting these two, we will instead look at the work of Lise Meitner and Otto Hahn. Both the work and life of Lise, in particular, were a fusion of those of Einstein and Curie. Not only was her work building on the work of both Einstein and Curie, but her life was an intersection of their identities. Though Einstein and Curie became famous in part for their identities, Meitner's existence was almost erased (both scientifically and literally) because of the intersection of those identities.

Meitner was the second woman to receive a doctorate in physics from the University of Vienna, and then tried to find ways of staying connected to scientific work at a high level. She was able to receive a position as an assistant to Otto Hahn at the Friedrich Wilhelm University. This position was unpaid for several years, and many at this University did not believe women should be there at all! In fact, Meitner was only allowed to enter into the lab space where she worked through an external door, and was not even provided a restroom to use while she was there. But she loved doing scientific research so much that she was willing to accept these conditions and was fortunate to have family who could support her financially in this endeavor.

Otto Hahn and Lise Meitner were about the same age, and she found him less intimidating than some of the other potential work colleagues at the University. Some have attributed Hahn's willingness to accept a woman co-worker as an equal to his time spent in Montreal where gender roles were not as rigid as they were in Germany at the time. Hahn was a chemist, and was attempting to do the sort of work that Marie Curie had done about a decade before: discover new element based on their radioactive properties. As a physicist, Meitner was more interested in the radiation itself than the new elements. As has happened many times throughout the history of science, this combination of diverse perspectives led to a great scientific breakthrough. Particularly the combination of Hahn's experimental work and Meitner's theoretical work. But Meitner would face additional personal challenges before this breakthrough occurred.

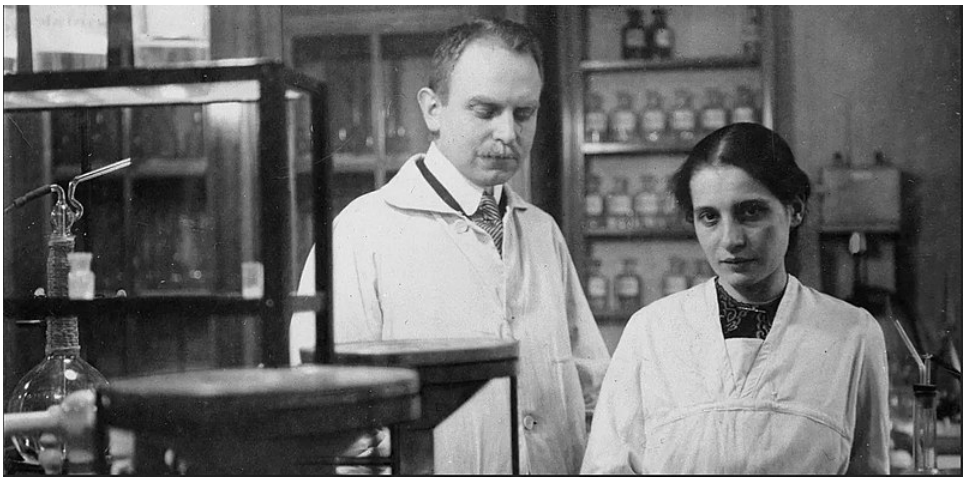


Figure **14.11.1**: A photograph of Otto and Lise from early in their career. The exact data and credit for the photographer are unknown. Public Domain. Wiki Commons.

Hahn and Meitner had a thirty year working relationship. They were considered for Nobel Prizes in both chemistry and physics for their work on radioactive elements at many points during this time. They had a break from this work during World War I, when Hahn was involved in making chemical weapons and Meitner was involved as an X-Ray nurse-technician for Austria. Meanwhile, Marie Curie was creating X-Ray equipment on the other side of the fox holes in France. Meitner returned to the lab before Hahn

did, and she followed Hahn's instruction via correspondence to continue the chemical analysis that had been underway. It was not the last war that would separate them, nor the last time they would need to work via correspondence.

Many prominent Jewish scientists, such as Albert Einstein, were able to flee Germany in the mid 1930s as the Nazis came to power. But as a woman of Jewish ancestry, Lise Meitner was in a much more precarious position. She had enjoyed working with Hahn for all these years and thought that a life without being able to do science was not worth living. She did not know where else she could go to do the sort of fulfilling work that she had partnered with Otto Hahn for the past three decades. When her home country of Austria was invaded in 1938, she realized that she had to leave, but it was much more dangerous now than it had been earlier. Neils Bohr and other Danish physicists attempted to find a position for her somewhere in Northern Europe, but the political situations were changing on almost a monthly basis and she didn't have the right papers now that Austria had been taken over by the Nazis. Eventually they came up with a plan to smuggle her out, but kept it secret from all of her colleagues, including Hahn, until the day before she was to leave. When Hahn found out, he gave her a diamond ring that had been his mothers "in case she needed it." She kept the ring until the day she died, despite what would happen later. She rode the train from Germany to the Netherlands with Danish physicist Dirk Coster. When the German border guards boarded the train and asked for papers, Coster presented his papers and Lise sat quietly beside him until they left. Fortunately, she was not asked. It was perhaps the only time in her life that being overlooked was beneficial to her. Meanwhile Hahn initially told his colleagues at Friedrich Wilhelm University that she was on vacation, and then, after the summer break, that she had retired.

Lise eventually ended up in Stockholm, but was quite miserable. There was nobody to collaborate with in the type of work she had been doing. When her belongings were shipped to her several months later they had been destroyed. And eventually she got word that her family members were being sent away to concentration camps.

Lise and Otto continued their collaboration via correspondence throughout the rest of that year. It was quite productive for Hahn, and the only avenue for continued meaningful work for Meitner. Otto was attempting to make elements larger than uranium by bombarding it with slow moving neutrons, but was instead detecting elements smaller than uranium. In a letter to Meitner, Hahn said "Perhaps you can come up with some fantastic explanation. We ourselves realize that it can't actually burst apart into barium." Meitner's initial response was skeptical as well, but also reminded Hahn that "in nuclear physics we have experienced so many surprises, that one cannot unconditionally say: 'It is impossible.'" Hahn had felt urgent about publishing the results and sent the article for publication before the Christmas break and just days after his letter to Meitner. He would revise it twice over the course of the next month. The results were quite puzzling. How could barium be created from uranium? And yet that is exactly what the experimental results were telling him again and again.

Meanwhile, Lise and her nephew Otto Robert Frisch were spending the holidays in Sweden and discussing Hahn's latest letter. Frisch was also a nuclear physicist, and was among the many Jewish scientists who left Germany when Hitler came to power in 1933. It was a smaller than usual gathering for the Meitner family, as many of their family members were displaced or imprisoned at this point. Otto was skiing and Lise walking through the snow as they discussed what to make of the bizarre results. The existence of barium suggested that the nucleus must have split in two if the experimental results were to be trusted, which Meitner assured her cousin that they should be. As physicists, they had to consider what might cause this. They calculated that the electrical repulsions within the nucleus of uranium 235 were very close to the strong force holding it together. As such, a very small disturbance would likely be sufficient to break such a nucleus apart. But they also needed to determine what energy source would power the rapid acceleration of the two new nuclei away from each other. Meitner realized that Einstein's equation $E=mc^2$ might explain the results. When they did the calculations they found that the amount of energy necessary for the process could be explained by the difference in mass between the products and the reactant. Frisch was able to conduct an experiment in his lab to confirm this, and the two of them published their results at the beginning of 1939.

