CONCEPTUAL PHYSICS



Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett
College of DuPage



Conceptual-Physics-1736961871.epub

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 borne 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 were 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.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 10/16/2026

TABLE OF CONTENTS

Licensing

About This Book

Disclaimer

I: Introduction

Chapter 1: About science

II: Kinematics

Chapter 2: Newton's first law

Chapter 3: Linear motion

Chapter 4: Newton's second law

Chapter 5: Newton's third law

Chapter 6: Momentum

Chapter 7: Work and energy

Chapter 8: Rotational Motion

Chapter 9: Gravity

Chapter 10: Projectile and satellite motion

III: Matter

Chapter 11: The atomic nature of matter

Chapter 12: Solids

Chapter 13: Liquids

Chapter 14: Gases

IV: Thermodynamics

Chapter 15: Temperature and Heat

Chapter 16: Heat transfer

Chapter 17: Phase change

Chapter 18: Thermodynamics

V: Sound and waves

Chapter 19: Vibrations and Waves

Chapter 20: Sound

Chapter 21: Musical sounds

VI: Electricity and magnetism

Chapter 22: Electrostatics

Chapter 23: Electric circuits

Chapter 24: Magnetism

Chapter 25: Electromagnetic induction

Chapter 26: Light

Chapter 27: Color

Chapter 28: Reflection and refraction
Chapter 29: Light waves
Chapter 30: Light emission

VII: Quantum and nuclear physics

Chapter 31: Light Quanta
Chapter 32: Quantum physics
Chapter 33: Nuclear physics and radioactivity
Chapter 34: Fission and fusion

Appendices

Appendix A: Vectors and vector arithmetic
Appendix B: Units and unit conversions
Appendix C: Answers to selected questions

Index

Glossary

Detailed Licensing

Glossary (1)

This page titled Conceptual Physics (Pasquale, Fazzini, and Bennett) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Licensing

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

This page titled Licensing is shared under a license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

About This Book

This textbook and ancillary materials were created as part of the Illinois Support for the Creation of Open Educational Resources (SCOERs) project, facilitated by the Consortium of Academic and Research Libraries in Illinois (CARLI) and funded by a $2 million Open Textbooks Pilot Program grant from the Department of Education and the Fund for the Improvement of Post-Secondary Education.[1]

Illinois SCOERs is a true statewide enterprise fundamentally changing the open educational resources landscape in Illinois by providing a new holistic support model that promotes student success through OER awareness, implementation, growth, and adoption. This project features a collaboration between educational systems in Illinois, the library community, and workforce representatives. Developed OER course materials will be released under a license that permits their free use, reuse, modification, and sharing with others., the library community, and workforce representatives. Developed OER course materials will be released under a license that permits their free use, reuse, modification, and sharing with others.

This book was made possible through an Illinois SCOERs Subgrant awarded to College of DuPage. This textbook was created for Physics 1100: Physics, a high enrollment course for undergraduate students at College of DuPage.

Contributors

Contributors to the development of this OER textbook include librarians and staff from College of DuPage:

- Lauren Kosrow, Subgrant Coordinator
- Lara Tompkins, Accessibility Coordinator and Instructional Designer
- Danielle Oakes, 3D Printing Coordinator
- Amanda Crowley, Proofreader

Illinois SCOERs Project Staff

The Illinois SCOERs Project Staff helped coordinate the management of the Open Textbook Pilot grant that made this OER possible. Topical experts shared their knowledge on open educational resources, incorporating technologies including 3D printing, accessibility, author rights, instructional design, publishing, and participated with the review and proofreading processes.

- Anne Craig, Principal Investigator, CARLI Senior Director
- Michele Leigh, Project Director, CARLI Senior Coordinator Open Illinois
- Margaret Chambers, CARLI Director and Advisor, Membership and Communications
- Michelle Haake, CARLI Senior Coordinator Membership and Communications
- Diane Day, CARLI Fiscal Officer
- Elizabeth Clarage, CARLI Director, Collections Services and OER Committee Liaison
- Nicole Swanson, CARLI Senior Coordinator, Library Services and Outreach and OER Committee Liaison
- Sara Benson, Copyright Training Coordinator, Assistant Professor, Copyright Librarian, University of Illinois Urbana-Champaign
- Emily Boles, Instructional Design Coordinator and Trainer, eLearning Specialist, University of Illinois Springfield
- Elisandro (Alex) Cabada, 3D Printing Coordinator Trainer, Assistant Professor and Medical & Bioengineering Librarian, University of Illinois Urbana-Champaign
- Ann Fredricksen, Disability Specialist, Coordinator of Accessible Media Services, University of Illinois Urbana-Champaign, Disability Resources and Educational Services (DRES)
- Kathleen Ricker, Proofreader, ProofreadingPal.com Editing Services

Advisory Committee

The Illinois SCOERs Advisory Committee provided input on the subgrant applications, and the materials produced and contains industry members, librarians, graduate students, and representatives from our project partners:

- Illinois Community College Board
- Illinois Board of Higher Education
- Illinois Department of Commerce & Economic Opportunity
- Illinois Heartland Library System
- Reaching Across Illinois Library System
- Illinois State Library
- Illinois Library Association
- Health Sciences Librarians of Illinois
- University of Illinois Urbana-Champaign's Division of Disability Resources and Educational Services
- Laurie Brown, Executive Director, Hickory Point Christian Village (assisted living) and Illinois Workforce Investment Board Local 1949 Chair
- Lisa Jones, DCEO, Manager, Office of Employment & Training (state workforce agency)
- Suzanne Kinsey, Executive Director, Fox Valley Community Services, (in-home service agency) in DeKalb, LaSalle, Kendall, and Kane Counties
- Danna Williamson, Nursing Education Specialist, Carle Foundation Hospital Clinical Education

Customization/Terms of Use

This textbook is licensed under a Creative Commons Attribution-Non-Commerical 4.0 (CC-BY-NC 4.0) license unless otherwise indicated, which means that you are free to:

- SHARE – copy and redistribute the material in any medium or format
- ADAPT – remix, transform, and build upon the material for any purpose

Licensing Terms

The licensor cannot revoke these freedoms as long as you follow the license terms.

- Attribution: You must give appropriate credit, provide a link to the license, and indicate if any changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- NonCommercial: You may not use the material for commercial purposes.
- No Additional Restrictions: You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.
- Notice: You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable United States copyright exception or limitation.
- No Warranties are Given: The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Suggested Attribution Statement for this Book

Content that is taken from the above OER should include the following attribution statement:

Copyright © Alyssa J. Pasquale, Ph.D., David R. Fazzini, Ph.D., and Carley Bennett, Ph.D. (2024). Illinois SCOERs Conceptual Physics is licensed under a Creative Commons Attribution-Non-Commercial 4.0 (CC BY-NC 4.0) License.

References

This Preface chapter is a derivative of Ernstmeyer, K., & Christman, E. (Eds.). (2022). Nursing Mental Health and Community Concepts by Chippewa Valley Technical College is licensed under CC BY 4.0

---

1. The contents of this publication were developed under a grant from the Fund for the Improvement of Postsecondary Education, (FIPSE), U.S. Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

This page titled About This Book is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Disclaimer

The authors of this textbook are not responsible for any injuries you may sustain, or any property damage that may occur, while performing these experiments. Please use caution and wear appropriate personal protective equipment (PPE).

This page titled Disclaimer is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

I: INTRODUCTION

Chapter 1: About science

This page titled I: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 1: About science

Summary

This chapter will explain how we will interpret and come to understand physics in this class. It focuses on the terminology, processes, and procedures that will form the foundation for a solid understanding of conceptual physics.

What is science?

Science is the systematic pursuit and organization of knowledge that comes about through asking questions, making careful observations, and carrying out controlled experiments about how the universe works. Science has helped us answer questions as diverse as: how large is our planet? How old is the universe? What are the fundamental building blocks of matter?

The scientific method is the process by which scientists come to answer questions about the universe. It all starts with asking a question. What is light? What elements make up stars? How far away is the nearest galaxy to ours?

Science does a really great job answering questions we have that ask: “What…?” and “How…?” For example, how big is our planet? That is a question we can use the scientific method to solve. What is all matter made of? That is another question we can investigate using scientific principles.

Unfortunately, science is not always well suited to answering questions that ask: “Why…?” Why does our universe exist? Why do humans have consciousness? These are really great questions. Science may help us inch closer to an answer, but likely will not be able to provide any definitive answers. In these cases, philosophy and religion can help us to interpret and understand these deep questions.

Science is best used to answer questions that can be tested and validated experimentally. Consider the statement, “the Earth has a mass of 6×10^{24} kilograms.” That statement is only a hypothesis until it is experimentally validated by other scientists. If this hypothesis is verified by many other scientists, then it becomes a fact.

Experiments that seem to disprove a well-known fact can lead to revolutions in our understanding of the physical world. Up until about the late 19th century, it was commonly accepted that light was a wave. All of our observations and experiments verified that light acted exactly like a wave. At this time, scientists realized that the photoelectric effect, which had been observed for many decades, led to some conclusions that could only be explained if light also acts like a particle! This discovery helped usher along the newly formed quantum physics, and has led to technology such as light sensing photomultiplier tubes and some of the sensors used in digital cameras and night vision goggles.

However, experiments must be repeatable and verified by more than one scientist. Experimental errors, or misunderstandings of experimental results, can lead to setbacks in science. Whether these mistakes are malicious or not, scientists must remain skeptical of experimental evidence until it is verified by others.

From all of this, we can conclude that scientists must be not only open-minded and curious, but also exhibit healthy skepticism and fastidiousness!

Scientific terminology

One thing you may notice in the physical sciences is the use of proper scientific terminology. It is important to convey information in such a way that multiple readers can come to the same correct interpretation and understanding of a concept. This is important not only in the physical sciences, but in engineering and technology fields as well.

Sometimes you may find that there is a disconnect between the colloquial or casual use of a term, and the physics use of a term. This textbook will aim to point out these cases when they occur and give you a note of caution in using these terms in a classroom

setting!

A hypothesis is a descriptive statement about how the world works, based on observations. In order for a statement to be a scientific hypothesis, there has to be a way to prove the statement wrong. This hypothesis can be tested to determine whether or not it is valid. An example of a hypothesis would be: “The Sun is more massive than the Earth.” Any scientist can devise an experiment to test whether or not this hypothesis is true. If shown to be untrue, this hypothesis is discarded and a new one formulated to explain an observation or experimental result.

A scientific theory uses well-tested hypotheses as an explanation of how the universe works and is supported by multiple, repeated experiments. A good theory makes predictions that can be verified as true or false by experiment. If a theory offers no predictions that can be proven or disproven by experimental observations, then we have no way to determine whether or not it is valid.

There are many theories that explain our physical world. Plate tectonics, general relativity, and quantum mechanics are just three examples. These three theories have held up to much scientific rigor and scrutiny over many years. We may revise and edit theories as we learn more about the universe. Often an individual or a group of individuals dismiss them as “just a theory” (see Figure 1.1 for a humorous example), which discounts the vast amount of evidence that points to their accuracy. However, in doing so, the individual is generally confusing the term “theory” with “hypothesis”.



Figure 1.1 – Gravity is a theory that has been tested by multiple scientists and held up to much scientific rigor. This meme highlights the absurdity of calling gravity “just” a theory. Gravity, by the mad LOLscientist, is licensed under CC BY 2.0.

A scientific law is a statement that describes relationships among the quantities that we observe or measure. These laws can be used to predict the behavior of objects. Many times, laws use mathematical equations to explain and quantify these behaviors and phenomena.

Newton’s law of universal gravitation explains the motion of objects such as planets, stars, and satellites. We use an equation to quantify the degree of gravitational attraction between different objects: for example, to determine the gravitational force between yourself and the planet Earth.

The scientific method

How do we actually do science? How can we go from an observation to an understanding of how the universe works?

The scientific method that you may have learned about in school is usually presented as a straightforward set of tasks that, when completed, will result in an updated understanding of our physical world. That is not always the case!

Regardless of the set of steps we go through in our pursuit of science, the scientific method always starts with a question, or an observation that leads to a question. Why is the sky blue? What is an atom made of? How fast is light?

After asking our question, we then conduct background research. Has this question already been answered? If so, do we agree with the results? If the question has not already been answered, are other scientists working on answering this question? If not, are there similar questions that have been asked that might give us an idea of how to formulate a hypothesis or conduct an experiment?

Then we form a hypothesis. This can be an educated guess, or a more reasoned statement that attempts to answer our question. Our hypothesis should be testable using experimental methods. We may find ourselves reformulating or revising our hypothesis as we go through the next few steps in the scientific method.

Determining how to test a hypothesis can be a challenge. Some experiments may be simple and straightforward to conduct. Others may require expensive equipment or a specialized research laboratory to conduct.

Once the experimental setup and testing has been devised, we then conduct the experiment. This may require collaboration with other scientists. For example: a scientist may have a hypothesis that can be tested by using a particle collider, say, the one at Argonne National Laboratory, or at CERN.

After performing the experiment, the data that was generated by the experiment must be analyzed. If we're lucky, the data points to a clear conclusion that supports our hypothesis. Sometimes the data may be vague or unclear, which may require us to modify the experiment to collect more data. Maybe the data disproves our hypothesis! Some very important experiments have disproved incorrect scientific hypotheses. Famously, the Michaelson-Morley experiment in 1887 disproved the existence of some type of medium that permeates through all of space. We now know that light does not require a medium to propagate.

What if, instead of having clear data after our experiment, we got weird results instead? We need to determine if we made an experimental error, or if perhaps there was a problem with our hypothesis. We troubleshoot the problem – sometimes by conducting additional research – and then either reformulate the hypothesis, or create a new experiment to gather more data.

After we've completed our experiment, we communicate our results to the scientific community. Frequently this will be through a peer-reviewed scientific publication. Peer review means that other scientists in our field read through our work to double check our experimental procedures and data analysis. We may also decide to present our work to others at scientific conferences. These conferences can lead to new questions being asked, and collaborations formed to help answer them!

Once we've concluded our experiment, we can create technology based on our results. This does not always happen right away; sometimes it may take years to create technology based on scientific results.

Many times, the introduction of a new technology may generate new problems that require solutions using the scientific method. The internal-combustion engine is a wonderful technology, but it leads to greenhouse gas emissions. Nuclear power provides lots of energy to our homes, but generates hazardous waste. Even technologies such as computers, smart phones, and the Internet have led to issues involving the widespread propagation of misinformation and disinformation.

Sometimes, our knowledge of the universe is expanded upon by observations that occur serendipitously or by accident. For example: Penicillin was first discovered when a scientist's research was left out, causing mold to grow. It was discovered that this mold was capable of killing bacteria. Through the use of the scientific method, and the collaboration and communication between many scientists, the life-saving antibiotic drug was developed.

Science and technology

Science, as mentioned, is the pursuit of knowledge about how the universe works. Physics is a branch of science that focuses on the fundamentals of the workings of our universe. While physics sometimes overlaps with other fields, such as chemistry, biology, and even mathematics and materials science, we consider physics to be the fundamental science, as it provides a foundation for the other sciences.

Two contributors to this textbook, Dr. David R. Fazzini and Dr. Carley Bennett, are physicists. They received their Ph.D. degrees in physics. Much of Dr. Fazzini's research was conducted at Brookhaven National Laboratory and focused on condensed matter physics. Dr. Bennett's high energy physics research was conducted at the Large Hadron Collider and focused on dark energy.

Engineering (in contrast to physics) is the application of the science that is generated in physics, chemistry, biology, and other fields into the creation of new technology.

One of the authors of this textbook, Dr. Alyssa J. Pasquale, has a Ph.D. degree in electrical engineering. She is an engineer.

This means that she learned about physics, especially electricity and magnetism. This is science: explaining how the universe works, specifically as it pertains to electricity and electronics. She also took a lot of math classes such as calculus and differential equations to quantify those topics. In addition, she learned about and focused on the application of those scientific and mathematical principles to solving problems and creating new technology.

When you think about technology, what is the first thing that comes to mind? Computers? Electric vehicles? Spaceships and satellites? Those are all great examples of technology. Technology is what we get when we use scientific principles to solve problems. There are a lot of examples of technologies all around us: some of them are more high tech than others. Everything from the homes we live in, to the clothes we wear, to things as mundane as pencils and paper are technology. Over time, technologies improve as we learn more about the science behind them.

At times, science very quickly leads to technological innovations. Semiconductors were found to emit light, and this led to the invention of the light-emitting diode, or LED. These days, LED lighting has become affordable and practical enough for most of us to have LED lighting in our homes, flashlights, and even smartphones.

Other times, science may take years or even decades to evolve into practical technology. It took several decades between Einstein's formulation of the theory of relativity and a technology that uses it: GPS. General relativity is required to coordinate the timing between moving satellites and the humans on the Earth's surface who use them to accurately determine their position on the planet.

The scientific mind

There are many traits that a good scientist should possess. First and foremost: a scientist should be curious! Many people in science and engineering fields have been excited about learning how and why things work. This drive to learn more about the universe helps us to come up with new questions to solve, and new techniques that we can use to solve them.

Scientists should also be open-minded about how the universe works. One of the difficulties in teaching a physics class is that many students come to the class with many misconceptions about physics. And that's okay, you're still learning! However, be prepared to learn that the universe may work differently from how you currently think it does. It's possible that you just don't have all of the information you need to understand the full picture of how our universe works. This happens all the time in science!

On the other hand, a good scientist remains skeptical of scientific data until it has been held up to rigorous experimentation and scrutiny. You should trust, but verify, that information is correct. Hold things to a high standard of rigor. Do not believe what people in authority tell you just because they are in a position of authority: demand experimental validation and verification!

Scientists should also have integrity. They should not falsify their experimental data or observations. They repeat their experiments to confirm that they are reproducible. Good scientists should never mislead people about their discoveries or their impacts. Unfortunately, there have been many scientists throughout history who have purposefully misled the public about their results, who have plagiarized the work of others, or who have fabricated information to boost their careers.

Dishonesty and falsified research

Unfortunately, dishonest scientists have falsified research over the years, sometimes due to the high-stakes nature of scientific research funding.

Notably, in the late 1990s and early 2000s, physicist Jan Hendrik Schön falsified research on organic superconductors. His claims about the extraordinary conducting properties of this material could have had enormous applications in the development of lasers and other technology. However, his data could not be reproduced by other scientists. Eventually, Jan Hendrik Schön admitted that his data was incorrect. He had his PhD degree revoked, lost his job, and had all of his falsified research papers revoked by the journals that initially published them.

Other examples of falsified data include the 1989 claim by scientists Fleischmann and Pons to have discovered "cold fusion" (the generation of more energy from the fusion process than went into starting the fusion process in the first place), and the 2023 discovery that work by behavioral scientist Francesca Gino on dishonesty was either fabricated or falsified.

Many instances of falsified or fabricated data are discovered when the scientific discoveries cannot be reproduced by other scientists. This points to the importance of confirming the work of others (not to mention, the importance of scientific integrity in the first place)!

Further reading

- Cassuto, L. (2002, September 18). Big trouble in the world of "Big Physics." *The Guardian*. https://www.theguardian.com/education/2002/sep/18/science.highereducation
- Cold fusion. (2024, January 26). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Cold_fusion&oldid=1199313242
- Leif, U., Joe, &. (2023, June 17). *[109] Data Falsificada (Part 1): "Clusterfake."* Data Colada. https://datacolada.org/109
- Reich, E. S. (2009). *Plastic fantastic: How the Biggest Fraud in Physics Shook the Scientific World*. St. Martin's Press.
- Schön scandal. (2024, February 29). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Sch%C3%B6n_scandal&oldid=1211092898

Not only should scientists be careful in how they present their data, they should also think through the possible repercussions of their work. Being able to split the atom was an outstanding scientific innovation, but it also led to the creation of atomic bombs. It is irresponsible to society to disconnect science from its possible uses and outcomes.

When people misuse scientific phrases and terminology – usually to mislead others into purchasing a product that sounds "high tech" – we call this pseudoscience. This is scientific-sounding jargon that is usually meant to get you to part with your money to purchase something that has not been verified to work using the scientific method. Or, pseudoscience can relate to an idea about how the world works. Unfortunately, there are individuals and companies that exploit people who lack the scientific literacy to see through their claims.

Examples of pseudoscience

The following are all examples of pseudoscience. None of these concepts have been experimentally validated. Note that this is not an exhaustive list of pseudoscientific claims or concepts.

Ancient astronauts, aromatherapy, astrology, chromotherapy, climate change is a hoax, conversion therapy, creationism, crop circles designed by aliens, crystal healing, dowsing, flat Earth, ghosts, homeopathy, levitation, the Moon landing was fake, numerology, perpetual motion, phrenology, quantum mysticism, and scientific racism.

To identify pseudoscience, ask a few questions: Does the idea or product make use of vague, exaggerated, or unverifiable claims? Is the data behind the idea or product open and has it been tested by other scientists? Does the idea or product ignore compelling scientific evidence? Does the person or people behind the idea or product attempt to suppress others who dissent or show evidence to the contrary?

For example, astrology, the belief that we can make predictions about our lives based on the positions and movements of planets and other celestial bodies, has no basis in science. No evidence has been found to support astrology, and all falsifiable predictions made with astrological methods have been falsified using the scientific method.

Unfortunately, there are many areas of pseudoscience in areas ranging from physics to medicine, history, and psychology and sociology. At best, these products and ideas cost us money. At worst, they are harmful, toxic, perpetuate racism, and cause other societal problems.

The mathematics of physics

In physics, we frequently need to quantify our discussions of physical phenomena. This means we want to be able to express different values numerically. We use math to put numbers to our discussions.

Each physical quantity that we discuss will have a symbol associated with it. Every time we learn something new, that symbol will be defined so that you know what it means. For example: mass has a symbol of the lowercase letter m. Temperature has a symbol of the capital letter T.

Sometimes the symbols use the English alphabet. Sometimes they will use Greek letters. For example, the concept of wavelength tells us the distance between two identical points on a wave. The symbol for wavelength is the lowercase Greek letter lambda (λ).

Many quantities just convey a numerical quantity: a value or a strength. We call these quantities scalars. Sometimes a quantity needs to express both a strength and a direction. These quantities are known as vectors. Vectors have an arrow above the symbol (example: $\vec{F}$), or are expressed using a bold font (example: $\boldsymbol{F}$).

All this is to say that understanding that symbols relate to actual physical quantities is important. Different physical quantities relate to each other by means of mathematical equations. These equations are frequently derived from a physical law, for example, Newton's laws of motion. We will learn about each equation as it is defined in this textbook, and what that equation tells us about the physical world.

Sometimes we may find that a few physical quantities use the same symbol. Specific heat capacity is defined by the lowercase letter c. But so is the speed of light! Capital V can stand for either volume or voltage. How do we know which is which? The same way we learned the meaning of the word "bat." Does that refer to the animal? Or does that refer to the stick you use to hit a baseball? It all depends on context. When we see a symbol in an equation, we look at the other symbols nearby to help us determine the meaning, just as we read other words in a sentence to determine the meaning of the word "bat."

If all of this seems overwhelming right now, that's okay. Just as you did not learn to speak or read the English language overnight, learning the mathematics of physics is also a process. Be patient with yourself, but also know that all things improve with practice.

One last thing – all of the physical quantities we use in this textbook are not only defined by a symbol, but also have units that we use to express those values. One great example is temperature. Temperature has a symbol of the capital letter T. But if you are told that the temperature is 10, that's not enough information to determine what to wear when you go outside. You also need to know the units (in that example: Celsius or Fahrenheit).

In physics, we will use SI units, known as the international system of units. These units are internationally used in physics and engineering. They differ from the English units you may be used to. We use meters as the SI unit of distance, not feet or inches. We use kilograms as the SI unit of mass, not pounds or slugs. Whenever possible, this textbook will attempt to give you some context for these units, as they may be new for you.

Why are units important? Just as scientists need to use precision in our language and terminology, so do we need to use precision in our quantities. It would be inconvenient to be told that it's 10 degrees outside, so you decide to put on several layers of sweaters and jackets, only to realize that it's 10 degrees Celsius outside, not 10 degrees Fahrenheit! That example may not be a big deal, but incorrect units used in dosing medication or medical therapies could have a fatal outcome. In 1999, an orbiter was sent to Mars. Rather than landing on the surface, it crashed, creating a very expensive crater on the surface of the red planet. The crash occurred because some of the team members used English units (pounds), while other members used SI units (newtons). Units matter! Get into the habit of using units all the time in your practice.

Practice questions

Conceptual comprehension

1. What is the difference between a scientific theory and a law?
2. Einstein's theory of gravity (commonly known as general relativity), formulated in the early 20th century, found some conceptual flaws with Newton's law of universal gravitation. Does this mean that Newton's law of universal gravitation is wrong and should be discarded? Why or why not?

This page titled Chapter 1: About science is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

II: KINEMATICS

Chapter 2: Newton’s first law
Chapter 3: Linear motion
Chapter 4: Newton’s second law
Chapter 5: Newton’s third law
Chapter 6: Momentum
Chapter 7: Work and energy
Chapter 8: Rotational Motion
Chapter 9: Gravity
Chapter 10: Projectile and satellite motion

This page titled II: Kinematics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 2: Newton's first law

Summary

Newton's first law of motion describes the property of inertia, which defines the resistance objects have to changes in their motion, whether that change is to speed up or slow down the object, or to change its direction. All objects of mass will continue to move at a constant speed, in the same direction, unless acted upon by an outside force. If an object of mass is not moving, it will stay at rest unless pushed or pulled by an outside force.

Inertia

Newton's first law is sometimes referred to as the law of inertia, as it describes this property of matter. Newton's first law says that an object in motion tends to stay in motion, and an object at rest tends to stay at rest, unless acted upon by an outside force.

Inertia describes how difficult it is to change the motion of an object. If an object is at rest, inertia relates to how hard it is to get the object to start moving. If an object is moving in a straight line at a constant speed, inertia relates to how hard it is to get the object to speed up, slow down, or change its direction.

Consider the three objects shown in Figure 2.1. They appear to have similar shapes and sizes. How could we determine the inertia of each object? If inertia relates to the difficulty of getting something to change its motion, then we can figure out which, if any, of these objects has the most inertia by picking each one up and measuring the difficulty in getting each object to move.

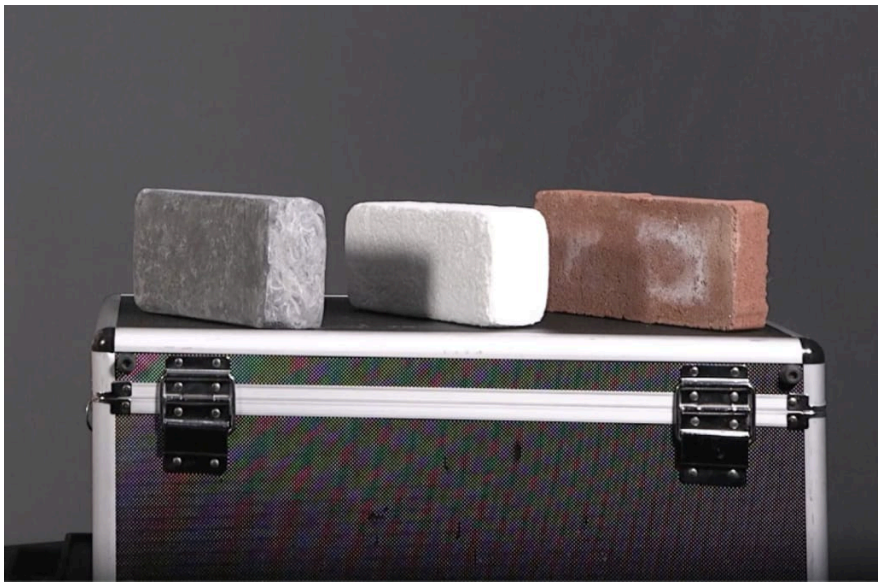


Figure 2.1 – From left to right: a lead, styrofoam, and clay brick sitting on top of a chest. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale holds a piece of styrofoam and shakes it back and forth as many times as she can in five seconds. This object was not too difficult for her to move around: Dr. Pasquale was able to move it back and forth 14 times in five seconds.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In the next video (below), Dr. Pasquale will move a clay brick back and forth as many times as she can in five seconds. This brick was more difficult to move than the styrofoam. It had a greater resistance to changes in motion, meaning that it has more inertia. Dr. Pasquale was able to move it back and forth seven times in five seconds.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Finally, in the third video (below), Dr. Pasquale will repeat the experiment one more time with the last object, a piece of lead. This object was very difficult to move. In fact, it was hard even for Dr. Pasquale to pick it up from the cart it was sitting on. She was able to move the lead brick back and forth only two times in five seconds.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

What is the difference between the three objects? What property of matter makes the styrofoam easy to move and the lead hard to move? It can't be volume, because the sizes are all the same. The difference is in the mass of the objects.

The tablecloth trick is a great demonstration of inertia. In the video below, you can watch Dr. Pasquale perform this demo. A table setting consisting of a plate, bowl, glass, vase with a flower, napkin, and silverware is placed on a silk tablecloth. This setup, viewed from top-down, is shown in Figure 2.2. Dr. Pasquale very quickly pulls the tablecloth out from under the table setting.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

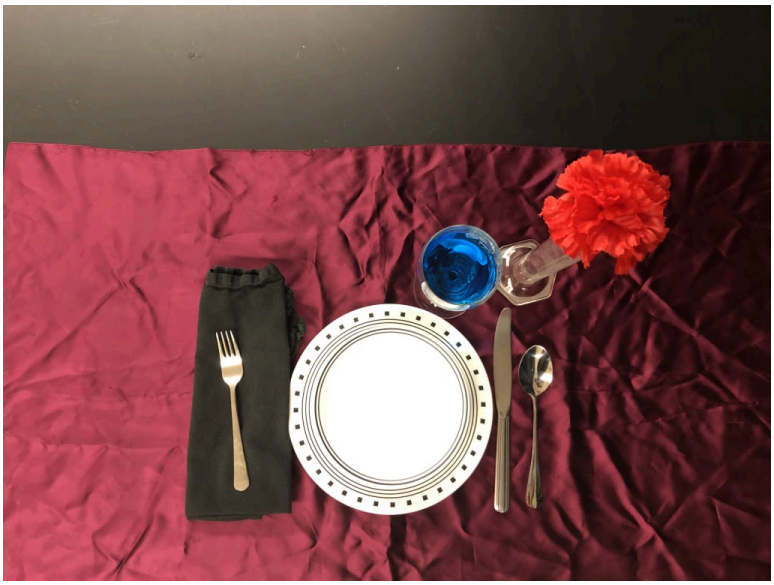


Figure 2.2 – A table setup prior to removal of the tablecloth. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

After Dr. Pasquale quickly pulled the tablecloth out from under the objects, most of the objects stayed put. The top-down view of the table setting after the tablecloth was removed is shown in Figure 2.3. The plate, the flower, and the glass of water pretty much stayed where they were. However, there was a little bit of force present that made the silverware move around a little bit. Inertia relates to the mass of an object. The silverware has the least amount of mass, so it makes sense that they would be the most affected by this demonstration.



Figure 2.3 – A table setup after the abrupt removal of the tablecloth. The smaller items have slightly shifted. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Inertia tells us that unless a force acts on an object to change its motion, it will keep doing what it was already doing. While we have not yet formalized the concept of forces, for the purposes of this discussion, a force is nothing more than a push or a pull. Perhaps you are aware of the force of gravity. Gravity acts on each of the objects sitting on the table: the plate, the silverware, the flower in the vase, and the glass of blue water. A support force counteracts gravity and holds each object up.

When the tablecloth is very quickly removed, no new forces are being applied to any of the objects, other than the tablecloth itself. Gravity and the support forces are still present, and there is not much, if any, horizontal force present when the tablecloth is pulled out. Therefore, most of the objects stay put.

Inertia does not only apply to objects that are at rest. It also applies to objects in motion. An object in motion will continue moving in a straight line at a constant speed unless acted upon by an outside force.

In the video below, a ballistic cart contains a small yellow ping-pong ball. As the cart moves past a trigger on the track, it launches the ball into the air. Because the ball and the cart were both moving at the same horizontal speed when traveling together, they will continue to move at the same horizontal speed even if they are separated. Unless a force acts on either object to slow it down or speed it up horizontally, the object will continue moving at the same horizontal speed. After launching into the air, the ball and cart have the same horizontal speed. That means that once the ball comes back down again, it will land right back into the cart. While the ball is in the air, the downward force of gravity does act to change the motion of the ball, but only in the vertical direction. On the way up, gravity acts to slow the ball's vertical motion. On the way down, gravity acts to speed up the ball's vertical motion.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This relates to the motion of any objects that are moving together. If you are riding in a train car on a flat straight track, you could throw something straight up in the air and expect to catch it straight back down, as long as the train doesn't speed up or slow down while the object is in the air.

However, if the train speeds up after you throw a ball into the air, the ball will continue moving at its initial horizontal speed as the train goes faster. The ball will land behind where you initially threw it. If the train slows down after you throw a ball into the air, the ball will land ahead of where you initially threw it.

Inertia also helps to explain why we wear seatbelts in cars. If we are riding or driving in a car, then we are moving at the same speed as the car. If the car is stopped abruptly by an outside force, our bodies naturally want to continue moving at the same speed in the same direction. Without a seatbelt, a collision with another object could cause us to be thrown out of the windshield of a car. Even braking a car very forcefully could otherwise cause us to move forward dangerously out of our seats if we didn't have a seatbelt on. Seatbelts are important safety devices that prevent that from happening. They apply a force to us to ensure that we slow down or stop moving when the car abruptly stops.

Mass

Mass is a property of physical objects that relates to relates to inertia (resistance to changes in motion). Mass relates to how much "stuff" is in an object. Mass has a symbol of the lowercase letter m, and we use the unit kilograms (kg).

It's important to understand that, for most physical objects, mass is a characteristic that does not change. It does not change over time, and it does not change based on where the object is located: the Earth, the Moon, Jupiter, or in the vacuum of space.

The styrofoam brick shown in Figure 2.1 and demonstrated in the video above has a mass of 0.04 kg, the clay brick has a mass of 2 kg, and the lead brick has a mass of 12 kg. Those masses would be exactly the same if they were to be measured on another planet. The inertia of these objects is independent of their location.

The situation for humans is a little different. Our mass can change, but that's because we eat, exercise, cut our hair, and otherwise do things that may cause our mass to fluctuate. However, if we could somehow keep our mass at a constant value, that number would not change if we then traveled to the Moon, or to Mars, or to any planet or celestial body with a different gravitational force than the Earth.

In other words, while mass is related to a concept called weight, it is a fundamentally different thing! Mass has to do with our resistance to changes in our motion – our inertia. Weight is a force that comes about due to gravity.

Force

Simply put, a force is a push or a pull that causes an object to change its motion. The symbol we use for force is the capital letter F. Because force is a vector, it will either appear symbolically with bold font ($\boldsymbol{F}$) or with an arrow over the top ($\vec{F}$). A vector is a quantity that contains both a magnitude and a direction. That means that the direction in which a force is applied is just as important as how strong the force is.

Read more!

Appendix A in this textbook contains a full description of vectors, including how they are represented and drawn. It also describes how to add vectors both graphically and numerically. If you are not already familiar with using vectors, read Appendix A before moving along in this textbook.

The units used for force are newtons (N). A newton is a unit that has other units "hidden" inside it. A newton is equal to a kilogram times a meter divided by a second-squared:

$$1\ \mathrm{N} = 1\ \mathrm{kg\ m/s^2}. \tag{2.1}$$

It is possible to use either of these units to express force, but it's much easier to use the unit of newton, which is why it is used so often in physics.

Forces are used to overcome inertia and make an object change its motion. The amount that an object changes its motion due to a force is the topic of Newton's second law.

It can be difficult to get a full appreciation for the concept of inertia, as we experience so many forces while living on the planet Earth. Gravity and friction constantly impact the motion of objects on this planet. When you push something, it eventually comes to rest. That may seem to go against the law of inertia, but it doesn't. By accounting for all of the forces acting on an object's motion, we can really start to appreciate Newton's first law.

Right now, as you are reading or listening to this textbook, there are forces acting on you. One of those forces is gravity. Gravity is a force that pulls you toward the surface of the Earth. Another force that is acting on you is a support force from the floor, or chair, or bed, or whatever surface you are sitting, lying, or standing upon. The support force holds you up and prevents you from falling to the center of the Earth. These two forces are likely in perfect balance with each other, which is why you are probably not accelerating either up or down right now (unless you happen to be jumping at the moment!).

Mechanical Equilibrium

By analyzing the sum total of all forces acting on an object, we can determine whether or not that object will experience a change in its motion. When the sum of all of the forces acting on an object is equal to zero, then that object is said to be in mechanical equilibrium.

In physics, we use the capital Greek letter Sigma (Σ) to define "the sum of." Therefore, $\Sigma\vec{F}$ means "sum of all forces." We can add together each force to calculate that sum. Sometimes we call this quantity the net force.

Mechanical equilibrium defines a condition in which the sum of all forces acting on an object is zero. So by saying that the net force is zero, we are saying that the sum of all forces is zero. When the net force on an object is zero, that object is in mechanical equilibrium. As a result, the object will have a constant velocity. That is, the object will have a constant speed along a straight line.

Mathematical representation of mechanical equilibrium

Mechanical equilibrium is achieved when an object has no net force acting on it. That can be mathematically represented using the notation in the following equations:

$$\sum \vec{F} = 0$$

$$\vec{F}_{net} = 0$$

There are two types of mechanical equilibrium: static equilibrium and dynamic equilibrium. Static equilibrium defines a condition of mechanical equilibrium where an object is at rest. The velocity remains constant at zero.

Static equilibrium

In Figure 2.4, a cart sits at rest on a flat, low-friction track. There are forces acting on the cart, but they are completely balanced, and because the cart is not moving, it will continue to not move unless that condition of equilibrium is disturbed somehow by a push or a pull.

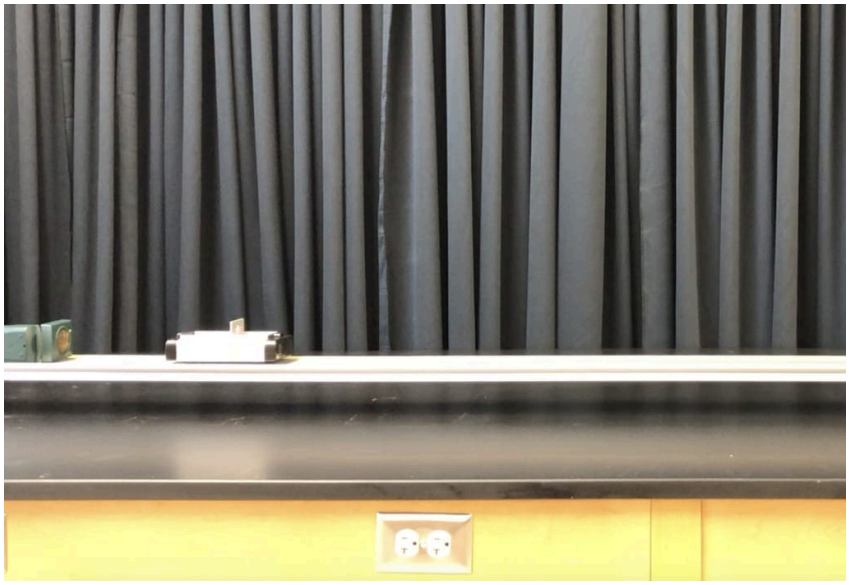


Figure 2.4 – A low-friction demonstration cart sits at rest on a low-friction track. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The two forces acting on the cart are gravity, which pulls the cart to the Earth, and a supporting force from the track, keeping it in position on the track. These two forces are equal to each other, and point in opposite directions. Therefore, the net force is zero and the cart is in static equilibrium.

Dynamic equilibrium

Dynamic equilibrium defines a condition of mechanical equilibrium where an object moves at a constant nonzero velocity.

In the video below, a cart has been pushed until it travels at a constant speed along a low-friction track. After Dr. Pasquale's hand stops pushing the cart, the only two forces acting on the cart are gravity and the support force, which remain equal and opposite, creating a net force of zero. This means that the cart will continue to move at the same speed, in the same direction. (In reality, a very small friction force between the wheels and track, as well as a force of air drag as the cart pushes through the air molecules surrounding it, will cause the cart to slow down over time.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Another example of dynamic equilibrium would be an airplane flying in a straight line at cruising altitude at a steady speed (by using autopilot, for example). Forces in the vertical direction (gravity and the lift) are equal in magnitude and opposite in direction and cancel out. Forces in the horizontal direction (an applied force from the engine and the forces of friction and air drag opposing the motion of the airplane) will also cancel out.

Without a force that propels an object to keep it in motion, it can be difficult to experience a condition of pure dynamic equilibrium on Earth. That's because there are usually always other forces we have to contend with, notably air drag and friction. By minimizing those forces as much as possible, we can get close to a situation of dynamic equilibrium.

Practice questions

Conceptual comprehension

1. Let's say you were to push against a parked car. The push exerts a large force, but not enough to make the car move. Because the car is not moving, it is in mechanical equilibrium. Is your push the only force acting on the car? If not, what other forces can you identify that are acting on the car?
2. Look around you and identify at least five items that are in static equilibrium. What are all of the forces acting on those items?
3. As you go throughout your day, identify as many objects that you can that are in dynamic equilibrium. What are all of the forces acting on those objects?

Numerical analysis

1. A box sitting on a flat surface has a gravitational force of 10 N and a support force of 10 N. What is the magnitude of the net force acting on the box? Is the box in mechanical equilibrium?
2. A person rowing a boat in water exerts a force of 45 N to get the boat to speed up. Drag from the water exerts a force in the opposite direction of 15 N. What is the magnitude of the net force acting on the boat and its occupants? Is the boat in mechanical equilibrium?
3. An object has three forces acting on it. A force of 8 N pushes to the right, a force of 3 N pushes to the left, and a force of 2 N pulls upward. Draw a picture of all of the forces. What is the overall horizontal component of the net force? What is the overall vertical component of the net force? What is the magnitude of the net force?
4. An airplane has four forces acting on it. There is 4500 N of lift (up), 4000 N of weight (down), 500 N of thrust (forward) and 100 N of drag (backward). Draw a picture of all of the forces. What is the overall horizontal component of the net force? What is the overall vertical component of the net force? What is the magnitude of the net force?

Hands-on experiments

1. Place a marble in the center of a plate on a flat surface. Try to tilt the plate to make the marble move. Observe how the marble tends to stay in place until you apply an external force to the plate.
2. Place a piece of cardstock paper (or cardboard) over the opening of a cup and put a coin on top of it. Swiftly flick the paper out from under the coin. The coin should fall into the glass because of its inertia. This is demonstrated in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

3. Place a piece of paper (or a dollar bill) over the opening of a glass bottle and put a coin on top of it. Remove the paper as quickly as possible from under the coin. The coin should stay on the top of the bottle because of its inertia. This is demonstrated in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This page titled Chapter 2: Newton's first law is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 3: Linear motion

Summary

Linear motion allows us to start exploring kinematics, which is the study of the motion of objects. The topics in this chapter will allow us to answer the questions of how far something goes, how fast it goes, and how quickly it speeds up or slows down in the process. These topics will enable us to study the motion of objects traveling at a constant velocity and also objects that are moving under the influence of gravity or some other force.

Path and displacement

The concepts of path and displacement enable us to analyze how far an object moves as it travels. Path describes the total length of the motion of an object as it moves from one point to another, measured along its trajectory. It is a scalar quantity. The symbol that we use could be the lowercase letter d, the lowercase letter x, or the lowercase letter y. We typically use the letter d when discussing the concept of distance. We use the letter x when discussing a horizontal motion. We use the letter y when discussing a vertical motion. Either way, the units we use for path are meters (m).

To demonstrate the concept of path, as depicted in Figure 3.1, a person starts at a position of 0 m (known as the origin). They walk at a mostly constant pace until they reach a position 4 m to the right of the origin. The person's path describes the distance that they walked, which is equal to 4 m.

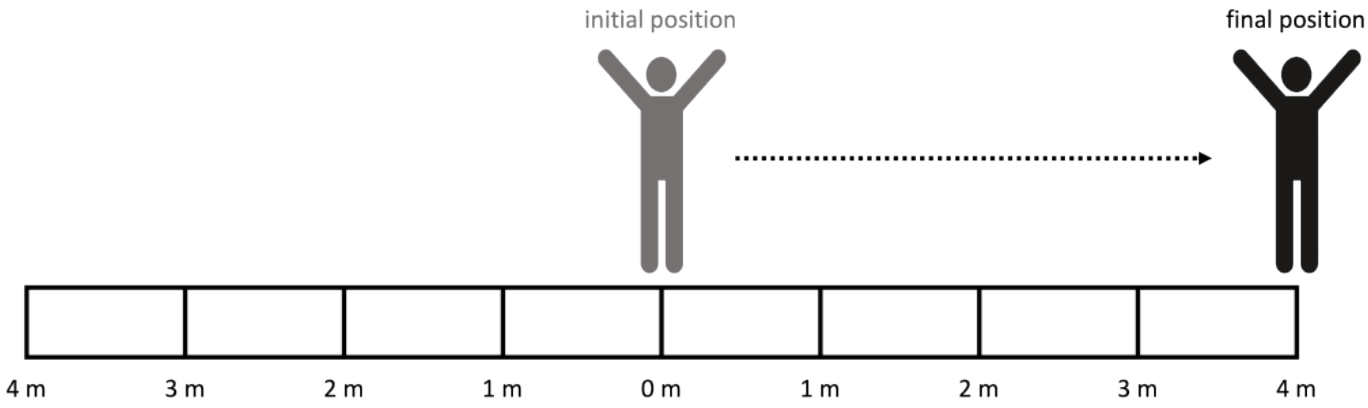


Figure 3.1 – A person starting at position 0 moves to a position 4 m to the right of the origin. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Displacement describes the straight-line distance from one point to another. It is a vector quantity, so it must include direction. The symbol that we use could be $\Delta\vec{d}$, $\Delta\vec{x}$, or $\Delta\vec{y}$. The units of displacement are meters (m).

The uppercase Greek letter Delta (Δ) means "change in." Any physical quantity that can change has a final value and an initial value. A quantity starts at its initial value, and after a change is applied, ends up at the final value:

$$\text{initial value} + \Delta = \text{final value}.$$

This equation can be rearranged by subtracting the initial value from both sides of the equation to allow us to solve for the change:

$$\Delta = \text{final value} - \text{initial value}.$$

Therefore, every time an expression contains the symbol Δ, it can be solved by subtracting the initial value of a quantity from the final value of that same quantity. Because Δ is used in the symbols of displacement, it indicates that displacement is equal to an

overall change in position between two points (a start and an end) and is equal to the final position minus the initial position.

Consider Figure 3.1 again. After starting at a position of 0 m and moving 4 m to the right of the origin, the person then moves to a position 2 meters to the left of the origin. This is shown in Figure 3.2.

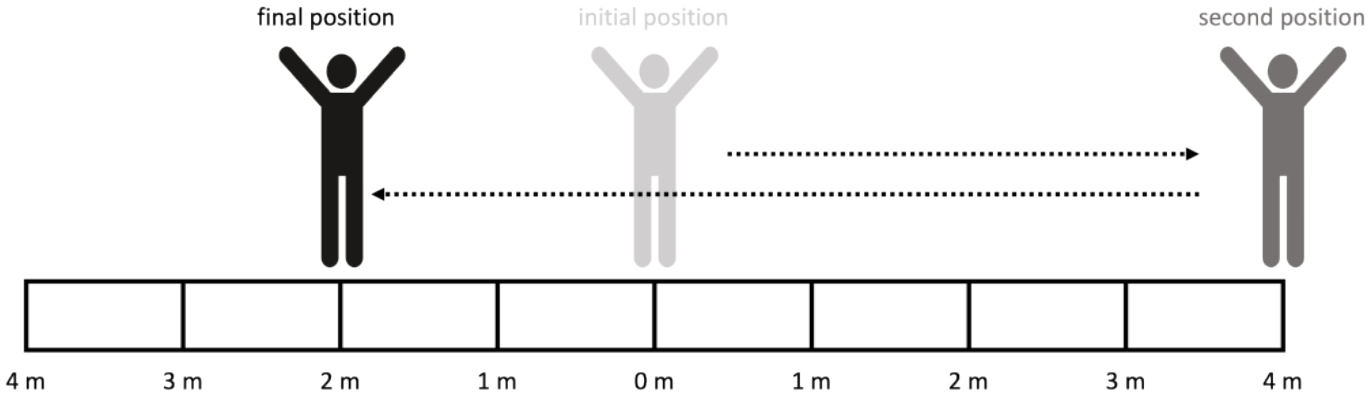


Figure 3.2 – After starting at the origin and moving 4 m to the right of the origin, a person then moves 2 m to the left of the origin. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The path of the person after moving between these locations is the total distance that the person walked. From the origin to 4 m the person walked 4 m. From 4 m to the right of the origin to 2 m to the left of the origin, the person walked an additional 6 m. In this example, the total path is 10 m.

Displacement, on the other hand, only considers the change in position from the initial to final locations. The initial position is 0 m. The final position is -2 m (negative because it is to the left of the origin).

$$\begin{aligned}\Delta\vec{x} &= x_{final} - x_{initial} \\ &= -2\text{ m} - 0\text{ m} \\ &= -2\text{ m}\end{aligned}$$

If, after all of this walking around, the person were to return to the origin, the total path would be 12 m (the first 4 m walking to the right, 6 m walking to the left, and then 2 m walking back to the origin). The total displacement would be zero. This is because the initial and final locations are the same.

To further demonstrate the concepts of path and displacement, consider driving in a car from Naperville, Illinois, to the College of DuPage (COD). When driving a car, it isn't always possible to go in a direct straight line. Instead, somebody driving a car has to follow the roads. This route describes the path of the car, and is shown in Figure 3.3.

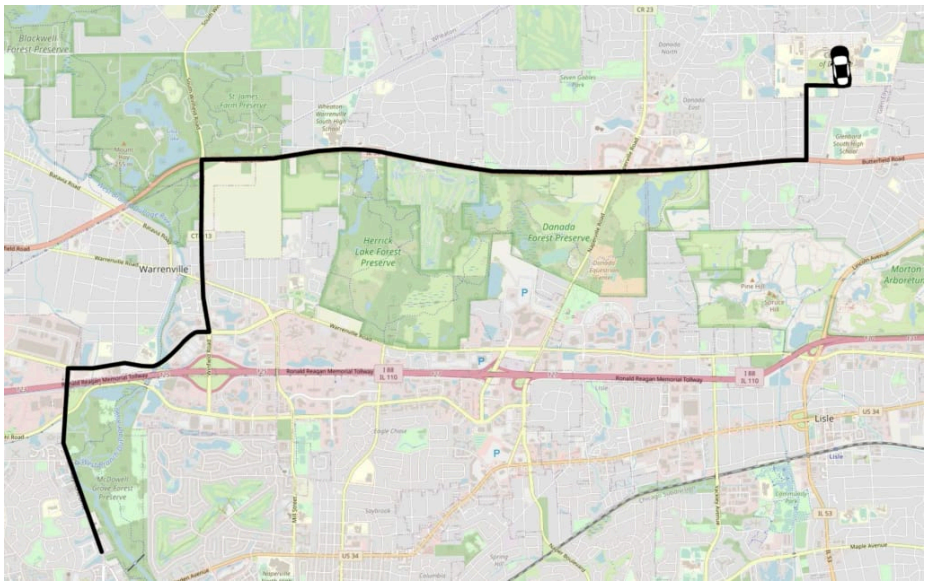


Figure 3.3 – A path from Naperville, Illinois, to the College of DuPage, along roadways. This figure, created by Alyssa J. Pasquale, Ph.D. is licensed under CC BY-NC-SA 4.0. Map attribution: OpenStreetMap, licensed under the Open Database License.

The displacement between Naperville and the college only considers the straight-line distance between the starting and ending locations. Regardless of the path that's taken to move between the two points, the displacement remains the same. It is always equal to the vector pointing from the starting point to the finishing point. This displacement vector is depicted in Figure 3.4. The displacement vector points to the northeast.

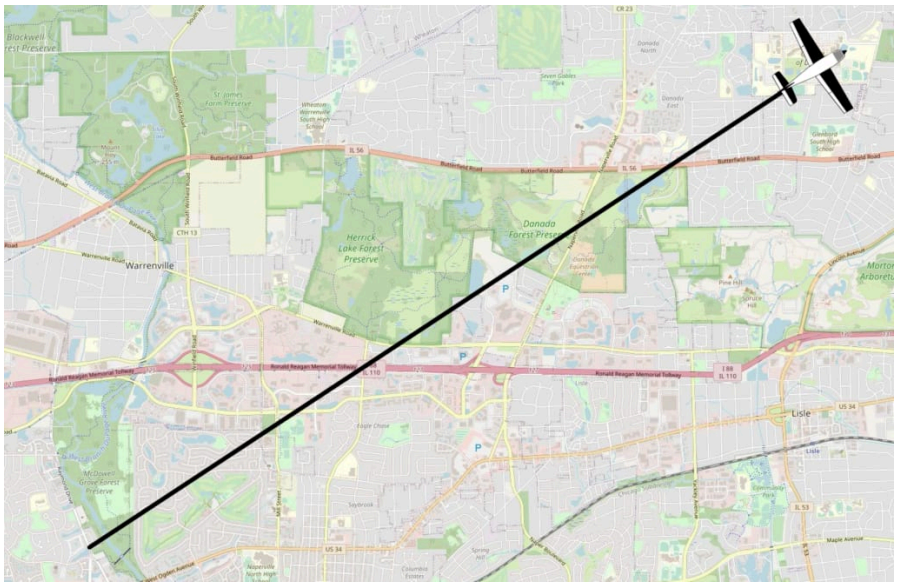


Figure 3.4 – The overall displacement from Naperville, Illinois, to the College of DuPage. This figure, created by Alyssa J. Pasquale, Ph.D. is licensed under CC BY-NC-SA 4.0. Map attribution: OpenStreetMap, licensed under the Open Database License.

When driving in a car, you can determine your path by taking a look at your odometer. Unless you are able to drive in a completely straight line, it's slightly more difficult to calculate the displacement as you drive from one place to another. However, if you know the coordinates of the starting and ending location, you can draw a straight line between two points. By splitting the straight line into horizontal and vertical components, you can use the Pythagorean theorem to calculate the displacement. The direction of the displacement can be determined by using a compass.

Speed and velocity

The concepts of speed and velocity enable us to analyze how fast (or slowly) an object moves as it travels. Speed is the scalar

quantity that describes the rate at which an object is covering distance. The symbol used can be the lowercase letter s, or the lowercase letter v in absolute value sign: $|\vec{v}|$, indicating that we are looking at the magnitude of the velocity. The unit of speed is meters per second (m/s).

Velocity is the vector quantity that describes the rate at which an object changes its position. Because velocity is a vector, it's either denoted in bold font or with an arrow over the top. The symbol is the lowercase letter $\vec{v}$. The unit of velocity is also meters per second (m/s).

Both speed and velocity can be expressed as averages or as instantaneous values. That is, we can express how fast an object moves on average from the beginning to the end of its motion. That would be the average speed or velocity. Average speed has an equation of path divided by time. Average velocity has an equation of displacement divided by time. Note that the average speed is not merely the magnitude of the average velocity because path and displacement generally have different values (unless the motion is a straight line in one direction).

The instantaneous value tells us how fast an object is going at any given snapshot in time. This might be the same as the average value, or it might be different. Instantaneous speed and velocity can either be measured or calculated using an equation.

Average speed

Consider the drive (path) from Naperville to COD, shown in Figure 3.3. The average speed is going to be equal to the path divided by the amount of time it took to drive that distance. Dr. Pasquale drove this distance in her car and recorded the path distance using her odometer and measured the time using a stopwatch. The path was 15.6 km, and the time was 20 min (1/3 hour).

The average speed is 46.8 kph (kilometers per hour). If you're not comfortable with units of kph, this is equal to an average speed of 29 mph (miles per hour). However, as Dr. Pasquale drove the path from Naperville to COD, her speed was not always 46.8 kph! At times the speed was faster, at times slower, and at times she was stopped at a traffic light or stop sign. Average speed accounts for all of the time spent driving at all speeds (including time spent stopped, where speed is zero).

Instantaneous speed

This is where we see a big difference between average speed and instantaneous speed. The instantaneous speed at any time during the journey can be determined by the readout on the car's speedometer at that moment. At times it could be as high as 72 kph, and at times it could be as low as 0 kph. The average and instantaneous speeds are not necessarily going to be equal!

Average velocity

The displacement of the drive from Naperville to COD was 11.1 km northeast. (This value was determined by using an online distance-mapping website.) The amount of time it took to drive between the two points was 20 min. Therefore, the average velocity of this trip was

Average velocity tells us, on average, what the velocity was throughout the motion of an object. The displacement is always going to be less than or equal to the path of an object, as the shortest distance between two points is a straight line. Therefore, the average velocity will always be less than or equal to the average speed of an object. That hopefully makes sense intuitively, as the shortest distance you can get from one point to another would be to travel in a straight line. Any other way you travel, that is not a direct path, is going to add extra time to your motion and slow you down.

Instantaneous velocity

The average velocity for the trip from Naperville to COD is the average over all of the motion. At any given snapshot in time, the instantaneous velocity would be indicated by the current speed and direction. Not all time is spent traveling at 33.3 kilometers per hour northeast! Sometimes the car is stopped and has an instantaneous velocity of zero. Sometimes the car is moving faster, or may be pointed in a different direction.

In general, as you drive or ride around in a car, you can determine your instantaneous speed by taking a look at your speedometer. If you want to know instantaneous velocity, look at your speedometer and a compass to determine what direction you're moving in.

After driving from Naperville to the College of DuPage, Dr. Pasquale eventually drove back to her starting location in Naperville. Then, she drove the same exact route back home and the return trip took 30 minutes. We can analyze the path, displacement, average speed, and average velocity of the round trip.

path	31.2 km
displacement	0 km
average speed	37.44 kph
average velocity	0 kph

Note that the vector quantities, displacement and average velocity, are zero. This is because the starting and ending locations are the same!

Acceleration

Acceleration quantifies the rate at which an object changes its velocity. The symbol for acceleration is the lowercase letter $\vec{a}$. The units for acceleration are meters per second-squared (m/s^2). The acceleration of an object is equal to the change in velocity (final velocity minus initial velocity) divided by time:

$$\vec{a} = \frac{\Delta \vec{v}}{t}.$$

We have to be very careful when using the word acceleration in a physics class, as it differs from the definition of "acceleration" that you may be used to using. Since the velocity of an object can change in different ways (get faster, get slower, change direction), the physics definition of acceleration can mean three things: speeding up, slowing down, and changing direction.

You could therefore look at something that is slowing down, and say "that object is accelerating," and you would be correct. However, in colloquial English, people tend to use the word acceleration only to refer to objects that are speeding up. We have to be careful and precise with our words in physics.

For the most part, this textbook will deal with situations where the acceleration is constant throughout an object's motion. That means that instantaneous and average acceleration will be equal to each other. (It is important to point out that this is not true for Dr. Pasquale's drive between Naperville and COD, but it would be too complicated to consider an analysis of her acceleration in that scenario, so we will not do so.) We will consider simpler examples (where instantaneous and average acceleration are equal) to learn about acceleration.

Linear motion when acceleration is zero

In the video below, a cart travels along a flat, low-friction track. After it is initially pushed, the cart moves without any net force acting on the cart. (The vertical forces of gravity and the support force cancel out.) Because there is no force acting on the cart to

change its speed, it has a constant speed throughout its motion.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A motion detector was used to record the position, velocity, and acceleration of the cart throughout its motion while it was moving on its own with no net force (after Dr. Pasquale finished pushing it). Position, velocity, and acceleration graphs for this motion are shown in Figure 3.5. (Download this data [XLSX, 11 kB])

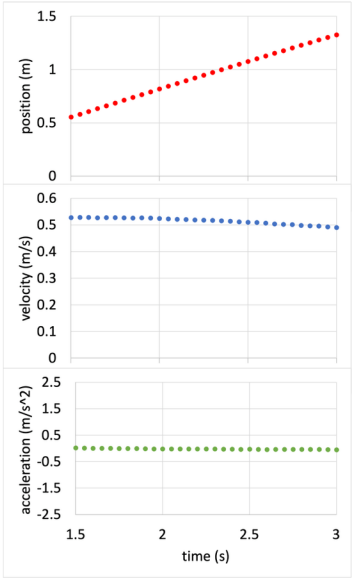


Figure 3.5 – Graphs demonstrating the change in position, velocity, and acceleration of a cart over time as the cart moves along a flat track. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The position graph shows a fairly straight line pointing upward, indicating that the cart increases its position at a steady rate. The velocity graph is also a fairly straight, (approximately) horizontal line, indicating that the velocity stays (about) the same throughout the cart's motion. The speed at the start is slightly above 0.5 m/s, and the speed at the end is 0.5 m/s. (The speed is not completely constant due to a small amount of friction and air drag.) The acceleration is therefore very close to zero m/s^2. This is confirmed by the acceleration graph, which shows a flat line at zero.

Accelerated motion

Consider an example that shows all three aspects of acceleration: speeding up, slowing down, and changing direction. As shown in the video below, Dr. Fazzini pushes a cart so that it starts moving up a ramp. As the cart travels up the ramp it slows down. At the point farthest up the ramp it stops for a brief instant, changes direction, and travels down the ramp, speeding up as it gets back to the end of the track.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Position, velocity, and acceleration graphs for this motion are shown in Figure 3.6. (Download this data [XLSX, 12 kB])

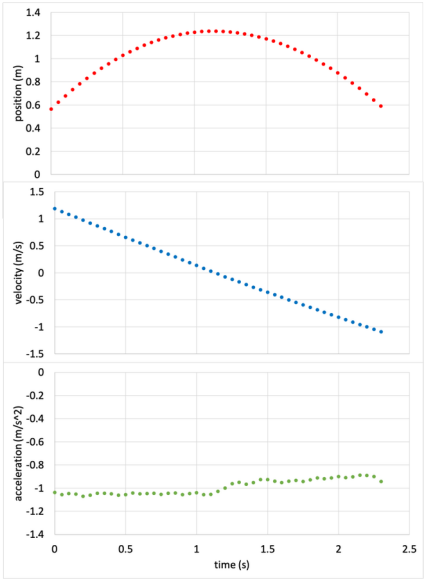


Figure 3.6 – Graphs demonstrating the change in position, velocity, and acceleration of a cart over time as the cart moves up and then down an inclined track. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The position graph is a curved trend, resembling a kind of bowl shape, called a parabola. This indicates that the cart changes its position at a decreasing rate as it moves up the ramp, then changes its position at an increasing rate as it moves down the ramp.

The velocity graph is a straight line pointing downward, showing that the velocity decreases at a constant rate. The velocity starts out as large positive numbers, decreases in value, becomes zero briefly when the cart is at the top of the ramp, and then decreases and becomes larger negative values. The speed at the start, after being pushed, is 1.2 m/s, and the speed before being caught after 2.3 s is -1.1 m/s.

Using the equation for acceleration defined above, we can calculate the expected value of acceleration from the velocity data as

$$\begin{aligned}\vec{a} &= \frac{\vec{v}_{final} - \vec{v}_{initial}}{t} \\ &= \frac{-1.1 \text{ m/s} - 1.2 \text{ m/s}}{2.3 \text{ s}} \\ &= \frac{-2.3 \text{ m/s}}{2.3 \text{ s}} \\ &= -1 \text{ m/s}^2.\end{aligned}$$

In Figure 3.6, the acceleration graph shows a (mostly) straight horizontal line at approximately -1 m/s^2, confirming that the acceleration is (mostly) a constant value throughout the motion of the cart. It is worth noting that there is a small change in the average acceleration value after the cart turns around and heads back down the inclined track again. This is because there is a small amount of friction involved. While the cart travels up the ramp, the forces of gravity and friction both act to slow the motion of the cart up the ramp. That is, gravity and friction act together to slow the cart. However, when the cart travels down the ramp, gravity tends to speed up the cart while friction tends to slow the cart, so those two forces are now competing against each other.

The result is that the total acceleration of the cart will be slightly smaller in magnitude when traveling down the ramp compared to when the cart travels up the ramp.

The position, velocity, and acceleration graphs tell us a lot about the motion of the cart. The first part, where velocity is negative and acceleration is positive, shows us the period of time when the cart is moving up the ramp and slowing down. **An object will slow down when the velocity and acceleration vectors point in opposing directions.**

For just an instant of time, the cart stops. This is because the cart must change direction before it rolls back down the hill. The velocity is zero, but the acceleration is still a constant value of -1 m/s^2.

For the remainder of the graph, the velocity is positive and the acceleration is positive. This tracks the time when the cart is rolling down the ramp and speeding up. **An object will speed up when the velocity and acceleration vectors point in the same direction.**

Equations of accelerated motion

The equations to calculate distance and velocity when there is an acceleration present are the most general equations to use in kinematics. They can be used all the time, even when acceleration is zero. But when acceleration is not zero, these equations of motion **must** be used. The equations are

$$\begin{aligned}\vec{d} &= \frac{1}{2}\vec{a}t^2 + \vec{v}_i t, \text{ and} \\ \vec{v}_f &= \vec{v}_i + \vec{a}t,\end{aligned}$$

where $\vec{d}$ is the displacement, $\vec{a}$ is the acceleration, t is the time, $\vec{v}_i$ is the initial velocity, and $\vec{v}_f$ is the final velocity of the object.

When acceleration is zero, these equations simplify to

$$\begin{aligned}\vec{d} &= \vec{v}_i t, \text{ and} \\ \vec{v}_f &= \vec{v}_i,\end{aligned}$$

where the first equation is simply a rearranged version of $v = d/t$. (Note that when velocity is constant, it does not make sense to refer to final or initial velocity, but simply velocity.)

The equations of accelerated motion are summarized in the table below:

how far?	distance	$\vec{d} = \vec{v}t$	$\vec{d} = \frac{1}{2}\vec{a}t^2 + \vec{v}_i t$
how fast?	velocity	$\vec{v}_f = \vec{v}_i$	$\vec{v}_f = \vec{v}_i + \vec{a}t$

Position and velocity graphs

Position, velocity, and acceleration can be plotted in a graph to determine information about the motion of an object. When discussing graphs, we use the terminology “y vs. x” to describe the variable plotted on the vertical axis (y) as it relates to the variable plotted on the horizontal axis (x). Therefore, we would call a graph of position (vertical, or y-axis variable) as it is graphed with respect to time (horizontal, or x-axis variable) a “position vs. time” graph.

Position vs. time graphs

Consider again the person from Figure 3.2, who started at the origin. Say the person moved to the position 4 m to the right of the origin at a constant speed, taking 2 s to arrive at that position. Then they stop and remain still at that position for 2 s. Finally, they move at a constant speed to a position 2 m to the left of the origin at a constant speed, taking 2 s to arrive at the final location. This motion can be graphed with position on the y-axis and time on the x-axis, as shown in Figure 3.7.

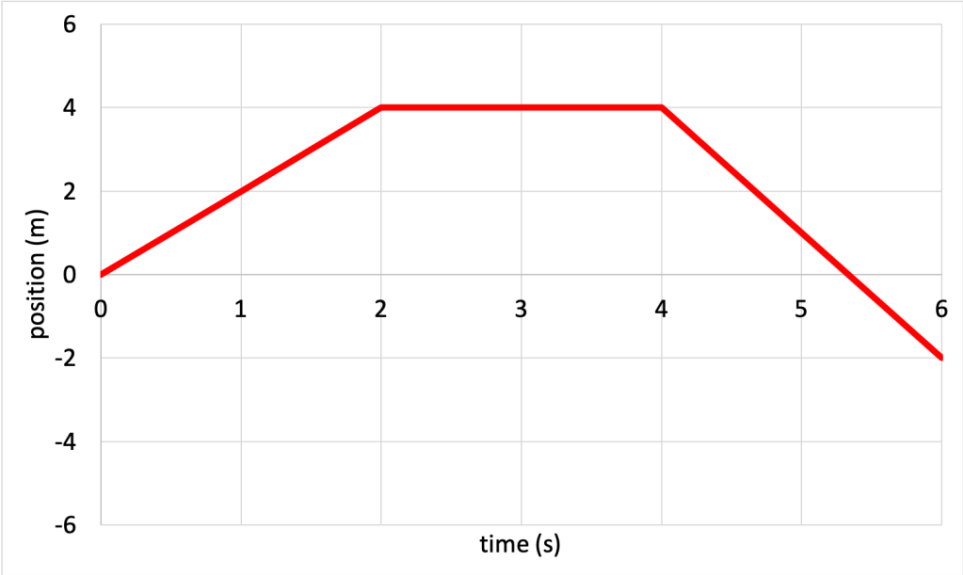


Figure 3.7 – A position vs. time graph for a person who starts at the origin, moves to +4 m at a constant speed, stops for 2 s, then moves to -2 m at a constant speed. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Assuming that the position vs. time graph is composed of straight lines (which is true in Figure 3.7), the slope of each section of the graph is equal to the velocity of the object at that time. For the first 2 s of motion, the person has a velocity of

$$\frac{4\ \text{m}}{2\ \text{s}} = 2\ \text{m/s}. \tag{3.1}$$

For the next 2 s of motion, the slope and velocity of the person is zero. For the final 2 s of motion, the person has a velocity of

$$\frac{-6\ \text{m}}{2\ \text{s}} = -3\ \text{m/s}. \tag{3.2}$$

Velocity vs. time graphs

Just as position can be plotted with respect to time, so can velocity. The velocities calculated above are plotted and shown in Figure 3.8.

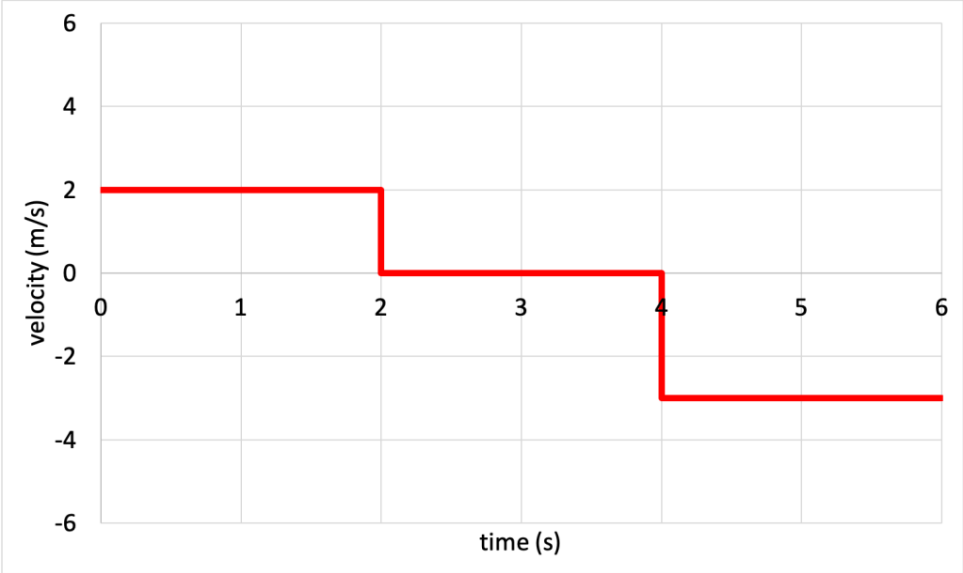


Figure 3.8 – A velocity vs. time graph for a person who starts at the origin, moves to +4 m at a constant speed, stops for 2 s, then moves to -2 m at a constant speed. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The slope of a velocity vs. time graph is equal to acceleration.

Free fall

The most prevalent source of acceleration in our lives is gravity. Gravity is a force of attraction between masses. It is this attractive force that pulls us down toward the surface of the Earth. That pull produces an acceleration due to gravity, which is equal to 9.8 m/s^2 pointing downward. This acceleration is a (mostly) constant value along the surface of the Earth. Because this quantity is used so frequently in physics, it has a special symbol, the lowercase letter g. This symbol (g) refers to the acceleration due to gravity. At times in this textbook this quantity may be referred to as “little g.” This distinguishes it from another quantity, “big g” (G), which is the universal gravitational constant.

Free fall is defined as a situation where gravity is the only force acting on an object. For example, if an object were dropped off the top of a very tall building, free fall would define its motion assuming that there are no other forces such as friction and air drag to slow it down. Free fall is not always a completely accurate way to discuss the motion of an object, but it does a good job in some scenarios, giving us a straightforward way to analyze position, velocity, and acceleration when air drag and friction are negligible.

When analyzing a free-fall situation, we use the equations for distance and velocity when acceleration is present. The value used for acceleration is -9.8 m/s^2, as that is the value that comes from gravity. (It is customary to take “up” as positive and “down” as negative. The minus sign signifies that the direction of the acceleration is downward.)

The equations of accelerated motion defined above are used to describe the motion of an object in free fall, using $g = -9.8\ \mathrm{m/s^2}$ as acceleration.

Consider a ball or other object thrown vertically upward and then caught again at the end of its motion. At the start, the acceleration of the ball points downward and the velocity points upward. Because the two vectors are pointing in opposite directions, the ball will slow down. The ball slows down until it reaches the top of its motion, where it stops just for an instant in time, before moving downward again. Now the acceleration and velocity vectors both point in the same direction, downward. Because the two vectors point in the same direction, the ball will speed up. The ball speeds up until it is caught and ends its motion. The acceleration is constant throughout the motion of the ball.

The last sentence of the previous paragraph bears repeating: **the acceleration is constant throughout the motion of the ball**. Gravity is constant, and it never turns off. The ball may instantaneously stop at the top of its motion, but that’s because gravity is pushing down on the ball, and as it’s moving upward it slows down and must change its direction before coming back down to the Earth. The acceleration due to gravity is -9.8 m/s^2 the entire time the ball is moving. It is a constant value that never changes.

People often find it difficult to accept that the acceleration is not zero at the very instant that the velocity is zero. However, during the flight, the velocity is always changing. Even when the ball reaches the top of its flight, the velocity is still changing. It just happens to be changing through zero (from positive to negative). If the acceleration was zero at any time during the flight, there would have to be at least two points on the velocity graph that have the same value. Since this never occurs, the acceleration cannot be zero at any time while the ball is in the air. This was also the case for the cart that was pushed up the ramp and then came back down. In Figure 3.6, it is clearly seen that the acceleration is never zero (even when the cart momentarily stops at the top of its climb).

Further reading

- ***Penny Drop* (MythBusters season 1, episode 4, 2003)** – This episode of MythBusters debunks the myth that dropping a penny from the top of the Empire State Building would destroy the sidewalk and kill any pedestrians unlucky enough to stand at the base of the building.

Practice questions

Conceptual comprehension

1. If you go for a drive, recall that the odometer on your car measures the path that the car has driven. Using a piece of string and a map, describe how you could go about calculating the displacement of the car for that excursion.

Numerical analysis

1. A person walks 30 m east, then turns around and walks 15 m west. Calculate the person's path. Then, calculate the person's total displacement.
2. An elevator starts at the ground floor and goes up to the fifth floor, then comes back down to the third floor. Calculate the elevator's path. Then, calculate the elevator's total displacement. (Use units of floors in your calculations.)
3. A skier starts at the top of a hill, descends 500 m down the slope, climbs 100 m back up, then skis an additional 400 m down. Calculate the skier's path. Then, calculate the skier's total displacement.
4. A boat moves 10 km downstream in a river flowing at 5 km/hr. Later, the boat returns to its starting point by moving 10 km upstream. Calculate the boat's path. Then, calculate the boat's total displacement.
5. A car travels east to west from point A to point B, covering a distance of 200 km in 2 hr. Calculate the car's average speed. Then, calculate the car's average velocity. (Use units of m/s in your calculations.)
6. A bird flies 2 km north in 20 min and then returns 2 km south in 15 min. Calculate the bird's...
    1. ...instantaneous speed as it flies north.
    2. ...instantaneous speed as it flies south.
    3. ...average speed for its entire journey.
    4. ...average velocity for its entire journey.
7. Draw a position vs. time graph and a velocity vs. time graph for an object moving with a steady pace in the positive direction, and then stopping and staying still for several moments.
8. Draw a position vs. time graph and a velocity vs. time graph for an object that starts at rest for several moments, then moves at a constantly increasing speed for the next several moments.
9. Draw a position vs. time graph and a velocity vs. time graph for an object that speeds up at an increasing rate, travels at a constant rate for a few moments, and then slows down until it stops.
10. Describe the motion needed to generate the position vs. time graph shown in Figure 3.9.

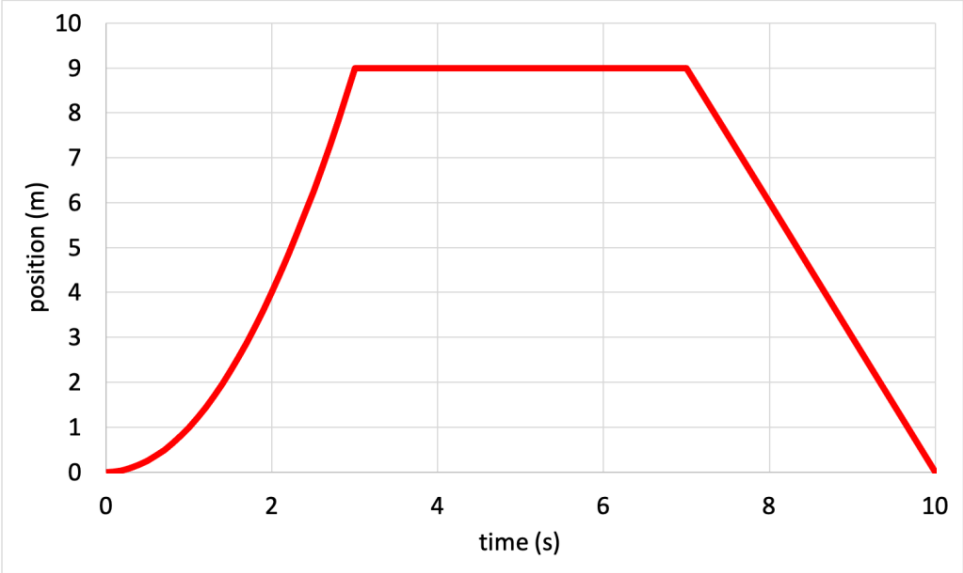


Figure 3.9 – A graph showing position versus time for an object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

11. Describe the motion needed to generate the velocity vs. time graph shown in Figure 3.10.

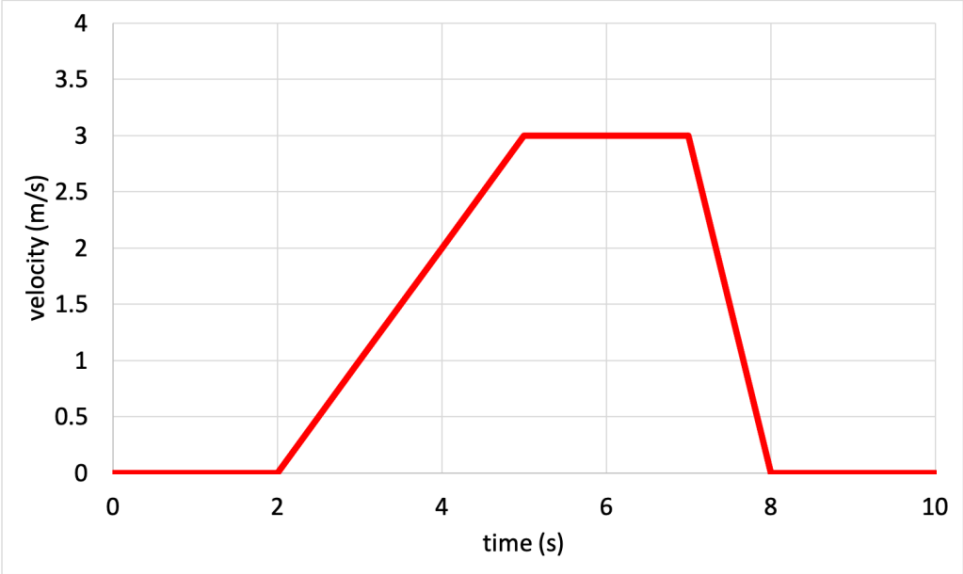


Figure 3.10 – A graph showing velocity versus time for an object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

12. A train travels a distance of 500 m in 20 s at a constant speed. Calculate the velocity of the train.
13. A car starts from rest and accelerates at a rate of 2 m/s^2 for 5 s. Calculate the car's final velocity. Then, calculate the distance the car traveled during the 5 seconds.
14. An object is thrown vertically upward with an initial velocity of 20 m/s and travels in free fall. Calculate how high the object will go.
15. An object falls freely from a height of 40 m. What is its velocity just before it hits the ground?

This page titled Chapter 3: Linear motion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 4: Newton's second law

Summary

Newton's second law of motion describes the relationship between force, mass, and acceleration. This will expand upon our current understanding of motion, and open us up to discussions about momentum, energy, and more in the next several chapters.

Force

Simply put, a force is some type of push or pull that causes an object to change its motion. The symbol for force is the capital letter $\vec{F}$. The unit for force is the newton (N).

There are many manifestations of forces. These include gravity, support force, friction, viscous drag, applied forces, tension, elastic/spring forces, electrical, magnetic, and nuclear forces.

Gravity

Gravity is one of the most prevalent forces in our lives. At this stage, it will suffice to say that gravity is a force of attraction between masses. Because the Earth has an enormously large mass, it creates a large attractive force, causing us to constantly experience a downward force pulling us toward the center of the Earth.

If you've ever dropped something, or rolled something downhill or down a ramp, then you've seen gravity's effect on the motion of an object. In addition, our weight comes from gravity.

Support force

A support force occurs whenever something is in contact with a surface such as the ground, the floor, a wall, or something else. If you are standing as you read this, then the floor is supporting your weight and preventing gravity from accelerating you to the center of the Earth. The support force is sometimes called a normal force as the force always acts perpendicular to the supporting object. (In science and mathematics, the term "normal" means "perpendicular.")