Neils Bohr and the physics community at large immediately saw the veracity of these results and there was much interest in them. The correspondence between Meitner and Hahn became strained as word got out about the discovery of nuclear fission. Bohr was presenting Meitner's findings at a prominent meeting in Washington D.C. by the end of January. Bohr made very clear in this presentation that it was Meitner and Frisch who had correctly interpreted the experiments of Hahn and Strassmann, he feared that if he did not speak she would not get the credit she deserved. But Hahn was not acknowledging this in any of his public comments or publications. When Meitner expressed frustration with this to Hahn, his response was "I don't give these things much weight, of course, but I didn't want to confess to [a work colleague] that you were the only one who found out everything immediately." Hahn feared that it might jeopardize his position at the University if it came out that he was collaborating with a Jew, and asked Meitner

for loyalty in his correspondence. He thought the little bit he had done to aid in her escape was sufficient. Meanwhile, the Nazis had taken away Meitner's access to her pension and removed her name from the articles she had already published with Hahn. Hahn acknowledging her current contributions was the only way for her voice to be heard correctly within the scientific community, so it was troubling for her that he refused to do so. Meitner's time was also divided throughout 1939, as she was working to get the rest of her family to safety. Their correspondence throughout the early months of 1939 grew increasingly fraught.

By September, barely a year beyond when Hahn had given Lise the diamond ring the day before she escaped Germany, the Nazis were invading Poland and everything that had been strained between them was ratcheted up to a new level. Albert Einstein had written a letter to President Roosevelt, warning him that the Nazis were probably working on an atomic bomb. There were many obstacles to be overcome in the process of developing a nuclear bomb, but the Manhattan Project was able to convince Frisch, Bohr, and almost every other nuclear physicist of note (particularly those who had fled Germany already) to participate. The two notable exceptions were Einstein and Meitner. Einstein was not given clearance to participate in the project because of his political views, while Meitner simply said she wanted to have nothing to do with a bomb.

The Allied forces were highly concerned that the Nazis might be working on an atomic bomb themselves. In February 1944, they bombed the University where Hahn worked as an effort to disrupt this assumed work. A special allied mission arrested Hahn and several physicists including Werner Heisenberg in April 1945, transporting them across Europe several times, and eventually confining them in comfortable house arrest at Farm Hall, where their conversations were recorded for several months. When Hahn and the others heard about the bombing of Hiroshima in August, they initially expressed shock at how much farther along the Allied forces were in their development of the atomic bomb. And then they decided that it was better that they had not been successful in creating an atomic bomb, and even that they were morally superior to the Allied scientists for having made the choice to focus on nuclear energy as a source of power instead of as a weapon. They wrote a memorandum explaining their logic, and they doubled down on the narrative of Hahn and Strassmann as the discoverers of nuclear fission as part of this writing.

Throughout this time, Lise was largely in the dark about what was going on. Although she had some correspondence with Bohr and Frisch in the U.S., they were not allowed to even mention that they were working on a bomb. Lise found out about it after the bombs were dropped in Japan when a reporter came to her to ask about it. She was horrified to learn that anything she had done would be put to such use, but still she was reported to have been involved in the process by much of the media at that time, often being called the Mother of the Atomic bomb.

While he was still a prisoner of war, Hahn was awarded the Nobel Prize in Chemistry a few short months after the war had ended for his discovery of nuclear fission. His location was still top secret at this time, so he had to read about it in the newspaper. He was repatriated into British controlled Germany in early 1945, but it was difficult to get permission to travel until much later in the year. He was finally able to give his acceptance speech in December 1946, over a year after he had learned about winning the award. Lise attended the ceremony, hoping that he would mention her contributions at some point during his speech, but he did not. She was quite disappointed by this, and by Hahn's lack of "thoughtfulness and a certain strength of character," as she wrote to a colleague afterwards. Many of her colleagues joked that the crowning achievement of her career was winning a Nobel Prize for Otto Hahn. This erasure of Lise Meitner's work continues to this day with her absence in the recent popular film Oppenheimer.

They both tried to make amends in their own ways in the years that followed. Meitner became an advocate for nuclear peace, and when Hahn gave her some of the money from the Nobel Prize she donated it to this cause. Despite their differences, Meitner still had a fondness for Hahn and simply wished he had done more to acknowledge his complicity with all the horrors done in Nazi Germany even if he was not a Nazi himself. They both died in 1968 at the age of 89. Lise Meitner's gravestone reads "A physicist who never lost her humanity."

There were several transuranium elements discovered during the secrecy of the Cold War, and which remained officially unnamed because there were disputes about who had discovered them. This dispute was finally resolved in 1997, five years after the Farm Hall transcripts were released to the public. Although the United States team proposed an element named after Otto Hahn, the disputed element was eventually named Dubnium after the location in Russia where it had been isolated. The German team was awarded the rights to name three elements, and chose to name one of them after Lise Meitner. Lise Meitner and Marie Curie are the only non-mythological women with elements of the periodic table named after them (and technically Curium is named after both Marie and her husband Pierre). There are eight elements named after mythological women.



Figure **14.11.2**: A photograph of Meitner on the steps of a college talking to students in 1959. Public Domain. Wiki Commons.

Additional Resources:

an article by Meitner about her life.

a biography of Meitner written in the 1990s.

An article about Otto Hahn's role in Nazi Germany.

An article about the transcripts from Farm Hall.

A short biography of Meitner (and others).

A fairly recent podcast about Meitner's experience. (this links to the first of two parts)

A recent biography written by the person interviewed in this podcast.

14.11: Process of Science- Lise Meitner and Otto Hahn is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Index

B

band of stability
14.9: Nuclear Binding Energy
binding energy per nucleon
14.9: Nuclear Binding Energy

C

coordination complex
9.7: Our Extraordinary World of Colors
coordination complexes
9.7: Our Extraordinary World of Colors
crystal field theory
9.7: Our Extraordinary World of Colors

D

debriefing
1.4.2: What Should Science be Allowed to do?
deception
1.4.2: What Should Science be Allowed to do?
democritus
10.1: Introduction to Atoms and Elements

E

electromagnetic spectrum
9.7: Our Extraordinary World of Colors
electron volt
14.9: Nuclear Binding Energy
evaporation
12.5: Transitions between Phases

I

IACUC
1.4.2: What Should Science be Allowed to do?
Informed consent
1.4.2: What Should Science be Allowed to do?
Institutional Animal Care and Use Committee
1.4.2: What Should Science be Allowed to do?
Institutional Review Board
1.4.2: What Should Science be Allowed to do?
ionic bonding
11.5: Ionic Bonding
IRB
1.4.2: What Should Science be Allowed to do?