When an object is in contact with a flat surface, the normal force will point up. But when that surface is tilted, then the normal force no longer points up. It points perpendicular to the surface. This is shown in Figure 4.1.

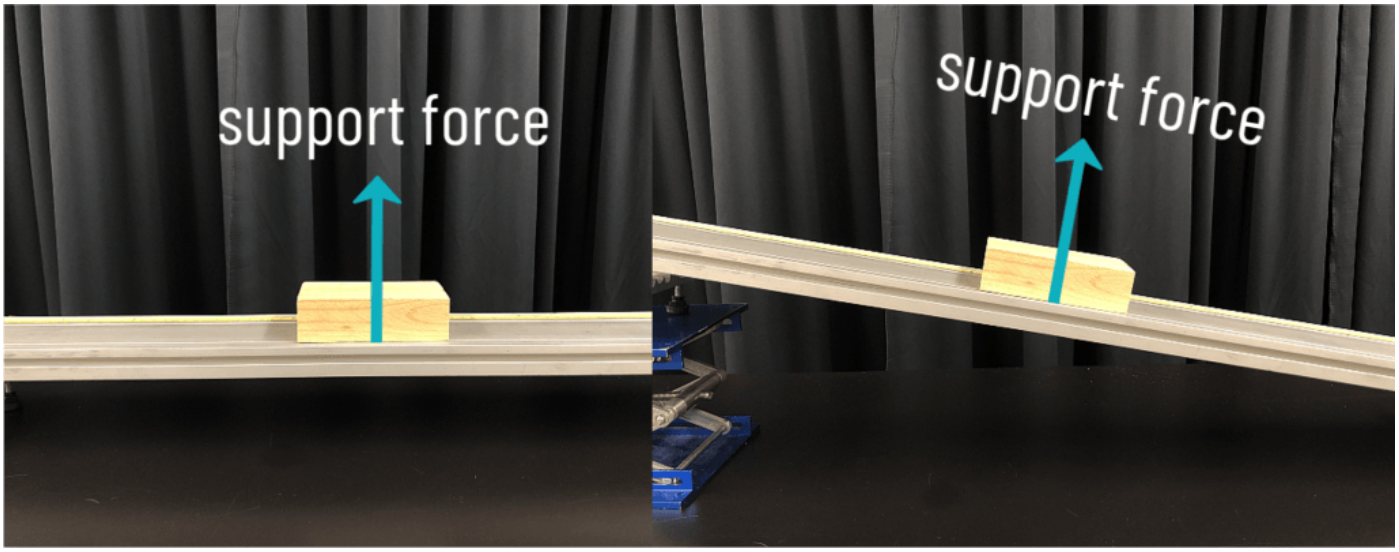


Figure 4.1 – The support force vector points up when an object is on a horizontal surface (left), and has an angle perpendicular to any tilted surface (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Friction

Friction is another force we experience in our daily lives. There are two types of friction: static friction and kinetic friction.

Static friction

Static friction is the force that must be overcome in order to get an object at rest to start moving. When static friction is involved, the surfaces in contact do not slide past one another. For example, if you've ever tried to move something heavy by pushing it, perhaps you've noticed that it takes more effort to get the object to start moving than it takes to get it to continue moving. That's because of static friction. Static friction is also the force we experience between our feet and the ground when we walk or run. It is also the force between wheels and pavement when we drive a car, ride a bicycle, or push a wheelchair.

Kinetic friction

Kinetic friction is the force that occurs between surfaces as they slide past each other. The amount of kinetic friction depends on the two materials that are in contact with each other as they slide. Something like steel and ice will experience relatively low kinetic friction, which you may have experienced if you've ever gone ice skating. On the other hand, wood and steel would experience much more kinetic friction.

Another factor that influences the amount of kinetic friction is the support force between the object and the surface that it slides along. The larger the support force, the higher the kinetic friction, all other things being equal. Interestingly, for dry non-lubricated surfaces, the force of friction is essentially independent of the relative speed of the sliding surfaces.

Viscous drag

Viscous drag is a force that occurs when an object travels through a fluid, such as air or water. Drag acts in the opposite direction of motion. Consider an object tossed into the air. When the object is moving up, the drag (air resistance) will point down. When the object is moving down, the air resistance will point up. Unlike friction, air resistance does depend on the speed of an object. The faster the object goes, the greater the air resistance will be.

An entire area of study, aerodynamics, has been created to determine what kinds of objects travel through the air without experiencing too much drag. You can get an idea about aerodynamic shapes by looking at airplane design, especially fast airplanes that would otherwise be subject to very high resistance at their speeds. In the case of airplanes and other aircraft, air resistance is something that needs to be minimized to keep the fuel costs of air travel as low as possible.

Air resistance will reduce the acceleration of an object traveling through the air. Consider a rock dropped from the top of a tall skyscraper. If that rock falls only under the influence of gravity, then its speed will increase linearly over time as gravity increases its speed by 9.8 m/s every second.

However, air resistance will act on the rock to reduce the rate at which the speed increases. At first, when the rock is initially dropped, air resistance will be zero because the rock is not moving yet. As the rock speeds up, the air resistance will increase. This limits the increase in speed. At a certain point, air resistance and gravity will become equal to each other, causing the object to fall at a constant speed. This speed is known as the terminal speed. The rock will continue moving at its terminal speed until it hits the ground. The velocity vs. time graph for each scenario (with and without air drag) is shown in Figure 4.2. (Download this data [XLSX, 9 kB])

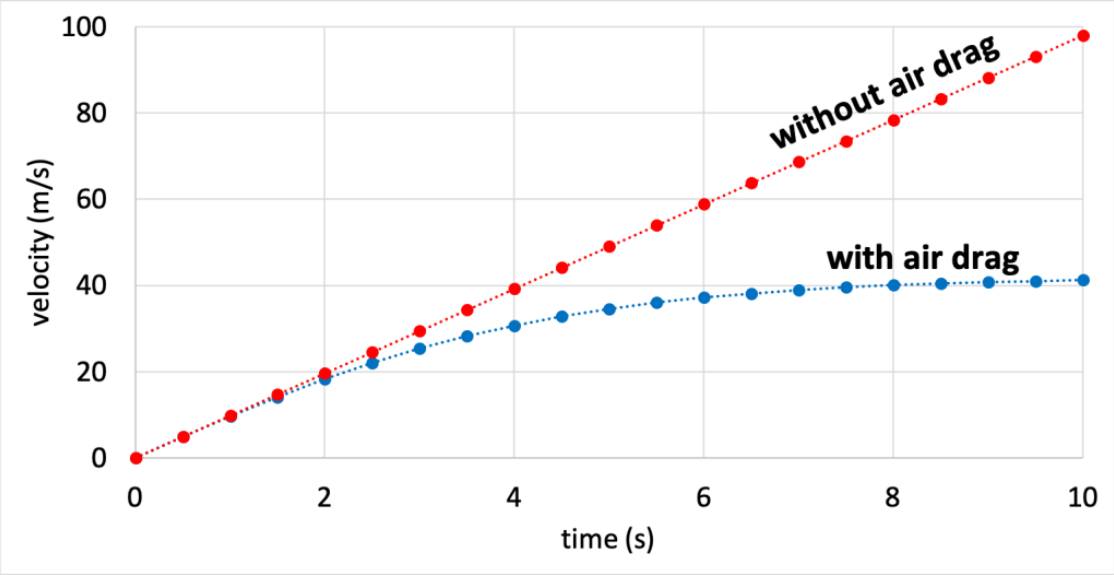


Figure 4.2 – A velocity vs. time graph for a rock falling both with and without air drag. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Every object will have a different terminal speed depending on its size, shape, density, and other factors. If we drop any object from a large enough height, it will speed up to its terminal speed, and then travel in dynamic equilibrium at that terminal speed until it hits the ground.

Air resistance is not always a bad thing. We can use air resistance to make falls from great heights not only survivable, but also fun. Skydivers use parachutes to decrease their terminal speed and allow them to survive after jumping out of an airplane.

Objects moving through the vacuum of space will not experience any air resistance. This is because there is no fluid such as air to act on the object to change its motion. This concept was demonstrated in 1971 on the Apollo 15 mission by astronaut Dave Scott. He dropped a hammer and a feather at the same time, and because there is no atmosphere on the Moon, there was no air resistance that would cause the feather to move slower than the hammer. Both objects hit the surface of the Moon at the same time.

Applied force

An applied force is simply a push or pull that is applied to an object by an outside influence. Whenever we move an object by pushing or pulling, we are applying a force. For example, if you push a car stuck in the snow or pull a child in a sled, you are applying a force. The magnitude and direction of that force depends on how hard and which way you push or pull.

Tension

Tension is the pulling force present in a string, cable, or other object as it holds something up. The direction of the tension force is along the string or cable. Tension is present in pendulums, pulley systems, and even in truss bridges.

Elastic and spring forces

Elasticity describes how much an object deforms when experiencing a stress or force. Springs are a great example of an object that can be deformed with a force. After the force is removed, as long as the force wasn't too large, the spring will restore itself to its original size and shape.

Electromagnetic force

Electromagnetism describes the interactions of charged objects and magnetic materials. In our daily lives, we experience electromagnetism any time we use an electronic appliance or gadget, and even when we do something as mundane as use a magnet to hang something on a refrigerator.

Nuclear forces: the strong and weak forces

The strong and weak nuclear forces are forces present inside the nucleus of an atom. They're not forces we think about in our daily lives, and yet they are important in that they literally keep atoms together. Electromagnetism tells us that objects with like charges repel. So what is it that keeps the nucleus of an atom, comprised of like-charged protons, together? The answer is the strong force. The strong force acts on a very small scale to keep the nucleus of an atom together. It also helps us to understand nuclear processes such as fission and fusion.

The weak force is responsible for radioactive decay. It is the force responsible for changing one type of quark into another, say, by turning a proton into a neutron.

Fundamental forces

While there are many different forces that can act on an object, there are only four fundamental forces. That is to say, all of the forces just described can be put into one of four categories. Those categories are gravity, electromagnetism, the strong force, and the weak force.

Pretty much all of the forces that we experience in our lives, other than gravity, actually fall into the category of electromagnetic forces. The support force, for example, occurs due to the repulsion of electrons in two atoms when they are placed near each other. Friction occurs due to electromagnetic forces interacting at the surfaces of two objects that are touching each other. And so on.

gravity	gravity
support force	electromagnetism
friction	electromagnetism
viscous drag	electromagnetism
applied force	electromagnetism

tension	electromagnetism
elastic and spring forces	electromagnetism
electromagnetic forces	electromagnetism
the strong force	strong force
the weak force	weak force

Of the four fundamental forces, the strongest is the strong force, followed by electromagnetism, the weak force, and gravity. Gravity is by far the weakest of the four fundamental forces.

Newton's second law

Acceleration defines the rate at which the velocity of an object changes. Acceleration can mean one of three things: that an object is speeding up, slowing down, or changing direction. Newton's second law defines the relationship between force, mass, and acceleration.

Newton's second law states that the net force on an object is equal to its mass times the acceleration. In other words,

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

The net force is the sum of all of the forces that act on an object, and it is a vector quantity. Acceleration is also a vector, and it will point in the same direction as the net force. Newton's second law can be rearranged to solve for acceleration, in which case

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

Mechanical equilibrium occurs when the net force on an object is equal to zero. If the net force is zero, we can see from Newton's second law that the acceleration is also zero.

Consider pushing on a very heavy box to try to get it to move. As you push on the box, the box does not move. Friction is the force acting against your push, but how could we calculate the magnitude of that force? Is friction greater than, less than, or equal to the push?

We can use Newton's second law to find an answer. Let's say you push against the box with a force of 25 N, and the box does not move. Because the box is not moving, it is in static equilibrium (velocity is constant at zero and acceleration is equal to zero). We know from Newton's second law that the net force on the box must be zero. We can express the net force on the object as

$$\vec{F}_{net} = \vec{F}_{applied} + \vec{F}_{friction},$$

where $\vec{F}_{net}$ is the net force, $\vec{F}_{applied}$ is the applied force, and $\vec{F}_{friction}$ is the friction force. The sum of these three vectors is zero.

We can plug in 25 N for the applied force, and solve for the force of friction.

$$\vec{F}_{net} = 0 = 25\ \text{N} + \vec{F}_{friction}$$
$$-25\ \text{N} = \vec{F}_{friction}$$

This calculation states that friction is equal to -25 N. The negative sign indicates that it points in the opposite direction as the applied force. But the applied force and friction have equal magnitudes: 25 N!

Students frequently have a hard time with this concept. If you struggle with the idea that friction and the applied force are equal when the box is not moving, then consider an extreme example. Consider what would happen if the force of friction could somehow be greater than your applied force. In this case, as you push against an unmoving box, if friction somehow became larger than your applied force, it would cause a net force in the direction opposite of your push, causing the box to accelerate against the direction of your push! That's definitely not what happens when you push on an object.

This situation (applied force and friction forces are equal in magnitude but opposite in direction) is the same once you get the box moving at a constant speed. Once the box is moving at a constant velocity, the box is in dynamic equilibrium, and the net force is again equal to zero. As long as there are no other forces acting on the box, the force of friction must be equal and opposite to your push.

Consider a two-dimensional example. A cart moves down an inclined track, speeding up as it moves. Ignoring air drag, there are three main forces acting on the cart: friction (which points in the direction opposing motion), a support force from the track (which points perpendicular to the track), and gravity (which points toward the center of the Earth). These vectors are qualitatively shown in Figure 4.3, left.

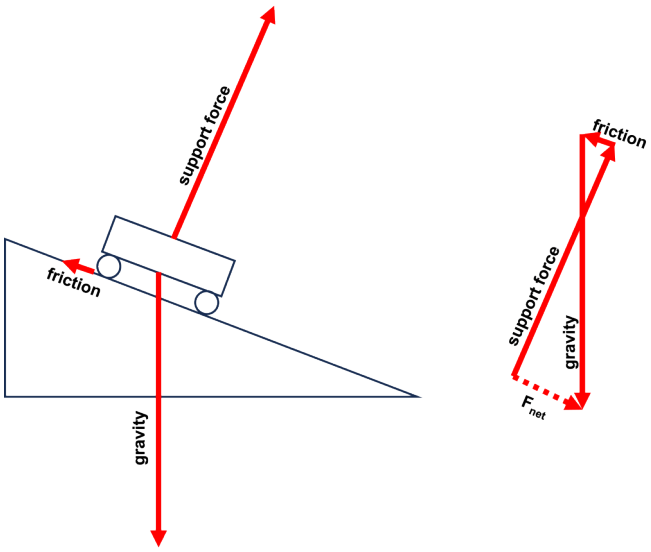


Figure 4.3 – A cartoon cart is depicted as rolling down a triangular ramp. The support, friction, and gravity force vectors are drawn. (left) A graphical sum of the three force vectors leads to the resultant net force vector (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The force vectors can be added graphically, shown in Figure 4.3, right. The resultant vector shows the net force, which points down and to the right. This is in the same direction as the cart's acceleration vector.

Mass and weight

It is important to review the concepts of mass and weight and how they differ from each other. This is another concept that is misused frequently, as people tend to (incorrectly, from a physics standpoint) use the terms interchangeably in casual conversation.

Mass comes from the "stuff" in an object: the atoms and molecules that it is made out of. Mass gives an object inertia, a resistance to changes in its motion. Because mass comes from the molecular configuration of something, it does not change based on position or location. While the situation for humans is admittedly complicated due to the fact that we eat, excrete, and exercise, if we could somehow hold our masses constant, that mass would not change if we were to be here on Earth, on the

Moon, on Saturn, or just floating in the vacuum of space.

Weight, on the other hand, is a force that comes about due to the force of gravity. While weight is a somewhat complicated topic, for now it is sufficient to say that weight is equal to our mass times the acceleration we experience due to gravity, “little g”. In other words, $\vec{W} = m\vec{g}$. “Little g” is equal to -9.8 m/s^2.

If we were to change our location, our mass would stay the same, but our weight might change. The gravitational force on the Moon is one sixth that of Earth. Therefore, our weight on the Moon will be one sixth the value of our weight on Earth. Astronauts visiting the Moon experienced this decreased weight as they conducted moonwalks.

Jupiter, however, has a much higher gravitational force than the Earth. If we could go to Jupiter (which actually has no solid surface upon which we could stand), we would find ourselves weighing 2.5 times more than we weigh on Earth. Meanwhile, if we could somehow safely transport ourselves to the Sun, the gigantic gravitational attraction of the Sun would make us weigh nearly 30 times our weight on Earth!

When astronauts visit the space station, or otherwise orbit the Earth, the lack of a support force causes them to be weightless. They are still under the influence of Earth’s gravity, which is what causes the space station to orbit the planet, but the astronauts and other objects traveling in orbit appear to be weightless. However, that does not mean that inertia has stopped impacting the motion of an object. Mass, and inertia, exist everywhere.

There is a NASA podcast episode that takes a fascinating look at the impact of inertia in space. Even when astronauts and their bulky equipment are weightless, the inertia of the objects still affects their ability to move around, put on their spacesuits, and get things done in space. NASA’s spacesuit hardware manager Les Padilla describes the effect of inertia in microgravity (this quote is recorded between times 30:37 and 31:04 in the podcast):

> What can be our friend at times is momentum or inertia. That can be an enemy at times as well when you’re in that microenvironment. Crew members are trained specifically to go slow along a space station. If you get 300 pounds moving, it can be difficult to stop. Now, they can stop it, but it’s just wasted energy. So they go very slowly so they don’t get that large mass moving too quickly.

Practice questions

Conceptual comprehension

1. A small steel BB is inserted into coiled plastic tubing, as shown in Figure 4.4. It travels in several circles before exiting the tubing. What direction will the BB travel when it exits the tubing? Consider all of the forces acting on the BB at the time it exits the tube.

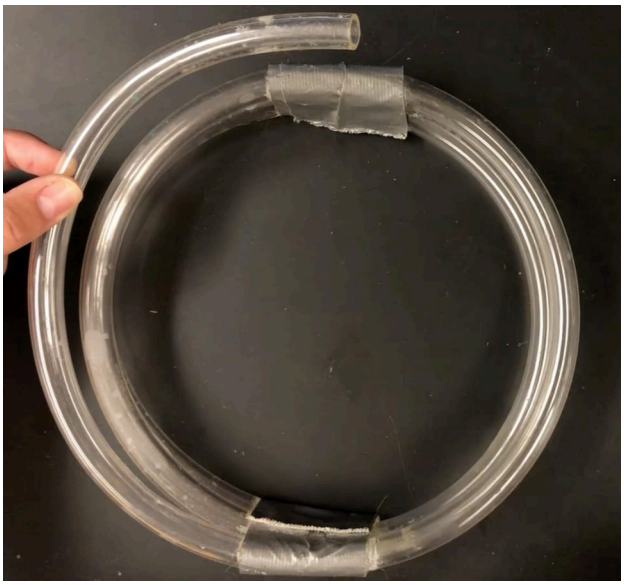


Figure 4.4 – A BB traveling through coiled tube will travel in a circular path while it's moving through the tubing. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Numerical analysis

1. If an object with a mass of 5 kg experiences an acceleration of 2 m/s^2, calculate the net force that must be acting on that object.
2. Calculate the acceleration that will be created by a net force of 0.2 N acting on an object with a mass of 10 g.
3. Calculate the net force acting on a 10 kg bicycle moving at a constant velocity of 6 m/s.
4. A spacecraft (with a mass of 2 million kg) is in deep space, far from any significant gravitational forces, and is subject to two thrusting forces: a "horizontal" force of 2,000 N and a "vertical" force of 1,000 N. (The designations of "horizontal" and "vertical" are arbitrary; what is important is that the two forces act at right angles to each other.) Calculate the net force acting on the spacecraft. Then, calculate the acceleration of the spacecraft while the thrusters are in operation.
5. If an object on the surface of the Earth has a mass of 10 kg, calculate the gravitational force acting on that object.
6. On the Moon, the acceleration due to gravity is approximately 1.6 m/s^2. Calculate the gravitational force (weight) of a person who has a mass of 70 kg if they are on the Moon.

Hands-on experiments

1. Take two sheets of otherwise identical paper. Hold one of them so that it is flat and parallel to the ground. Crumple the second sheet of paper and hold it up to the same height as the flat paper. Drop both pieces of paper simultaneously. Which sheet of paper experiences more air resistance? How do you know?
2. Using household items, how can you demonstrate that the electromagnetic force is stronger than gravity?

This page titled Chapter 4: Newton's second law is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 5: Newton's third law

Summary

Newton's third law describes the forces that exist between two objects when they interact. This helps us understand the physics of everyday motion: why we move when we walk, how rockets are capable of propelling through space, and more!

Newton's third law

Newton's third law states that for every action, there is an equal and opposite reaction. Forces always come in pairs: action-reaction pairs. As you read this book, you may be standing or sitting. You are pushing down on the floor (or ground). The reaction force is that the floor (or ground) is pushing up on you. How do you know the floor pushes up on you? It's possible that you feel that force on your feet (if you are standing), or your thighs (if you are sitting), or your side (if you are laying down).

Any time two objects interact with each other, or exert forces on each other, there will be two forces: an action force and a reaction force. The two forces are equal in magnitude and opposite in direction. Also note that action-reaction pairs of forces **never** act on the same object.

If you high five somebody, there's an action-reaction pair. Both hands exert equal forces on each other in opposite directions.

In sports, if you hit a ball with a bat, your head, your feet, or another object, there's an action-reaction pair. Does it really matter which force we call the action and which is the reaction? Not really. In softball and baseball, the bat and ball both have equal forces exerted in opposite directions on each other. The bat exerts a force on the ball. By Newton's third law, the ball exerts an equal and opposite force on the bat. Each represents forces on different objects (in this example, the ball and the bat).

Newton's third law is applicable not only when two objects are stationary, but also when one or both of the objects is moving or even accelerating. In the video below, a force-sensing probe has been connected to a stand to hold it stationary. It is connected to a second force-sensing probe with a rubber band, which Dr. Pasquale pushes and pulls to create action-reaction pairs. The forces of each probe are recorded in real time using Logger Pro, and shown on the computer readout in the background of the video, demonstrating equal and opposite forces.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Dr. Pasquale then repeats the demonstration while holding both force probes, causing one or both of the force probes to be in motion at different times. This is shown in the video below. Regardless of whether the system is stationary, accelerating in one direction, or moving at a constant velocity, the two force probes exert equal and opposite forces on each other (shown on the computer readout in the background of the video).

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=493#oembed-2

Any time two objects are in contact with each other, they will exert equal and opposite forces, regardless of the mass, volume, speed, or acceleration of either of the objects.

As shown in Figure 5.1, a wheelchair user walking a dog experiences many action-reaction forces. The dog exerts a force on the leash to the right and the leash exerts the same-sized force on the dog to the left. This means there will be an interaction between the person and the leash: the leash exerts force on the person to the right. If we let this be the "action," then the reaction force is the person exerting a force on the leash to the left. Other action-reaction pairs include:

- the person pushes down against the wheelchair, the wheelchair pushes up on the person,
- the wheelchair pushes down on the ground, the ground pushes up on the wheelchair,
- the Earth pulls down on the person, the person pulls up on the Earth, and
- the dog pushes on the ground to the left, the ground pushes on the dog to the right.



Figure 5.1 – A wheelchair user holds a leash as they walk their dog. There are many action-reaction pairs present in this scenario. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Tug of war

Consider the game of tug of war. In this game, two teams stand at opposite sides of a rope. Both teams pull on the rope, attempting to pull the other team over to their side. Both teams are in contact with the rope. (See Figure 5.2, below.)

If one team consists of lots of very muscular individuals, and the other team consists of just a few rather small individuals, which team exerts more force on the rope? The large group of large folks? The small group of small folks?

The answer is: both teams are exerting exactly the same force on the rope! How do we know this is true? If both teams are in contact with the rope, then Newton's third law tells us that their forces will be equal and opposite. This is exactly the same situation we saw with the dog and the person walking the dog, as well as with the two force probes pulling against each other.

So who wins the game of tug of war? The answer is: whichever team exerts the most force on the ground with their feet will win the game. Why? The answer is that the team that pushes harder on the ground (toward the other team) will have the ground exert an equal force back on that team in the opposite direction away from the other team. If the team of very large people is standing on a slick floor wearing socks, they will not be able to exert a lot of force on the ground, and will be moved around very easily. If the team of very small people are standing on asphalt wearing rubber soled shoes, they will be able to exert a lot of force on the ground, and will be harder to move.

Next time you play a game of tug of war, wear shoes that give you a lot of traction with the floor, and make good contact. Note the footwear and position of the schoolchildren playing tug of war in Figure 5.2, indicating the importance of the force against the ground in winning the game. It doesn't matter how hard you pull on the rope, as long as your feet exert a larger force on the ground than your opponents' do!



Figure 5.2 – Two teams of schoolchildren compete in a game of tug of war. Note the position of their feet on the ground, showing that the team that exerts the most force against the ground will win the game. Inter-house sports- tug of war, by Uzoma Ozurumba, is licensed under CC BY-SA 4.0.

Finding action-reaction pairs

If you can identify an action force, you can determine the reaction force by switching the order of the nouns describing the force. For example: if the action force is that a baseball bat pushes against a ball, the reaction force is that the ball pushes against the baseball bat. Figure 5.3 demonstrates how to switch the order of each noun. The table below lists different action-reaction pair scenarios.

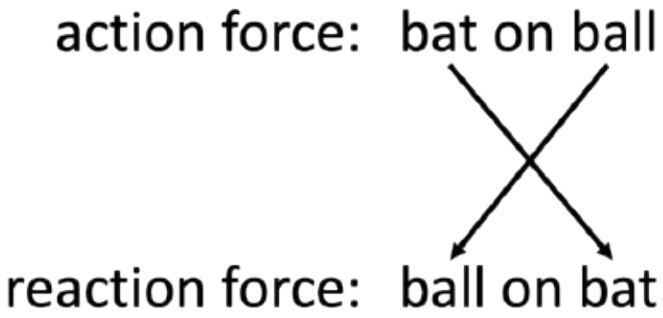


Figure 5.3 – To find the reaction force, simply switch the position of each noun in the force. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

feet pushing down on the ground	ground pushing up on feet
airplane propeller pushing backward on air molecules	air molecules pushing forward on airplane propeller

bicycle tire pushing backward on pavement	pavement pushing forward on bicycle tire
canoe paddle pushing backward on water	water pushing forward on canoe paddle
hammer pushes left on nail	nail pushes right on hammer

Gravity is a force where the Earth pulls you (and other objects) downward. What is the reaction force? Is the reaction force the support force between yourself and the floor? Consider the support force to be a push or a pull. The support force is the floor pushing up on you. These are NOT action-reaction pairs because the nouns are different. Furthermore: **two forces that act on the same object cannot be action-reaction pairs**.

What is the reaction force to gravity? The reaction force is that you are pulling up on the Earth. What is the reaction force to the support force? You are pushing down on the floor.

Newton's third law can be considered to be the reason why things move. A rocket works by creating a chemical reaction that causes exhaust gases to be pushed out of the bottom of the engine. The reaction force is thrust provided by the gases acting on the rocket that causes the rocket to move upward. A jet engine also produces thrust by moving gases backward through a turbine engine.

Systems and external forces

A system is any collection of objects that we define. If there are two carts moving on a track, we can define a system as consisting of just one cart, or just the other cart, or both carts. In other words, a system can be defined however we choose. This concept is important in understanding how external forces can act on an object to change its motion. What do we consider to be an external force? While the concept of a system can be somewhat arbitrary, it can be useful when determining how to analyze the properties of an object or collection of objects.

Consider a horse pulling a cart along a road, as shown in Figure 5.4. If we consider the system to be the cart, then the horse is considered external to that system. From the perspective of the system, the horse exerts an external force on the cart which causes the cart to start moving.



Figure 5.4 – A horse pulls a cart carrying people and supplies along a road. Arbeit in Ungarn, by zause01, is licensed under CC BY 3.0.

What if we expand our view and define the system as consisting of both the cart and the horse? What is the external force that causes the system to move? In this case, we have to look outside of the horse-cart system for the answer. In this case, the force that causes the system to start moving forward is the reaction force that the Earth exerts on the horse's hooves when the horse pushes backward on the ground.

But what if we continue to expand our view of the system to include everything: the cart, the horse, the Earth, everything? How can anything move if the definition of a system is arbitrary, and forces that may be external in one system are not external in another? How can things move if every action has an equal and opposite reaction force? Why don't all forces just cancel out and cause everything to be in equilibrium all the time? Keep reading to find out!

Using Newton's second and third laws

If every force causes an equal and opposite counter-force, why isn't everything in equilibrium? The reason is that while all forces have equal and opposite reaction forces, the effect of those forces depends on the object they act on. We can use Newton's second and third laws together to get an idea of why this is.

As you walk, run, or roll around, you are able to change your motion. It is possible to change your motion by pushing off the floor. The reaction force is that the floor pushes back on you. It is this reaction force that causes us to move. (This is true both for feet and wheels.)

Why is it that we move when we walk, run, or roll, but the floor does not move? Think about the floor: it is attached to a building that's probably extremely massive and hard to move. It has a lot of inertia. Compared to the building, we are relatively small and don't have much mass. While we experience equal forces, we do not experience equal accelerations, or changes in our motion, as a result of that force.

Consider a hypothetical example where you accidentally drop your smartphone on the ground. The approximate mass of a smartphone is 0.16 kg (160 g). The Earth's gravitational field causes the smartphone to have an acceleration of 9.8 m/s^2 (pointing downward). Using Newton's second law to calculate the net force on the smartphone, we find that

$$\begin{aligned}\vec{F}_{net} &= m\vec{a} \\ &= (0.16\text{ kg})(-9.8\text{ m/s}^2) \\ &= 1.568\text{ N}.\end{aligned}$$

The Earth pulls down on the smartphone with a net force of 1.568 N.

The reaction force is that the smartphone pulls up on the Earth with a force of 1.568 N: equal and opposite. Why can't we see the effect of this reaction force? The reason is that the Earth is extremely massive and its inertia makes it exceptionally difficult to

move. Using Newton's second law to calculate the acceleration of the Earth due to the smartphone, we find that

$$\begin{aligned}\vec{a} &= \frac{\vec{F}_{net}}{m} \\ &= \frac{1.568\ \text{N}}{6 \times 10^{24}\ \text{kg}} \\ &= 2.61 \times 10^{-25}\ \text{m/s}^2.\end{aligned}$$

This calculation concludes that the acceleration of the Earth is $2.61 \times 10^{-25}\ \text{m/s}^2$, which is immeasurably small, so it is essentially zero.

Even if you drop something with a mass of thousands of kilograms, or a million kilograms, or a billion kilograms, the acceleration of the Earth would still be essentially zero. The Earth is extremely massive; it's massive on literally a planetary scale. So it would take an object of comparable mass and size to make any noticeable effect on the Earth's motion.

Every object that the Earth pulls down, pulls the Earth up with equal magnitude. You are pulling up on the Earth yourself, and do so all the time! Just because there are equal and opposite forces does not mean there is equal and opposite motion. The Earth can cause us to accelerate very readily, whereas we do not make even a tiny budge to the Earth's motion.

Practice questions

Conceptual comprehension

1. List as many action-reaction pairs as you can for forces acting on yourself or objects that you can see right now.
2. An unfortunate mosquito flies into the windshield of a sports car driving 100 KPH down the highway. Which object (the squashed mosquito or the sports car), if any, experiences a greater force in this impact? How do you know?

Numerical analysis

1. A person (with a mass of 80 kg) standing on a frictionless surface kicks a 5 kg rock with a force of 10 N. Calculate the...
    1. ...magnitude of the force that the rock exerts on the person during the kick.
    2. ...acceleration of the person.
    3. ...acceleration of the rock.
2. The Saturn V rocket had a mass of 3,000,000 kg and generated 350,000 N of thrust. As the engine pushes against the Earth (which has a mass of approximately 6 x 10^{24} kg), calculate the...
    1. ...magnitude of the force the Earth exerted on the rocket.
    2. ...acceleration of the rocket.
    3. ...acceleration of the Earth.

Hands-on experiments

1. If you have a balloon, string, and straw, build a balloon rocket. Consider the action and reaction forces that contribute to the motion of the balloon.

This page titled Chapter 5: Newton's third law is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 6: Momentum

Summary

Momentum is inertia in motion. It is a property that quantifies the amount of motion of an object. Using momentum, we can start to better understand the motion of objects, and what happens when we have collisions between two or more things.

Momentum

Momentum is a quantity that describes the amount of motion of an object. It is a vector, indicating that it has both a magnitude and a direction. The symbol for momentum is the lowercase letter $\vec{p}$. The units for momentum are kilogram-meters per second (kgm/s), or newton-seconds (Ns). Both of those units are interchangeable, meaning that either one can be used.

The quantities that affect momentum are mass and velocity. Before we quantify this using an equation, consider how mass and velocity will affect the momentum of an object. Imagine that you have a friend standing a few meters away from you who gently throws a softball to you. Think about the amount of impact, or force, the softball will have when you catch it. Then consider what would happen if the softball were thrown faster. Momentum is directly proportional to the speed or velocity of an object.

Next, imagine that same friend throwing a bowling ball to you. The impact of the bowling ball, even thrown at the same speed as the softball, would definitely be higher. Momentum is also directly proportional to the mass of an object.

Putting this into an equation, we find that

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

where $\vec{p}$ is momentum, m is mass, and $\vec{v}$ is the velocity of the object. Momentum and velocity are both vectors that point in the same direction. In the video below, a blue cart sits at rest on a low-friction track. When it gets impacted by the red cart, the momentum transfers from the red cart to the blue cart, causing the blue cart to move. The direction of that motion is in the same direction as the red cart. Direction cannot be ignored in vector quantities!

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A pickleball with a mass of 25 g (0.025 kg) is served with a speed of 18 m/s. The magnitude of its momentum is

$$\begin{aligned}\vec{p} &= m\vec{v} \\ &= (0.025\ \text{kg})(18\ \text{m/s}) \\ &= 0.45\text{kgm/s}.\end{aligned}$$

Do two objects with the same mass traveling in different directions have equal momentum? Momentum is a vector quantity, so we know that direction is involved. If direction is not the same in the two cases, then even if the mass and speed are the same, the momenta of the two objects are not equal.

Impulse

Impulse is a measure of the change in momentum of an object or of a system of objects. Because impulse defines the change in a quantity, the uppercase Greek letter delta (Δ) will appear in the symbol. The symbol for impulse is $\vec{J}$, which is equal to the change in momentum $\Delta\vec{p}$.

The equation for impulse is that impulse is the change in momentum,

$$\begin{aligned}\Delta\vec{p} &= \vec{p}_{final} - \vec{p}_{initial}\\ \Delta\vec{p} &= m_{final}\vec{v}_{final} - m_{initial}\vec{v}_{initial}.\end{aligned}$$

If the impulse is calculated for an object or system of objects whose mass does not change, then the above equation can be simplified to

$$\begin{aligned}\Delta\vec{p} &= m\vec{v}_{final} - m\vec{v}_{initial}\\ &= m(\vec{v}_{final} - \vec{v}_{initial})\\ &= m\Delta\vec{v}.\end{aligned}$$

Because momentum is a vector, the change in momentum (impulse) is also a vector. Make sure you pay close attention to the sign of each momentum variable when calculating impulse.

Consider an object that is initially at rest that you'd like to get moving quickly. (Perhaps you're playing tennis and want to serve the tennis ball to your opponent, or maybe you are playing frisbee with your dog and want to get the frisbee moving nice and fast.) In order to get that object moving as quickly as possible, you will need to apply a force to the object. The change in momentum will be proportional to the force you apply. To maximize the impulse of the object, apply a large force to it.

The other property that affects the change in momentum will be the amount of time during which you apply the force to the object. The longer the time duration during which the force is applied, the larger the impulse will be. This can be seen, for example, in the game of baseball, where to maximize the change in motion of the baseball, a large force on the object is applied over a long period of time. This can be seen with a pitcher's "windup," shown in Figure 6.1. The pitcher "winds up" by holding the baseball as far back as arm span will allow. This provides the maximum amount of time to apply a force to the baseball, causing it to increase its momentum as much as possible.



Figure 6.1 – A pitcher holds a baseball as far back as possible before pitching. This leads to the maximum amount of time to apply a force, increasing the momentum of the ball as much as possible. Noah Syndergaard, by Keith Allison, is licensed under CC BY-SA 2.0.

Therefore, we can derive a second equation that we can use for impulse,

$$\Delta\vec{p} = \vec{F}\Delta t,$$

where $\vec{F}$ is the force acting on the object, and Δt describes the amount of time that force is applied. Force and impulse are both vector quantities, and these two vectors will point in the same direction.

To further demonstrate the concept of impulse, and the relationship with force and time, in the video below, Dr. Pasquale will throw a tennis ball to change its momentum. First, she will apply a small force over a short time (minimum impulse), a small force over a long time (medium impulse), a large force over a short time (medium impulse), and finally a large force over a long time (maximum impulse).

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Because impulse means change in momentum, and momentum is a vector, it doesn't only apply to when an object speeds up. It also applies when an object slows down or changes its direction. To speed an object up, the force applied to the object needs to be applied in the same direction as the velocity. To slow an object down, the force needs to be applied in the opposite direction as the velocity.

We can see this by looking at the impulse generated by two nearly identical balls falling down a track and hitting a block of wood, as shown in the video below. One ball bounces, and the other does not. This helps us to understand how impulse changes with and without a bounce involved in the collision.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When the non-bouncing ball is released from the track and hits the wooden board, the board is deflected slightly due to the force the ball applies to the wood during the interaction. Newton's third law tells us that the wood will apply an equal and opposite force to the ball, slowing it down to a stop.

When the bouncing ball is released from the track, the board completely falls over. This is because the board had to not only stop the ball, but give the ball additional impulse to send it back the way it came. If the board imparts a greater impulse to the ball to reverse the ball's direction, then the ball has to impart a greater impulse on the board to make it fall over. How can this be? It follows from Newton's third law of motion. We know that the time of impact has to be the same for both objects. (After all, how can the ball touch the block for more or less time than the block touches the board? Obviously it can't.) Since the force that the board exerts on the ball is equal and opposite to the force that the ball exerts on the board, the impulses must also be equal and opposite. Therefore, the bouncing ball imparts a greater impulse to the board than the non-bouncing ball.

Both balls have a mass of 10 g (0.01 kg). They both hit the piece of wood at a velocity of approximately 0.75 m/s.

The non-bouncing ball comes to a complete stop after it hits the force plate, so the final velocity is zero. The impulse of the non-bouncing ball can therefore be calculated as

$$\begin{aligned}\Delta\vec{p} &= (0.01\text{ kg})(0.75\text{ m/s} - 0\text{ m/s}) \\ &= (0.01\text{ kg})(0.75\text{ m/s}) \\ &= 0.0075\text{kgm/s}.\end{aligned}$$

The bouncing ball rebounds with a speed of approximately -0.6 m/s. (The value has a negative sign because it points to the left.) The impulse of the bouncing ball is

$$\begin{aligned}\Delta\vec{p} &= (0.01\text{ kg})(0.75\text{ m/s} - (-0.60\text{ m/s})) \\ &= (0.01\text{ kg})(1.35\text{ m/s}) \\ &= 0.0135\text{kgm/s}.\end{aligned}$$

This provides us with quantitative evidence that the impulse is indeed greater when an object bounces compared to when it does not bounce.

Conservation of momentum

Conservation of momentum means that, in the absence of any net external force acting on a system, the momentum in that system will not change. This is known as the law of conservation of momentum.

Starting with a very simple example of conservation of momentum, consider the red cart in the video below. It starts at rest, and is pushed by Dr. Pasquale. After she removes her hand and that force is no longer being applied to the cart, it moves in the same direction at the same speed until coming to a rest at the end of the track.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The velocity data from the experiment was collected from the motion detector and is plotted below in Figure 6.2. (Download this data [XLSX, 11 kB]) The time period between 3.5 and 5.5 seconds is when the cart is moving without any noticeable external forces, and a horizontal red dashed line is included on the graph to show how constant the velocity is during that time.

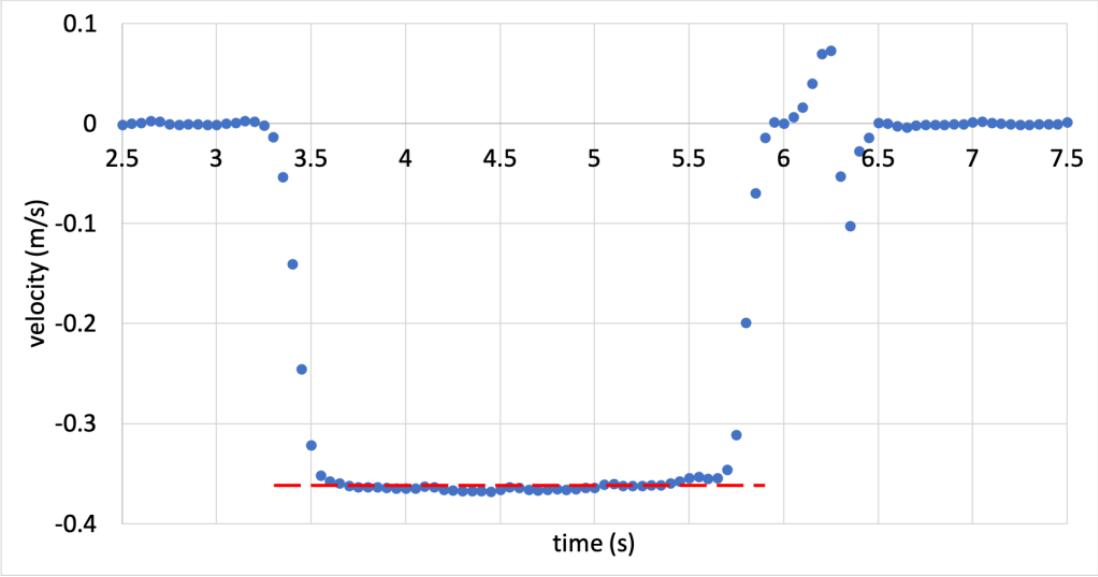


Figure 6.2 – The velocity data is plotted for a system (one blue cart) moving along a track with no external forces present. During the time the cart moves without external forces, the velocity is constant. A horizontal dashed line is included to reference the constant value of the velocity. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In this simple system consisting of just the cart, momentum is conserved in the absence of any net external force. (That changes, of course, when there are external forces, such as Dr. Pasquale's hand pushing on the cart, or the cart pushing into the motion

detector when it stops.)

When a high-friction pad is placed on the cart (demonstrated in the video below), there is a very large external force present on the system. The track (being external to the system) exerts a frictional force on the cart. The cart slows down greatly, and it's very clear that the momentum of the cart is not constant and therefore, not conserved.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The velocity data from this experiment was also collected, and is graphed in Figure 6.3. (Download this data [XLSX, 11 kB]) During the time period between about 4.25 and 6.75 seconds, the cart is moving on its own after having been pushed. The velocity gets smaller and smaller (the cart slows down) at a constant rate throughout this time period until it comes to a complete stop.

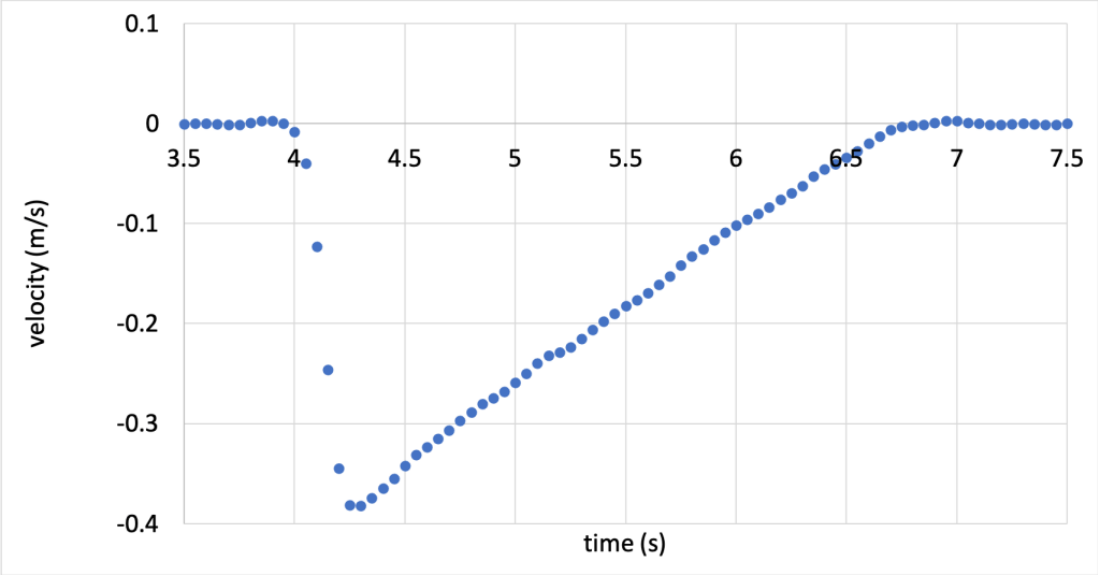


Figure 6.3 – The velocity data is plotted for a system (one blue cart) moving along a track with a frictional force present. During the time the cart moves along the track, the velocity slows down until the cart comes to a stop. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Next we can consider a slightly more complicated scenario. There are now two carts: a blue cart and a red cart. The blue cart is pushed into the red cart. When they come together, repelling magnets cause the blue cart to push the red cart, which subsequently moves down the track. This experiment is shown in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The velocity data during this experiment was recorded for both carts using both motion detectors, and is shown in Figure 6.4. (Download this data [XLSX, 12 kB]) Note that the velocity, while each cart is in motion, is not a flat horizontal line. This is due to the presence of friction on the track, causing each cart to slightly slow down over time. Due to this effect, we will consider the period of time immediately before (at approximately 5 s) and immediately after (at approximately 5.5 s) the collision in calculations, to reduce the influence of friction on the demonstration of conservation of momentum.

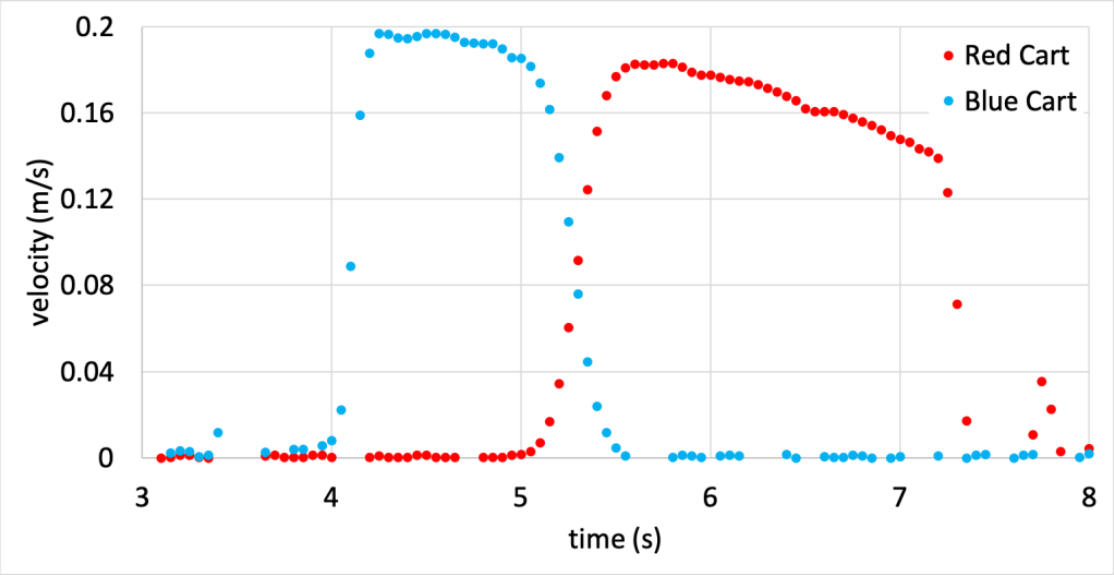


Figure 6.4 – The velocity data is plotted for two carts (one blue cart and one red cart) during a collision where the blue cart is pushed into the red cart. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

If we define the system as consisting of both the blue AND red carts, we can look at the momentum of each object before they impact and after they impact and compare. Both carts have a mass of approximately 0.4 kg.

Before the collision, the system consists of the momentum of the blue cart plus the momentum of the red cart. Both can be calculated using the momentum equation. The blue cart momentum is

$$\begin{aligned}\vec{p}_{\text{blue}} &= (0.4\text{ kg})(0.18\text{ m/s}) \\ &= 0.072\text{kgm/s}\end{aligned}$$

The red cart has zero momentum because it is not moving. Therefore the initial system momentum is 0.072 kgm/s.

After the collision, the system still consists of both carts. The momentum of the blue cart is now zero, because it is at rest. The red cart is moving at a velocity of 0.18 m/s, and therefore has a momentum of 0.072 kgm/s. The final system momentum is 0.072 kgm/s. These two momenta are equal, demonstrating the property of conservation of momentum in a system where there is no net external force.

Now, consider that the system consists of **only** the blue cart. We won't consider the red cart in our calculations any more. The momentum of the blue cart started at 0.072 kgm/s. After the collision, the blue cart has zero momentum, because it is stopped. Momentum definitely wasn't conserved in this system! This is because the red cart, which is not part of the system, now acts as an external object that exerts an external force that slows the blue cart down and changes its momentum.

Similarly, we can consider the scenario where the system only consists of the red cart. The momentum of the red cart starts at zero, and after the collision becomes 0.072 kgm/s. The momentum is not conserved. This is because the blue cart exerts an external force that speeds up the red cart and changes its momentum.

Notice that when we define the two carts together as "the system," the forces that they exert on each other are internal to the (two-cart) system and therefore cannot change the momentum of the (two-cart) system. Momentum is conserved.

These examples have provided us with evidence that momentum is conserved in one dimension, as all of the scenarios we considered took place along a straight line. What happens in two or even three dimensions? Is momentum still conserved in the absence of external forces? The answer is yes!

When a collision or interaction occurs in two dimensions, we can independently consider both the horizontal and the vertical directions, or x and y. All of the momentum in the x-direction will be conserved and all of the momentum in the y-direction will be conserved.

In three dimensions, we can consider a very similar outcome, but now with momentum being conserved in width, depth, and height. All three dimensions will have a conservation of momentum in the absence of external forces, as long as we consider the colliding objects together as “the system.”

Collisions

In short, a collision is when two objects interact for a very brief time. There are different types of collisions: inelastic and elastic. In all collisions that have no external forces, momentum is conserved within the system of interacting objects.

Inelastic collisions

An inelastic collision is defined by the presence of a lasting deformation of the objects that collide, or by the generation of heat during the collision (or both). A perfectly inelastic collision occurs when two objects collide and then stick together and travel as one unit after the collision.

An example of an inelastic collision would be dropping a coffee mug on the floor, the lasting deformation is the ceramic mug breaking into pieces after colliding with the floor.

Consider an inelastic collision where two carts (one red cart and one blue cart) collide. The blue cart has a mass that is twice that of the red cart, but exact mass values are unknown. Before the collision (shown in Figure 6.5, top), the red cart travels towards the blue cart (which is initially at rest) at a velocity of 2 m/s. The two carts collide and stick together. After the collision (shown in Figure 6.5, bottom), the two carts travel together at an unknown velocity.

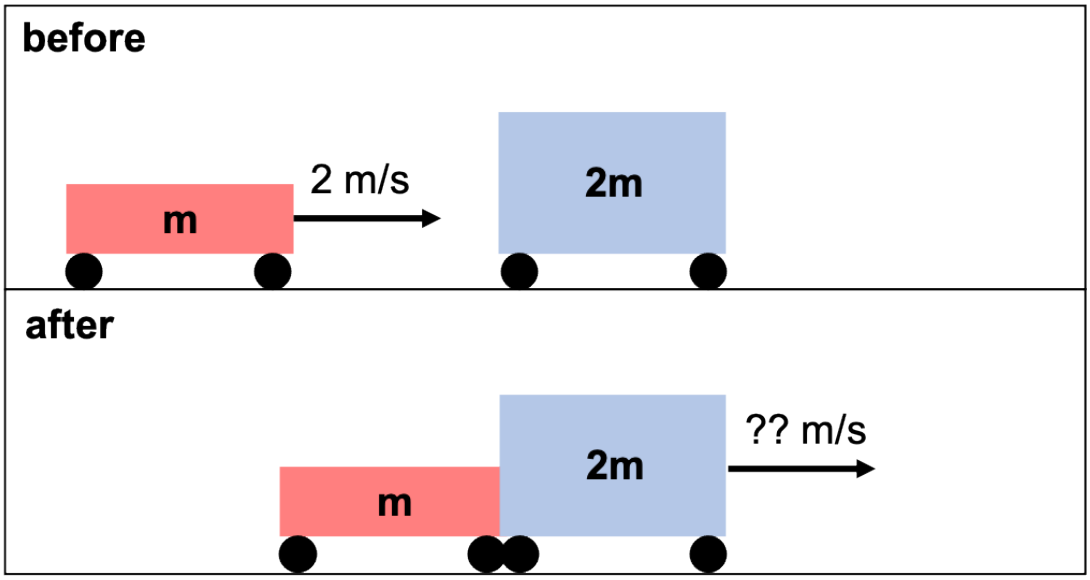


Figure 6.5 – Two carts before (top) and after (bottom) an inelastic collision. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

With this information, it is possible to calculate the final velocity of the system. Assuming there are no external forces (such as friction and air drag), the momentum of the system (consisting of both the red and blue carts) will be conserved throughout the collision. This means that the momentum of both carts before the collision will be equal to the momentum of both carts after the collision.

Note that both carts will travel at the same velocity after the collision ($\vec{v}_{final}$), as the carts travel together.

$$\begin{aligned}
\vec{p}_{red,initial} + \vec{p}_{blue,initial} &= \vec{p}_{red,final} + \vec{p}_{blue,final} \\
m_{red}\vec{v}_{red,initial} + m_{blue}\vec{v}_{blue,initial} &= m_{red}\vec{v}_{red,final} + m_{blue}\vec{v}_{blue,final} \\
(m)(2\text{ m/s}) + (2m)(0\text{ m/s}) &= (m)\vec{v}_{final} + (2m)\vec{v}_{final} \\
(m)(2\text{ m/s}) &= (3m)\vec{v}_{final} \\
\vec{v}_{final} &= \frac{(m)(2\text{ m/s})}{3m} \\
&= \frac{2}{3}\text{ m/s}
\end{aligned}$$

Elastic collisions

An elastic collision occurs when there is neither lasting deformation of objects in the collision, nor generation of heat during the collision. An example of an elastic collision is when two billiard balls collide together and then rebound afterward. (While in reality there will likely be a very small amount of heat generated when the two billiard balls collide, it is still a very good approximation of an elastic collision.)

The experiment shown in the video above (“Conceptual Physics: Conservation of momentum in a system” with data graphed in Figure 6.4) is an example of an (very nearly) elastic collision. The blue cart and red cart interact with repelling magnets, and because they don’t touch, they don’t experience any lasting deformation or generate any appreciable heat.

Collisions in two (or more) dimensions

In our everyday world, the collisions that occur do not always occur in one dimension (along a single straight line). Traffic collisions typically occur in two dimensions (vehicles approach each other at different angles along a flat surface). Collisions between molecules in our atmosphere occur in all three dimensions of space.

Regardless of how many dimensions are involved in the collision, the conservation of momentum will always apply in collisions in the absence of external forces. The vector sum of the momenta of all objects in the system before the collision is equal to the vector sum of the momenta of all objects in the system after the collision. In two dimensions, these vectors will have horizontal and vertical components. In the case of a collision on a roadway, perhaps more accurate descriptions would be north-south and east-west components. What each component is called (horizontal or east-west) is not as important as the fact that the components are perpendicular to each other. In three dimensions, each vector will have three perpendicular components. Each of the components before and after the collision will be conserved.

In Figure 6.6 left, two objects (a circle and a triangle) approach each other at right angles. The circle moves to the right and the triangle moves up. Figure 6.6 right shows the two objects after the collision. They travel together and have a velocity vector that points up and to the right.

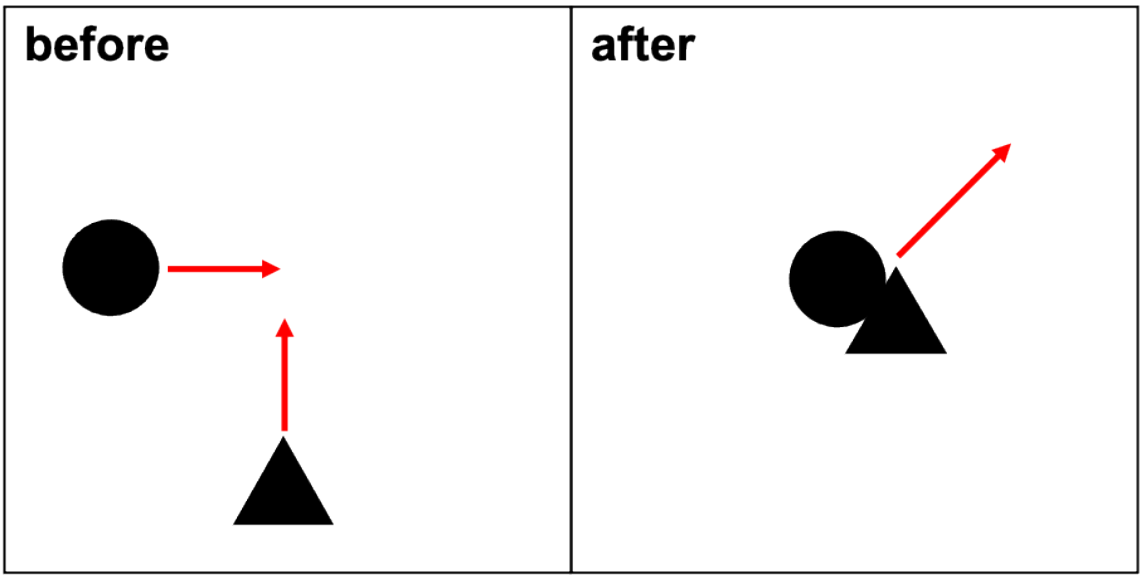


Figure 6.6 – A circle and triangle travel in perpendicular directions before a collision (left) then collide and travel together after the collision (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

If the circle and triangle in Figure 6.6 both have a mass of 1 kg, the circle has a velocity of (1 m/s, 0 m/s), and the triangle has a velocity of (0 m/s, 1 m/s), then the momentum of the circle is (1 kgm/s, 0 kgm/s) and the momentum of the triangle is (0 kgm/s, 1 kgm/s). The total momentum of the system (both before and after the collision) is (1 kgm/s, 1 kgm/s). The magnitude of the total momentum can be calculated using the Pythagorean theorem as 1.41 kgm/s.

If the circle were to have more momentum than the triangle (say, by having a greater speed or greater mass), then after the collision, the system will have more rightward motion than upward motion. If instead the triangle were to have more momentum than the circle, then after the collision, the system will have more upward motion than rightward motion.

Engineering safer collisions

Based on what we've learned in this lecture, we can find a way to engineer a safer collision. We've seen that the momentum of an object is going to create an impact force on an object when it collides. We've also learned that the change in momentum (impulse) of an object is equal to force times time.

We want to have safe collisions in cases when we're driving a car, or riding a bike, or any other situation where we could be injured when we collide with an object. In these scenarios, we want the force to be as minimal as possible in the collision. Because when we get into a collision, we're unable to change our mass, our vehicle's mass, or our velocity at the moment of the collision, and because we're going to come to a complete stop after the collision, the initial momentum, final momentum, and therefore impulse are going to be fixed. We'd ideally like the number to be small, but we can't control it after the fact. Because we know that impulse is equal to force times time, and we want the force to be as small as possible, we can see mathematically that extending the duration of a collision, which has the result of making Δt a large number, will cause the force to be small, given that impulse is fixed.

The video below demonstrates the effect of time on the force generated to cause a fragile object to stop once it's in motion. First, Dr. Pasquale throws an egg in such a way that the force acting on the egg is as small as possible. She throws the egg at a sheet, and the egg doesn't break. The time duration of the collision is extended for a long period of time and therefore the force is small enough for the egg not to break. (In this way, the sheet acts like an airbag in a car.)

Then, Dr. Pasquale throws the egg so that its motion will be stopped very quickly. By throwing it at a wall, the duration of the collision is very fast, the force is very large and the egg shell cracks and makes a giant mess.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This is why cars are engineered to have crumple zones and air bags. Anything we can do to increase the time of a collision in the case of a car crash, for example, will reduce the stopping force on the occupants of the car and keep them as safe as possible.

Further reading

- Car safety rating search: National Highway Traffic Safety Administration – This website allows you to search for your car by make and model to see how "crashworthy" it is.
- Understanding Car Crashes: It's Basic Physics – This YouTube video demonstrates the physics of car crashes, and why features such as airbags and seatbelts are important for traffic safety.

Practice questions

Conceptual comprehension

1. Show that the unit of Ns is equivalent to the unit of kgm/s.

Numerical analysis

1. A car with a mass of 1,200 kg is traveling at a velocity of 20 m/s. Calculate its momentum.
2. A soccer ball with a mass of 0.45 kg is kicked with a velocity of 10 m/s. If it comes to a stop in 2 seconds, calculate the change in momentum of the soccer ball.
3. A tennis player hits a 0.05 kg tennis ball with a force of 100 N, changing its velocity from 15 m/s to -10 m/s. Calculate the impulse applied to the tennis ball.
4. Two cars, A and B, have masses of 1,200 kg and 800 kg, respectively. Car A is moving at 15 m/s to the right, and car B is moving at 10 m/s to the left. They collide and stick together. Calculate...
    1. ...the initial momentum of the system.
    2. ...the final momentum of the system.
    3. ...the final velocity of both cars.
    4. Was this an inelastic or elastic collision? How do you know?
5. Suppose you have a 0.2 kg hockey puck initially at rest on an ice rink. It is then struck by a 0.3 kg hockey puck moving at 5 m/s to the right. After the collision, the two pucks travel together. Calculate...
    1. ...the initial momentum of the system.
    2. ...the final momentum of the system.
    3. ...the final velocity of both hockey pucks.
    4. Was this an inelastic or elastic collision? How do you know?
6. Two cars, C1 and C2, have masses of 1,200 kg and 900 kg, respectively. Car C1 is moving at 20 m/s to the right, and car C2 is moving at 10 m/s to the left. After the collision, car C1 has a velocity of 6 m/s to the left, and car C2 has an unknown velocity. Calculate...
    1. ...the initial momentum of the system.
    2. ...the final momentum of the system.
    3. ...the final velocity of car C2.
7. A car with a mass of 1,500 kg collides with a brick wall at an initial speed of 40 m/s, coming to a stop in 0.5 seconds. Calculate the magnitude of the force of the collision.
8. If the car above instead collides with a bale of hay, slowing from 40 m/s to a stop in 4 seconds, calculate the magnitude of the force of the collision.

Hands-on experiments

1. If you have a smartphone, download a free sensor app such as phyphox or Physics Toolbox Sensor Suite. Using the acceleration selection, jump up and down and notice that there is an acceleration on the phone generated when you jump while holding the phone. (From Newton's second law, we know that acceleration and force are directly proportional.) Jump up and down and land with bent knees (extending the time of impact as much as possible) and compare the acceleration in that case to jumping up and down and landing with straight knees (minimizing the time of impact). Try to control for as many other variables as possible. How does the time of your collision with the ground relate to the force generated?

This page titled Chapter 6: Momentum is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 7: Work and energy

Summary

Work and energy are important concepts that allow us to understand and quantify the ability of objects to do something useful. That utility might be moving a piston in an engine, turning a wheel about an axis, or spinning a turbine to generate electricity. Work and energy are so intimately related that they have somewhat circular definitions that describe how intertwined they are. Energy is defined as the capability of an object to do work. And work is defined as the transfer of energy by applying a force to that object over a distance.

Work and energy as concepts can allow us to understand simple machines such as pulleys and levers, and also enable us to analyze the motion of projectiles, planets, and roller coasters, among other things.

Forms of energy

Energy is defined as the capability of an object to do work. The symbol used for energy depends on the type of energy being quantified, but regardless of the type of energy, the unit used for energy is the joule (J). A joule is a unit that contains other units hidden inside of it.

$$\begin{aligned} 1\text{ J} &= (1\text{ N})(1\text{ m}) \\ &= (1\text{ kgm/s}^2)(1\text{ m}) \\ &= 1\text{ kg m}^2/\text{s}^2 \end{aligned}$$

All three of these units mean the same thing: J, Nm, and kg m^2/s^2. However, the unit of Nm is typically not used when discussing work and energy, as it's used as the unit for something called torque (which will be discussed in the next chapter of this textbook). Therefore, we will primarily use J as the unit for work and energy.

Energy and work are scalar quantities. That means that there is no direction associated with either of the two concepts, only a magnitude or strength.

There are many different forms of energy. In general, energy can be "kinetic," which is energy due to motion. Energy can be "potential," or stored energy, which is energy that exists based on the arrangement or configuration of something. For each form of energy, we have discovered a formula and have assigned a symbol to quantify that particular form of energy.

Kinetic energy

Kinetic energy is energy that objects have by virtue of their motion. Kinetic energy has a symbol of KE, units of J (joules), and an equation of

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

where m is the mass of the object and v is the velocity of the object.

Remember that kinetic energy, like all forms of energy, is a scalar quantity. This is an important concept. Energy has no direction associated with it. When calculating kinetic energy, be very careful to use parentheses in your calculator when squaring the velocity. When taking the square of any number, positive or negative, the result will be positive. However, calculators are not so smart and may give you a negative answer if you do calculations without parentheses. It's not possible to have negative kinetic energy. It would imply that it is possible to have less energy of motion than an object that isn't moving, which doesn't make sense at all!

Gravitational potential energy

Gravitational potential energy is stored energy that comes about due to the gravitational attraction between the Earth and other objects of mass. Gravitational potential energy has a symbol of GPE. (As previously mentioned, the unit for all types of energy is J.)

Consider where GPE comes from. An object sitting on a table has gravitational potential energy. If you push the object off the table, it will move downward, gaining kinetic energy. What aspects of the object sitting on the table can be changed to maximize GPE?

You can increase the gravitational potential energy by increasing the mass of the object. If you double the mass of the object, it will have twice the GPE. Tripling the mass gives three times the GPE. Therefore, GPE is directly proportional to mass.

You can also increase the GPE by raising the height of the table. The greater the height above our reference, the larger the GPE. Twice the height doubles the GPE and tripling the height triples the GPE. Therefore, GPE is directly proportional to height. For heights that are below the reference level, the GPE is negative. (Exactly what the height is measured in relation to – sea level, ground level, floor level, or something else – is up to you to decide. You just need to be consistent with that choice throughout the analysis.)

Finally, you can increase gravitational potential energy if you can somehow move the object of mass to a planet with a larger gravitational acceleration. The object would have more GPE on a planet like Jupiter or Saturn, and less GPE on a planet like Mars.

Tying this all together, we can look at the equation for gravitational potential energy

$$GPE = mgh,$$

where m is the mass of the object, g is the magnitude of the gravitational acceleration, and h is the height above the reference level.

Elastic potential energy

Elastic potential energy is stored in the configuration of some type of deformable object. We can see this in a spring. The following video shows a mass bouncing up and down on a spring. When the mass is pulled downward, the deformation of the spring causes it to start moving once Dr. Pasquale lets go of the mass. The up and down motion, known as an oscillation, shows us the conversion of energy from elastic potential energy, to kinetic energy, to gravitational potential energy, over and over again as the mass bounces up and down.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Electrical energy

Electric potential energy comes from placing charged particles some distance apart from each other. This is the type of energy that can be stored in components used in electric circuits. If you have ever used a battery, or powered an electronic device using a wall outlet, then you have used electrical energy.

Chemical energy

Chemical energy is a type of stored energy that's based on the configuration of atoms or molecules. For example, gasoline

contains stored energy that we can use when we ignite it. Chemical energy is also how we as humans obtain energy from the food we eat during the digestive process.

Nuclear energy

Nuclear energy is the energy stored in the nucleus of an atom. This energy can be released in radioactive decay, fission, or fusion.

Radiant energy

Radiant energy is potential energy that's stored in electromagnetic waves. There are many different categories of electromagnetic waves, each with a particular range of values for their energies. Visible light is the most common type of electromagnetic wave that we experience.

Heat (thermal) energy

Heat, or thermal energy, is energy that moves between objects that have different temperatures. The so-called "useful" types of energy are generally potential energy and kinetic energy. Heat can be utilized to do useful work, but is also where a lot of useful energy is wasted and cannot be used again. For example, when two objects rub together, the force of friction generates heat. We consider that heat to be wasted energy, as it cannot be utilized as useful work. The discussion of thermal energy and its transformations into other forms will be discussed later in this textbook.

Conservation of energy

Energy can be converted from one form to another. However, it cannot be created or destroyed. If we could somehow measure all of the energy in the entire universe today, and then measure it again tomorrow, those two numbers should be the same. This principle doesn't just apply to the entire universe – it applies to any isolated system (a system in which no energy enters or leaves). Generators, motors, and engines are examples of devices that perform energy conversions. If you start making observations, you'll probably notice energy conversions taking place all around you.

For example, an apartment or house in Illinois likely contains a gas furnace, which converts chemical energy to thermal energy. A microwave oven converts electrical energy to radiant energy (microwaves are a form of electromagnetic radiation). A coffee bean grinder converts electrical energy to kinetic energy. An electric stove converts electrical energy to thermal energy. (Many appliances in homes have an input form of electrical energy, as it is generally inexpensive and extremely convenient.) The table below lists some other common objects that convert energy from one form to another.

car (internal combustion engine)	chemical	kinetic
generator	electrical	kinetic
light bulb	electrical	radiant
the Sun	nuclear	radiant

human	chemical	kinetic, thermal
computer, smartphone	electrical	radiant, thermal, acoustic
plants	radiant	chemical
wind generator	kinetic	electrical
solar panel	radiant	electrical

In a rocket launch (such as that shown in Figure 7.1), chemical energy is converted to heat, sound, kinetic energy, and gravitational potential energy. Pretty impressive!

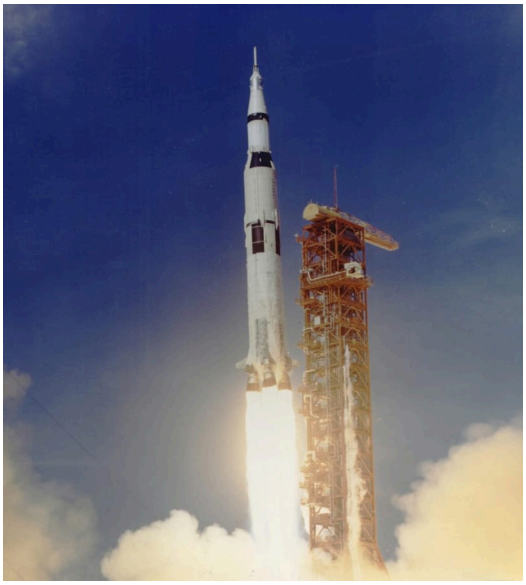


Figure 7.1 – In a rocket launch, chemical energy is converted to kinetic energy, thermal energy, and acoustic energy. Saturn Apollo Program, by NASA Marshall Space Flight Center, is licensed under Public Domain.

A rather smashing demonstration of conservation of energy comes from the bed of nails. Demonstrated in the video below, Dr. Pasquale is lying in between two large pieces of wood that are studded with nails. Each nail is individually very sharp. The reason the nails don't hurt (too much) is because there are enough nails to distribute her weight over a relatively large area. A cinder block is then placed over the upper piece of wood, and Dr. Fazzini swings a sledgehammer to break the brick. Dr. Fazzini raises the sledgehammer high above his head, giving it gravitational potential energy. As it swings downward, it gains kinetic energy. That kinetic energy is transferred to the cinder block and causes the block to break apart into many pieces. That is, energy from motion goes into separating some of the molecules in the block. Some of the energy goes to heat, and some of it goes to sound. The force from the blow is distributed along the upper layer of nails, which doesn't cause too much discomfort due to the large number of nails. (Disclaimer: do not try this experiment at home!)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Energy in collisions

As we learned in the chapter on momentum, collisions between objects can be defined as inelastic or elastic. These two types of collisions can be distinguished by how kinetic energy is affected during the collision. While momentum is always conserved in a system in the absence of external forces, kinetic energy is not necessarily conserved in all collisions.

Kinetic energy in inelastic collisions

In an inelastic collision, kinetic energy is not conserved in the collision. This is because some of the kinetic energy of the system is lost to deformation or heat. In a perfectly inelastic collision, the maximum amount of kinetic energy is lost. In such collisions, the colliding objects literally stick together.

Traffic collisions are also examples of inelastic collisions. Consider two cars that collide on the road. A red car with a mass of 1400 kg traveling at an initial velocity of 9 m/s collides with a blue car with a mass of 1100 kg traveling at an initial velocity of -18 m/s. After the collision, the red car travels at a velocity of -5.3 m/s and the blue car travels at a velocity of 0.2 m/s. (None of the occupants of the vehicles are harmed in this hypothetical example.)

The initial kinetic energy of the system is

$$\begin{aligned} KE_i &= \frac{1}{2}m_{red}v_{i,red}^2 + \frac{1}{2}m_{blue}v_{i,blue}^2 \\ &= \frac{1}{2}(1400\text{ kg})(9\text{ m/s})^2 + \frac{1}{2}(1100\text{ kg})(-18\text{ m/s})^2 \\ &= 56700\text{ J} + 178200\text{ J} \\ &= 234900\text{ J}. \end{aligned}$$

After the collision, the kinetic energy of the system is

$$\begin{aligned} KE_f &= \frac{1}{2}m_{red}v_{f,red}^2 + \frac{1}{2}m_{blue}v_{f,blue}^2 \\ &= \frac{1}{2}(1400\text{ kg})(-5.3\text{ m/s})^2 + \frac{1}{2}(1100\text{ kg})(0.2\text{ m/s})^2 \\ &= 19663\text{ J} + 22\text{ J} \\ &= 19685\text{ J}. \end{aligned}$$

Kinetic energy is not conserved in this collision. Some (215215 J) of the energy of motion was converted to sound, heat, and other forms of energy.

Kinetic energy in elastic collisions

In an elastic collision, kinetic energy is conserved in the collision. We can anticipate that a collision will be elastic if the two objects colliding won't create any heat from rubbing together, if they won't create any sound, and if the molecules inside the objects won't permanently deform when they impact together.

Once again, consider the red and blue cars (in which none of the occupants are harmed in a collision). If the two cars have the uncanny ability to collide without physically impacting each other, no physical deformation, sound, or heat will be lost in this collision. After the collision, the red car travels at a velocity of -14.76 m/s and the blue car travels at a velocity of 12.24 m/s.

The initial kinetic energy of the system is still 234900 J. After the collision, the kinetic energy of the system is

$$\begin{aligned} KE_f &= \frac{1}{2}m_{red}v_{f,red}^2 + \frac{1}{2}m_{blue}v_{f,blue}^2 \\ &= \frac{1}{2}(1400\text{ kg})(-14.76\text{ m/s})^2 + \frac{1}{2}(1100\text{ kg})(12.24\text{ m/s})^2 \\ &= 152500.32\text{ J} + 82399.68\text{ J} \\ &= 234900\text{ J}. \end{aligned}$$

Mechanical energy

Mechanical energy (ME) is equal to the sum of the potential energy and kinetic energy of an object. In other words,

$$ME = PE + KE.$$

In the absence of dissipative effects such as friction and drag, the mechanical energy will be conserved in a system. Conservation of mechanical energy can therefore be a very useful way to calculate properties of an object.

In the video below, a set of tracks is used to demonstrate conservation of mechanical energy. The tracks all have equal starting heights, and equal ending heights. That means that all of the GPE that the balls have at the start will be converted to equal amounts of KE at the end. Because each of the balls is identical, they will have the same velocity when they reach the end of the track. How far the ball goes as it launches off the end of the track is related to the velocity. When the balls are released from the tracks, while they might take different amounts of time to reach the end, they all hit the ground the same distance away, indicating that they all had the same amount of mechanical energy throughout their motion.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The amount of time it takes for the balls to reach the end of the tracks has to do with the average speed of each of the balls. If you go fast for a longer period of time, then you'll reach the end first. This is why the ball moving down the ramp with the earliest down-slope (the blue ramp in the video) reaches the end of the track the fastest. But regardless, the conservation of mechanical energy still holds up.

A roller coaster (an example of which is shown in Figure 7.2) provides a great example of conservation of mechanical energy. A chain driven by chemical or electrical energy will convey a cart to the top of the first hill. If there are no more chains to provide any other input energy, then the first hill must be the tallest part of the roller coaster. At the top of that hill, the cart has all GPE and no KE. As it moves downward, the GPE converts to KE. At the lowest point in the track, the KE of the coaster, and therefore its speed, is at its highest value. As long as the level of the track remains lower than the starting hill, the energy the roller coaster has at the start will keep it moving throughout the ride (assuming there is no friction).



Figure 7.2 – A roller coaster's initial hill must be tall enough to provide sufficient energy for the car to make it to the end of the track (unless there are other forms of input energy later in the ride). Nitro coaster, by Dusso Janladde, is licensed under CC BY-SA 3.0.

If the roller coaster were designed with a taller hill after the start, there would not be enough energy to get the roller coaster up that hill. In an actual ride, there would need to be a drive chain that provides that energy from chemical or electrical sources to get the cart up to the top of that next tall hill.

Solving kinematics equations using conservation of mechanical energy

When solving for properties of an object in free fall (ignoring forces such as friction and air drag), the conservation of mechanical energy can be used instead of using the equations of accelerated motion introduced earlier in this textbook.

Say a 0.1 kg apple is dropped from a height of 400 m, and that apple does not experience any air drag or friction as it falls through the atmosphere to the ground. There is no wind, and the apple has no horizontal velocity at any point in time throughout its journey. The initial form of energy of the apple is entirely GPE, as there is initially no motion. The GPE (and total ME) of the apple is

$$\begin{aligned} GPE &= (0.1\text{ kg})(9.8\text{ m/s}^2)(400\text{ m}) \\ &= 392\text{ J}. \end{aligned}$$

How fast will the apple be moving at the instant before it hits the ground? At that point, the height of the apple will be 0 m, and its mechanical energy will be entirely in the form of KE (GPE = 0 J). We can set the KE equal to 392 J and solve for the speed, which is

$$\begin{aligned} v &= \sqrt{\frac{2KE}{m}} \\ &= \sqrt{\frac{2(392\text{ J})}{0.1\text{ kg}}} \\ &= \sqrt{7840\text{ m}^2/\text{s}^2} \\ &= 88.5\text{ m/s}. \end{aligned}$$

The table below shows the values of GPE and KE at different points in the journey of the apple. The values of KE at any height can then be used to solve for the instantaneous speed of the apple.

h (m)	ME (J)	GPE (J)	% GPE	KE (J)	% KE
400	392	392	100	0	0
300	392	294	75	98	25
200	392	196	50	196	50
100	392	98	25	294	75
0	392	0	0	392	100

Alternatively, if we know the velocity of the apple at any point in its journey, we can use conservation of mechanical energy to solve for its height at that moment in time. Say the apple is known to have a speed of 20 m/s. What is the height? The value of KE can be calculated as

$$\begin{aligned} KE &= \frac{1}{2}(0.1\text{ kg})(20\text{ m/s})^2 \\ &= 20\text{ J}. \end{aligned}$$

Because the total mechanical energy of the apple is 392 J, if 20 J is in the form of KE, the remaining 372 J must be in the form of GPE. Set the value of GPE equal to 372 J and solve for the height, and we find that

$$\begin{aligned} h &= \frac{GPE}{mg} \\ &= \frac{372\text{ J}}{(0.1\text{ kg})(9.8\text{ m/s}^2)} \\ &= 379.6\text{ m} \end{aligned}$$

Systems with dissipative forces

The video below demonstrates a good example of a system that contains some dissipative forces like air drag and friction. Dr. Pasquale stands against a wall, holding a pendulum right up against her nose. The pendulum is not moving, so all of its mechanical energy is in the form of GPE. She then releases the pendulum by letting it fall, and the large mass swings away from her, gaining KE as it gets to the bottom of its motion. It gains GPE again on the upswing, and this conversion occurs in reverse as it swings back to her face.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=551#oembed-4

By the time the heavy mass reaches Dr. Pasquale's face again, the ball has lost some of its mechanical energy, which means that it is no longer able to get to the same height. This means that as long as she doesn't move her face away from the wall, she can be very confident that the ball will not hit her in the face when it returns. The heavy mass experiences air drag as it travels through the air, and the pin holding it to the ceiling experiences friction as it rubs against the ceiling supports.

In much the same way, a roller coaster, which we discussed earlier, must start out with a sufficient amount of mechanical energy to give the ride enough energy to continue throughout its entire motion without any more input work. If there's only one drive chain bringing the roller coaster cart up to the top of the tallest hill, that GPE must be enough to not only get the roller coaster to

the end of the track, but also be able to account for the energy converted to sound and heat due to friction and air drag.

Work

We need to be careful when discussing the concept of work in a physics context. You may use the word work to talk about something you do to make money. Or maybe you think about going to the gym or otherwise physically exerting yourself. Be careful! Work, when we're using it to refer to a physics phenomenon, has a very specific meaning that cannot be interchanged with our loose definition of "work" that we use in everyday language.

Work is energy that is transferred to an object by exerting a force on that object through a distance. Work has a symbol of the uppercase letter W. The units are joules (J). The equation for work is

$$W = F_{//}d, \tag{7.1}$$

where d is the displacement of the object and $F_{//}$ is the component of force that is parallel to the displacement of the object (the // symbol denotes the concept of "parallel"). In order to calculate the work done by the force, it is important to only use the component of the force that is parallel to the displacement.