L

law of definite proportions
10.3.2: Law of Definite Proportions

M

magic number
14.9: Nuclear Binding Energy
mass defect
14.9: Nuclear Binding Energy
metallic bonding
11.6: Metallic Bonding

N

nuclear binding energy
14.9: Nuclear Binding Energy
nuclear chemistry
14.9: Nuclear Binding Energy
nucleon
14.9: Nuclear Binding Energy

P

Phrenology
1.4.3: Pseudoscience
Pseudoscience
1.4.3: Pseudoscience

R

radioactivity
14.9: Nuclear Binding Energy
radioisotope
14.9: Nuclear Binding Energy
Reflexology
1.4.3: Pseudoscience
relativistic kinetic energy
14.8: The Energy Within the Nucleus and all Matter
rest energy
14.8: The Energy Within the Nucleus and all Matter

S

scientific method
1.3.1: Laws and Theories in Science
1.4.3: Pseudoscience
spectrochemical series
9.7: Our Extraordinary World of Colors
strong nuclear force
14.9: Nuclear Binding Energy

T

total energy
14.8: The Energy Within the Nucleus and all Matter

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Concepts of Physical Science

Webpages: 227

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 95.6% (217 pages)
- Undeclared: 3.1% (7 pages)
- CK-12 License: 0.9% (2 pages)
- CC BY 4.0: 0.4% (1 page)

By Page

- Concepts of Physical Science - *CC BY-NC-SA 4.0*
  - Front Matter - *CC BY-NC-SA 4.0*
    - Concepts of Physical Science - *CC BY-NC-SA 4.0*
    - About Libretexts - *CC BY-NC-SA 4.0*
    - Prelude - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY-NC-SA 4.0*
  - 1: Science Itself - *CC BY-NC-SA 4.0*
    - 1.1: Observing the World - *CC BY-NC-SA 4.0*
    - 1.2: Developing the Scientific Method - *CC BY-NC-SA 4.0*
      - 1.2.1: Careful Observations- Documentation, Measurements, and Patterns - *CC BY-NC-SA 4.0*
      - 1.2.2: Experiments - *CC BY-NC-SA 4.0*
      - 1.2.3: A Scientific Method - *CC BY-NC-SA 4.0*
    - 1.3: Processes of Science - *CC BY-NC-SA 4.0*
      - 1.3.1: Laws and Theories in Science - *CC BY-NC-SA 4.0*
      - 1.3.2: The Scientific Process - *CC BY-NC-SA 4.0*
    - 1.4: The Limits of Science - *CC BY-NC-SA 4.0*
      - 1.4.1: Other Ways of Knowing - *CC BY-NC-SA 4.0*
      - 1.4.2: What Should Science be Allowed to do? - *CC BY-NC-SA 4.0*
      - 1.4.3: Pseudoscience - *CC BY-NC-SA 4.0*
    - 1.5: Organization of Science - *CC BY-NC-SA 4.0*
      - 1.5.1: Pure and Applied Science - *CC BY-NC-SA 4.0*
      - 1.5.2: Fields of Science - *CC BY-NC-SA 4.0*
  - 2: Matter and Motion - *CC BY-NC-SA 4.0*
    - 2.1: Introduction to Motion - *CC BY-NC-SA 4.0*
    - 2.2: Modern Ideas about Motion - *CC BY-NC-SA 4.0*
    - 2.3: Direction - *CC BY-NC-SA 4.0*
    - 2.4: Vectors, Scalars, and Coordinate Systems - *CC BY-NC-SA 4.0*
    - 2.5: Speed and Velocity - *CC BY-NC-SA 4.0*
    - 2.6: Acceleration - *CC BY-NC-SA 4.0*
    - 2.7: Determining how to solve a motion equation problem - *CC BY-NC-SA 4.0*
    - 2.8: Motion Graphs - *CC BY-NC-SA 4.0*
      - 2.8.1: Motion graphs of position, velocity, and speed versus time - *CC BY-NC-SA 4.0*
      - 2.8.2: Motion graphs of acceleration, velocity, and position versus time. - *CC BY-NC-SA 4.0*
      - 2.8.3: Predicting Motion graphs - *CC BY-NC-SA 4.0*
    - 2.9: Types of Acceleration - *CC BY-NC-SA 4.0*
      - 2.9.1: Falling Objects - *CC BY-NC-SA 4.0*
      - 2.9.2: Centripetal Acceleration - *CC BY-NC-SA 4.0*
      - 2.9.3: Projectile Motion - *CC BY-NC-SA 4.0*
    - 2.10: Measuring the World - *CC BY-NC-SA 4.0*
    - 2.11: Development of Force Concept - *CC BY-NC-SA 4.0*
    - 2.12: Galileo and the Beginnings of Modern Science - *CC BY-NC-SA 4.0*
  - 3: Types of Motion and Their Causes - *CC BY-NC-SA 4.0*
    - 3.1: Introduction to Dynamics and Momentum. - *CC BY-NC-SA 4.0*
    - 3.2: Newton's First Law of Motion- Inertia - *CC BY-NC-SA 4.0*
    - 3.3: Newton's Second Law of Motion- Force and Acceleration - *CC BY-NC-SA 4.0*
    - 3.4: Newton's Third Law of Motion- Symmetry in Forces - *CC BY-NC-SA 4.0*

- 3.5: Examples of Force - *CC BY-NC-SA 4.0*
  - 3.5.1: Normal Force - *CC BY-NC-SA 4.0*
  - 3.5.2: Tension - *CC BY-NC-SA 4.0*
  - 3.5.3: Spring Force- Hooke's Law - *CC BY-NC-SA 4.0*
  - 3.5.4: Friction - *CC BY-NC-SA 4.0*
  - 3.5.5: Centripetal Force - *CC BY-NC-SA 4.0*
- 3.6: Summary of Newton's Laws - *CC BY-NC-SA 4.0*
- 3.7: Newton's Universal Law of Gravitation - *CC BY-NC-SA 4.0*
- 3.8: Linear Momentum and Force - *CC BY-NC-SA 4.0*
- 3.9: Conservation of Momentum - *CC BY-NC-SA 4.0*
- 3.10: Rotation - *CC BY-NC-SA 4.0*
- 3.11: Angular Momentum and Its Conservation - *CC BY-NC-SA 4.0*