While work is a scalar quantity, indicating that it has no direction associated with it, it is still possible to do positive or negative work. Positive work means that the parallel force points in the same direction as the displacement, and negative work means that the parallel force points in the opposite direction as the displacement. In this manner, we can discuss the concept of net work, which considers the effect of both applied forces and any dissipative forces (such as friction and air drag) that act on an object.

When a force acts on an object, Newton's second law states that the object will experience an acceleration (which could cause the object to speed up, slow down, or change direction). It's possible that the displacement, or motion of the object, and the force may not be pointed along exactly the same line. A great example of this is the motion of a wheeled suitcase, depicted in Figure 7.3.

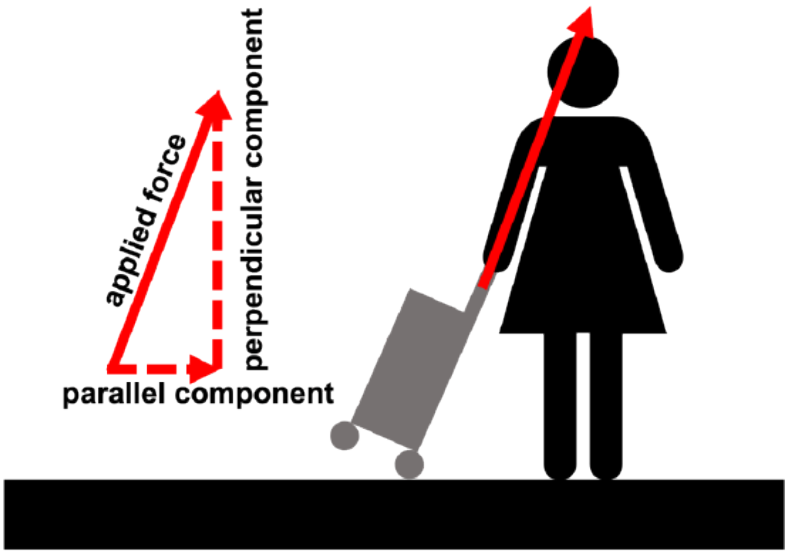


Figure 7.3 – A person applies a force to a wheeled suitcase to pull it along the floor (right). The parallel and perpendicular components of the applied force are shown (left). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A wheeled suitcase can be rolled along a floor, which is much easier than older styles of luggage that you had to carry. This way, most of the weight of the suitcase can be supported by the floor. The wheels allow the force we apply to the handle to cause the suitcase to move in the same direction as our body. We apply the force on the handle at an angle, but the suitcase does not move along the same line as our force.

The traveler pulls the suitcase in Figure 7.3 horizontally to the right. The horizontal distance that the suitcase moves is the displacement. The applied force (shown on the left of Figure 7.3) is broken down into components: a component that's parallel to the displacement, and a component that's perpendicular to the displacement.

Only the parallel component of force is doing work. And the work is equal to the parallel component of force multiplied by the displacement of the suitcase. The perpendicular component of the force is keeping the suitcase upright, but it isn't doing any actual work, from a physics perspective.

Say the traveler in Figure 7.3 applies 113 N of parallel force on the suitcase, with a horizontal component of 15 N and vertical component of 112 N. The traveler pulls the suitcase horizontally along the ground a distance of 2 m. The work done on the suitcase by the traveler is

$$\begin{aligned} W &= (15\text{ N})(2\text{ m}) \\ &= 30\text{ J}. \end{aligned}$$

This work is positive: the parallel component of the force and the displacement of the object both point in the same direction.

Suppose there's 8 N of friction opposing the motion of the suitcase. The work done by friction is

$$\begin{aligned} W &= (-8\text{ N})(2\text{ m}) \\ &= -16\text{ J}. \end{aligned}$$

Friction does -16 J of work (negative, because force and displacement oppose each other). The net work that is done on the suitcase is

$$\begin{aligned} W_{net} &= 30\text{ J} - 16\text{ J} \\ &= 14\text{ J}. \end{aligned}$$

What if a force is applied perpendicular to the motion of an object? Consider holding a heavy bowling ball at a constant height as you walk at a constant speed from one side of a room to another. The displacement vector will be a horizontal line in the direction of your motion. The force that you apply to the bowling ball, however, points up. The force and the displacement are perpendicular. While it may be hard for you to carry the bowling ball, and you will likely feel like you're exerting yourself, from a physics perspective, you are doing zero work on the bowling ball!

Work-energy theorem

As just described, when a force is applied on an object parallel to its displacement, work is done on that object. Newton's second law tells us that force equals mass times acceleration, implying that the object will speed up, slow down, or change direction. If the work done on the object is positive, it will speed up, gaining kinetic energy. If the work done on the object is negative, it will slow down, losing kinetic energy.

By how much? The work-energy theorem can be used to quantify this. The work-energy theorem states that work is equal to the change in kinetic energy of an object.

$$\begin{aligned} W &= \Delta KE \\ &= KE_{final} - KE_{initial} \end{aligned}$$

If 10 J of positive work is done on an object, then the kinetic energy of that object will increase by 10 J.

Consider the traveler from Figure 7.3 who does 30 J of work on a suitcase having a mass of 50 kg (in this case, there is no friction). If the suitcase starts at rest (which corresponds to an initial KE of 0 J), we can use the work-energy theorem to calculate the final kinetic energy of the suitcase: 30 J. The final velocity of the suitcase is therefore

$$\begin{aligned} v_{final} &= \sqrt{\frac{2(30\text{ J})}{50\text{ kg}}} \\ &= \sqrt{1.2\text{ m}^2/\text{s}^2} \\ &= 1.1\text{ m/s}. \end{aligned}$$

Power

Mechanical work is useful to us, especially when harnessed in a device such as a motor. Mechanical work is done in the engine of a car to move us around from one place to another. But work by itself isn't always enough information. How do we describe the

engine of a car? Usually we talk about how much horsepower it has. What is horsepower, and what does it measure?

Horsepower is one unit that's used to quantify the concept of power. While we will not use horsepower in this class, it is used frequently in the United States. It was originally used to relate the capability of motors to horses.

In short, power quantifies how quickly work is done. Power has a symbol of the uppercase letter P. We will use the SI unit watt (W), which is equal to joules per second (J/s). The equation for power is

$$P = \frac{W}{t},$$

where W is the work done by the person, object, or machine, and t is the amount of time it takes to do that work.

A golf cart and a sports car will both be able to move us from point a to point b. But a golf cart will speed up much more slowly than the sports car. The golf cart is therefore less powerful.

Consider a 500 kg golf cart that speeds up from rest to 9 m/s in the time span of 60 s. We can use the work-energy theorem to calculate the work done by the motor, and then divide by time to calculate the power. The work is equal to the change in kinetic energy. Because the golf cart was initially at rest, the initial KE is zero. The final KE (and work) is therefore

$$\begin{aligned} W &= \frac{1}{2}(500\text{ kg})(9\text{ m/s})^2 \\ &= 20250\text{ J}. \end{aligned}$$

To calculate power, divide by time. The power of this golf cart is

$$\begin{aligned} P &= \frac{20250\text{ J}}{60\text{ s}} \\ &= 337.5\text{ W}. \end{aligned}$$

For context, if you are familiar with the unit of horsepower, 337.5 W corresponds to approximately 0.5 hp. (The conversion is 1 hp = 746 W.)

Compare that golf cart to a sports car with a mass of 1600 kg that speeds up from rest to 27 m/s in 3 s. First, solve for work using the work-energy theorem. We find that work is equal to

$$\begin{aligned} W &= \frac{1}{2}(1600\text{ kg})(27\text{ m/s})^2 \\ &= 583200\text{ J}. \end{aligned}$$

The power of the sports car is

$$\begin{aligned} P &= \frac{583200\text{ J}}{3\text{ s}} \\ &= 194400\text{ W}. \end{aligned}$$

This sports car has a power rating of 194400 W (260 hp). In case we needed mathematical evidence, a sports car is much more powerful than a golf cart!

Simple machines

A simple machine is used to change either the direction or magnitude (or both) of an applied force. There are essentially six types of simple machines: levers, wheels and axles, pulleys, inclined planes, wedges, and screws. (Some would argue that there are really only four types by stating that the wedge is just two inclined planes placed back-to-back, and that the screw is an inclined plane wrapped around an axle.)

Some simple machines can multiply an input force. This force multiplication is called mechanical advantage. Mechanical advantage has a symbol of MA. The equation for mechanical advantage is

$$MA = \frac{F_{out}}{F_{in}},$$

where F_{out} is the force output generated by the machine (say, the weight of the heavy object that the machine is lifting) and where F_{in} is the force input by the user of the machine. Because in this equation, one force is divided by another, mechanical advantage is a unitless quantity.

What is the tradeoff that we have to deal with when using a simple machine? Hopefully your understanding of physics at this point makes you feel profoundly uncomfortable with the idea that something could come out of nothing. In fact, while force can be multiplied, the work put into the simple machine will be equal to the work output of the simple machine (ideally). Because work is equal to force times distance, the operator of the simple machine will apply less force than the machine will output, at the cost of applying that force over a longer distance than the output of the machine will generate.

Efficiency quantifies how well work input is converted to work output. The symbol for efficiency is the lowercase letter e. The equation for efficiency is

$$e = \frac{W_{out}}{W_{in}},$$

where W_{out} is the output of the machine, and W_{in} is the input to the machine. Efficiency is unitless, but we usually multiply the result by 100 to express it as a percentage. In ideal situations, efficiency will be 100%. However, in reality, friction, air drag, or other situations such as bending of a simple machine will cause the efficiency of a simple machine to be less than 100%.

Lever

Perhaps you've heard the Archimedes quote, "Give me a lever long enough and a fulcrum on which to place it, and I shall move the world." While literally moving the world would be impractical, not to mention inconvenient, levers can be used to lift a heavy object using less force than we get out of the system. A seesaw, which is an example of a lever, is shown in Figure 7.4.



Figure 7.4 – View of two children playing on a seesaw (lever) at a playground in Chicago, Illinois. A girl is hanging by her hands underneath the raised half of the seesaw, and a boy is holding down the lowered half of the seesaw. Two children playing on a seesaw at the Webster School playground by DN-0000074, Chicago Daily News collection, Chicago History Museum.

To use a lever to generate mechanical advantage requires a pivot that is not located in the center of the lever (as it is in Figure 7.4, where the mechanical advantage is equal to one). In this case, the mechanical advantage comes from lifting a large force over a short distance using a small force over a long distance, shown in Figure 7.5. The shorter the distance between the end of the lever containing the heavy object and the pivot, the larger the mechanical advantage, all other things being equal.

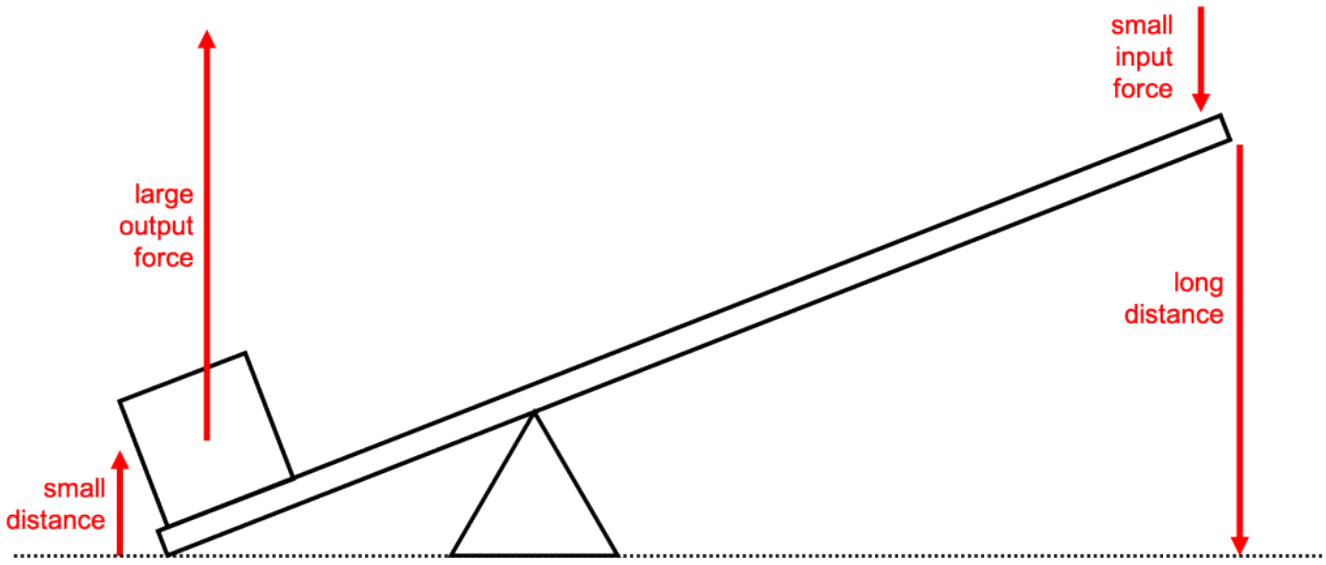


Figure 7.5 – Mechanical advantage is generated with a lever by placing the pivot away from the center. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Since energy is conserved, you cannot get more work output than the work input. Under ideal conditions (no friction, deformation, or other dissipative effects), the work output would be equal to the work input:

$$W_{out} = W_{in}$$
$$F_{out}D_{out} = F_{in}D_{in}.$$

Since there will always be some friction at the pivot and some bending of the lever, not all of the input work goes into lifting the load. Friction at the pivot dissipates some of the input work into thermal energy and the bending of the lever stores some of the input work as elastic potential energy. Thus, the work that actually goes into lifting the load is slightly less than the work input ($W_{out} < W_{in}$):

$$W_{out} + \text{Elastic Potential Energy Stored} + \text{Thermal Energy Dissipated} = W_{in}.$$

Wheel and axle

A wheel and axle is also a simple machine. A linear force can be used to cause rotation, and the force can be multiplied if the wheel and axle system is set up to have a mechanical advantage. Wheelchairs, bicycles (Figure 7.6), and cars are all examples of objects that use the wheel and axle simple machine.



Figure 7.6 – A bicycle is an example of a technology that uses the wheel and axle simple machine. Triumph Cycle bicycle, by Andrew Dressel, is licensed under CC BY-SA 3.0.

Pulley

Pulleys are capable of changing the direction of a force. They can also multiply a force. The mechanical advantage of a pulley system is related to the number of sections of the string (or rope) used in the pulley system that support the load.

A single fixed pulley (Figure 7.7, Scenario A) doesn't create any mechanical advantage, but allows us to pull down on an object to lift it, rather than pulling up, which may be more difficult for us to do, especially if we have bad knees or a bad back! However, a single pulley that connects to the load and is connected with two sections of string (Figure 7.7, Scenario B) will have a mechanical advantage of two.

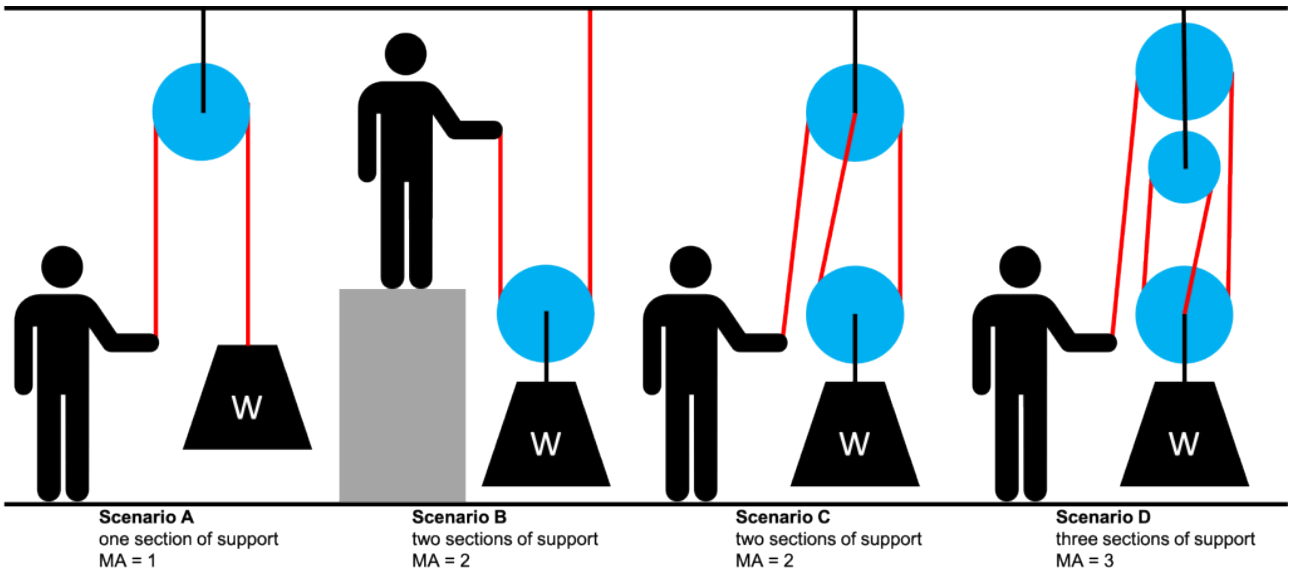


Figure 7.7 – The mechanical advantage of a pulley is equal to the number of sections of string (or rope) directly supporting the load. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The more sections of support there are in the system, the less input force we need to apply to lift an object. (Figure 7.7, Scenarios C and D.) However, the caveat is that we must apply the force over a longer distance. The concepts of mechanical advantage and efficiency still apply.

Pulleys are used in many different applications, including redirecting forces from a yoke or stick and rudder pedals on an airplane to flight controls (elevator, ailerons, rudder), cable machines at the gym, elevator cables, and hoists.

Inclined planes

Inclined planes are also known as ramps. You've probably seen them at building entrances, providing an accessible way to enter a building that would otherwise require steps (see Figure 7.8). Not only do ramps provide accessibility, but they also reduce the amount of force that's required to move along the ramp. The mechanical advantage of the inclined plane is related to the steepness of the slope of the ramp. As mentioned, the lower the force, the longer the distance that the force must be applied in order to obtain the desired amount of work output. As a result, the smaller the force you wish to apply, the greater the length of the ramp is required. This should make sense since if the ramp has a shallower angle, the ramp will need to be longer in order to reach the desired height.



Figure 7.8 – A wheelchair user utilizes a ramp (inclined plane) to enter a building. Access to the ramp in a wheelchair, by Gov39.ru, is licensed under CC BY 4.0.

Wedge

A wedge is a simple machine found in things such as axes, knives, scissors, and wood splitters (see Figure 7.9). A force applied to the blunt edge of the wedge is transmitted to the pointy edge, which in turn redirects the force in a direction that is perpendicular to the "inclined" surfaces of the wedge. As such the wedge can be used to split something apart.



Figure 7.9 – A wood splitting wedge. Wedge, by Shakespeare, is licensed under CC BY-SA 3.0.

Screw

A screw takes rotational motion and converts it to linear motion. You may have used screws or bolts (see Figure 7.10) when putting together furniture or building something. As stated earlier, a screw can be thought of as an inclined plane wrapped around an axle. Like a wedge, the force on the screw head is transmitted in a direction perpendicular to the “incline,” which then acts somewhat like a wedge to drive the screw into the object.



Figure 7.10 – A bolt (with a nut) and a screw. Bolt and screw, by Tomasz Sienicki, is licensed under CC BY 3.0.

Practice questions

Conceptual comprehension

1. What is the conceptual issue with the cartoon shown in Figure 7.11? Should the person at the bottom of the ladder flinch?

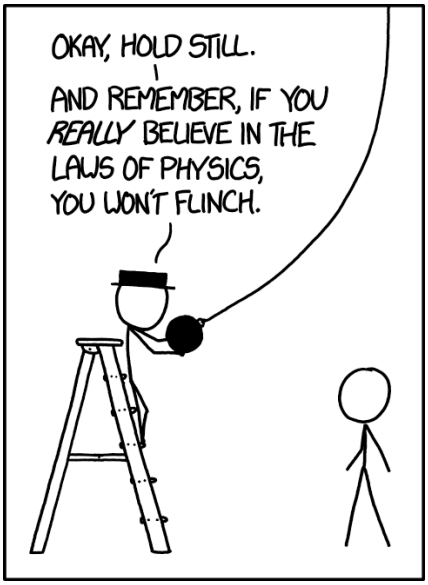


Figure 7.11 – Is this cartoon conceptually sound? Should the person at the bottom of the ladder flinch? Laws of Physics, by Randall Munroe, is licensed under CC BY-NC 2.5.

Numerical analysis

1. Calculate the gravitational potential energy of an object with a mass of 10 kg when it is positioned 5 m above the Earth's surface.
2. Calculate the gravitational potential energy of a 15 kg object that is situated 20 m above the Earth's surface.
3. Calculate the gravitational potential energy of a spaceship with a mass of 15,000 kg positioned 50 m above the surface of the Moon. (The Moon has an acceleration due to gravity of approximately 1.6 m/s^2.)
4. Calculate the kinetic energy of a 12 kg object when it is moving at a speed of 20 m/s.
5. Calculate the kinetic energy of an object with a mass of 3.2 kg when it is traveling at a velocity of -8.5 m/s.
6. A 2 kg object initially at rest is released from a height of 10 meters above the ground. Calculate its total mechanical energy at the starting point and when it's 5 meters above the ground. Calculate the KE at both heights.
7. A 500 g ball is dropped from a height of 15 meters. Determine its velocity just before it hits the ground.
8. A 5 kg object is sliding down a frictionless hill with a height of 8 meters. Calculate its speed at the bottom of the hill.
9. A 3 kg object is swinging back and forth on a pendulum. If its maximum height above the lowest point is 2 meters, calculate its velocity at the lowest point.
10. A car engine exerts a force of 5,000 N to move a car a distance of 100 meters. Calculate the work done by the engine.
11. A person pulls a sled with a force of 65 N (the force has a horizontal component of 33 N and a vertical component of 56 N), moving it a horizontal distance of 10 meters. Calculate the work done by the person.
12. A perfectly efficient pulley system is used to lift a 100 kg weight to a height of 5 meters in 10 seconds. Calculate the power output of the pulley system.
13. A ramp is used to lift a 500 kg object to a height of 4 meters in 20 seconds. Calculate the power applied to lift the object using the ramp. You may safely ignore forces such as friction and air drag.
14. A person pushes down on a lever. The distance the person pushes down is 50 cm. The load is lifted upward a distance of 10 cm. Calculate the mechanical advantage of the lever.
15. A person pulls a 100 N object up an inclined plane a length of 5 meters. The force the person exerts on the object is 80 N. Calculate...
    1. ...the distance the object moved upward.
    2. ...the mechanical advantage of the inclined plane.
16. A pulley system consists of two pulleys. Calculate the mechanical advantage of the pulley system.
17. A person uses a lever to lift an 80 N force a distance of 20 cm. The person exerts a force of 40 N over a distance of 50 cm to obtain this output. Calculate...
    1. ...the mechanical advantage of the lever.
    2. ...the work input to the simple machine.
    3. ...the work output from the simple machine.
    4. ...the efficiency of the lever.
18. A pulley system requires an input force of 60 N exerted over 10 m to lift a load, but due to friction and other losses, the pulley system only does 480 J of useful work. Calculate...
    1. ...the work input to the simple machine.
    2. ...the efficiency of the pulley system.

Hands-on experiments

1. Create a simple lever using a ruler or a wooden board and a small object as a load. Experiment with different positions for the fulcrum and observe how it affects the effort needed to lift the load. How does this relate to the mechanical advantage of the lever?
2. If you have a wind-up toy, wind it up, release it, and measure how far it travels. Try different amounts of winding, and see how the results change. What does this experiment tell you about work, energy transfer, and efficiency? What external forces may be affecting the efficiency of this toy?

This page titled Chapter 7: Work and energy is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 8: Rotational Motion

Summary

Rotational motion is all about things that rotate. So far, our discussion has been focused on translational motion: motion along a straight line. But translational motion can't help us fully understand the motion of things like bicycle wheels, bowling balls, figure skaters, propellers, or even the motion of stars and planets. We need new concepts to help us understand rotational motion.

Rotational speed

Rotational speed quantifies how quickly (or slowly) something rotates about an axis. The symbol for rotational speed is the lowercase Greek letter omega (ω). There are many different units that we could use for rotational speed, depending on how we want to measure a rotation, including

- revolutions (or rotations) per unit of time (e.g., RPM, rev/s),
- degrees per unit of time (e.g., degrees per second), and
- radians per unit of time (e.g., rad/s).

One revolution consists of 360 degrees or 2π radians. These are the ratios used to convert between the different types of rotational speed units.

All objects on the same rotating platform will experience identical rotational speeds. For example, on a playground merry-go-round (as seen in Figure 8.1), all of the people sitting on the merry-go-round as it spins will be traveling at exactly the same rotational speed. They will all move 360 degrees (or 2π radians) as the disk makes a single rotation in the same interval of time.



Figure 8.1 – A child sits on a merry-go-round as it spins. All objects on the spinning disk are rotating with the same rotational velocity. Playing on the Merry go round, by Charles Knowles, is licensed under CC BY 2.0.

A geostationary satellite orbits the Earth at a distance such that it appears not to move with respect to objects on the ground. An observer standing directly underneath the satellite would not be able to see any motion of that satellite.

We can calculate the rotational speed of a geostationary satellite, knowing that it makes a complete orbit of the Earth every 24

hours. First, convert 24 hr to units of minutes and seconds

$$\begin{aligned} 24 \text{ hr} &= (24 \text{ hr})(\frac{60 \text{ min}}{1 \text{ hr}}) = 1440 \text{ min} \\ &= (1440 \text{ min})(\frac{60 \text{ s}}{1 \text{ min}}) = 86400 \text{ s}. \end{aligned}$$

Using units of revolutions per minute (RPM), one revolution of the satellite occurs every 1440 min, giving a rotational speed of

$$\begin{aligned} \omega &= \frac{1 \text{ rev}}{1440 \text{ min}} \\ &= 6.94 \times 10^{-4} \text{ RPM}. \end{aligned}$$

We can convert this rotational speed to use units of degrees per second

$$\begin{aligned} \omega &= (6.94 \times 10^{-4} \text{ RPM})(\frac{360^\circ}{\text{rev}})(\frac{1 \text{ min}}{60 \text{ s}}) \\ &= 0.004 \ ^\circ/\text{s}. \end{aligned}$$

Converting to rotational speed with units of radians per second, we find that

$$\begin{aligned} \omega &= (6.94 \times 10^{-4} \text{ RPM})(\frac{2\pi \text{ rad}}{\text{rev}})(\frac{1 \text{ min}}{60 \text{ s}}) \\ &= 7.27 \times 10^{-5} \text{ rad/s}. \end{aligned}$$

Tangential velocity

Have you ever ridden on a merry-go-round or a carousel? The way we experience speed on a rotating platform has to do not only with the rotational speed of the ride, but also with how far we are from the center of rotation. This quantity is called tangential velocity. While rotational speed is constant everywhere along a rotating object, the tangential velocity is not.

Tangential velocity is very similar to the velocity that was defined earlier in this textbook. It is the linear distance that something travels as it rotates per unit of time. Instead of quantifying the rotation in terms of revolutions, degrees, or radians, the actual distance traveled (with units of meters) is used. The symbol is a lowercase v with a subscript of t to make it clear that it's a tangential velocity (v_t). The unit is m/s.

The direction of the tangential velocity is tangent to the motion of the rotation: hence the name tangential velocity. As an object rotates, the tangential velocity will continue to rotate in order to point tangent to the direction of the rotational motion. Consider using a sling to rotate an object around in circles before letting it go. The tangent is the direction the object will go when it is released and stops rotating.

To calculate tangential velocity, determine the physical distance that an object moves as it travels along the spinning object per unit of time.

The distance around a circle is known as the circumference. It is equal to

$$C = 2\pi r,$$

where r is the radius of the circle. Therefore, the magnitude of the tangential velocity will be equal to

$$v_t = \frac{2\pi r}{t},$$

where r is the distance (in meters) that the object is from the center of rotation and t is the time (in seconds) to complete one full rotation. Note that 2 times π divided by time is equal to the rotational speed of an object in units of rad/s. Therefore if the rotational speed is known with units of rad/s, tangential velocity can be calculated using the equation

$$v_t = \omega r.$$

The closer an object is to the center of rotation, the smaller the tangential speed. The farther away it is from the center of rotation, the higher the tangential speed. The reason that we feel like we're moving faster when we sit in the outer part of a carousel or

merry-go-round is because our tangential velocity is higher.

The altitude of a satellite in geostationary orbit is 36 million meters, and the radius of the Earth is 6.4 million meters. Therefore, a geostationary satellite is 42.4 million meters away from the center of rotation. The time of one rotation, as discussed above, is 86400 s. The tangential velocity of a geostationary satellite is therefore

Rotational acceleration

Acceleration, as we have seen earlier in this textbook, means one of three things: speeding up, slowing down, or changing direction. It's possible for the rotational speed of an object to change. Rotational acceleration therefore quantifies how rapidly the rotational speed of an object changes (speeds up, slows down, or changes direction). We will not quantify rotational acceleration in this textbook; it is only being mentioned to give you a more complete understanding of rotational motion.

Newton's second law states that force generates changes in linear motion (which are resisted by inertia). In order to generate changes in rotational motion, a new quantity called torque is introduced. Below, we'll see that torque generates such changes. Furthermore, just as mass resists changes in linear acceleration, a quantity called rotational inertia resists changes in rotational motion.

Rotational inertia

Rotational inertia is a measure of how much an object resists changes to its rotational speed. If an object is not rotating, rotational inertia is related to how difficult it is to get that object to start rotating. If it's already rotating, rotational inertia is related to how hard it is to make that rotation speed up or slow down.

Rotational inertia comes from two things: the mass of an object and how that mass is distributed with relation to the center of rotation of the object. The more massive an object is, the harder it is to get it to start rotating. The distribution of that mass is important as well. The farther away the mass is distributed from the object's center, the harder it is to get it to start rotating.

In the video below, there are two objects that are very similar to each other: a hoop and a disk. The hoop is hollow on the inside, and the disk is solid. But they have the same radius and both have a mass of 0.5 kg (see Figure 8.2).

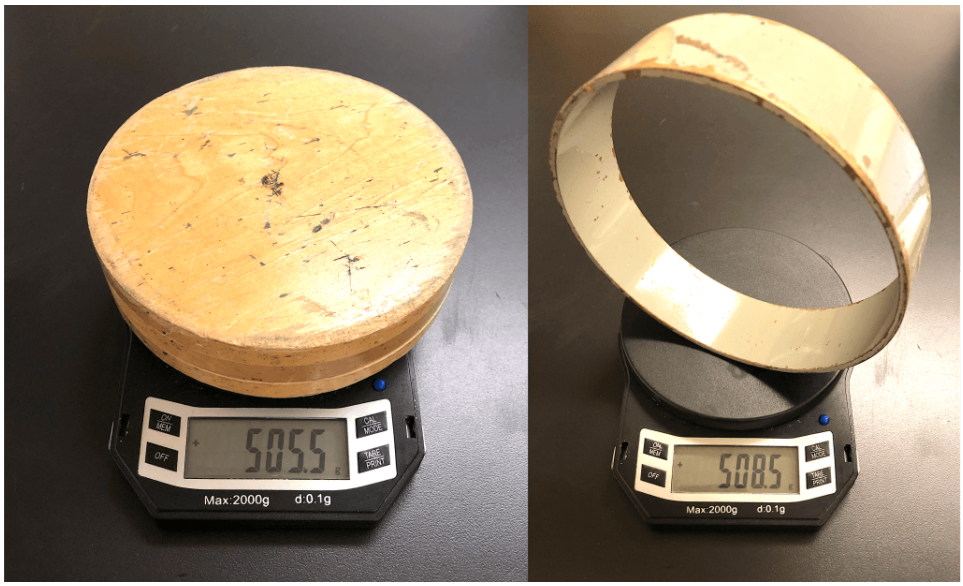


Figure 8.2 – A solid wooden disk has a mass of 505.5 g (0.5 kg), left. A metallic hoop has a mass of 508.5 g (0.5 kg), right. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Both objects have the same mass and therefore the same translational inertia (in other words, it would take the same amount of effort to shake either object back and forth). Because the disk is solid, its mass is uniformly distributed throughout. The hoop,

however, has its mass concentrated at the edges of the circle, far from the center of the hoop. This means that the hoop will have more rotational inertia than the solid disk. Therefore, we can predict that the hoop will take longer to roll down the track than the solid disk. We can see that this is the case by watching the video below.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=578#oembed-1

It is possible to quantify rotational inertia using equations depending on the shape and distribution of mass of an object. However, the mathematics behind rotational inertia is beyond the scope of this textbook, so it will not be covered.

Torque

Torque is a twist or a turn that causes an object to change its rotational motion. Just as force creates acceleration and makes something moving in a straight line speed up, slow down, or change direction, torque creates rotational acceleration and causes the rotation to speed up, slow down, or change direction. Torque has a symbol of the lowercase Greek letter tau (τ). The unit of torque is called the newton-meter (Nm). Because torque is so different from concepts such as work and energy, we do not simplify this as joules, but just use newton-meters (Nm), even though they share common units.

Torque is equal to the perpendicular force times the distance between the applied force and the center of rotation, or

$$\tau = F_{\perp} r.$$

The concept of perpendicular force is important, as not all directions of force contribute to making an object rotate.

Consider using a wrench to tighten a bolt in a piece of furniture. Simply pushing or pulling along the handle of the wrench will not cause the wrench to rotate. It is necessary to push or pull perpendicular to the handle. Figure 8.3 shows three scenarios. In scenario (a), the applied force is completely parallel to the center of rotation. This applied force will not cause any rotation of the wrench, and therefore the torque will be zero. In scenario (b), the applied force is neither all parallel or all perpendicular, but can be broken into parallel and perpendicular components. Only the perpendicular component is applying torque to the wrench, and only that force component will cause the wrench to rotate. In scenario (c), the applied force is all perpendicular. The entirety of the applied force is contributing to the torque. In all three scenarios, the dashed line indicates the distance between the applied force and the center of rotation.

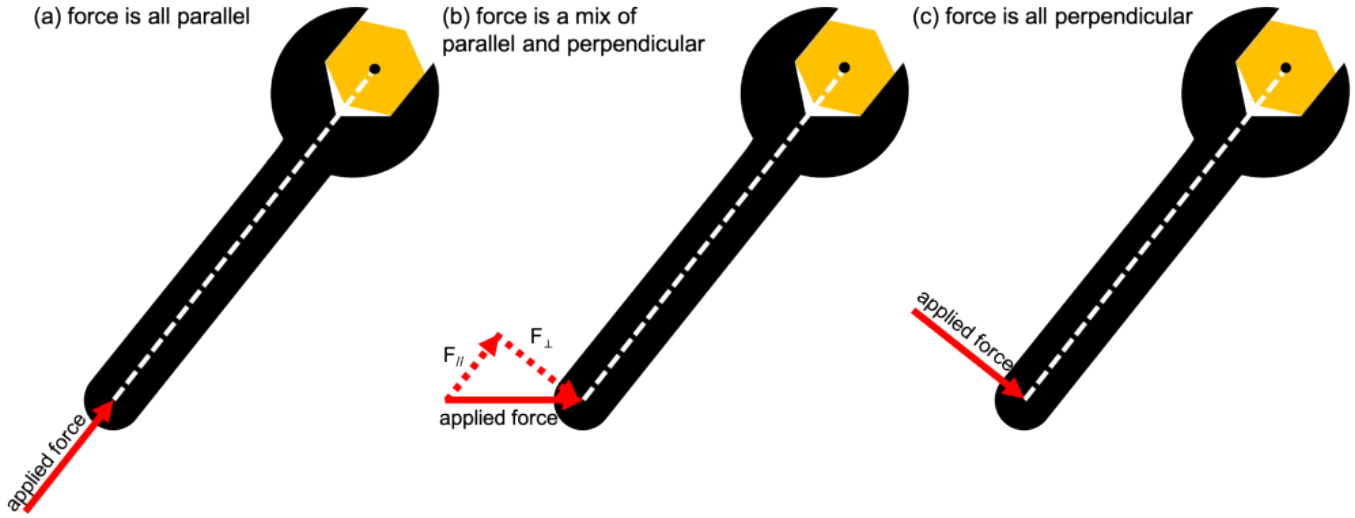


Figure 8.3 – A force is applied to a wrench to cause a rotation. (a) The applied force is parallel to the center of rotation. (b) The applied force is neither parallel nor perpendicular to the center of rotation, and can be split into components. (c) The applied force is perpendicular to the center of rotation. The dashed line indicates the distance between the applied force and the center of rotation. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

If you're familiar with the term "lever arm," it has to do with that distance between the force and the center of rotation. The longer the lever arm, the easier it is to rotate something. This is why there are extensions that we can put on a socket wrench. If

something is really tight, we need to apply a lot of torque to loosen a bolt (for example). There's only so much force we can apply, so we can use the r in the torque equation to make up for our relative lack of strength.

Mechanical equilibrium

Previously in this textbook, we learned that mechanical equilibrium occurs when an object is traveling at a constant velocity and occurs when the net force on an object is equal to zero. We now have to expand this definition.

Mechanical equilibrium occurs when an object is traveling at a constant velocity **and** at a constant rotational speed. This occurs when the net force on an object is equal to zero **and** the net torque on an object is equal to zero.

We learned that forces are positive when they go up or to the right, and negative when they go down or to the left. How do we give sign to torque? We do that by looking at the direction that it would make an object rotate. By common convention, if an object would rotate counterclockwise, we call that a positive torque. A clockwise rotation is generated by a negative torque.

Center of mass

Center of mass is another important concept in rotational motion. The center of mass is the point in an object around which it rotates freely. And if you could somehow pick an object up at that exact point, it would balance perfectly.

Another way to define center of mass is this: the center of mass in an object is the point where, on average, the mass is distributed. In an object that has a uniform shape and density, the center of mass will be in the center of the object.

A meter stick made of wood has a relatively uniform density. In Figure 8.4, Dr. Pasquale balances a meter stick. The center of mass is located at (approximately) the center of the meter stick.

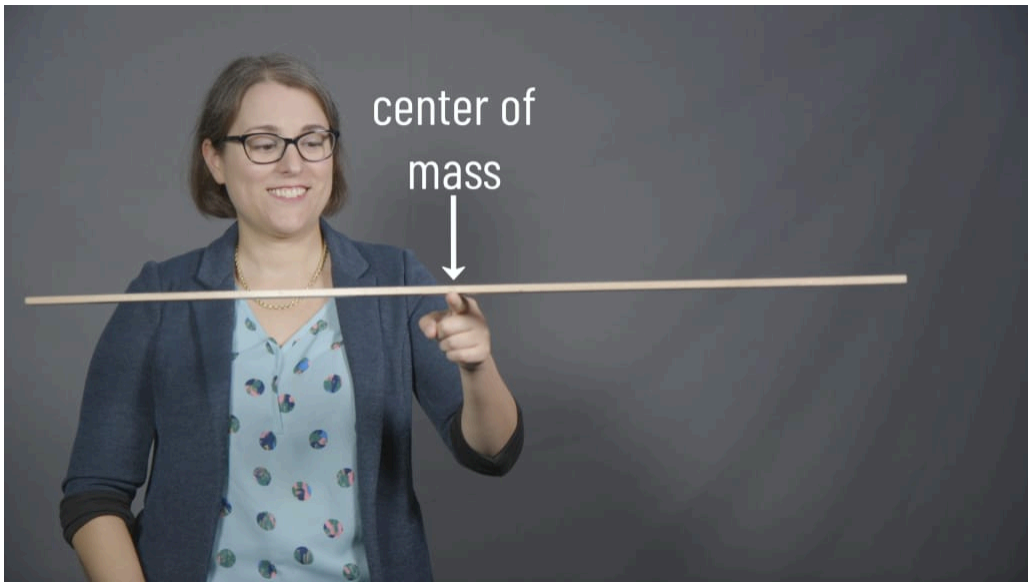


Figure 8.4 – Dr. Pasquale balances a meter stick by holding it with one finger directly under its center of mass, which is located at the center of the stick. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In an object like a hammer, which is made of different materials, the center of mass is no longer in the center of the object. It's closer to where most of the mass of the object is concentrated. In this case, it will be closer to the head of the hammer. In Figure 8.5, Dr. Pasquale balances a hammer using a single finger. Note the position of the center of mass is much closer to the head of the hammer than the handle.



Figure 8.5 – Dr. Pasquale balances a hammer by holding it with one finger directly under its center of mass, which is located closer to the metal end of the hammer. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

It's possible for an object to have a center of mass located where there is no actual mass. For example, a donut (assuming it has uniform density) will have the center of mass located in the center (see Figure 8.6), even though no actual mass is located in that spot.



Figure 8.6 – A donut of uniform density will have a center of mass in the center: where no actual mass exists. Woman holding chocolate donut closeup, by Nenad Stojkovic , is licensed under CC BY 2.0.

The hoop from the rotational inertia video (Figure 8.2) also has a center of mass located in the center, even though no actual mass exists at that spot.

Center of gravity

Maybe you've heard the term center of gravity (CG). What's the difference between that and the center of mass (CM)? The answer is: they're almost the same thing. Center of mass has to do with how the mass of an object is distributed. Center of gravity has to do with how the weight of an object is distributed.

If the force of gravity is equal at every spot in an object, then CM and CG are equal. That's true for most objects most of the time.

The only time we might see a difference is in really tall objects, like skyscrapers.

As of the time this textbook was written, the Burj Khalifa (Figure 8.7) was the tallest building in the world, standing at an impressive 830 meters tall. Assuming that the Burj Khalifa is a rectangular cuboid of uniform density (which is not true, but we'll make the assumption for purposes of illustrating this concept), then the center of mass would be 415 meters from top and bottom, in the exact center of the building.



Figure 8.7 – The Burj Khalifa is the tallest building in the world (as of the writing of this textbook). Burj Khalifa, by Donaldytong, is licensed under CC BY-SA 3.0.

Because the force of gravity changes with respect to the distance from the center of the Earth, the parts of this building that are closer to the ground experience more gravitational force than the parts of this building that are higher in the air.

In that case, even though the mass is distributed in the center of the building, the weight distribution is going to be slightly closer to the bottom and will not be collocated with the center of mass. By how much? Again, assuming that the Burj Khalifa were a solid rectangle of uniform density (which it is admittedly not), the center of gravity would be only a couple of millimeters below the center of mass. That's not much of a difference. In practice the Burj Khalifa has much of its mass located closer to the base of the building than the top (see Figure 8.7), which would cause the offset between the center of mass and center of gravity to be even smaller than this idealized calculation.

So for the most part, it's perfectly valid to say that center of mass and center of gravity are the same thing. But we should understand that there may be exceptions.

Stability

Stability has to do with how easily an object will topple over. More accurately, stability has to do with this: If you take an object that is in equilibrium and disturb it, how will its motion change? Will it return to its state of equilibrium? Or, will the object move away from that state of equilibrium toward another state?

An object will stay put if it has a center of mass over the support base. Furthermore, an object is stable if the support base is wide enough to ensure that any motion that the object might make won't upset that condition.

In the video below, Dr. Fazzini holds a trapezoid-shaped block. At the center of mass of the block, there is a pointer that uses gravity to point downward. When the block is placed such that the pointer points to a spot over the support base, the block remains stable and stays put. In one configuration, the pointer points to a spot that is not over the support base, and when placed down, the block falls over.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This concept can also be demonstrated with the “leaning tower” in the video below. At first, the tower sits in its leaning configuration without toppling over. Because the composition of the tower is unclear, it’s not entirely certain where the center of mass is. But because the tower is staying put, the center of mass must be situated somewhere over the support base.

When the top piece of the tower is rotated by 180 degrees, it topples over to the left. That means that by rotating the top of the tower, the center of mass of the object was moved so that it was no longer over the support, but to the left of it.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Stability can also be demonstrated with a soda can. When a soda can is placed on a table (in the typical orientation that you would place a soda can), the center of mass is over the base, and it stays nice and steady. After drinking just the right amount of liquid, the center of gravity is moved downward closer to the base of the can. Now there’s a second point on which you can rest the can. Dr. Pasquale demonstrates this in the video below, and you can also see still images of both orientations in Figure 8.8.



Figure 8.8 – A soda can with just the right amount of liquid is stable on its base (left) as well as on the lip between the base and side of the can (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

While the soda can is capable of balancing on its base and on the small lip between the base and the side of the can, the latter configuration is not terribly stable. The application of even a very small torque on either the top or sides of the can will cause the soda can to topple over and disturb its equilibrium.

Stability is a very important consideration in aircraft such as airplanes. Pilots have to be careful to distribute the weight of the passengers and baggage in an airplane in such a manner that the airplane remains stable. (This is known as weight and balance.) This is a routine consideration that pilots make in every flight. If you've ever flown on an airline and been asked to move your seat more forward in the airplane, or to a seat farther to the back, it's probably because the pilot needed to adjust the balance of the airplane.

Angular momentum

Momentum is inertia in motion. Angular momentum is rotational inertia in rotational motion. Angular momentum is equal to the rotational inertia multiplied by the rotational speed. While we will not do calculations with that equation in this textbook, you do not need to know the equation to understand the topic.

Just as momentum is conserved in a system where there is no net external force, angular momentum will be conserved as well, as long as there is no external torque acting on a system. That means that as long as there's nothing to speed up or slow down a rotating object, its angular momentum will stay the same. Not necessarily the rotational speed, but the angular momentum.

In the video below, Dr. Pasquale has gotten herself rotating in a spinning chair while holding some dumbbells. As long as she doesn't move, and as long as friction doesn't slow her down too much, her rotational speed will stay the same because her rotational inertia will not change, and because there are no external torques.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

As Dr. Pasquale extends her arms out, she increases her rotational inertia. Because momentum is conserved, when her rotational inertia is increased, her speed will have to slow down to compensate, and it does. When she pulls her arms back in, she reduces her rotational inertia. Her speed will increase to compensate for that change, and to keep the angular momentum the same.

This is the same concept that you may have seen if you have watched or participated in figure skating. In a move called a scratch spin, a figure skater will start spinning, then pull their arms in tight and cross their legs to reduce their rotational inertia as much as possible. This has the effect of causing the skater to speed up, sometimes spinning at a very fast rate. At the conclusion of a spin, a figure skater will expand their arms and free leg, causing their rotation to slow down again due to the increase in their rotational inertia.

Centripetal force

Now we're going to talk not about rotation, but about motion along a circular path. When an object moves along a circular path at a constant speed, its direction is constantly changing. If the direction changes, there must be an acceleration. When there is acceleration, we know that a force must be acting on that object. We call that force centripetal force. The symbol for centripetal force is F with a subscript of C (F_c). The unit for centripetal force is the newton (N). The equation for centripetal force is

$$F_c = \frac{mv_t^2}{r},$$

where m is the mass of the object, v_t is the magnitude of the tangential velocity of the object, and r is the radius of the circular path. Centripetal force always points toward the center of the circular path.

Centripetal force isn't a separate force in and of itself, but is a force that must be supplied by some means to keep something moving in that circular path. Some common forms of circular motion, and the source of the centripetal force used to achieve that, are listed in the table below.

orbiting satellite	gravity
driving in a circular path on a flat road	friction
airplane turning	lift
rotating an object in a sling (see Figure 8.9)	tension
roller coaster loop	support (normal) force and gravity

When you drive a car around a circular path, the centripetal force is supplied by friction between the tires and the pavement. The faster the car is going, the higher that force needs to be. That's why we usually slow down to turn when we drive. Also, the tighter the turn we want to take, the more force that's needed. If the roads are slippery, then it's harder to get enough friction to move in a circular path.

Centrifugal force (not a real force)

It is often difficult for students first learning physics to accept the fact that the force required to make something move in a circle (the centripetal force) is a "center-seeking" force. That is, the force is directed toward the center of the circle. If you have ever ridden a merry-go-round on a playground or have ridden on a spinning ride at an amusement park, you may have noticed that it feels like you are being pulled outward. This "center-fleeing" sensation is called a centrifugal force. However, the centrifugal force is not a real force. Rather, this sensation is merely due to inertia and your perspective.

According to Newton's first law of motion, a moving object tends to continue moving in a straight line at constant speed unless acted upon by a force. Consider a car driving around a sharp curve. Suppose you are a passenger in the car while it makes a sharp turn to the left. If you chose not to wear your seat belt (not a good idea), you would "feel" like you are being pulled to the right toward the door, which is away from the center of the circular motion. As the car rounds the curve to the left, you, the passenger, want to keep on moving in a straight line. However, you are not being pulled outward. Rather, you are being pushed inward. Here's why: as the car turns to the left, the passenger door of the car moves into you and pushes you to the left, forcing you to round the curve with the car. You are sensing that you are being pulled outward due to Newton's third law of motion. The action of the door pushing inward on you (forcing you to turn in a circle) produces the reaction of you pushing against the door, giving you the sensation that you are being pulled outward when you actually are not.

Centripetal acceleration

Newton's second law states that forces create acceleration. Centripetal force causes an object to change its direction, which is a type of acceleration. We call that centripetal acceleration.

Centripetal force is equal to mass times centripetal acceleration. Using Newton's second law to calculate the equation for centripetal acceleration, we find that

$$a_c = \frac{mv_t^2/r}{m}$$
$$= \frac{v_t^2}{r}$$

Just as with centripetal force, centripetal acceleration always points to the center of the circular path.

In the video below, Dr. Pasquale rotates a cup of water vertically around in a circle, swinging it around over her head. The tension she applies to the string (which is apparent in Figure 8.9, left, with the string being held very taut) keeps the platform and cup moving in that circular path. The centripetal force points toward the center of the circle. So what keeps the water in the cup? What stops it from spilling out over Dr. Pasquale's head (see Figure 8.9, right)?

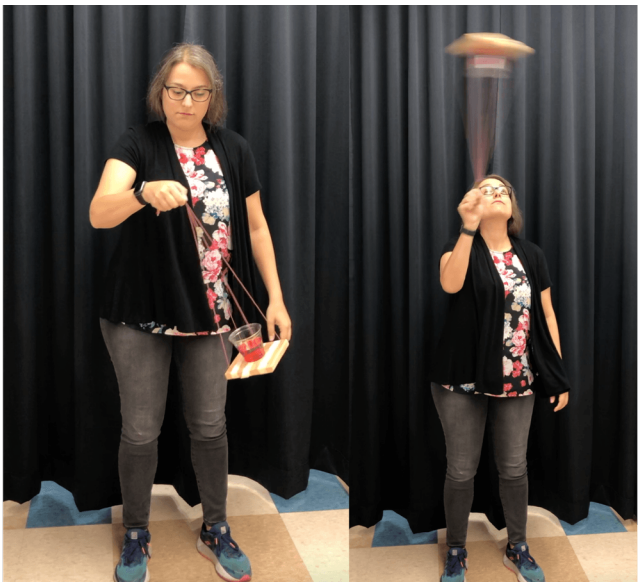


Figure 8.9 – Dr. Pasquale spins a cup of dyed water on a platform in a vertical circular path. The applied force to the string creates the rotational motion (left), and the inertia of the cup prevents the water from spilling out when the cup is upside-down (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Why doesn't the water spill onto the floor and all over Dr. Pasquale's head? The answer is inertia. The position of the water in the cup wants to stay where it initially was. So as long as the apparatus is rotated quickly enough, the inertia is going to keep the water firmly planted in the cup.

How fast is fast enough? We need to make sure that the centripetal acceleration of the water is greater than the acceleration due to gravity (9.8 m/s^2). In this example, the string and platform have a length of 0.46 m, which is equivalent to the radius of the rotational motion. One rotation is completed in 0.63 s. The tangential velocity is

$$v_t = \frac{2\pi(0.46\text{ m})}{0.63\text{ s}}$$
$$= 4.59\text{ m/s}.$$

Using this tangential velocity, we can calculate the centripetal acceleration as

$$a_c = \frac{(4.59\ \text{m/s})^2}{0.46\ \text{m}} = 45.75\ \text{m/s}^2.$$

This is indeed larger than the acceleration due to gravity. In fact, it is 4.67 times stronger than the acceleration due to gravity (4.67 “g’s”).

Further reading

- “Dancing” t-handle in zero g – Watch a video of a T-handle as it rotates about its intermediate axis, causing the rotation to be unstable. (An object that spins about its longest or shortest axis is stable, rotation along the intermediate axis is unstable.)
- Dorothy Hamill scratch spin – Watch as Dorothy Hamill starts spinning slowly and speeds up when she decreases her angular inertia. This demonstrates the conservation of angular momentum.

Practice questions

Conceptual comprehension

1. Why is a doorknob typically located as far as possible from the hinges of a door (see Figure 8.10)?



Figure 8.10 – A photograph of an office door. On the left side of the door are three hinges, and on the right side of the door is the doorknob. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Numerical analysis

1. A blade of a ceiling fan completes 120 rotations in one minute. Calculate the rotational speed of the fan in units of...
    1. ...RPM (rotations per minute).
    2. ...rotations per second.
    3. ...degrees per second.
    4. ...radians per second.
2. Vinyl records were commonly used to play music in the mid-20th century. A long-playing vinyl record would rotate at 33 ⅓ revolutions per minute, and would typically have a diameter of 25 cm. What is the rotational speed of this record? What is the tangential velocity at the outer rim of this record?
3. A bicycle wheel has a radius of 0.35 m and is rotating at 60 RPM (revolutions per minute). Calculate the tangential velocity of a point on the outer rim of the wheel.
4. Compact discs (CDs) were commonly used to play music in the late 20th century. A CD has a diameter of 12 cm, and rotates such that the tangential velocity remains constant at 1.2 m/s. What is the rotational speed of the CD when playing from its

outer rim? What is the rotational speed when playing from a point halfway between the center and outer rim?

5. A wrench is used to tighten a bolt with a force of 40 N at a distance of 0.2 m from the bolt. Calculate the torque applied to the bolt.
6. A seesaw is 4 m long and has a person with a mass of 60 kg sitting 2 m from one end. How far from the other end should a person with a mass of 80 kg sit to balance the seesaw?
7. A force of 13 N is applied at an angle to a wrench. If the lever arm is 0.4 m, calculate the torque produced, if 5 N of the force is applied parallel to the lever arm, and 12 N of the force is applied perpendicular to the lever arm.
8. A car with a mass of 1,200 kg is traveling around a curve with a radius of 50 m at a speed of 20 m/s. Calculate the...
    1. ...centripetal force required to keep the car traveling in that circular path.
    2. ...centripetal acceleration of the car.
9. A 0.5 kg ball is attached to a string and swung in a horizontal circle with a radius of 1 m. If the tension in the string is 10 N, calculate the...
    1. ...centripetal force acting on the ball.
    2. ...centripetal acceleration of the ball.

Hands-on experiments

1. If you have a hammer, hold the hammer by the head and rotate it back and forth as many times as you can in 5 seconds. The rotational inertia is low because the mass is concentrated at the center of rotation (your hand). Then, hold the hammer by the end of the handle and rotate it back and forth as many times as you can in 5 seconds. The rotational inertia is much higher now because the mass is concentrated away from the center of rotation. Which orientation was harder to rotate? Why is that?
2. Use the concept of torque to find the most and least efficient ways to open a door. Start by applying forces as close to the rotational axis as possible (near the hinge), and then try applying forces in the center of the door and finally as far from the rotational axis as possible (near the outer edge of the door, presumably near the doorknob). Which force application was the easiest to open the door? Which was the hardest? Why? Then try applying the force at different angles. Apply a force directly perpendicular to the door's motion. Then try applying the force at different angles. Finally, open the door and apply a force parallel to the door and observe if it is possible to cause the door to rotate with a force applied in that manner.
3. Note: the following experiment should only be attempted if you have a space that's easy to clean, or won't be damaged by the possible spill of soda. It's very easy to accidentally spill soda during this demonstration! If you have access to a can of soda, try to reproduce the demonstration where the can is able to be tilted on the lip between the base and sides without falling over (see Figure 8.8). (If you don't like drinking soda, or are worried about drinking too much of it, pour the soda into a different container so you can pour some back in the can if the CG becomes too low.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This page titled Chapter 8: Rotational Motion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 9: Gravity

Summary

Gravity is a force we experience all the time in our lives. It allows us to understand concepts such as weight, acceleration, and the motion of planetary objects. Newton's law of universal gravitation will allow us to understand how gravity works on both the Earth and other planets, and where the value 9.8 m/s^2 ("little g") comes from.

Apparent weight

Weight is a topic that we've discussed several times already in this textbook. We've learned that weight is a force that comes about due to our gravitational attraction with the Earth, and that it's related to our mass. While none of that is wrong, it's also not the full story of what weight is.

As you read this textbook, consider where you feel your weight in your body. Where does the sensation of weight come from right now? If you're standing, you can probably feel your weight in your feet. If you are sitting, you likely feel your weight in your bottom and the backs of your thighs. If you are lying down, you likely feel your weight along your side or back.

We have weight because of gravity, but our apparent weight comes from the support force we get from whatever surface we're in contact with. Our apparent weight is equal to the value of the support force that is acting on us.

In the video below, Dr. Pasquale jumps up and down on a force plate. The force that is measured by the plate is recorded by LoggerPro on the laptop in the background. This recording is also shown in Figure 9.1.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When Dr. Pasquale is standing still, her weight records as being constant. However, when she jumps up and down, her weight changes. When she is in the air, she is no longer in contact with the force plate, and her weight is zero. When she is coming back to the surface again, her weight increases. This demonstration shows that weight is not necessarily constant, even while being located in pretty much the same position on the Earth's surface.

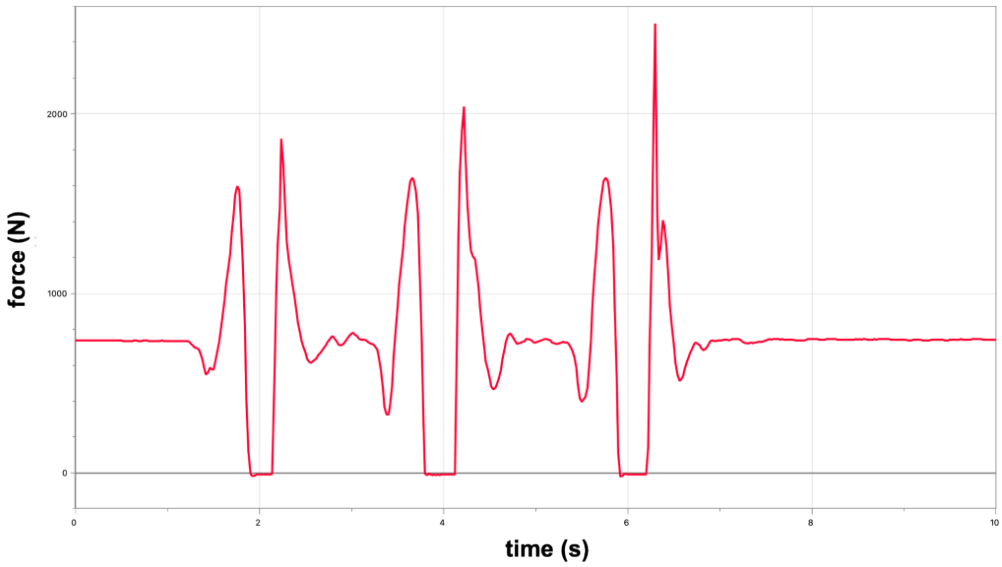


Figure 9.1 – The graph of force recorded from a force plate while Dr. Pasquale jumps up and down on it. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Any time our motion contains a vertical component of acceleration, whether it be from jumping up and down, riding a roller coaster, or traveling in an elevator, our support force will change, and our apparent weight will change as well.

Consider a hypothetical person (let's call her Jane Physics) riding in an elevator. Any time the elevator moves at a constant speed, there is no acceleration in the vertical direction, and Jane's apparent weight will be equal to the force of gravity that is acting on her mass. Because the net force acting on Jane is zero, her support force must be equal and opposite to her gravitational force. The support force is therefore equal to her mass times gravitational acceleration.

In general, we can relate the net force, support force, and force of gravity by using an equation. Remember that net force is equal to the sum of all forces (in this case, there are two forces: gravity and the support force). Therefore,

$$\begin{aligned}\vec{F}_{net} &= \vec{F}_{support} + \vec{F}_{gravity} \\ &= \vec{F}_{support} + m\vec{g}.\end{aligned}$$

Gravitational acceleration is -9.8 m/s^2. Because we are dealing with vectors in the above equation, that negative sign on $\vec{g}$ is very important!

Now consider Jane Physics riding in an elevator that is speeding up as it moves upward. Because the elevator is speeding up in the upward direction, there is an upward acceleration, and a net force pointing up. Her mass hasn't changed, so the $m\vec{g}$ term in the above equation remains the same. The support force now becomes greater than her weight. This is because $m\vec{g}$ (gravitational force) plus the support force equals the resultant force, that net upward force causing the elevator to speed up. This is why apparent weight increases as we speed up in an elevator moving upward.

If Jane Physics has a mass of 75 kg, her weight in an unaccelerated condition is

$$\begin{aligned}\vec{F}_{gravity} &= (75\text{ kg})(-9.8\text{ m/s}^2) \\ &= -735\text{ N}.\end{aligned}$$

If the elevator accelerates upward at a rate of 4 m/s^2, then Jane's net force is equal to her mass times the acceleration of the elevator,

$$\begin{aligned}\vec{F}_{net} &= (75\text{ kg})(4\text{ m/s}^2) \\ &= 300\text{ N}.\end{aligned}$$

Solving the equation above to solve for the support force, we can calculate Jane's apparent weight to be

$$\begin{aligned}\vec{F}_{support} &= \vec{F}_{net} - \vec{F}_{gravity} \\ &= (300\text{ N}) - (-735\text{ N}) \\ &= 1035\text{ N}.\end{aligned}$$

Jane Physics reaches the top floor of the building and decides to ride back down to the lower level. At this time, the elevator will speed up moving downward. There is a downward pointing acceleration, and therefore, a net force pointing down. The support force becomes less than Jane's weight. This is why apparent weight decreases as an elevator speeds up moving downward.

Jane's unaccelerated weight is -735 N. If the elevator accelerates downward at a rate of 3 m/s^2, then Jane's net force is

$$\begin{aligned}\vec{F}_{net} &= (75\text{ kg})(-3\text{ m/s}^2) \\ &= -225\text{ N}.\end{aligned}$$

We can solve for Jane's apparent weight by using the above equation, solved for the support force.

$$\begin{aligned}\vec{F}_{support} &= (-225\text{ N}) - (-735\text{ N}) \\ &= 510\text{ N}.\end{aligned}$$

Note that Jane's apparent weight is now less than her unaccelerated weight.

Free fall

The extreme condition of this would be a situation where the elevator cable snaps and the elevator moves downward at 9.8 m/s^2: free fall. In that case, Jane's apparent weight would go to zero! While this should never happen in an elevator, it's possible that you may have experienced such a condition of weightlessness in an amusement park ride.

An object in free fall, moving only under the influence of gravity with no other forces present, will have an apparent weight of zero. If you were to go skydiving, before the parachute deploys, and before air resistance slows you down, your apparent weight would be zero. Think about using a scale to measure your weight as you fall: there is nothing to give the scale a force reading. In those conditions you are truly weightless.

For the same reasons, astronauts in Earth orbit (whether they be in a space capsule, the Space Shuttle, or the space station) are also weightless, as shown in Figure 9.2. They are still under the influence of Earth's gravitational field, and their acceleration due to gravity is still rather high (about 9 m/s^2). The reason that astronauts experience weightlessness is that they are in free fall around the Earth during their orbit. To emphasize: astronauts in Earth's orbit do not experience a zero-gravity condition, only a zero-support force condition.



Figure 9.2 – Astronaut Sally Ride, along with other unsecured objects, experiences weightlessness on Space Shuttle mission STS-7. Mission Specialist (MS) Ride at forward flight deck pilots stations controls, by NASA Marshall Space Flight Center, is licensed under Public Domain.

The inverse-square law

The inverse-square law describes physical properties that decrease when the distance between two objects, or the distance away from a single object, increases. As distance increases, the physical property decreases. Furthermore, the property doesn't just decrease proportionally with distance. Rather, it decreases with the square of the distance. This means that changes in distance have a very powerful effect on the physical property. Mathematically, we can represent the inverse-square law as a statement that a property is proportional to one divided by the distance squared,

$$\frac{1}{d^2}.$$

A graph of a property that has an inverse-square law relationship is shown in Figure 9.3. (Note that there are no units on either axis of the graph in Figure 9.3 because the relationship between the physical property and distance is arbitrary in this case.)

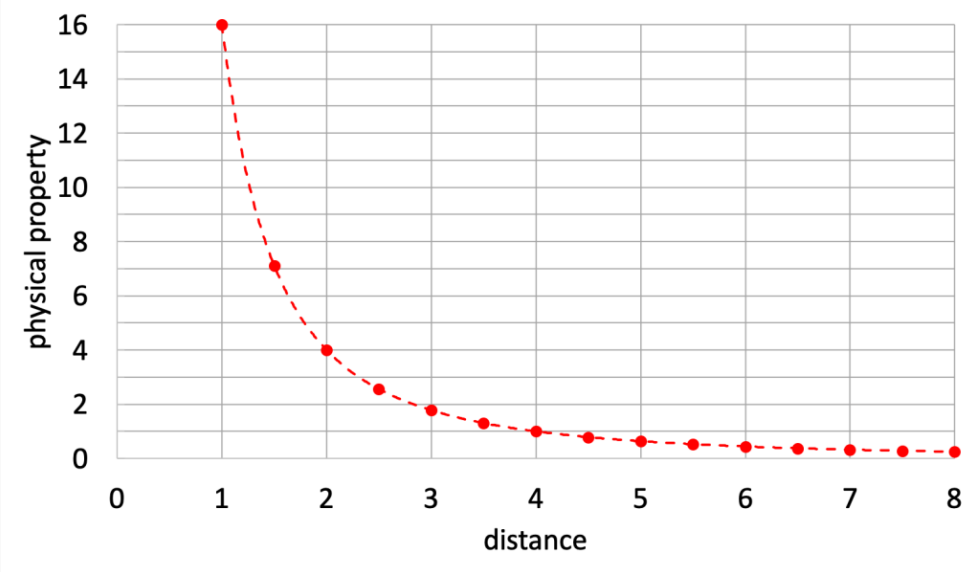


Figure 9.3 – A graph of the relationship between a physical property and distance for a property that obeys the inverse-square law. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The graph in Figure 9.3 can be used to help conceptualize the effect of distance on a property that obeys the inverse-square law. When the distance between the objects is equal to one, the physical property in this case is 16. When the distance doubles (becomes equal to 2), the physical property decreases by 2^2 and becomes 16/4 = 4. When the distance doubles again (from 2 to 4), the physical property decreases by 2^2 again, and becomes 4/4 = 1.

Many physical properties obey the inverse-square law: light intensity, sound intensity, electromagnetic and gravitational forces. On a conceptual level, this is because the effect of any single point of sound, light, electric charge, or mass dilutes by the distance squared as it moves outward in three-dimensional space.

Newton's law of universal gravitation

Newton's law of universal gravitation describes how we can quantify the force of gravity acting between any two objects in the universe. The equation states that

$$F = G\frac{m_1 m_2}{d^2}.$$

In other words: gravitational force (F) is equal to the universal gravitational constant (G), times the mass of one object (m_1), times the mass of the second object (m_2), divided by the distance between the two objects (d) squared.

The universal gravitational constant, G (sometimes known as "big G," as opposed to g, "little g"), is

$$G = 6.67 \times 10^{-11}\ \mathrm{Nm^2/kg^2}. \tag{9.1}$$

The term universal indicates that this number is the same anywhere in the universe. It's a very small number, indicating that the force of gravity is relatively weak. In fact, of the four fundamental forces, gravity is the weakest! In order for the force of gravity to be noticeably strong, one or both of the objects must have a lot of mass.

Note that this equation obeys the inverse-square law. If the distance between two objects were to double, the gravitational force would decrease by one fourth. If the distance between two objects were to be cut in half, the force would increase four times.

The denominator of this equation tells us another interesting thing about gravity: the force of gravity between two objects cannot become exactly equal to zero unless the distance between the two objects is infinite. Therefore, it can be concluded that all objects of mass that exist in the universe exert gravitational forces on each other. A planet millions of light-years away from you is exerting a gravitational force on you. However, the value of that force is infinitesimally small. We can effectively treat it as being zero, but it is not exactly zero.

The distance between two objects is more accurately stated as the distance between the center of mass of two objects. Although the density of the Earth is not constant, if we could approximate the Earth as being perfectly spherical and having a uniform density throughout, we would expect the gravitational force to not change as we move along the Earth's surface.

However, the Earth is not a perfect sphere. (The shape of the Earth is better described as an oblate spheroid.) The distance between the center of the Earth and the north and south poles is smaller than the distance between the center of the Earth and the equator (shown in Figure 9.4). Therefore, the force of gravity would be larger at the poles.

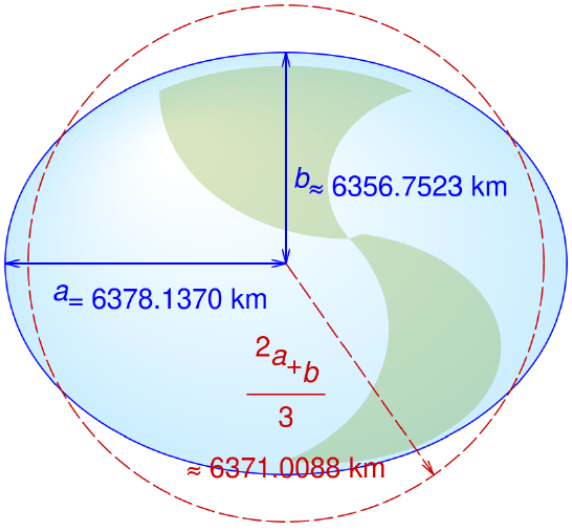


Figure 9.4 – A diagram of the Earth, showing that the radius to the North pole is smaller than the radius to the equator. A red circle is shown in the background to compare with a perfect sphere. Note that this image is not to scale. WGS84 mean Earth radius, by Cmglee, is licensed under CC BY-SA 4.0.

In addition, as we move farther away from the center of the Earth, say, by climbing a mountain, the gravitational force would decrease as well. It is important to point out that these differences (due to the oblateness of the Earth, and due to distance from sea level) are very small, and would not be perceptible unless you had a very accurate scale to stand on to weigh yourself.

These effects are further complicated by the fact that the Earth does not have a uniform density throughout its volume. For the most part, we can treat the gravitational force at all points on the surface of the Earth as approximately equal, but it is important to note that it is not.

In fact, astronauts visiting the space station are still under the influence of gravity – that is why the space station is able to orbit the Earth. Without gravity, no force would be able to create a centripetal force to cause objects to orbit, and they would not be able to move in a circular path. The astronauts experience weightlessness because of the lack of a support force.

One last interesting thing about gravity: because mass can only come in positive numbers, gravity is only an attractive force. When we learn about electricity later in this textbook, we'll see that electric charges are capable of both attracting and repelling.

However, the force of gravity is only an attractive force, causing two objects to accelerate toward each other, but never away.

Newton's law of universal gravitation can be used to calculate the gravitational force acting between the Sun and the Earth. The Earth has a mass of 6×10^{24} kg. The Sun has a mass of 2×10^{30} kg. The Earth and Sun are separated by a distance of 148×10^{9} m. Using the equation above, we calculate that

$$\begin{aligned} F &= (6.67 \times 10^{-11}\ \mathrm{Nm^2/kg^2}) \frac{(6 \times 10^{24}\ \mathrm{kg})(2 \times 10^{30}\ \mathrm{kg})}{(148 \times 10^{9}\ \mathrm{m})^2} \\ &= 3.65 \times 10^{22}\ \mathrm{N}. \end{aligned}$$

It's also possible to use Newton's law of universal gravitation to determine the effect that a change in properties will have on the force of gravity between two objects. For example: consider two planets with masses of m_1 and m_2 that are separated by a distance of d. (The exact values do not matter in this case.)

If the mass of planet one (m_1) were to be cut in half, what would the effect be on the force? Force and mass are directly proportional. Therefore if the mass decreases by a factor of two, the force will decrease by a factor of two as well. The new force would be one half of the original force.

If instead, the masses of both planets stay the same, but the distance between the two planets is cut in half, what is the effect on the gravitational force? Force and distance obey the inverse-square law. If the distance between the two planets decreases by a factor of two, the force will increase by a factor of two-squared: four. The new force would be equal to four times the value of the original force.

Gravitational acceleration

In this textbook, we have frequently encountered "little g," the acceleration due to gravity on the Earth's surface. This quantity has a value of -9.8 m/s^2. (Sometimes, when we do not need great accuracy in our calculations, we can round "little g" to -10 m/s^2.) Where does this number come from?

From Newton's second law, we know that the net force on an object is equal to its mass times its acceleration. If the net force on an object comes from gravity, then we can use Newton's second law and Newton's law of universal gravitation together, and solve for the acceleration. In this case,

$$\begin{aligned} m_{object} a &= G \frac{m_{object} m_{Earth}}{r_{Earth}^2} \\ a &= G \frac{m_{Earth}}{r_{Earth}^2}, \end{aligned}$$

where a is the acceleration of any object on the Earth's surface, G is the universal gravitational constant, m_{Earth} is the mass of the Earth, and r_{Earth} is the radius of the Earth. The distance between the surface of the Earth and the center of mass of any object on Earth is infinitesimally small compared to the radius of the Earth itself and will be ignored in this textbook.

When an object is, say, orbiting the Earth, then the distance between the surface of the Earth and the center of mass of the object h will become an important quantity to be considered in the equation. In this case, the equation will become

$$g = G \frac{m_{Earth}}{(r_{Earth} + h)^2},$$

where m_{Earth} is the mass of the Earth, r_{Earth} is the radius of the Earth, and h is the distance between the surface of the Earth and the center of mass of the object (satellite, space station, etc.).

Gravitational acceleration, being a particularly important quantity, is given its own symbol: g. In the equation, the mass of the object canceled out on both sides of the equation. This means that gravitational acceleration acts equally on all objects on the Earth's surface: you, an airplane, a butterfly, a tree, or a house.

This equation can be solved to find the value that we have seen so many times so far in this textbook. The mass of the Earth is 5.97×10^{24} kg, and the radius of the Earth is 6.37×10^{6} m. Therefore,

$$g = (6.67 \times 10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{5.97 \times 10^{24}\ \text{kg}}{(6.37 \times 10^{6}\ \text{m})^2}$$
$$= 9.8\ \text{m/s}^2.$$

Gravitational acceleration at different heights above the Earth's surface

Gravitational acceleration at sea level is 9.81 m/s^2. What is the gravitational acceleration at the top of Mount Everest?

When using the equation in this manner, the mass of the Earth is treated as a constant, but the distance will be equal to the radius of the Earth plus whatever distance away the object is from sea level.

The summit of Mount Everest is located approximately 8,850 m above sea level. Plugging that in to our equation, we find that

$$g = (6.67 \times 10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{(5.97 \times 10^{24}\ \text{kg})}{((6.37 \times 10^{6}\ \text{m}) + (8850\ \text{m}))^2}$$
$$= 9.79\ \text{m/s}^2.$$

The result, 9.79 m/s^2, is not a big difference from sea level! This goes to show that we can treat "little g" as a constant of 9.8 m/s^2 for the entire surface of the Earth if precision is not important in our calculations (which it will not be in this textbook).

To calculate the value of the gravitational acceleration on a different celestial body (the Moon, the Sun, or another planet), simply use Newton's law of universal gravitation and Newton's second law. In general,

$$g = G\frac{m_{celestialbody}}{r^2_{celestialbody}},$$

where m is the mass of the celestial body, and r is the radius of the celestial body. The universal gravitational constant ("big G") is equal everywhere in the universe.

The gravitational acceleration on Mars is less than that of Earth. Mars has a mass of 6.4×10^{23} kg and a radius of 3.4×10^{6} m. We can plug these values into the equation and find that

$$g = (6.67 \times 10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{6.4 \times 10^{23}\ \text{kg}}{(3.4 \times 10^{6}\ \text{m})^2}$$
$$= 3.7\ \text{m/s}^2.$$

If Mars were to magically double in mass, what would the effect be on the gravitational acceleration? Gravitational acceleration is directly proportional to mass. Therefore, if mass doubles, "little g" doubles as well. If Mars were to magically double in mass, its gravitational acceleration would become 7.4 m/s^2.

What if instead of Mars's mass magically changing, its radius expands until the planet has double the radius? What effect would that have on the gravitational acceleration? The relationship between "little g" and the radius obeys the inverse-square law. Therefore if the radius increases by a factor of two, the gravitational acceleration will decrease by a factor of two-squared: four. In that case, "little g" on Mars would decrease to 0.9 m/s^2.

Gravitational fields

A gravitational field is a model that we can use to describe how objects interact as a result of gravitational forces. We can draw a gravitational field by showing where any object placed around a massive object would move due to gravity.

Consider the gravitational field of the Earth. Any object of mass is going to want to move toward the center of the Earth. This is depicted by showing the field using arrows pointing toward the center of the Earth.

In Figure 9.5, the gravitational field is depicted around the Earth. (In reality, the field is three-dimensional, but it is difficult to

convey that in a two-dimensional format such as a computer screen or printed sheet of paper.)

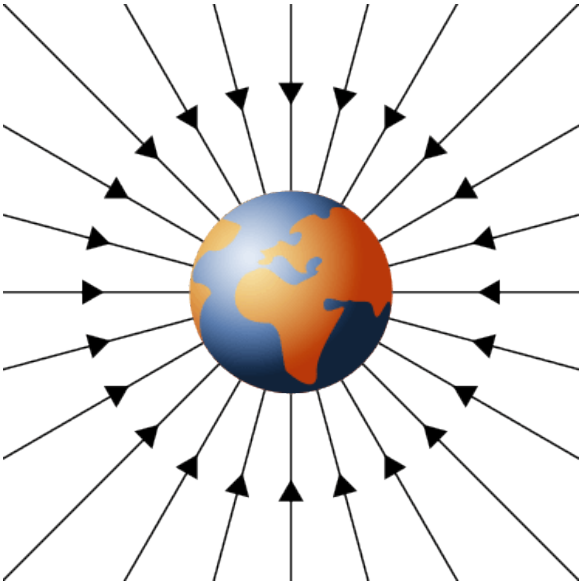


Figure 9.5 – The gravitational field of the Earth is depicted by radial arrows pointing toward the center of the planet. Gravitational field Earth lines, by MikeRun, is licensed under CC BY-SA 4.0.

At any given distance away from the center of the Earth, the distance between each of the field lines is proportional to the relative strength of the force. Close to the center of the Earth, the arrows are much closer together than they are far away from the center of the Earth. This means that gravitational forces are stronger closer to the center of the Earth, and weaker farther away. This makes sense based on our understanding of the inverse-square law.

The limits of Newton's laws

Newton's laws do a really great job of explaining the motion of so many objects. But there starts to be a discrepancy that becomes apparent when objects move fast – close to the speed of light. In these cases, Newton's laws break down and we require another physical theory to take over. Einstein's theory of general relativity can be used to accurately describe the forces and motions of objects in these cases.

The theory of general relativity explains the motion of objects due to curves in spacetime, the fabric of the universe. A massive object causes a large curving of spacetime, and anything moving near that massive object is going to travel along spacetime in a curved direction.

It's possible to conceptualize this by using a model of the universe. Here, the fabric of spacetime is literally a piece of fabric stretched along a circular frame. When no massive objects exist to cause any curves in spacetime, marbles that move through space travel in straight lines, shown in the first part of the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A large piece of mass is placed in the center of the fabric. Viewed from above (Figure 9.6, left), it doesn't seem like much has changed, but viewed from the side (Figure 9.6, right), we can see that the fabric is warped. This is similar to how largely massive objects: stars, planets, and so on, are able to curve spacetime.



Figure 9.6 – A piece of stretched fabric over a circular frame is used to depict the warping of space-time. A large object of mass is placed in the center of the fabric. Viewed from above (left), the fabric does not appear to be warped. Viewed from the side (right), it is clear that the fabric is stretched and warped. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Smaller objects can then be put in motion, and we can observe the effect of this spacetime warping, shown in the video above. Instead of traveling in straight lines like they did before, the marbles now travel in circular paths. This describes the orbits of planets around suns, and the orbits of moons and satellites around planets.

Further reading

- The Elegant Universe: Part 1 – Start watching the video at 9:52 to see the chapters of The Elegant Universe about squaring up Newton's law of universal gravitation and what we know about the "universal speed limit" from Einstein's theory of relativity.

Practice questions

Conceptual comprehension

1. Make a list of scenarios (other than what's provided in this textbook) where your apparent weight is either less than or greater than your gravitational force.

Numerical analysis

1. Imagine a person standing on a bathroom scale inside an elevator. The elevator is accelerating upward at 2 m/s^2. The magnitude of the person's actual weight is 600 N. What is their apparent weight in the elevator?
2. A roller coaster car goes through a loop-the-loop with a radius of 10 m. At the top of the loop, passengers experience a sensation of weightlessness. Calculate their apparent weight at the top of the loop if the coaster is moving at a speed of 20 m/s.
3. Calculate the gravitational force between two 50 kg masses separated by a distance of 10 meters.
4. Determine the gravitational force between the Earth (6 x 10^{24} kg) and a satellite (200,000 kg) in orbit 500 km above the Earth's surface.
5. Calculate the magnitude of the gravitational force that the Sun (2 x 10^{30} kg) exerts on Neptune (1 x 10^{26} kg). The Sun and Neptune are separated by a distance of 4.5 billion km.
6. Calculate the gravitational acceleration on the surface of Jupiter. Jupiter has a mass of roughly 2 x 10^{27} kg and a radius of approximately 70,000 km.
7. Calculate the gravitational acceleration on the surface of Venus. Venus has a mass of about 5 x 10^{24} kg and a radius of

approximately 6,000 km.