4: Fluid Mechanics and Waves - *CC BY-NC-SA 4.0*

- 4.1: Introduction to Fluids and Waves - *CC BY-NC-SA 4.0*
- 4.2: What Is a Fluid? - *CC BY-NC-SA 4.0*
- 4.3: Density - *CC BY-NC-SA 4.0*
- 4.4: Pressure - *CC BY-NC-SA 4.0*
  - 4.4.1: Pressure Due to the Weight of Fluid - *CC BY-NC-SA 4.0*
  - 4.4.2: Boyle's Law - *CC BY-NC-SA 4.0*
- 4.5: Fluid Principles - *CC BY-NC-SA 4.0*
  - 4.5.1: Archimedes' Principle - *CC BY-NC-SA 4.0*
  - 4.5.2: Pascal's Principle - *CC BY-NC-SA 4.0*
  - 4.5.3: Bernoulli's Principle - *CC BY-NC-SA 4.0*
- 4.6: Wave Mechanics - *CC BY-NC-SA 4.0*
  - 4.6.1: Period and Frequency in Oscillations - *CC BY-NC-SA 4.0*
  - 4.6.2: Simple Harmonic Motion- A Special Periodic Motion - *CC BY-NC-SA 4.0*
  - 4.6.3: Waves - *CC BY-NC-SA 4.0*
- 4.7: Properties of Waves - *CC BY-NC-SA 4.0*
  - 4.7.1: Transverse and Longitudinal Waves - *CC BY-NC-SA 4.0*
  - 4.7.2: Forced Oscillations and Resonance - *CC BY-NC-SA 4.0*
  - 4.7.3: Wave Interference- Standing Waves and Beats - *CC BY-NC-SA 4.0*
  - 4.7.4: When Waves Sources Move- Doppler effects and Bow Wakes - *CC BY-NC-SA 4.0*
- 4.8: Sound - *CC BY-NC-SA 4.0*
  - 4.8.1: Speed of Sound, Frequency, and Wavelength - *CC BY-NC-SA 4.0*

5: Work and Energy - *CC BY-NC-SA 4.0*

- 5.1: Introduction to Work and Energy - *CC BY-NC-SA 4.0*
- 5.2: Work- The Scientific Definition - *CC BY-NC-SA 4.0*
- 5.3: Kinetic Energy and the Work-Energy Theorem - *CC BY-NC-SA 4.0*
- 5.4: Potential Energy- Gravity and Springs - *CC BY-NC-SA 4.0*
  - 5.4.1: Gravitational Potential Energy - *CC BY-NC-SA 4.0*
  - 5.4.2: Spring Potential Energy - *CC BY-NC-SA 4.0*
- 5.5: Conservative Forces, Potential Energy, and Conservation of Energy - *CC BY-NC-SA 4.0*
- 5.6: Forms of Energy - *CC BY-NC-SA 4.0*
- 5.7: Simple Machines - *CC BY-NC-SA 4.0*
- 5.8: Power - *CC BY-NC-SA 4.0*
- 5.9: Energy and Momentum - *CC BY-NC-SA 4.0*

6: Heat and Energy - *CC BY-NC-SA 4.0*

- 6.1: Introduction to Thermodynamics - *CC BY-NC-SA 4.0*
- 6.2: Temperature - *CC BY-NC-SA 4.0*
  - 6.2.1: Temperature and Thermometers - *CC BY-NC-SA 4.0*
  - 6.2.2: The Molecular Basis for Temperature - *CC BY-NC-SA 4.0*
- 6.3: Heat - *CC BY-NC-SA 4.0*
- 6.4: Heat Transfer Methods - *CC BY-NC-SA 4.0*
- 6.5: Temperature Change and Heat Capacity - *CC BY-NC-SA 4.0*
- 6.6: The First Law of Thermodynamics - *CC BY-NC-SA 4.0*
- 6.7: Heat Engines and the Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 6.7.1: Efficiency and the Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 6.7.2: Carnot's Perfect Heat Engine- The Second Law of Thermodynamics Restated - *CC BY-NC-SA 4.0*
- 6.8: Entropy and the Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 6.8.1: Statistical Interpretation of Entropy and the Second Law of Thermodynamics- The Underlying Explanation - *CC BY-NC-SA 4.0*
  - 6.8.2: Entropy, Life, the Universe, and Everything - *CC BY-NC-SA 4.0*
- 6.9: Energy Sources for Our Lives - *CC BY-NC-SA 4.0*

- 7: Fundamentals and Applications of Electricity - *CC BY-NC-SA 4.0*
  - 7.1: Introduction to Electricity - *CC BY-NC-SA 4.0*
  - 7.2: Static Electricity and Charge- Conservation of Charge - *CC BY-NC-SA 4.0*
  - 7.3: Coulomb's Law - *CC BY-NC-SA 4.0*
  - 7.4: Representing Electric Fields - *CC BY-NC-SA 4.0*
  - 7.5: Electric Potential and Potential Energy - *CC BY-NC-SA 4.0*
  - 7.6: Conductors and Current - *CC BY-NC-SA 4.0*
  - 7.7: Ohm's Law - *CC BY-NC-SA 4.0*
  - 7.8: Circuits - *CC BY-NC-SA 4.0*
  - 7.9: Applications- Power and Safety - *CC BY-NC-SA 4.0*
    - 7.9.1: Electric Power and Energy - *CC BY-NC-SA 4.0*
    - 7.9.2: Electric Hazards and the Human Body - *CC BY-NC-SA 4.0*
- 8: The Amazing Connections Between Magnets and Light Switches - *CC BY-NC-SA 4.0*
  - 8.1: Introduction to Magnetism - *CC BY-NC-SA 4.0*
  - 8.2: Magnets - *CC BY-NC-SA 4.0*
  - 8.3: Natural Magnetism - *CC BY-NC-SA 4.0*
  - 8.4: The Source of Magnetism - *CC BY-NC-SA 4.0*
  - 8.5: Magnetic Fields and Magnetic Field Lines - *CC BY-NC-SA 4.0*
  - 8.6: The relationship between electric current, magnetic fields, and magnetic force - *CC BY-NC-SA 4.0*
  - 8.7: Force from Electricity and Magnetism- Motors and Meters - *CC BY-NC-SA 4.0*
  - 8.8: Faraday's Law- Electricity from Magnetism and Force - *CC BY-NC-SA 4.0*
    - 8.8.1: Applications of Electromagnetic Induction - *CC BY-NC-SA 4.0*
  - 8.9: Magnetism from Electricity - *CC BY-NC-SA 4.0*
  - 8.10: Maxwell's Equations- Summarizing the Connections Between Electricity and Magnetism - *CC BY-NC-SA 4.0*
  - 8.11: Sources of Voltage and the Electrification of America - *CC BY-NC-SA 4.0*
    - 8.11.1: Batteries - *CC BY-NC-SA 4.0*
    - 8.11.2: Induced Voltage and Electrical Generators - *CC BY-NC-SA 4.0*
    - 8.11.3: Alternating Current versus Direct Current - *CC BY-NC-SA 4.0*
    - 8.11.4: Transformers - *CC BY-NC-SA 4.0*
- 9: Electromagnetic Radiation and Optics - *CC BY-NC-SA 4.0*
  - 9.1: Introduction to Light - *CC BY-NC-SA 4.0*
  - 9.2: Maxwell's Equations- Electromagnetic Waves Predicted and Observed - *CC BY-NC-SA 4.0*
  - 9.3: Producing Electromagnetic Radiation - *CC BY-NC-SA 4.0*
  - 9.4: The Electromagnetic Spectrum- an Overview - *CC BY-NC-SA 4.0*
  - 9.5: The Electromagnetic Spectrum- Application Notes - *CC BY-NC-SA 4.0*
  - 9.6: Interaction of Light with Matter - *CC BY-NC-SA 4.0*
    - 9.6.1: Reflection - *CC BY-NC-SA 4.0*
    - 9.6.2: Refraction - *CC BY-NC-SA 4.0*
    - 9.6.3: Dispersion- The Rainbow and Prisms - *CC BY-NC-SA 4.0*
  - 9.7: Our Extraordinary World of Colors - *CC BY-NC-SA 4.0*
  - 9.8: Lenses, Mirrors, and Polarization - *CC BY-NC-SA 4.0*
    - 9.8.1: Image Formation by Lenses - *CC BY-NC-SA 4.0*
    - 9.8.2: Image Formation by Mirrors - *CC BY-NC-SA 4.0*
    - 9.8.3: Polarization - *CC BY-NC-SA 4.0*
  - 9.9: The Greenhouse Effect and Climate Change - *CC BY-NC-SA 4.0*
- 10: Elements Are Made of Atoms - *CC BY-NC-SA 4.0*
  - 10.1: Introduction to Atoms and Elements - *CC BY-NC-SA 4.0*
  - 10.2: Modern Understandings of Elements - *CC BY-NC-SA 4.0*
  - 10.3: The Development of Atomic Theory - *CC BY-NC-SA 4.0*
    - 10.3.1: Law of Conservation of Mass - *CC BY-NC-SA 4.0*
    - 10.3.2: Law of Definite Proportions - *CC BY-NC-SA 4.0*
    - 10.3.3: Law of Multiple Proportions - *CC BY-NC-SA 4.0*
    - 10.3.4: Dalton's Atomic Theory - *CC BY-NC-SA 4.0*
  - 10.4: Organizing the Elements - *CK-12 License*
    - 10.4.1: Mendeleev's Periodic Table - *CC BY-NC-SA 4.0*
    - 10.4.2: Periodic Law - *CC BY-NC-SA 4.0*