8. Calculate the gravitational acceleration on the surface of Saturn's moon, Titan. Titan has a mass of roughly 1.3 x 10^{23} kg and a radius of approximately 2,600 km.
9. A satellite orbits the Earth at an altitude of 400 km above the surface. Its centripetal force is supplied by its gravitational force with the Earth. Calculate the velocity of the satellite.

Hands-on experiments

1. Drop various objects of different sizes and masses (make sure they are safe and won't break) from the same height and use a stopwatch to time how long it takes for each object to reach the ground. Does the amount of time vary based on mass? Try to control for as many variables as you can.
2. If you have a bathroom scale, bring it into an elevator and watch your apparent weight change as you accelerate upward or downward. (Is your apparent weight different from your unaccelerated weight when the elevator is moving at a constant velocity?)

This page titled Chapter 9: Gravity is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 10: Projectile and satellite motion

Summary

Projectile motion describes the motion of an object in free fall, and allows us to analyze the motion of baseballs, rockets, missiles, satellites, planets, stars, and other celestial bodies. This chapter builds on the topics of motion and Newton's laws that have been introduced in previous chapters, and brings the section on kinematics to a close.

Projectile motion

Projectile motion describes the motion of an object that moves in free fall. That is, the only force acting on the object is gravity. Remember that free fall is an ideal scenario that ignores dissipative forces such as friction and air drag. The study of objects that travel like this is known as ballistics.

Ballistics (the physics of projectile motion) allows us to understand and quantify the motion of things such as baseballs after they leave the bat, rockets that are only powered internally very briefly before coming back to the Earth's surface again, as well as the motion of planets, other celestial bodies, and artificial satellites.

In the simplest scenario, an object can be thrown straight up so that it moves in a completely vertical path. That's the scenario we've already discussed in chapter three of this textbook. Alternatively, an object can be thrown upward, but at an angle to the horizon. This gives the object both vertical and horizontal motion.

In the absence of air drag and friction (which will be assumed unless otherwise stated), the only force that acts on an object during projectile motion is gravity. Gravity only works in the vertical direction. To understand how the horizontal motion of the projectile works, we must remember Newton's first law: inertia. The projectile will continue to move at the same speed in the horizontal direction, even while the vertical speed changes due to gravity. Fortunately, we can do an analysis of the horizontal motion separately from the analysis of the vertical motion.

In the video below, two identical steel balls are launched from a platform at exactly the same time. One is dropped in a completely vertical manner, and the other is pushed, giving it horizontal motion in addition to the vertical drop.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Both of the balls hit the surface of the lab table at exactly the same time, indicating that gravity acted on the objects in an identical manner. The ball that had horizontal motion did not exhibit any effects in its vertical motion, which means that the horizontal motion for this ball did not affect its vertical motion.

Fortunately, we are able to independently calculate the horizontal and vertical motion of a projectile in free fall. This will enable us to determine what equations we can use to analyze the motion in each direction. Each of the forces and related kinematic constants and equations (acceleration, velocity, and distance) are listed in the table below.

horizontal	none	0 m/s^2	$\vec{v}_{final} = \vec{v}_{initial}$	$\vec{d} = \vec{v}t$
vertical	gravity	-9.8 m/s^2	$\vec{v}_{final} = \vec{v}_{initial} + \vec{a}t$	$\vec{d} = \frac{1}{2}\vec{a}t^2 + \vec{v}_{initial}t$

In the horizontal direction, there are no forces at work on the object. Because of this, the law of inertia tells us that the object will continue moving with the same speed in the horizontal direction. As an equation, this is expressed as: $\vec{v}_{final} = \vec{v}_{initial}$, the velocity of the projectile at any moment in time is equal to the initial velocity. Because there is no accelerating force, the distance the object moves horizontally can be analyzed using the equation $\vec{d} = \vec{v}t$: displacement equals velocity times time.

In the vertical direction, the force acting on the object is gravity. The acceleration due to gravity is -9.8 m/s^2. (Remember that the negative sign indicates that the force and acceleration point down.) Because there is acceleration at work, the equations of motion in the vertical direction will be different from the equations in the horizontal direction.

The vertical velocity of the projectile at any given moment in time is equal to $\vec{v}_{final} = \vec{v}_{initial} + \vec{a}t$, which is the initial velocity plus acceleration (due to gravity) times time. The displacement that the object moves vertically can be analyzed using the equation $\vec{d} = \frac{1}{2}\vec{a}t^2 + \vec{v}_{initial}t$. It is important to note that the equations of non-accelerated motion cannot be used when there is a force acting on an object!

In the video below, a ballistic cart launches a ping-pong ball directly upward with no horizontal motion. The ball moves upward as it slows down (vertically) under the influence of gravity. It stops for a brief instant of time as it turns around and moves back down to the cart, speeding up (vertically) as it does so. This is a special case of projectile motion: where there is no horizontal motion at all. This type of motion was introduced and discussed in chapter three.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

On the other hand, the ballistic cart can launch the ball while it moves with a horizontal velocity. We learned in our discussion of inertia that the ping-pong ball will continue moving at the same horizontal speed, even while the vertical motion of the ball is influenced by gravity. This is why the ball is caught by the cart again as it moves back down to its initial height, shown in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The shape of the motion of a projectile moving with horizontal and vertical motion is known as a parabola. We can therefore say that a projectile has parabolic motion. A graph of the position of a projectile (qualitatively similar to the ball launched by the ballistic cart) with respect to time is shown in Figure 10.1. Understanding this motion enables scientists and engineers to design rockets, satellites, and other projectiles that have parabolic motion.

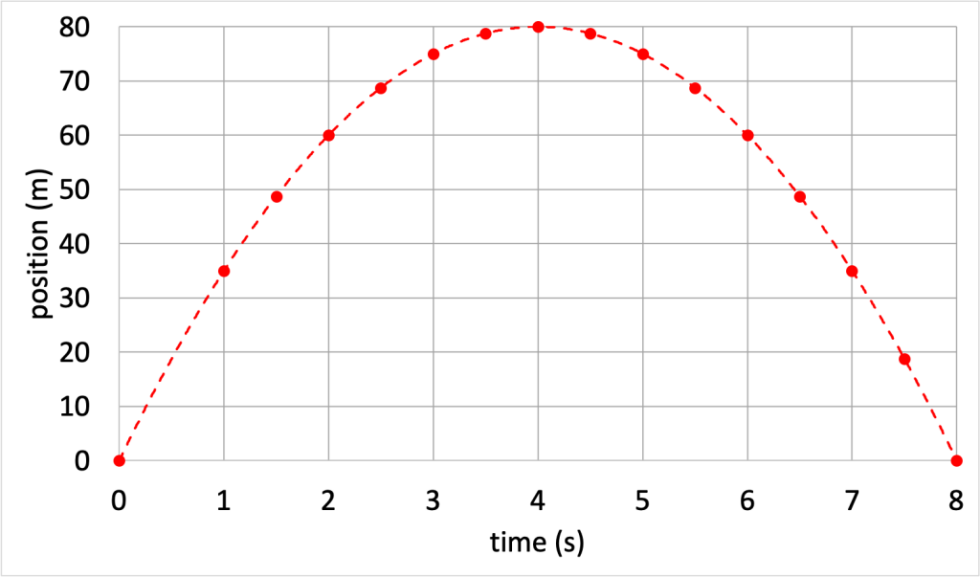


Figure 10.1 – A graph of parabolic motion with position plotted as a function of time. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

An important takeaway from the video of the steel balls dropping, and the videos of the ballistic cart launches, is that **the horizontal and vertical components of motion can be calculated and analyzed independently**. This greatly simplifies the study of projectile motion!

When air drag cannot be ignorede

It should be noted that when air drag cannot be ignored, the equations describing the trajectory of the projectile are not as simple. This is because the drag force always points in a direction tangent to the trajectory and opposite to the instantaneous velocity at every instant. This means that the vertical and horizontal forces are continuously changing. Therefore, the vertical and horizontal analyses are not independent and must be analyzed simultaneously. Such an analysis is beyond the scope of this textbook.

Maximizing horizontal range

Because a projectile can be launched at so many different angles with respect to the horizon, what angle would we choose if we wanted a projectile to have the maximum horizontal range?

If we conceptualize the motion of a projectile, an object launched directly upward will spend the most time in the air, but will have the least horizontal motion as it travels in a completely vertical line. On the other extreme, if something were to be launched at just a few degrees with respect to the horizon it would not spend as much time in the air, minimizing its horizontal range. To maximize the horizontal range, we must choose an angle somewhere in between zero and ninety degrees.

In the video below, Dr. Pasquale launched a projectile at fifteen degrees, thirty degrees, forty-five degrees, sixty degrees, and seventy-five degrees with respect to the horizon. Each of the projectiles has an identical initial velocity. In each case, we can look at the parabolic motion of the projectile. Because the initial height of the projectile changes when the launcher is tilted at different angles, we will only consider the motion back to the same starting height, rather than considering the motion back to the height of the table.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

As expected, very steep angles maximize the height of the ball, but don't move much horizontally. Very shallow angles spend less time in the air, causing them to lose out on horizontal motion. At forty-five degrees, we maximize the horizontal range by creating the best tradeoff between time in the air and horizontal motion.

Figure 10.2 shows a graph of the horizontal and vertical positions for projectiles launched at the same angles as in the video. These positions were calculated using the equations of motion. The graph demonstrates the vertical and horizontal ranges of each launch angle. (This data is qualitatively similar to the data shown in the video above, but is not identical due to the small amount of air drag that is present.)

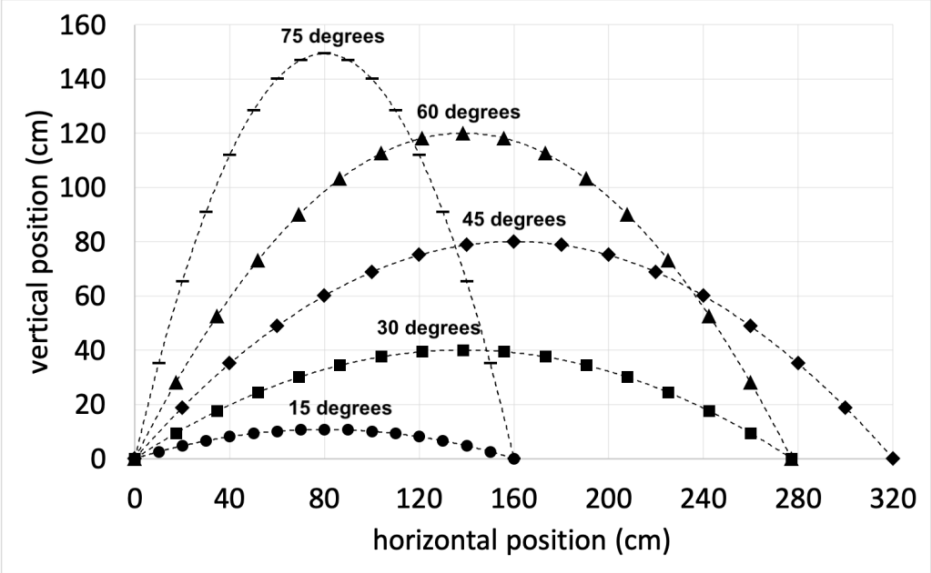


Figure 10.2 – A graph of the parabolic motion of projectiles launched at different starting angles. The projectile launched at 45 degrees has a maximum horizontal range. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the absence of air drag, a launch angle of forty-five degrees will maximize the horizontal range of a projectile.

Symmetry of motion

In the absence of forces such as air drag, the motion of a projectile is symmetric both in time and horizontally in space. This means that

- it takes a projectile the same amount of time to go from ground level to the top of its path as it does to go from the top of its path back down to ground level again,
- the vertical velocity of a projectile at the end of its motion when it reaches the ground again is equal and opposite to the initial vertical velocity of the projectile at launch, and
- a projectile moves the same horizontal distance between ground level and the top of its path as it does from the top of its path back down to ground level again.

From the symmetry of motion arises three key moments in the motion of a projectile, which are

1. the initial time a projectile is launched,

2. the moment the projectile reaches the top of its path, and
3. the time the projectile reaches ground level again.

These three key moments act as important times to calculate for different kinematic properties (height or vertical velocity) because there will always be a known variable, simplifying the equations that are used. This information is outlined in the table below.

key moment	time	vertical position (y)	horizontal position (x)	vertical velocity ($\vec{v}_{y,final}$)
initial launch	0	0	0	$\vec{v}_{y,initial}$
top of the path	½ of total time	maximum height	½ maximum range	0
return to ground level	total time	0	maximum range	$-\vec{v}_{y,initial}$

Solving projectile motion equations

The motion of projectiles can be analyzed using the information presented in this and previous chapters of this textbook.

For example, say a projectile is launched from ground level with an initial horizontal velocity of 10 m/s and an initial vertical velocity of 30 m/s. We want to solve for three different quantities:

1. How high will the projectile go?
2. How long will the projectile stay in the air?
3. How far will the projectile travel in the horizontal direction?

The first question asks about height. Because we are interested in the maximum height of the projectile, the key moment is the top of the path of the projectile. We can see from the table above (and recall from many examples in this textbook) that the vertical velocity at that key moment is 0 m/s. The two vertical kinematics equations are

$$\begin{aligned} \vec{d} &= \frac{1}{2}\vec{a}t^2 + \vec{v}_{initial}t \\ \vec{v}_{final} &= \vec{v}_{initial} + at \end{aligned}$$

While we eventually want to use the distance equation to solve for height, at this moment we don't know the amount of time that the projectile travels, so we can't use the equation yet. However, we can use the velocity equation to solve for the amount of time it will take the projectile to reach the top of its motion (which is, incidentally, equal to one-half of the total time of motion). Because the vertical velocity at this key moment is 0 m/s, we can solve the velocity equation for time.

$$\begin{aligned} \vec{d} &= \frac{1}{2}\vec{a}t^2 + \vec{v}_{initial}t \\ \vec{v}_{final} &= \vec{v}_{initial} + at \end{aligned}$$

It takes 3.06 s for the projectile to reach the peak of its motion. Now that we know the time, we can plug that value into the vertical distance equation to determine the height, which is

$$\begin{aligned} d &= \frac{1}{2}(-9.8\ \text{m/s}^2)(3.06\ \text{s})^2 + (30\ \text{m/s})(3.06\ \text{s}) \\ &= -45.88\ \text{m} + 91.8\ \text{m} \\ &= 45.92\ \text{m}. \end{aligned}$$

Now we can determine how long the projectile will remain in the air. Based on the calculation of time at the key moment of being

at the top of the path, we know that the projectile takes 3.06 s to complete one half of its total motion. Because the motion is symmetric, if it takes 3.06 s to get up, it will take another 3.06 s to get back down again. So we know that the projectile will spend 6.12 s in the air.

If you are asked to solve for how long the projectile will remain in the air without needing to first calculate the initial height, it's possible to determine this value by using the key moment of the return to ground level. At that time, the vertical position is zero, and the final vertical velocity is equal to the opposite value of the initial vertical velocity. Use the velocity equation to solve for time.

$$\begin{aligned} -30\ \text{m/s} &= 30\ \text{m/s} + (-9.8\ \text{m/s}^2)t \\ -60\ \text{m/s} &= (-9.8\ \text{m/s}^2)t \\ 6.12\ \text{s} &= t \end{aligned}$$

Finally, we can answer the question of how far the projectile moves horizontally. The position of maximum horizontal range occurs at the key moment of the return to ground level. We know from previous calculations that it takes the projectile 6.12 s to reach that key moment. Now we can use the equations of horizontal motion (no acceleration) to calculate the horizontal range.

$$\begin{aligned} d &= (10\ \text{m/s})(6.12\ \text{s}) \\ &= 61.2\ \text{m} \end{aligned}$$

Satellites and orbits

A satellite is a projectile that falls around the Earth, rather than falling back down to its surface. A thought experiment may help to conceptualize what this means. Consider that Jason Q. Physics builds an extremely tall tower, possibly several kilometers high, and stands at the top with several baseballs. Assuming that Jason has enough oxygen to survive this experiment, he can throw the baseballs at various speeds and see what happens.

This thought experiment is animated with a video, below. (Note: the video is not to scale!) To start, Jason drops a baseball. It falls straight down to the Earth's surface in a vertical line. If he now gives the ball a gentle throw in the horizontal direction, it will still fall to the Earth's surface, but the motion gives it a gentle curve. As he throws the remaining baseballs faster and faster, the baseball will travel a greater distance along the Earth's surface. At some point, we can no longer ignore the curvature of the Earth's surface. Eventually, the trajectory of the motion of the baseball will match the curvature of the planet. That trajectory is known as an orbit. An orbit describes the curved path that a satellite takes as it moves around an object of mass. In that case, Jason Physics has turned the baseball into a satellite!

When we were dealing with baseballs near the surface of the Earth, the force of gravity (both magnitude and direction) was considered constant: 9.8 m/s^2, vertically downward. For the baseballs launched from Jason Physics' tower, the gravitational force varies in magnitude (weaker at higher altitudes) and direction (points toward the center of the Earth). When these factors are considered, the resulting trajectory of the orbit becomes elliptical in shape.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

At what speed must the baseball be thrown to obtain this satellite motion? We can turn to rotational motion to calculate an exact answer. The centripetal force of the satellite must be supplied by gravity. Set the centripetal force equation equal to Newton's law of universal gravitation, and solve for velocity.

$$m_{satellite}\frac{v^2}{r} = G\frac{m_{satellite}m_{Earth}}{r^2}$$
$$v^2 = G\frac{m_{Earth}}{r}$$
$$v = \sqrt{G\frac{m_{Earth}}{r}}$$

In this equation, v is the orbital velocity (the minimum velocity required to obtain orbit around the Earth). r is the radius of the satellite's orbit (this is equal to the radius of the Earth plus the orbital height of the satellite). G is the universal gravitational constant, and m_{Earth} is the mass of the Earth. (If you wanted to place a satellite in orbit around a different celestial body, you would simply use the mass and radius of the celestial body you wish to orbit.)

This equation can be used to calculate the orbital velocity of the international space station, which orbits approximately 400 km above the Earth's surface. The mass of the Earth is 6×10^{24} kg, and the radius of the Earth is approximately 6400 km. The orbital velocity therefore equals

$$v = \sqrt{(6.67 \times 10^{-11}\ \mathrm{Nm^2/kg^2})\frac{6 \times 10^{24}\ \mathrm{kg}}{400000\ \mathrm{m} + 6400000\ \mathrm{m}}}$$
$$= 7654\ \mathrm{m/s}$$

The orbit of a satellite describes its path around a planet or star. All closed orbits are elliptical. A special case of an elliptical orbit is the circular orbit. The above baseball example, and calculations for the international space station, described circular orbits.

In a circular orbit (shown in Figure 10.3, left), the height above the Earth's surface remains constant. Considering the conservation of mechanical energy, because the orbital height remains constant, the GPE is constant throughout the satellite's motion. Because mechanical energy is conserved, the KE must also remain constant, causing the speed of the satellite to remain the same throughout its motion.

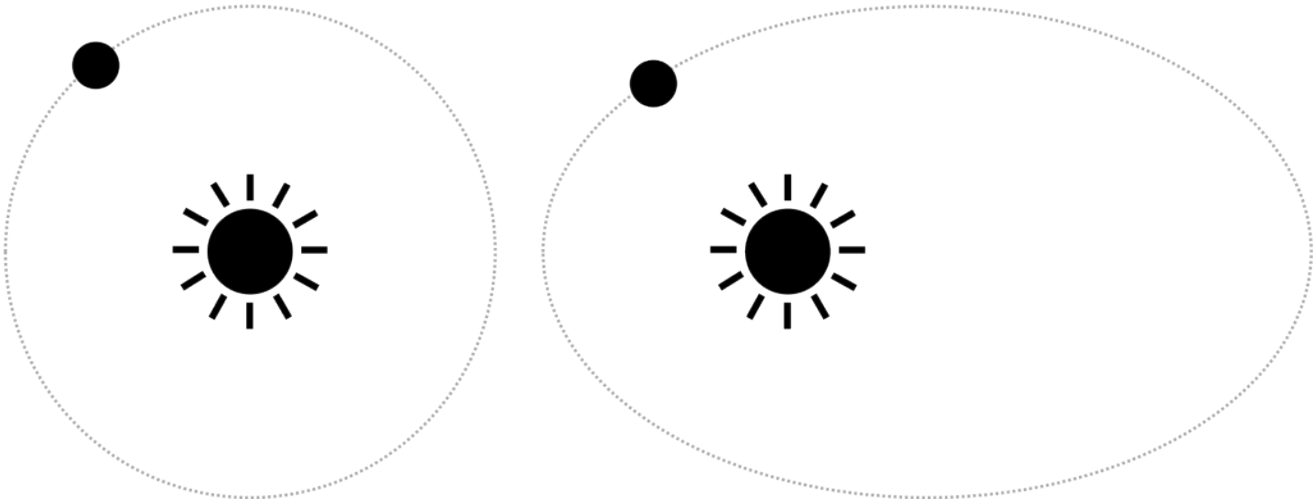


Figure 10.3 – Two possible orbits of a planet around a star. A circular orbit (left) has an equal distance away from the star at any point in the planet's orbit. An elliptical orbit (right) has different distances away from the star at different points in the planet's orbit. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In an elliptical orbit, a satellite will orbit around a planet or a star, with the planet or star at one of the focal points of the ellipse. As the satellite moves, the distance away from the planet or star changes (Figure 10.3, right). Mechanical energy is conserved. This means that as the GPE decreases, when the satellite is closest to the planet, the KE will increase, and vice versa. Therefore, when the satellite is closest to the planet, it will be moving at the fastest speed. When the satellite is farthest from the planet, it will be moving at the slowest speed.

Kepler's laws of planetary motion

The discovery of elliptical orbits came about in the early 1600s thanks to Johannes Kepler, before Newton published his theory on universal gravitation. Kepler was an apprentice of Tycho Brahe, a Danish astronomer, who made very accurate observations of

the positions of stars and planets over a span of decades in the late 1500s (without the aid of a telescope). Kepler used Brahe's data and plotted the orbit of Mars from the vantage point of someone looking down on the solar system from above. Kepler tried to use the equation of a circle to fit the points along the orbit and found that he couldn't. He was only able to fit the orbital path if he used the equation for an ellipse rather than a circle. The discovery that planets orbit in ellipses rather than circles is the first of what became known as Kepler's three laws of planetary motion.

Kepler's first law

Kepler's first law of planetary motion (the law of orbits) states that planets orbit their sun in ellipses with the sun at one focus of the ellipse (such as the orbit shown in Figure 10.3, right). This law is valid for any satellite in a closed orbit around a host body, whether it be a planet around a star or a moon around a planet. It should be noted that circular orbits are possible. A circle is just a special case of the ellipse in which the two foci are at the same point (Figure 10.3, left).

Kepler's second law

Kepler's second law of planetary motion (the law of areas) states that the position vector of the planet relative to its sun will sweep out equal areas in equal time intervals. As seen in Figure 10.4, the planet takes a certain time (Δt) to move from point A to point B. When the planet is at point C, it will move to point D in the same time interval (Δt). The areas of region ABO and region CDO are the same.

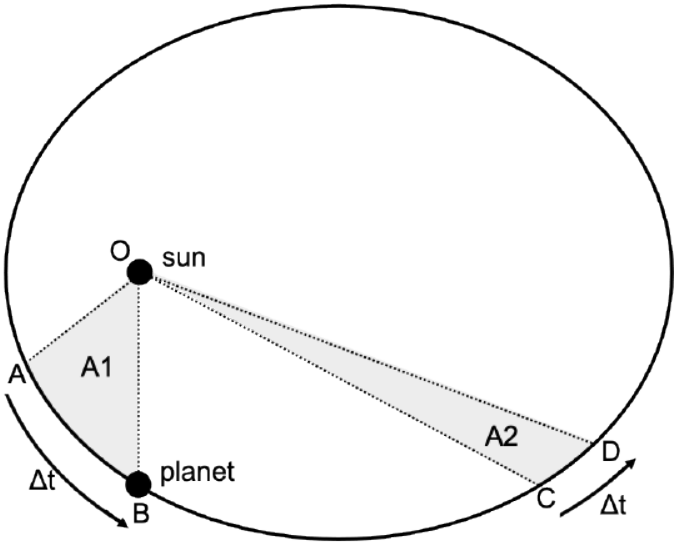


Figure 10.4 – An elliptical orbit showing two equal areas (ABO and CDO) where the planet will have equal transit times from point to point (A to B and C to D), leading to different orbital speeds. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This turns out to be a consequence of the conservation of angular momentum. While the mathematical details are too involved for this textbook, a qualitative description is warranted. Like the figure skaters who pull in their arms and legs to minimize their rotational inertia to make themselves rotate faster, the planet's rotational inertia about the sun is smaller when it is close to the sun. When the planet is farther away, its rotational inertia is greater because its mass is farther from the center of rotation (the sun). This increase in rotational inertia causes the rotational speed around the sun to decrease. Thus, the planet moves faster when closer to its sun and slower when farther away from its sun.

Kepler's third law

Kepler's third law of planetary motion (the law of periods) states a mathematical relationship between the time it takes for the planet to complete one orbit (the orbital period) and the semi-major axis of the elliptical orbit (see Figure 10.5).

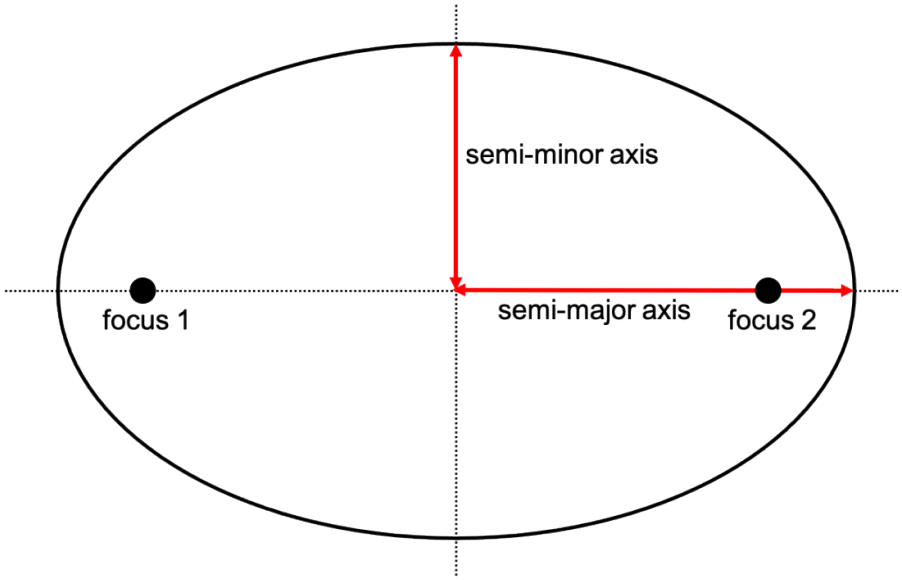


Figure 10.5 – An ellipse, showing the two focal points and the semi-minor and semi-major axes. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Specifically, the relationship states that the square of the orbital period is directly proportional to the cube of the semi-major axis: $T^2 \propto a^3$. For a circular orbit, the semi-major axis just becomes the radius of the circle. With just a little effort, Kepler's third law can be easily shown for a circular orbit (see the textbox below for the derivation). We start by recognizing that the gravitational force is responsible for the centripetal force that produces the circular motion.

Since the motion is circular, the gravitational force is producing a centripetal force as described in chapter 8. We then equate the gravitational force between the planet (m) and its sun (M) separated by a distance r (GmM/r^2) to the centripetal force (mv^2/r), where v is the orbital speed of the planet. Recalling that speed is the path divided by time, the orbital speed is just the circumference of the circle divided by the orbital period. That is,

$$v = \frac{2\pi r}{T}.$$

After plugging the above equation for speed into the expression for the centripetal force, we can solve for T^2. The result is

$$T^2 = \frac{4\pi^2}{GM} r^3.$$

Since π, G, and M are all constants, this proves the relationship that $T^2 \propto a^3$. Note that the mass of the planet (m) doesn't affect this relationship because it cancels out in the derivation.

Derivation of Kepler's third law

$$\begin{aligned} F_{gravity} &= F_{centripetal} \\ \frac{GmM}{r^2} &= \frac{mv^2}{r} \\ \frac{GM}{r} &= v^2 \end{aligned}$$

Orbital speed, v, is equal to path divided by time (T). For a circular orbit, the path is equal to the circumference of the circle: $2\pi r$. This value will be plugged into the above equation for v to continue the derivation.

$$\begin{aligned} \frac{GM}{r} = (\frac{2\pi r}{T})^2 &= \frac{4\pi^2 r^2}{T^2} \\ T^2 GM &= 4\pi^2 r^3 \\ T^2 = (\frac{4\pi^2}{GM}) r^3 & \end{aligned}$$

As an example, a planet that is four times farther from its host star will have an orbital period that is eight times greater, since $8^2 = 4^3$.

Practice questions

Numerical analysis

1. A ball is launched vertically upward from the ground with an initial speed of 10 m/s. Calculate...
    1. ...the time it takes the ball to reach its maximum height.
    2. ...the maximum height of the ball.
2. A soccer ball is kicked horizontally from a height of 2 meters above the ground with an initial horizontal speed of 15 m/s. Calculate...
    1. ...how far the ball travels horizontally before hitting the ground.
    2. ...the vertical speed of the ball immediately before impact.
    3. ...the horizontal speed of the ball immediately before impact.
    4. ...the total magnitude of the ball's velocity immediately before impact.
3. A baseball is thrown with an initial horizontal speed of 20 m/s. It lands on the ground 4 seconds later. Calculate...
    1. ...how far the ball travels horizontally before hitting the ground.
    2. ...the height the ball falls in that time.
    3. ...the vertical speed of the ball immediately before impact.
    4. ...the horizontal speed of the ball immediately before impact.
    5. ...the total magnitude of the ball's velocity immediately before impact.
4. A rock is dropped from a bridge into a river below. If it takes 3 seconds to hit the water, how high is the bridge above the water's surface?
5. A rocket is launched from the ground with an initial vertical velocity of 30 m/s upward and a horizontal velocity of 40 m/s. Calculate...
    1. ...the time it takes for the rocket to hit the ground.
    2. ...the horizontal distance traveled by the rocket.
6. A satellite with a mass of 500 kg is in a circular orbit around Earth at an altitude of 500 km above the surface. The tangential speed of the satellite is 7615 m/s. Calculate its kinetic energy, gravitational potential energy, and total mechanical energy.
7. A communication satellite (1,000 kg) is in a geostationary orbit around Earth at an altitude of approximately 35,786 km. (That is, the satellite orbits at the same speed as the rotation of the Earth around its axis.) Calculate the kinetic energy, gravitational potential energy, and total mechanical energy of the satellite.
8. A satellite (2,500 kg) is in an elliptical orbit around the Earth. As it moves from an altitude of 200 km to an altitude of 400 km, calculate...
    1. ...the GPE of the satellite at both altitudes.
    2. ...the change in KE as the satellite moves from low to high altitude.
    3. ...the change in ME as the satellite moves from low to high altitude.

Hands-on experiments

1. Crumple a piece of paper into a ball and set up a target (e.g., a trash can or a box) on the floor. Stand a fixed distance away from the target and try to toss the paper ball into the target. Experiment by changing the angle and force of your toss to observe how it affects the distance and accuracy of your throw.

This page titled Chapter 10: Projectile and satellite motion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

III: MATTER

Chapter 11: The atomic nature of matter
Chapter 12: Solids
Chapter 13: Liquids
Chapter 14: Gases

This page titled III: Matter is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 11: The atomic nature of matter

Summary

In order to understand the physics behind material objects (including phenomena such as flotation, elasticity, and aerodynamic lift), we need to know the composition of matter. We will now answer several questions, including: how do we know what matter is made of? How can we categorize matter? Are atoms the only type of matter that exist in the universe?

Matter

It's important to be able to categorize matter. From this categorization, we can learn about the various properties of different substances. We can also use this to determine what all matter has in common. Rather than being overwhelmed with the diversity of all of the different individual types of matter, we can realize that all matter can be placed into one of just a few categories. This categorization is outlined in Figure 11.1, and will be explained below.

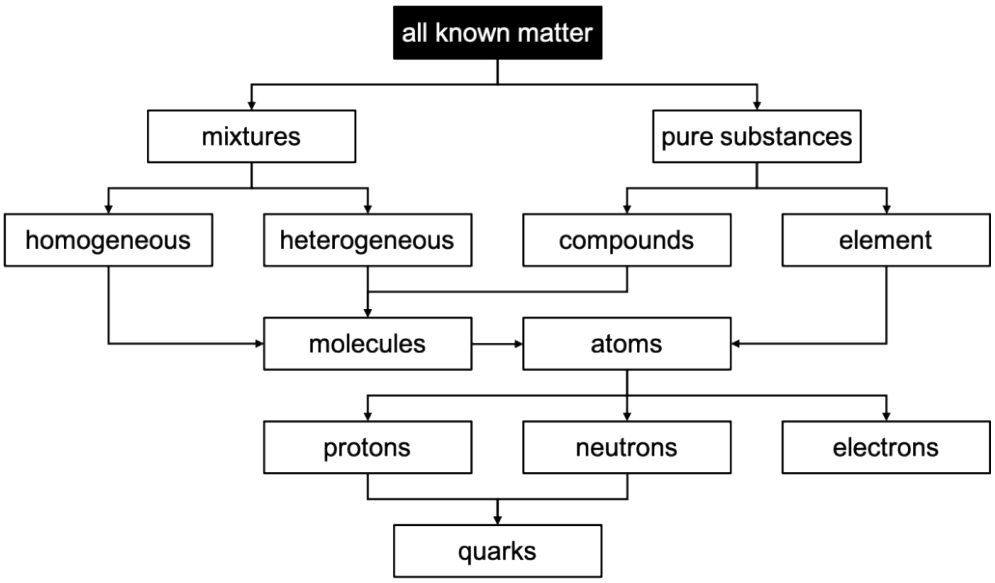


Figure 11.1 – A method of categorizing matter based on the properties of a substance. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Pure substances

All matter (that we are currently aware of) can be divided into two categories: pure substances and mixtures. In a pure substance, the properties of a substance such as density, melting point, specific heat capacity, and so on, are inherent to the material. A pure substance has a chemical composition that is known and does not change. For example: table salt (NaCl, or sodium chloride) is a pure substance. Water (H_2O) is another pure substance with known chemical composition. However, salt water (a combination of water and an unknown amount of table salt) is a mixture (defined below), as the amount of salt added to the water can be varied.

Pure substances can be elements. Gold, aluminum, sulfur, and lead are examples of solid elements. Mercury is a liquid element. Oxygen, nitrogen, and neon are all gaseous elements. An element is a substance that consists of atoms whose nuclei have the same atomic number. An element cannot be broken down further by chemical means.

Pure substances can also be compounds, which are substances made out of many identical molecules (defined below). Water is a compound consisting of molecules. A water molecule itself consists of two hydrogen atoms and one oxygen atom.

Mixtures

In a mixture, the properties of the substance vary based on concentration. As mentioned above, salt water is a mixture. The boiling point, the density, and the freezing point are all examples of properties that vary depending on how much (and what type) of salt was used to create that salt water. A mixture consists of two or more pure substances that do not react chemically to form a new substance. As a result, the mixture can be separated into its constituent substances by physical means (such as filtering or distillation).

Mixtures can be either homogeneous or heterogeneous. Homogeneous mixtures, such as air, salt water, or even iced tea, are all uniform throughout their volume. When drinking a glass of iced tea, it doesn't matter if you drink from the top or bottom of the glass, the iced tea will taste the same (it is homogeneous). Air is another homogeneous mixture. It doesn't matter what part of the room you breathe in: the air is essentially the same everywhere in a room.

A heterogenous mixture does not have a uniform consistency throughout. Granite, for example (see Figure 11.2), is composed of different minerals. If you were to cut two pieces of the same size out of a piece of granite, each one might have different amounts of each mineral.



Figure 11.2 – A piece of granite is an example of a heterogeneous mixture. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A more delicious example of a heterogeneous mixture is a burrito. Maybe you've had a burrito where one side is mostly lettuce, and in the middle is mostly sour cream, and the meat is all the way over to the other side. Depending on where you bite, you're going to get a different taste! An example of a heterogeneous burrito mixture is shown in Figure 11.3.



Figure 11.3 – A burrito is a delicious example of a heterogeneous mixture. Burrito madness, by Robby Van Arsdale, is licensed under CC BY 2.0.

Molecules

Molecules consist of two or more atoms that are held together by atomic bonds. A molecule is the smallest unit of a compound that still retains the identity of that compound. For example, water (H_2O) is a molecule. If the water molecule is broken apart, it is no longer water – it will be hydrogen and oxygen.

Molecules can also be the building blocks of homogeneous and heterogeneous mixtures. Mixtures can be separated into two or more chemical substances. And those substances are made of molecules, which are made of atoms.

Atoms

Every solid, liquid, gas, and plasma is composed of atoms. For a long time, scientists thought that atoms were the smallest unit of matter. (The term “atom” comes from the Greek term “atomos,” meaning indivisible.) Although we now know that atoms are made out of smaller entities (protons, neutrons and electrons), the term atom has been retained.

Protons and neutrons are made of smaller entities called quarks. Electrons belong to a class of particles known as leptons. As of our current understanding, quarks and leptons are elementary particles, as described by the standard model of particle physics. The standard model is the theory we use to describe all known forces and elementary particles in physics. While far from being a complete and unified theory of everything, the standard model does a fantastic job explaining what we do know. The standard model predicted the existence of a particle called the Higgs boson. The discovery of this particle in 2012 gave further support for the validity of the standard model.

Atomism

Throughout time, humans have attempted to answer the question: what are the fundamental building blocks of matter? If we were to cut something in half, then in half again, and again, until we could do so no more, what would we find? Do all objects of matter share common building blocks, or is everything different?

Something of particular interest was whether matter was fundamentally discrete or continuous. That is: if we conducted the experiment where we continued to cut something in half until we could do so no more, would we find that there is a discrete particle, a building block of matter? This idea is known as atomism. Or would we find that there aren’t any discrete particles, but instead a continuum, some type of fluid or substance that fills the space it occupies?

Aristotle believed that the continuum model must be true of nature. The phrase “nature abhors a vacuum” is attributed to him. Aristotle believed that if matter was made of particles, between those particles must exist a vacuum – nothing. He argued that a vacuum would violate physical principles. Of course, Aristotle had no evidence to back up this philosophical assertion.

Scientists have since come to understand that atomism is the correct answer to the question of whether or not matter is fundamentally continuous or discrete. There are fundamental building blocks of matter (described by the standard model of particle physics), and those building blocks create atoms, out of which all matter is made.

Atomic models (discussed below) have provided scientific evidence supporting atomism. Before discussing them, it is possible to demonstrate atomism with an empty aluminum soda can and some gallium (a metal that has a very low melting point). Gallium becomes a liquid at 30 ^{o}C, which is slightly warmer than room temperature. In the demonstration, liquid gallium is poured on top of a soda can to see what happens after time passes.

Initially, the aluminum soda can is just a regular empty can. If you've ever drunk out of a soda can, you can get a feel for its strength. While it can be crushed with a sufficient amount of force, it's a pretty strong container given how lightweight it is. In the video, before pouring gallium onto the can, Dr. Pasquale tests the strength of the can by pushing down on the top of the can with her fingers. Then, she pours liquid gallium onto the top of the can, placed near a space heater to keep the gallium in its liquid form. Dr. Pasquale then lightly scores the top of the can with a razor blade to expose the pure aluminum. After 20 minutes, the volume of gallium decreases as it diffuses into the can.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The can is now extremely weak and breaks apart as if it were tissue paper when Dr. Pasquale presses lightly with her fingers. The gallium inserted itself between the aluminum atoms and completely destroyed the bonds between each atom. If aluminum were made of a continuum, there would be no way for this to happen. Atomism is the correct interpretation of matter: all matter is made of atoms!

Another simple demonstration that supports atomism is to combine known volumes of water and alcohol. While the mass of this mixture is equal to the sum of the individual masses of the water and alcohol, the volume of the mixture will actually be less than the sum of the individual volumes! This is demonstrated in the video below.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=660#oembed-2

Dr. Pasquale mixed 50 mL of water (dyed blue) with 50 mL of alcohol (dyed red), and the mixture of both liquids (a dark green color resulting in the mixtures of the food dyes) had a volume of 97 mL. These results are shown in Figure 11.4.