        - 10.4.3: Modern Periodic Table- Periods and Groups - *CC BY-NC-SA 4.0*
    - 10.5: Understanding the Periodic Table - *CC BY-NC-SA 4.0*
        - 10.5.1: Metals, Nonmetals, and Metalloids. - *CC BY-NC-SA 4.0*
        - 10.5.2: Families of the Periodic Table - *CC BY-NC-SA 4.0*
    - 10.6: Inside the Atom - *CC BY-NC-SA 4.0*
    - 10.7: Electrons and Light - *CC BY-NC-SA 4.0*
    - 10.8: Quantum Mechanics - *CC BY-NC-SA 4.0*
    - 10.9: Electron Shells - *CC BY-NC-SA 4.0*
- 11: Chemical Bonds - *CC BY-NC-SA 4.0*
    - 11.1: Chemical Bond - *CC BY-NC-SA 4.0*
    - 11.2: Elements, molecules, and compounds - *CC BY-NC-SA 4.0*
    - 11.3: The Octet Rule - *CC BY-NC-SA 4.0*
    - 11.4: Ions and Ion Formation - *CC BY-NC-SA 4.0*
    - 11.5: Ionic Bonding - *CC BY-NC-SA 4.0*
    - 11.6: Metallic Bonding - *CC BY-NC-SA 4.0*
    - 11.7: Covalent Bonds and Molecules - *CC BY-NC-SA 4.0*
    - 11.8: Bond Polarity - *CC BY-NC-SA 4.0*
    - 11.9: Intermolecular Forces - *CC BY-NC-SA 4.0*
    - 11.10: Naming Compounds - *CC BY-NC-SA 4.0*
- 12: Physical Properties - *CC BY-NC-SA 4.0*
    - 12.1: Introduction to Physical Properties - *CC BY-NC-SA 4.0*
    - 12.2: Physical and Chemical - *CC BY-NC-SA 4.0*
    - 12.3: Classifying Matter - *CC BY-NC-SA 4.0*
    - 12.4: The Physical States of Matter - *CC BY-NC-SA 4.0*
    - 12.5: Transitions between Phases - *CC BY-NC-SA 4.0*
    - 12.6: Energy and Phase Transitions - *CC BY-NC-SA 4.0*
    - 12.7: Moles- Counting Atoms - *Undeclared*
        - 12.7.1: Moles and Avogrado's Number - *CC BY-NC-SA 4.0*
        - 12.7.2: Moles and Mass - *Undeclared*
    - 12.8: Solutions and Solubility - *CC BY-NC-SA 4.0*
    - 12.9: Concentrations of Solutions - *CC BY-NC-SA 4.0*
    - 12.10: Soap- an Application of Physical Properties - *CC BY-NC-SA 4.0*
- 13: Chemical Reactions - *CC BY-NC-SA 4.0*
    - 13.1: Introduction to Chemical Changes - *CC BY-NC-SA 4.0*
    - 13.2: Aspects of a Chemical Change - *CC BY-NC-SA 4.0*
    - 13.3: Chemical Equations- Representing Chemical Changes - *CC BY-NC-SA 4.0*
    - 13.4: Balancing Equations - *CK-12 License*
    - 13.5: Energy and Chemical Reactions - *CC BY-NC-SA 4.0*
    - 13.6: Entropy and Chemical Reactions - *CC BY-NC-SA 4.0*
    - 13.7: Classifying Reactions - *CC BY-NC-SA 4.0*
    - 13.8: Acid Base Reactions - *CC BY-NC-SA 4.0*
    - 13.9: Redox Reactions- The Transfer of Electrons - *CC BY-NC-SA 4.0*
    - 13.10: Polymerization Reactions - *CC BY-NC-SA 4.0*
- 14: Nuclear Radiation - *CC BY-NC-SA 4.0*
    - 14.1: Introduction to Radioactivity - *CC BY-NC-SA 4.0*
    - 14.2: The Atomic Nucleus - *CC BY-NC-SA 4.0*
    - 14.3: Discovery of Radioactivity - *CC BY-NC-SA 4.0*
    - 14.4: Nuclear Decay Processes - *CC BY-NC-SA 4.0*
    - 14.5: Nuclear Equations - *Undeclared*
    - 14.6: Half-Life - *CC BY-NC-SA 4.0*
    - 14.7: What Holds the Nucleus Together - *CC BY-NC-SA 4.0*
    - 14.8: The Energy Within the Nucleus and all Matter - *CC BY-NC-SA 4.0*
    - 14.9: Nuclear Binding Energy - *CC BY 4.0*
        - 14.9.1: Nuclear Fission - *CC BY-NC-SA 4.0*
        - 14.9.2: Nuclear Fusion - *CC BY-NC-SA 4.0*
    - 14.10: Nuclear Radiation in Our Lives - *CC BY-NC-SA 4.0*
    - 14.11: Process of Science- Lise Meitner and Otto Hahn - *Undeclared*
- Back Matter - *CC BY-NC-SA 4.0*
    - Index - *CC BY-NC-SA 4.0*
    - Glossary - *CC BY-NC-SA 4.0*
    - Detailed Licensing - *CC BY-NC-SA 4.0*
    - Detailed Licensing - *Undeclared*

Detailed Licensing

Overview

Title: Concepts of Physical Science

Webpages: 227

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 95.6% (217 pages)
- Undeclared: 3.1% (7 pages)
- CK-12 License: 0.9% (2 pages)
- CC BY 4.0: 0.4% (1 page)

By Page

- Concepts of Physical Science - *CC BY-NC-SA 4.0*
  - Front Matter - *CC BY-NC-SA 4.0*
    - Concepts of Physical Science - *CC BY-NC-SA 4.0*
    - About Libretexts - *CC BY-NC-SA 4.0*
    - Prelude - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY-NC-SA 4.0*
  - 1: Science Itself - *CC BY-NC-SA 4.0*
    - 1.1: Observing the World - *CC BY-NC-SA 4.0*
    - 1.2: Developing the Scientific Method - *CC BY-NC-SA 4.0*
      - 1.2.1: Careful Observations- Documentation, Measurements, and Patterns - *CC BY-NC-SA 4.0*
      - 1.2.2: Experiments - *CC BY-NC-SA 4.0*
      - 1.2.3: A Scientific Method - *CC BY-NC-SA 4.0*
    - 1.3: Processes of Science - *CC BY-NC-SA 4.0*
      - 1.3.1: Laws and Theories in Science - *CC BY-NC-SA 4.0*
      - 1.3.2: The Scientific Process - *CC BY-NC-SA 4.0*
    - 1.4: The Limits of Science - *CC BY-NC-SA 4.0*
      - 1.4.1: Other Ways of Knowing - *CC BY-NC-SA 4.0*
      - 1.4.2: What Should Science be Allowed to do? - *CC BY-NC-SA 4.0*
      - 1.4.3: Pseudoscience - *CC BY-NC-SA 4.0*
    - 1.5: Organization of Science - *CC BY-NC-SA 4.0*
      - 1.5.1: Pure and Applied Science - *CC BY-NC-SA 4.0*
      - 1.5.2: Fields of Science - *CC BY-NC-SA 4.0*
  - 2: Matter and Motion - *CC BY-NC-SA 4.0*
    - 2.1: Introduction to Motion - *CC BY-NC-SA 4.0*
    - 2.2: Modern Ideas about Motion - *CC BY-NC-SA 4.0*
    - 2.3: Direction - *CC BY-NC-SA 4.0*
    - 2.4: Vectors, Scalars, and Coordinate Systems - *CC BY-NC-SA 4.0*
    - 2.5: Speed and Velocity - *CC BY-NC-SA 4.0*
    - 2.6: Acceleration - *CC BY-NC-SA 4.0*
    - 2.7: Determining how to solve a motion equation problem - *CC BY-NC-SA 4.0*
    - 2.8: Motion Graphs - *CC BY-NC-SA 4.0*
      - 2.8.1: Motion graphs of position, velocity, and speed versus time - *CC BY-NC-SA 4.0*
      - 2.8.2: Motion graphs of acceleration, velocity, and position versus time. - *CC BY-NC-SA 4.0*
      - 2.8.3: Predicting Motion graphs - *CC BY-NC-SA 4.0*
    - 2.9: Types of Acceleration - *CC BY-NC-SA 4.0*
      - 2.9.1: Falling Objects - *CC BY-NC-SA 4.0*
      - 2.9.2: Centripetal Acceleration - *CC BY-NC-SA 4.0*
      - 2.9.3: Projectile Motion - *CC BY-NC-SA 4.0*
    - 2.10: Measuring the World - *CC BY-NC-SA 4.0*
    - 2.11: Development of Force Concept - *CC BY-NC-SA 4.0*
    - 2.12: Galileo and the Beginnings of Modern Science - *CC BY-NC-SA 4.0*
  - 3: Types of Motion and Their Causes - *CC BY-NC-SA 4.0*
    - 3.1: Introduction to Dynamics and Momentum. - *CC BY-NC-SA 4.0*
    - 3.2: Newton's First Law of Motion- Inertia - *CC BY-NC-SA 4.0*
    - 3.3: Newton's Second Law of Motion- Force and Acceleration - *CC BY-NC-SA 4.0*
    - 3.4: Newton's Third Law of Motion- Symmetry in Forces - *CC BY-NC-SA 4.0*

- 3.5: Examples of Force - *CC BY-NC-SA 4.0*
  - 3.5.1: Normal Force - *CC BY-NC-SA 4.0*
  - 3.5.2: Tension - *CC BY-NC-SA 4.0*
  - 3.5.3: Spring Force- Hooke's Law - *CC BY-NC-SA 4.0*
  - 3.5.4: Friction - *CC BY-NC-SA 4.0*
  - 3.5.5: Centripetal Force - *CC BY-NC-SA 4.0*
- 3.6: Summary of Newton's Laws - *CC BY-NC-SA 4.0*
- 3.7: Newton's Universal Law of Gravitation - *CC BY-NC-SA 4.0*
- 3.8: Linear Momentum and Force - *CC BY-NC-SA 4.0*
- 3.9: Conservation of Momentum - *CC BY-NC-SA 4.0*
- 3.10: Rotation - *CC BY-NC-SA 4.0*
- 3.11: Angular Momentum and Its Conservation - *CC BY-NC-SA 4.0*

4: Fluid Mechanics and Waves - *CC BY-NC-SA 4.0*

- 4.1: Introduction to Fluids and Waves - *CC BY-NC-SA 4.0*
- 4.2: What Is a Fluid? - *CC BY-NC-SA 4.0*
- 4.3: Density - *CC BY-NC-SA 4.0*
- 4.4: Pressure - *CC BY-NC-SA 4.0*
  - 4.4.1: Pressure Due to the Weight of Fluid - *CC BY-NC-SA 4.0*
  - 4.4.2: Boyle's Law - *CC BY-NC-SA 4.0*
- 4.5: Fluid Principles - *CC BY-NC-SA 4.0*
  - 4.5.1: Archimedes' Principle - *CC BY-NC-SA 4.0*
  - 4.5.2: Pascal's Principle - *CC BY-NC-SA 4.0*
  - 4.5.3: Bernoulli's Principle - *CC BY-NC-SA 4.0*
- 4.6: Wave Mechanics - *CC BY-NC-SA 4.0*
  - 4.6.1: Period and Frequency in Oscillations - *CC BY-NC-SA 4.0*
  - 4.6.2: Simple Harmonic Motion- A Special Periodic Motion - *CC BY-NC-SA 4.0*
  - 4.6.3: Waves - *CC BY-NC-SA 4.0*
- 4.7: Properties of Waves - *CC BY-NC-SA 4.0*
  - 4.7.1: Transverse and Longitudinal Waves - *CC BY-NC-SA 4.0*
  - 4.7.2: Forced Oscillations and Resonance - *CC BY-NC-SA 4.0*
  - 4.7.3: Wave Interference- Standing Waves and Beats - *CC BY-NC-SA 4.0*
  - 4.7.4: When Waves Sources Move- Doppler effects and Bow Wakes - *CC BY-NC-SA 4.0*
- 4.8: Sound - *CC BY-NC-SA 4.0*
  - 4.8.1: Speed of Sound, Frequency, and Wavelength - *CC BY-NC-SA 4.0*

5: Work and Energy - *CC BY-NC-SA 4.0*

- 5.1: Introduction to Work and Energy - *CC BY-NC-SA 4.0*
- 5.2: Work- The Scientific Definition - *CC BY-NC-SA 4.0*
- 5.3: Kinetic Energy and the Work-Energy Theorem - *CC BY-NC-SA 4.0*
- 5.4: Potential Energy- Gravity and Springs - *CC BY-NC-SA 4.0*
  - 5.4.1: Gravitational Potential Energy - *CC BY-NC-SA 4.0*
  - 5.4.2: Spring Potential Energy - *CC BY-NC-SA 4.0*
- 5.5: Conservative Forces, Potential Energy, and Conservation of Energy - *CC BY-NC-SA 4.0*
- 5.6: Forms of Energy - *CC BY-NC-SA 4.0*
- 5.7: Simple Machines - *CC BY-NC-SA 4.0*
- 5.8: Power - *CC BY-NC-SA 4.0*
- 5.9: Energy and Momentum - *CC BY-NC-SA 4.0*

6: Heat and Energy - *CC BY-NC-SA 4.0*

- 6.1: Introduction to Thermodynamics - *CC BY-NC-SA 4.0*
- 6.2: Temperature - *CC BY-NC-SA 4.0*
  - 6.2.1: Temperature and Thermometers - *CC BY-NC-SA 4.0*
  - 6.2.2: The Molecular Basis for Temperature - *CC BY-NC-SA 4.0*
- 6.3: Heat - *CC BY-NC-SA 4.0*
- 6.4: Heat Transfer Methods - *CC BY-NC-SA 4.0*
- 6.5: Temperature Change and Heat Capacity - *CC BY-NC-SA 4.0*
- 6.6: The First Law of Thermodynamics - *CC BY-NC-SA 4.0*
- 6.7: Heat Engines and the Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 6.7.1: Efficiency and the Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 6.7.2: Carnot's Perfect Heat Engine- The Second Law of Thermodynamics Restated - *CC BY-NC-SA 4.0*
- 6.8: Entropy and the Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 6.8.1: Statistical Interpretation of Entropy and the Second Law of Thermodynamics- The Underlying Explanation - *CC BY-NC-SA 4.0*
  - 6.8.2: Entropy, Life, the Universe, and Everything - *CC BY-NC-SA 4.0*
- 6.9: Energy Sources for Our Lives - *CC BY-NC-SA 4.0*

- 7: Fundamentals and Applications of Electricity - *CC BY-NC-SA 4.0*
  - 7.1: Introduction to Electricity - *CC BY-NC-SA 4.0*
  - 7.2: Static Electricity and Charge- Conservation of Charge - *CC BY-NC-SA 4.0*
  - 7.3: Coulomb's Law - *CC BY-NC-SA 4.0*
  - 7.4: Representing Electric Fields - *CC BY-NC-SA 4.0*
  - 7.5: Electric Potential and Potential Energy - *CC BY-NC-SA 4.0*
  - 7.6: Conductors and Current - *CC BY-NC-SA 4.0*
  - 7.7: Ohm's Law - *CC BY-NC-SA 4.0*
  - 7.8: Circuits - *CC BY-NC-SA 4.0*
  - 7.9: Applications- Power and Safety - *CC BY-NC-SA 4.0*
    - 7.9.1: Electric Power and Energy - *CC BY-NC-SA 4.0*
    - 7.9.2: Electric Hazards and the Human Body - *CC BY-NC-SA 4.0*
- 8: The Amazing Connections Between Magnets and Light Switches - *CC BY-NC-SA 4.0*
  - 8.1: Introduction to Magnetism - *CC BY-NC-SA 4.0*
  - 8.2: Magnets - *CC BY-NC-SA 4.0*
  - 8.3: Natural Magnetism - *CC BY-NC-SA 4.0*
  - 8.4: The Source of Magnetism - *CC BY-NC-SA 4.0*
  - 8.5: Magnetic Fields and Magnetic Field Lines - *CC BY-NC-SA 4.0*
  - 8.6: The relationship between electric current, magnetic fields, and magnetic force - *CC BY-NC-SA 4.0*
  - 8.7: Force from Electricity and Magnetism- Motors and Meters - *CC BY-NC-SA 4.0*
  - 8.8: Faraday's Law- Electricity from Magnetism and Force - *CC BY-NC-SA 4.0*
    - 8.8.1: Applications of Electromagnetic Induction - *CC BY-NC-SA 4.0*
  - 8.9: Magnetism from Electricity - *CC BY-NC-SA 4.0*
  - 8.10: Maxwell's Equations- Summarizing the Connections Between Electricity and Magnetism - *CC BY-NC-SA 4.0*
  - 8.11: Sources of Voltage and the Electrification of America - *CC BY-NC-SA 4.0*
    - 8.11.1: Batteries - *CC BY-NC-SA 4.0*
    - 8.11.2: Induced Voltage and Electrical Generators - *CC BY-NC-SA 4.0*
    - 8.11.3: Alternating Current versus Direct Current - *CC BY-NC-SA 4.0*
    - 8.11.4: Transformers - *CC BY-NC-SA 4.0*
- 9: Electromagnetic Radiation and Optics - *CC BY-NC-SA 4.0*
  - 9.1: Introduction to Light - *CC BY-NC-SA 4.0*
  - 9.2: Maxwell's Equations- Electromagnetic Waves Predicted and Observed - *CC BY-NC-SA 4.0*
  - 9.3: Producing Electromagnetic Radiation - *CC BY-NC-SA 4.0*
  - 9.4: The Electromagnetic Spectrum- an Overview - *CC BY-NC-SA 4.0*
  - 9.5: The Electromagnetic Spectrum- Application Notes - *CC BY-NC-SA 4.0*
  - 9.6: Interaction of Light with Matter - *CC BY-NC-SA 4.0*
    - 9.6.1: Reflection - *CC BY-NC-SA 4.0*
    - 9.6.2: Refraction - *CC BY-NC-SA 4.0*
    - 9.6.3: Dispersion- The Rainbow and Prisms - *CC BY-NC-SA 4.0*
  - 9.7: Our Extraordinary World of Colors - *CC BY-NC-SA 4.0*
  - 9.8: Lenses, Mirrors, and Polarization - *CC BY-NC-SA 4.0*
    - 9.8.1: Image Formation by Lenses - *CC BY-NC-SA 4.0*
    - 9.8.2: Image Formation by Mirrors - *CC BY-NC-SA 4.0*
    - 9.8.3: Polarization - *CC BY-NC-SA 4.0*
  - 9.9: The Greenhouse Effect and Climate Change - *CC BY-NC-SA 4.0*
- 10: Elements Are Made of Atoms - *CC BY-NC-SA 4.0*
  - 10.1: Introduction to Atoms and Elements - *CC BY-NC-SA 4.0*
  - 10.2: Modern Understandings of Elements - *CC BY-NC-SA 4.0*
  - 10.3: The Development of Atomic Theory - *CC BY-NC-SA 4.0*
    - 10.3.1: Law of Conservation of Mass - *CC BY-NC-SA 4.0*
    - 10.3.2: Law of Definite Proportions - *CC BY-NC-SA 4.0*
    - 10.3.3: Law of Multiple Proportions - *CC BY-NC-SA 4.0*
    - 10.3.4: Dalton's Atomic Theory - *CC BY-NC-SA 4.0*
  - 10.4: Organizing the Elements - *CK-12 License*
    - 10.4.1: Mendeleev's Periodic Table - *CC BY-NC-SA 4.0*
    - 10.4.2: Periodic Law - *CC BY-NC-SA 4.0*

        - 10.4.3: Modern Periodic Table- Periods and Groups - *CC BY-NC-SA 4.0*
    - 10.5: Understanding the Periodic Table - *CC BY-NC-SA 4.0*
        - 10.5.1: Metals, Nonmetals, and Metalloids. - *CC BY-NC-SA 4.0*
        - 10.5.2: Families of the Periodic Table - *CC BY-NC-SA 4.0*
    - 10.6: Inside the Atom - *CC BY-NC-SA 4.0*
    - 10.7: Electrons and Light - *CC BY-NC-SA 4.0*
    - 10.8: Quantum Mechanics - *CC BY-NC-SA 4.0*
    - 10.9: Electron Shells - *CC BY-NC-SA 4.0*
- 11: Chemical Bonds - *CC BY-NC-SA 4.0*
    - 11.1: Chemical Bond - *CC BY-NC-SA 4.0*
    - 11.2: Elements, molecules, and compounds - *CC BY-NC-SA 4.0*
    - 11.3: The Octet Rule - *CC BY-NC-SA 4.0*
    - 11.4: Ions and Ion Formation - *CC BY-NC-SA 4.0*
    - 11.5: Ionic Bonding - *CC BY-NC-SA 4.0*
    - 11.6: Metallic Bonding - *CC BY-NC-SA 4.0*
    - 11.7: Covalent Bonds and Molecules - *CC BY-NC-SA 4.0*
    - 11.8: Bond Polarity - *CC BY-NC-SA 4.0*
    - 11.9: Intermolecular Forces - *CC BY-NC-SA 4.0*
    - 11.10: Naming Compounds - *CC BY-NC-SA 4.0*
- 12: Physical Properties - *CC BY-NC-SA 4.0*
    - 12.1: Introduction to Physical Properties - *CC BY-NC-SA 4.0*
    - 12.2: Physical and Chemical - *CC BY-NC-SA 4.0*
    - 12.3: Classifying Matter - *CC BY-NC-SA 4.0*
    - 12.4: The Physical States of Matter - *CC BY-NC-SA 4.0*
    - 12.5: Transitions between Phases - *CC BY-NC-SA 4.0*
    - 12.6: Energy and Phase Transitions - *CC BY-NC-SA 4.0*
    - 12.7: Moles- Counting Atoms - *Undeclared*
        - 12.7.1: Moles and Avogrado's Number - *CC BY-NC-SA 4.0*
        - 12.7.2: Moles and Mass - *Undeclared*
    - 12.8: Solutions and Solubility - *CC BY-NC-SA 4.0*
    - 12.9: Concentrations of Solutions - *CC BY-NC-SA 4.0*
    - 12.10: Soap- an Application of Physical Properties - *CC BY-NC-SA 4.0*
- 13: Chemical Reactions - *CC BY-NC-SA 4.0*
    - 13.1: Introduction to Chemical Changes - *CC BY-NC-SA 4.0*
    - 13.2: Aspects of a Chemical Change - *CC BY-NC-SA 4.0*
    - 13.3: Chemical Equations- Representing Chemical Changes - *CC BY-NC-SA 4.0*
    - 13.4: Balancing Equations - *CK-12 License*
    - 13.5: Energy and Chemical Reactions - *CC BY-NC-SA 4.0*
    - 13.6: Entropy and Chemical Reactions - *CC BY-NC-SA 4.0*
    - 13.7: Classifying Reactions - *CC BY-NC-SA 4.0*
    - 13.8: Acid Base Reactions - *CC BY-NC-SA 4.0*
    - 13.9: Redox Reactions- The Transfer of Electrons - *CC BY-NC-SA 4.0*
    - 13.10: Polymerization Reactions - *CC BY-NC-SA 4.0*
- 14: Nuclear Radiation - *CC BY-NC-SA 4.0*
    - 14.1: Introduction to Radioactivity - *CC BY-NC-SA 4.0*
    - 14.2: The Atomic Nucleus - *CC BY-NC-SA 4.0*
    - 14.3: Discovery of Radioactivity - *CC BY-NC-SA 4.0*
    - 14.4: Nuclear Decay Processes - *CC BY-NC-SA 4.0*
    - 14.5: Nuclear Equations - *Undeclared*
    - 14.6: Half-Life - *CC BY-NC-SA 4.0*
    - 14.7: What Holds the Nucleus Together - *CC BY-NC-SA 4.0*
    - 14.8: The Energy Within the Nucleus and all Matter - *CC BY-NC-SA 4.0*
    - 14.9: Nuclear Binding Energy - *CC BY 4.0*
        - 14.9.1: Nuclear Fission - *CC BY-NC-SA 4.0*
        - 14.9.2: Nuclear Fusion - *CC BY-NC-SA 4.0*
    - 14.10: Nuclear Radiation in Our Lives - *CC BY-NC-SA 4.0*
    - 14.11: Process of Science- Lise Meitner and Otto Hahn - *Undeclared*
- Back Matter - *CC BY-NC-SA 4.0*
    - Index - *CC BY-NC-SA 4.0*
    - Glossary - *CC BY-NC-SA 4.0*
    - Detailed Licensing - *CC BY-NC-SA 4.0*
    - Detailed Licensing - *Undeclared*