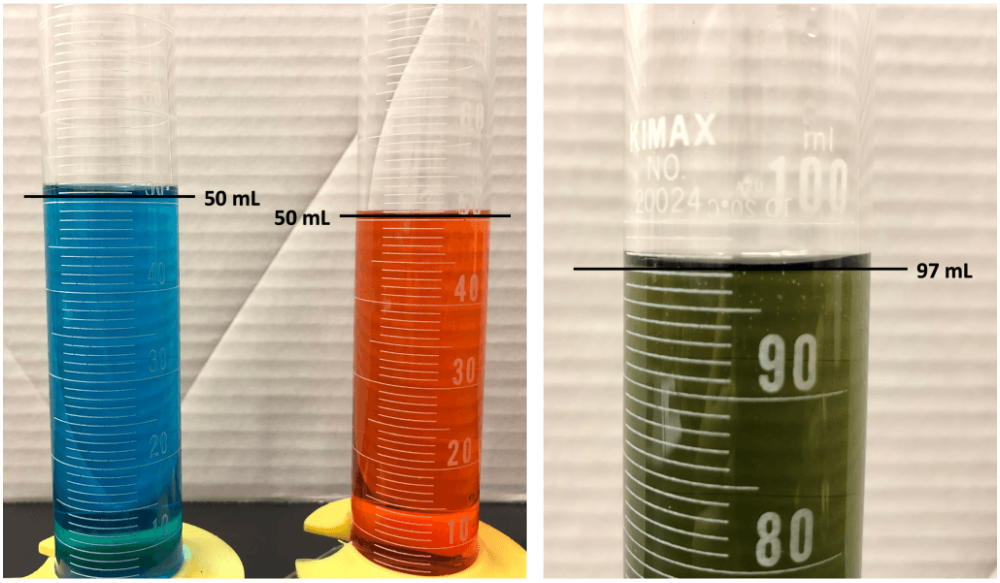


Figure 11.4 – Two cylinders, one filled with 50 mL of water (dyed blue) and one filled with 50 mL of alcohol (dyed red) (left) when combined create 97 mL of mixture (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

To understand why this is so, imagine a beaker filled halfway with marbles and another beaker filled halfway with BBs. When the marbles and BBs are mixed into a third beaker, the volume of the combination does not fill the entire beaker. This is demonstrated by Dr. Pasquale in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This result occurs because the smaller BBs can fill the gaps and voids between the larger marbles (see Figure 11.5). In a similar fashion, the smaller water molecules fill gaps and voids between the larger alcohol molecules.

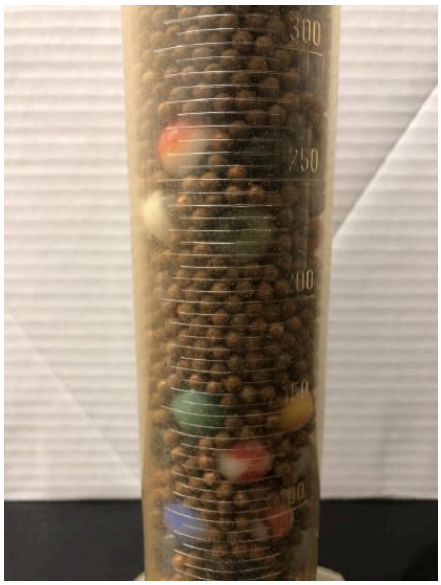


Figure 11.5 – Small BBs take up space in between large marbles. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Atomic theories

Many atomic theories have been proposed to explain the building blocks of matter: the atom.

Democritus, who lived around 400 BCE, theorized that everything was composed of atoms. Between atoms was empty space, and there were an infinite number of different types of atoms. He proposed that the properties of a substance related to the properties of the atoms that made it up. Iron would be composed of very strong interlocking atoms. Water atoms would be smooth and slippery. Salt atoms would be sharp and pointed. Many theories from this time period came about due to philosophical discussions and reasoning. It would be almost two thousand years before observations and experimental evidence were used in a systematic pursuit of scientific truth.

The scientist John Dalton, in the early 1800s, noticed that chemical compounds consisted of different elements in ratios of whole numbers. That is, things combined together in such a way that very consistent units of masses were involved. We now know this to be due to atomic bonds. A single oxygen atom can combine with two hydrogen atoms to create water. Water cannot be created with half an oxygen atom. There is no such thing as a half an atom, or a quarter of an atom. This seemed like a very convincing argument in favor of atomism.

Plum pudding model

In the late 1800s, J.J. Thomson discovered that the particles emanating from cathode ray tubes were much smaller than hydrogen, the lightest atom known. These particles are known to us today as electrons. Electrons were known at the time to create electric current. This discovery of electrons by themselves was evidence that atoms are not indivisible, but can be broken down into even smaller particles.

J.J. Thomson's model, known as the plum pudding model, attempted to explain how atoms worked based on this knowledge. On the whole, it was known that atoms were electrically neutral. If they weren't neutral, then all atoms would be attracted together or repelled apart due to electric forces. It was also known from the cathode ray tube experiments that electrons could move around by themselves. The plum pudding model hypothesized that an atom was made of a positively charged substance (pudding) that contained mobile electrons (plums). This is depicted in Figure 11.6. This model was eventually shown to be incorrect by Ernest Rutherford in the early 1900s.

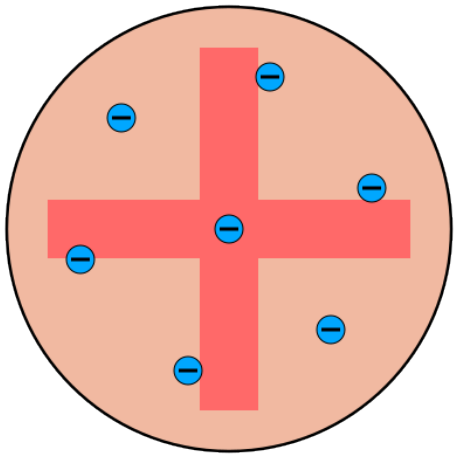


Figure 11.6 – The plum pudding model depicts an atom as a positively charged substance filled with mobile negatively charged electrons. Plum pudding model, by Kurzon, is licensed under CC BY-SA 4.0.

Rutherford model

In the year 1911, Ernest Rutherford and his colleagues conducted an experiment where they shot alpha particles (electrically charged particles discovered by Rutherford in 1899) at a thin piece of gold foil to probe the structure of matter. They observed that the vast majority of the alpha particles passed through the thin piece of foil undeflected and only a small number were deflected. This outcome implied that most of the volume of gold foil is composed of empty space.

Of the small number of alpha particles that were deflected, the scattering angles were much larger than expected and a few bounced backward completely. To this, Rutherford stated: “It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.” This was a remarkable result because it meant that the alpha particles were rebounding off of a very massive yet very tiny object.

These unexpected experimental results disproved the plum pudding model proposed by J.J. Thompson. If the Thomson model had been correct, the expectation would be that all of the alpha particles would simply pass through the gold foil with some slight deflections (Figure 11.7, left).

Rutherford’s experiment led to the discovery of the atomic nucleus. This revised atomic model consisted of a very small positively charged nucleus surrounded by mostly empty space in which electrons reside (Figure 11.7, right). Alpha particles, we now know, are a helium nucleus consisting of two protons and two neutrons and having a positive electric charge. When they came to interact with the nuclei in the gold foil, they bounced back due to the force of electrical repulsion between the protons in the gold nucleus and the protons in the alpha particles.

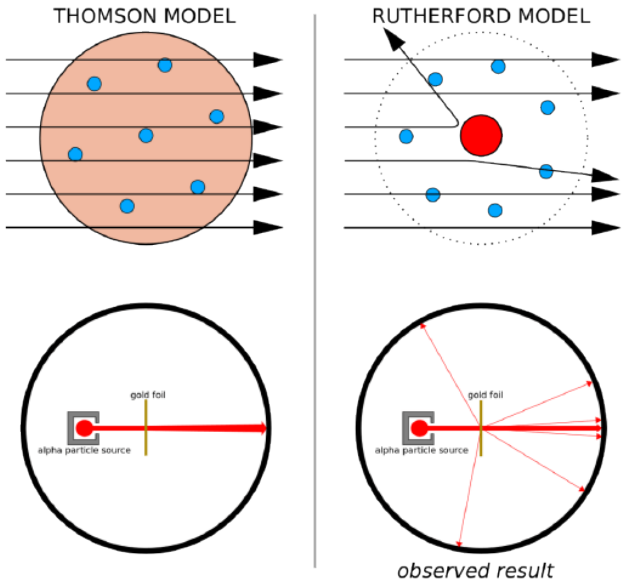


Figure 11.7 – The Thomson model (left) consists of a positively charged substance filled with mobile electrons, and predicts that alpha particles would pass through gold foil undeflected. The Rutherford model (right) consists of a positively charged nucleus surrounded by mobile electrons, and was discovered by the gold foil experiment where alpha particles were deflected and scattered by the atoms in gold foil. Geiger-Marsden experiment expectation and result, by Kurzon, is licensed under CC BY-SA 3.0.

Bohr (planetary) model

In 1913, Niels Bohr proposed that electrons orbit the nucleus of an atom much in the same way that planets orbit a star. These orbits were well defined, and only a finite number of orbits could exist. (In other words, only certain orbits were possible; an electron could not move around in any arbitrary orbit.) These orbits were defined by the amount of energy that electrons contained, which were measured by experiment as existing in very finite, discrete quantities.

The Bohr model is depicted in Figure 11.8, which shows an aluminum atom containing a nucleus and 13 electrons that orbit the nucleus in fixed orbits.

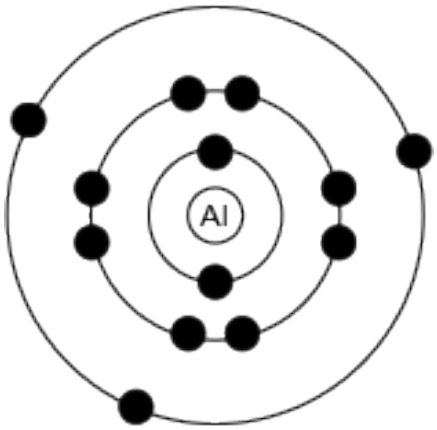


Figure 11.8 – An aluminum atom depicted by the Bohr model: a nucleus surrounded by 13 electrons orbiting in fixed orbits around the nucleus. Bohr2, by Stannered, is licensed under Public Domain.

Current atomic model

What do we know about the atom today? Does the Bohr model stand the test of time? Based on the advances of quantum physics, we know that atomic models are not so straightforward. Electrons do not simply orbit the nucleus of an atom in circular or elliptical paths. Instead, they exist probabilistically. That means that we cannot know the exact location of an electron but can model the probability of where we expect an electron to be, given its energy level. In this way, we can describe electrons, and all subatomic particles, as “smeared out” rather than as single particles existing at a single point in space.

Eventually the neutron was discovered, completing our atomic picture. Neutrons are neutral, meaning they do not influence the charge of an atom, but do contribute to its mass. Because each atom is defined by the number of protons in its nucleus, atoms that have the same number of protons but different masses must contain different numbers of neutrons.

The periodic table

The periodic table of the elements (shown in a simplified form in Figure 11.9) is a table into which all elements are placed. The periodic table was developed over a span of many years and was proposed when scientist Dmitri Mendeleev noticed periodic repetitions of chemical properties of atoms.

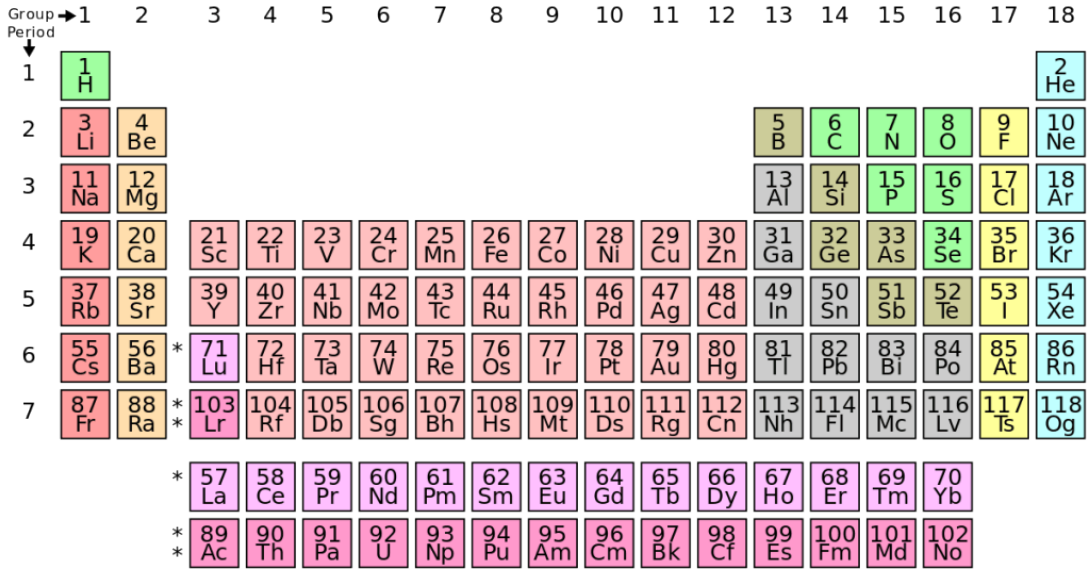


Figure 11.9 – The periodic table of the elements. Simple Periodic Table Chart, by Offnfopt, is licensed under Public Domain.

The columns on the periodic table represent elements that have similar chemical properties. Today we know that the columns correspond to the configuration of electrons in each element. For example, the column all the way to the right consists of noble gases. These gases do not readily react with other substances, and today we know this is because each of these elements contains a full shell of valence electrons.

The periodic table provides a lot of information about each element. Inside each box (for example, Oxygen is shown in Figure 11.10) is a symbol that is used to define the element in a chemical formula. Frequently, the name of the element will be spelled out as well. Oxygen has the symbol O. Lead has the symbol Pb.

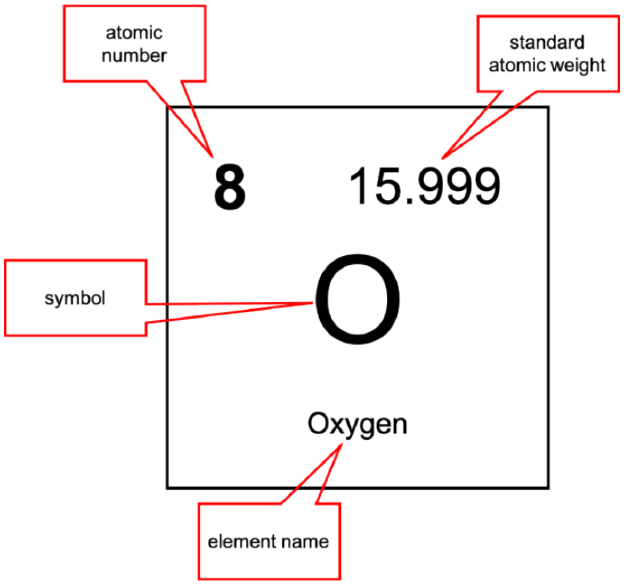


Figure 11.10 – The box depicting oxygen on the periodic table of the elements contains a symbol, element name, atomic mass, and standard atomic weight. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The number at the top left corner of each box is the atomic number of each element. The atomic number uniquely defines the element and is equal to the number of protons inside each atom's atomic nucleus. We can therefore say that the number of protons is what defines each atom. Any atom that contains two protons is, by definition, a helium atom, regardless of how many neutrons or electrons are present.

Isotopes

While an element is defined by the number of protons (Z) in the nucleus, the number of neutrons (N) in the nucleus can vary. An atom with six protons is carbon. A large sample consisting only of carbon atoms will contain atoms that all have exactly six protons. Most of those carbon atoms also have six neutrons, but some have seven neutrons, and a few may even have eight neutrons. Elements that have different numbers of neutrons are known as isotopes.

Many elements have isotopes that occur naturally. Some are stable, and some are radioactive. Carbon-12 is the most abundant isotope of carbon. It is a stable isotope and is present in all organic matter. About 99% of all carbon atoms are carbon-12. Carbon-13 makes up most of the rest of carbon atoms. Carbon-14 is very rare (about 1 out of every one trillion carbon atoms) and is radioactive. This isotope of carbon is used in carbon dating. The ratio of carbon-12 and carbon-14 in an organic substance can be used to determine how old that substance is.

The standard atomic weight of an element is also given on the periodic table. The standard atomic weight of an element is the weighted average of the mass of all the naturally occurring isotopes of that element. This just means that the standard atomic weight takes into consideration the different isotopes of an element that exist, and the probability of finding each of those isotopes. This explains why the standard atomic weight given on the periodic table is usually not a whole number. When averaged over all of the natural isotopes, the result will generally not be an integer value.

Sometimes the standard atomic weight is provided in parenthesis, or not at all. This is when the number or abundance of isotopes is not yet known.

Other properties of an element are not going to be provided on the periodic table, but can be given to define an element. For example: $^{3}\mathrm{He}$ describes helium-3. The number three is the atomic mass number (A) of the isotope. Atomic mass number is equal to the number of nucleons (protons and neutrons) in the element. Because the element is helium, we know that it has two protons. Two protons plus the number of neutrons is equal to 3. Therefore, we can conclude that helium-3 has one neutron.

The atomic mass number can therefore be used to determine the number of neutrons in an isotope using the equation

$$A = Z + N.$$

Remember that the number of protons is given by the atomic number. The table below has several examples of different isotopes,

including the number of protons and the number of neutrons in each isotope.

$^{12}\mathrm{C}$	6	12	6
$^{14}\mathrm{C}$	6	14	8
$^{29}\mathrm{Na}$	11	29	18
$^{104}\mathrm{Zr}$	40	104	64
$^{98}\mathrm{Pd}$	46	98	52
$^{136}\mathrm{Xe}$	54	136	82

Ions

Atoms, on average, are electrically neutral. This means that the number of positively charged protons in an atomic nucleus is completely balanced by an equal number of negatively charged electrons orbiting that nucleus. When the number of protons and the number of electrons is no longer equal, due to an excess or absence of electrons, then the atom is charged. We call a charged atom an ion. The number of protons minus the number of electrons tells us the charge. A hydrogen atom has one proton. If a particular hydrogen atom has two electrons, then we know the charge of that atom is -1. The charge is represented as a superscript after the atomic symbol. This particular ion is represented as H^-.

For example, an atom of $^{40}\mathrm{Ca}^{2+}$ is an atom of calcium. The number of protons (Z) is defined by the atomic number: 20. The atomic mass number is 40. To calculate the number of neutrons (N), subtract the atomic number from the atomic mass number and obtain 20. There are 20 neutrons in the atom. Finally, the charge is +2. 20 protons minus the number of electrons equals +2. Therefore, there are 18 electrons in this ion of calcium.

What is the universe made of?

Is that all there is to the universe? Protons, neutrons, and electrons? Not exactly.

Antimatter

Antimatter is composed of particles included in the standard model of particle physics, and they have been observed by scientists. Antimatter particles have the same mass as normal matter, but the opposite charge. Protons have an antiparticle known as the antiproton, which is a negatively charged particle with the same mass as a proton. Electrons have an antiparticle known as the positron, which is a positively charged particle with the same mass as an electron. Neutrons also have an antiparticle known as the anti-neutron. When a particle and its antimatter partner come into contact, they annihilate each other and transform into radiant energy. The amount of energy transformed can be calculated using Einstein's famous equation $E = mc^2$. E is the energy generated, m is the mass of the particles, and c is the speed of light.

Some antimatter is naturally produced, as the result of radioactive decay, for example. Antimatter is also artificially produced in particle colliders. These antiparticles take a lot of energy to create and only exist for a short time before getting annihilated by ordinary matter.

It appears pretty clear to scientists that our universe is composed almost entirely of what we consider to be normal matter, and not antimatter. Why is this? We don't know. This asymmetry of matter and antimatter is one of the biggest unsolved questions in physics.

Dark matter and dark energy

Many observations of the motions of galaxies indicate that they cannot be composed entirely of particles represented in the standard model of particle physics. In addition, observations about the rate of expansion of the universe cannot account for the amount of energy known to exist in the universe. In fact, the standard model probably only accounts for about 5% of the matter and energy that exist in the universe. Dark matter and dark energy are hypothesized concepts that enable the mathematics of the universe to work out correctly based on our observations. These concepts are called "dark" due to the fact we cannot "see" them directly. Rather, we infer their existence based upon the behavior of objects that we can see.

For instance, dark matter does not interact with electromagnetism. In other words, we cannot see dark matter because it does not interact with light. However, it does interact gravitationally, so we can observe the effects of dark matter from how it affects the motion of objects with mass that do interact with light.

The existence of dark energy has been proposed to explain the observation that the rate of expansion of the universe is increasing. The first evidence that the universe is expanding was presented by Edwin Hubble in 1929. This incredible discovery indicated that the universe appears to have had a "birth." With this new picture of the universe, it was believed that the expansion of the universe should gradually slow down due to the gravitational attraction of all the matter in the universe. However, in the late 1990s, it was discovered that the rate of expansion was actually increasing! This means that either Newton's law of universal gravitation breaks down at large-scale distances or that there is some other influence causing this observed "antigravitational" effect. The term dark energy has been assigned to explain this increasing rate of expansion of the universe.

It is hypothesized that 27% of our universe is composed of dark matter, and the remaining 68% is dark energy (see Figure 11.11). This is another area of intense research in physics today.

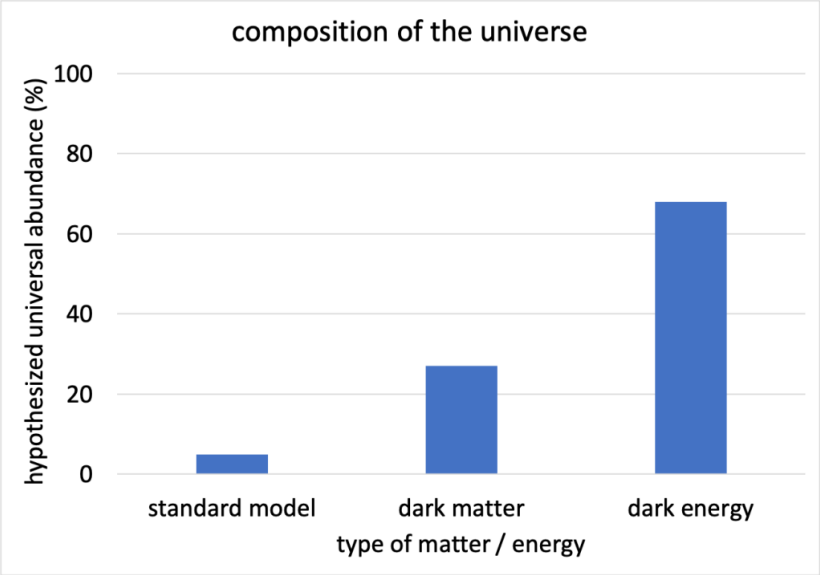


Figure 11.11 – A graph depicting the hypothesized composition of the universe: 5% accounted for by the standard model, 27% dark matter, and 68% dark energy. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Further reading

- Gallium-induced structural failure of an aluminum baseball bat – Watch as the structure of aluminum is destroyed by gallium

atoms that dissolve the aluminum bonds. This is a pretty good indication that matter is atomic; gallium would be unable to diffuse inside of the aluminum if it were a continuum.
- What and where is antimatter? – This short article from CERN explains some basics of antimatter and brings up some of the big questions about antimatter that physicists are trying to answer.
- Ptable.com – This website is a great interactive periodic table of the elements.
- Where Is Half of the Universe? – This article from *The Physics Teacher* discusses the imbalance of matter and antimatter in the universe.

Practice questions

Conceptual comprehension

1. List five different mixtures. Which of those are homogeneous and which are heterogeneous? How do you know?
2. List five different pure substances.
3. What do the following atomic nuclei, $^{18}\mathrm{F}$, $^{20}\mathrm{Na}$, $^{21}\mathrm{Mg}$, $^{23}\mathrm{Si}$, have in common? (Hint: These are all the same “isotones.” Look up: what is an isotone.)

Numerical analysis

1. How many protons are in the following elements?
    1. Hydrogen
    2. Helium
    3. Carbon
    4. Oxygen
    5. Iron
2. Identify the number of protons in each of the following molecules.
    1. H_2O (two hydrogen and one oxygen atom)
    2. CH_4 (one carbon and four hydrogen atoms)
    3. N_2 (two nitrogen atoms)
    4. CO_2 (one carbon and two oxygen atoms)
    5. $C_6H_{12}O_6$ (six carbon, twelve hydrogen, and six oxygen atoms)
3. Identify the number of protons and the number of neutrons in each of the following atoms.
    1. Carbon-12
    2. Helium-4
    3. Oxygen-16
    4. Silicon-28
    5. Uranium-235
4. Identify the number of protons, neutrons, and electrons in the following atoms.
    1. $^{23}\mathrm{Na}^{+}$
    2. $^{36}\mathrm{Cl}^{-}$
    3. $^{24}\mathrm{Mg}^{2+}$
    4. $^{80}\mathrm{Br}^{-}$

Hands-on experiments

1. Experiment with different substances like salt (sodium chloride) and sugar (sucrose). Add a small amount of each substance to a glass of water and stir until it dissolves. Note how the substances seem to disappear into the water, illustrating the concept of solubility and the dispersion of particles at the atomic level. You can then measure the density of each mixture to verify that the density varies with different concentrations of salt or sugar.
2. Try the alcohol/water experiment for yourself. If you don’t have a graduated cylinder, use a measuring cup or measuring spoons to measure equal volumes of each liquid. Verify that the volume of the combined mixture is less than the individual volumes of each liquid.

This page titled Chapter 11: The atomic nature of matter is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 12: Solids

Summary

Solids are one of the four most common states of matter. A solid is a substance where the molecules or atoms are very tightly bound together. This gives a solid a very rigid volume and shape. Solid objects do not change their shape to fit into a container, as a liquid does. Solids also do not change their volume to take up all available space, as a gas does.

The structure of solids

The atoms in a solid can be arranged in one of a few ways. When the atoms or molecules of a substance are arranged in an orderly, repeating manner, that solid is called a crystalline solid. In a crystal, knowing the location of one set of atoms allows us to determine the location of all other atoms in the solid, because that atomic pattern, known as a unit cell, will repeat itself regularly. Examples of crystalline solids include diamonds, quartz, and snowflakes.

Figure 12.1 (top left) shows an example of the arrangement of molecules in a crystal; these molecules are orderly and the arrangement is predictable. On the macro scale, crystalline solids are notable for appearing to have orderly structures. Figure 12.1 shows three examples of crystalline solids: bismuth (top right), pyrite (bottom left), and a snowflake (bottom right).

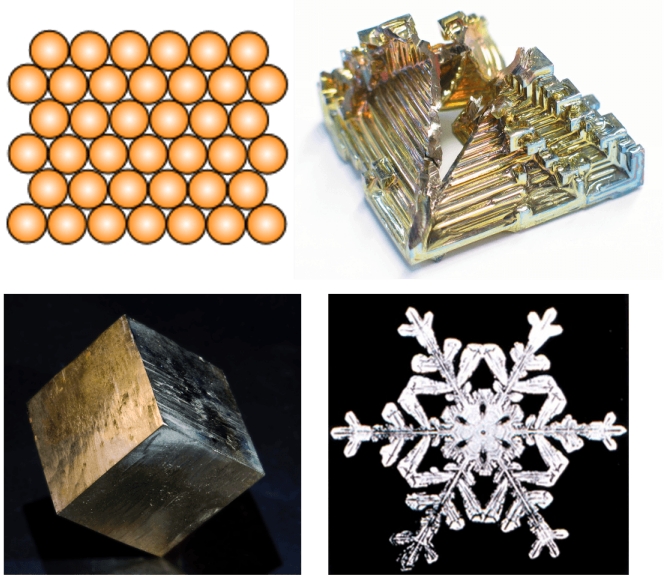


Figure 12.1 – Crystalline solids have an orderly arrangement of molecules (top left). Examples of crystals include bismuth (top right), pyrite (bottom left), and snow (bottom right), which all appear to have orderly structures, even on the macro scale. This figure is adapted from Crystalline polycrystalline amorphous2 (top left) by Cdang (CC BY-SA 3.0); Bismuth crystal macro (top right) by Dschwen (CC BY 2.5); Pyriteespagne (bottom left) by DidierDescouens (CC BY-SA 4.0); and Snowflake8 (bottom right) by Wilson Bentley (Public Domain).

Some crystalline solids have special properties. Silicon, for example, is grown in labs into large crystals called boules. These silicon boules are sliced into wafers. The crystalline nature of silicon makes it a high-quality semiconductor, and it is used predominantly in computer chip manufacturing.

Other solids have atomic or molecular arrangements that are disordered, or that otherwise lack a regularly repeating arrangement. These solids are known as amorphous solids. Examples of amorphous solids include plastics, wax, and glass. Figure 12.2 (left) shows an example of the arrangement of molecules in an amorphous solid. A large category of amorphous solids is plastic items (Figure 12.2, right).

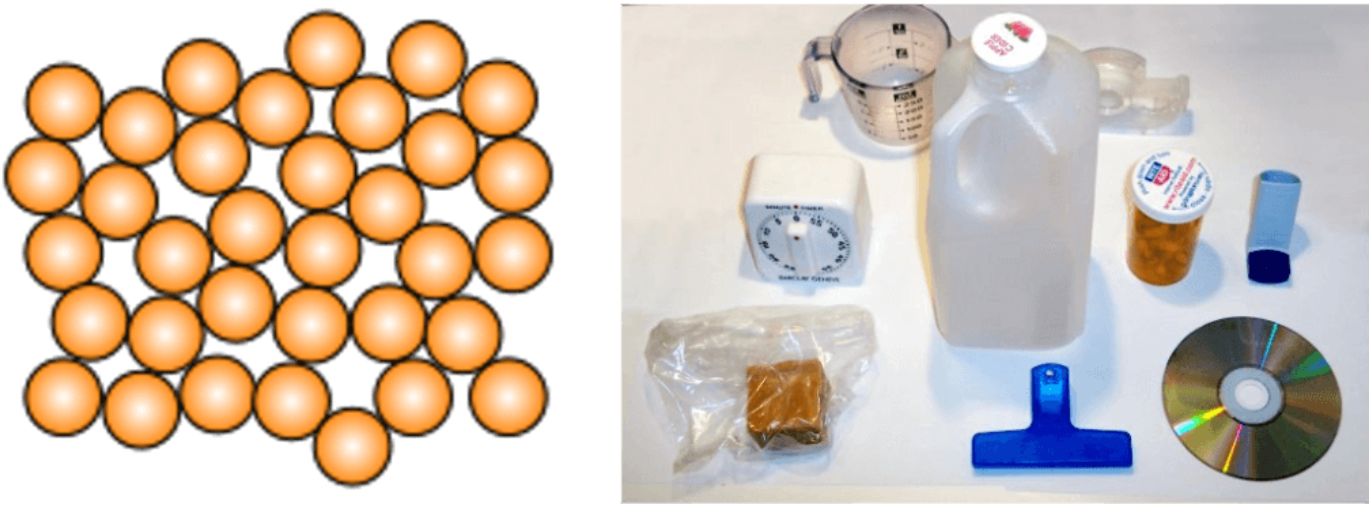


Figure 12.2 – Amorphous solids have a disorderly arrangement of molecules (left). Plastic items are amorphous solids (right). This figure is adapted from Crystalline polycrystalline amorphous2 (left) by Cdang (CC BY-SA 3.0); and Plastic household items (bottom right) by ImGz (CC BY-SA 3.0).

In between amorphous and crystalline materials is a class of material called polycrystalline. Polycrystalline materials have some order to them, but are not made of a single crystal. You can think of polycrystals as being composed of many individual crystals. Most metals and ceramics are polycrystalline.

The properties of crystalline, polycrystalline, and amorphous solids are a subject of interest to materials scientists. Materials scientists determine the properties of materials, and design new materials, for use in everything from industry and manufacturing to biomaterials and building materials.

Density

Density defines how compact the matter in an object is. The symbol for density is the capital letter D. (Note that many higher-level physics textbooks will use the lowercase Greek letter rho (ρ) to define density. In this book, we will use D for simplicity's sake.) The equation for density is mass divided by volume,

$$D = \frac{m}{V}.$$

Because the SI unit of mass is the kg and the SI unit of volume is m^3, the SI unit of density is kg/m^3. At times, the metric unit of g is used for mass and cm^3 for volume. In this case, volume can also be calculated with the unit g/cm^3.

Either of these sets of units can be used, but they are not interchangeable. Furthermore, when working with quantities such as force with the unit N, it is important to ensure that density is measured using the SI unit (kg/m^3)! To convert from g/cm^3 to kg/m^3, multiply by one thousand. To convert from kg/m^3 to g/cm^3, divide by one thousand. (You should try to convince yourself of this by performing the conversion given that 1 kg = 1000 g and that 1 m = 100 cm.)

Density is a property inherent to pure substances. That is, the density of gold (for example) is always going to be 19.3 g/cm^3. The densest naturally occurring element is osmium, which has a density of 22.6 g/cm^3.

To measure the density of a substance, first measure the mass of the object using a balance. To measure volume, there are two methods that can be used. The first is used with regularly shaped objects (such as a rectangular block) where we can measure a length, width, and height. Use a ruler or calipers to measure each dimension and multiply them together to find the volume. In other words, $V = lwh$.

In Figure 12.3, the density of a piece of acetate (a regularly shaped object) is calculated. The mass is measured using a balance, and is 7.5 g (Figure 12.3, top left). Then, with calipers, the length (top right) is measured (1.866 cm), the width (bottom left) is measured (1.845 cm), and height (bottom right) is measured (1.860 cm). The volume of the piece of acetate is therefore

$$V = (1.866\ \text{cm})(1.845\ \text{cm})(1.860\ \text{cm})$$
$$= 6.404\ \text{cm}^3,$$

and the density of the piece of acetate is

$$D = \frac{7.5\ \text{g}}{6.404\ \text{cm}^3}$$
$$= 1.2\ \text{g/cm}^3.$$

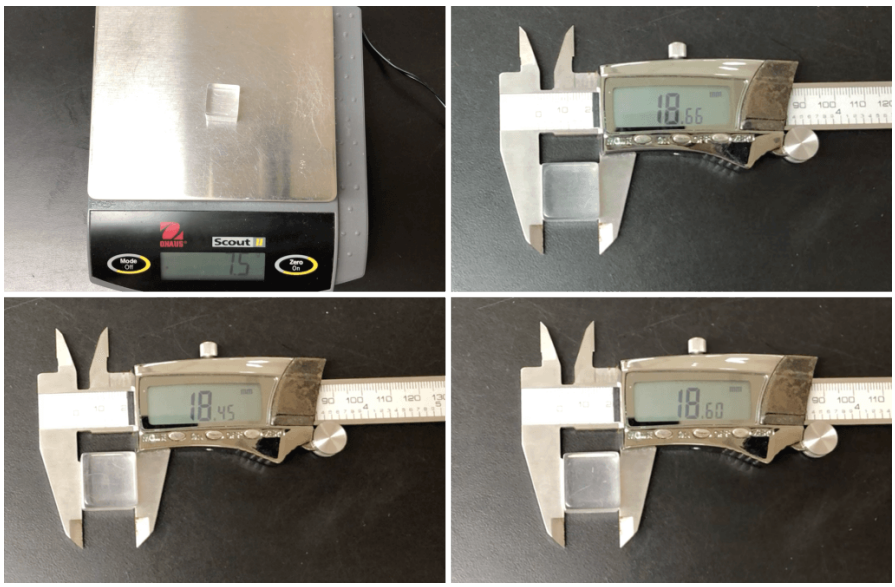


Figure 12.3 – A piece of acetate has its density calculated. First, the mass is measured using a balance (top left). Then, calipers are used to measure the length (top right), width (bottom left), and height (bottom right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

An irregularly shaped object does not have a length, width, or height that can be easily measured with a ruler. We need another way to measure the volume. To measure the volume of an irregularly shaped object, use the liquid displacement method. Fully submerge the object in a volume of water and see how much the water level rises. Every milliliter of water rise equals a volume of one cubic centimeter.

In Figure 12.4, the water displacement method is used to calculate the volume of a piece of granite. The mass is measured using a balance, and is 44.6 g (Figure 12.4, left). Then, the piece of granite is placed into a graduated cylinder with 200 mL of dyed water (Figure 12.4, middle) causing the water level to rise to 220 mL (Figure 12.4, right). The volume of the piece of granite is therefore 20 mL (20 cm^3), and the density of the piece of granite is

$$D = \frac{44.6\ \text{g}}{20\ \text{cm}^3}$$
$$= 2.2\ \text{g/cm}^3.$$

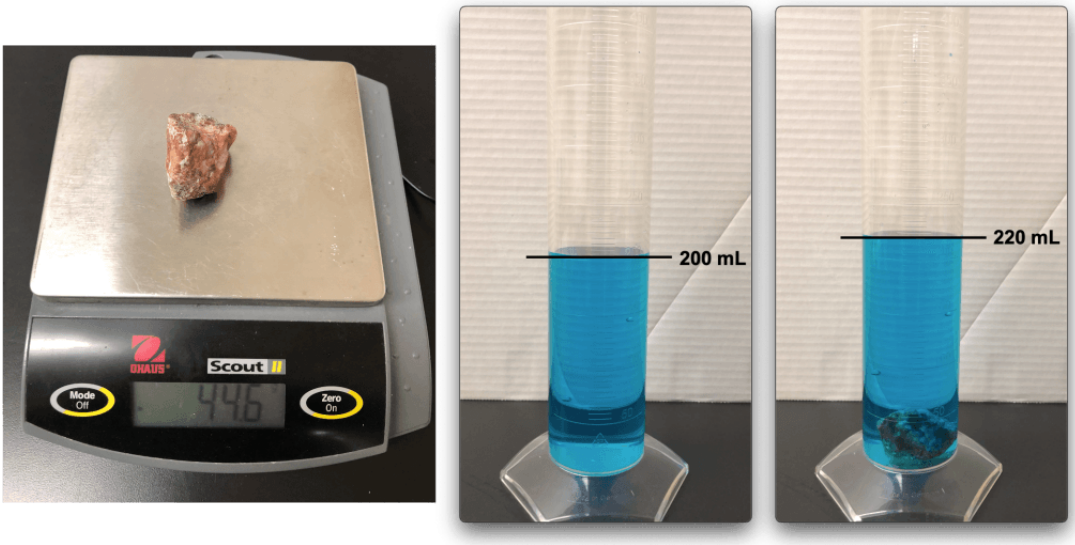


Figure 12.4 – An irregularly shaped object (piece of granite) has its density calculated. First, the mass is measured using a balance (left). Then, the water displacement method is used to measure the volume (middle and right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Elasticity

Elasticity describes how much an object deforms when experiencing a stress or force. If you've played with a Slinky™ (Figure 12.5) before, you've probably seen this property. If a Slinky™ is deformed by stretching it out lightly, once that force is removed, it will return to its original shape and configuration.



Figure 12.5 – A Slinky™ is a fun toy exhibiting the property of elasticity. Metal slinky, by Roger McLassus, is licensed under CC BY-SA 3.0.

The application of force causes an object to deflect. When the force is removed, the object will go back to its original shape. However, if too much force is applied, then that object will no longer return to its original configuration. This happens when we exceed the yield strength of an object. If you've ever seen a spring that won't quite go back to its coiled up position, then that spring has been stretched too far. A permanently deformed spring is shown in Figure 12.6.

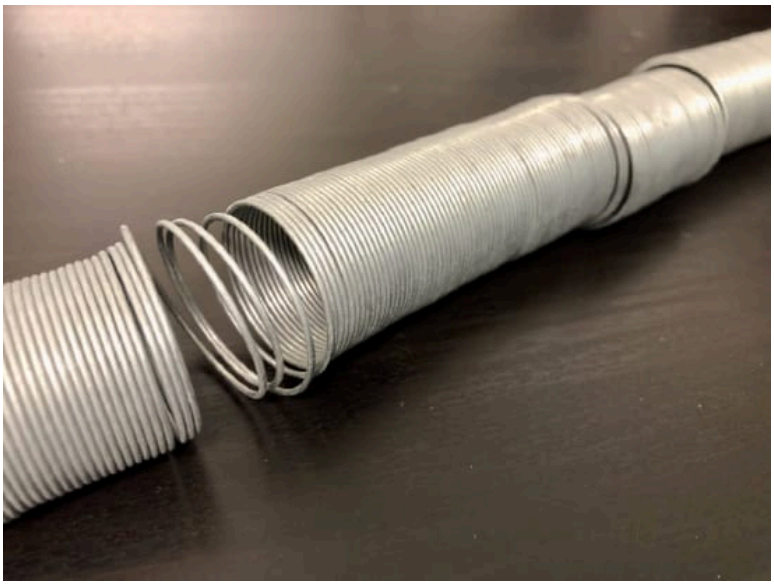


Figure 12.6 – A spring that has several sections that have been deformed due to the yield strength being exceeded when the spring was stretched out. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Adding even more force will eventually cause an object to break. This happens when we exceed the ultimate strength of an object. Applying enough force will cause any elastic object to break: a Slinky™, a rubber band, or a spring.

Elasticity is defined mathematically for springs using Hooke's law. Hooke's law states that $F = k\Delta x$, where F is the force applied to the spring and Δx is the distance that the spring stretches as a result of that force. The variable represented by the lowercase letter k is called the spring constant. The unit of the spring constant is N/m (newtons per meter). You can think of the spring constant as describing the stiffness of the spring: the larger the spring constant, the more force that needs to be applied to make it stretch.

The spring shown in Figure 12.7 deflects 0.105 m (10.5 cm) when a mass of 0.1 kg (100 g) is placed on it. The weight of the mass is 1 N. The spring constant is therefore 9.52 N/m.

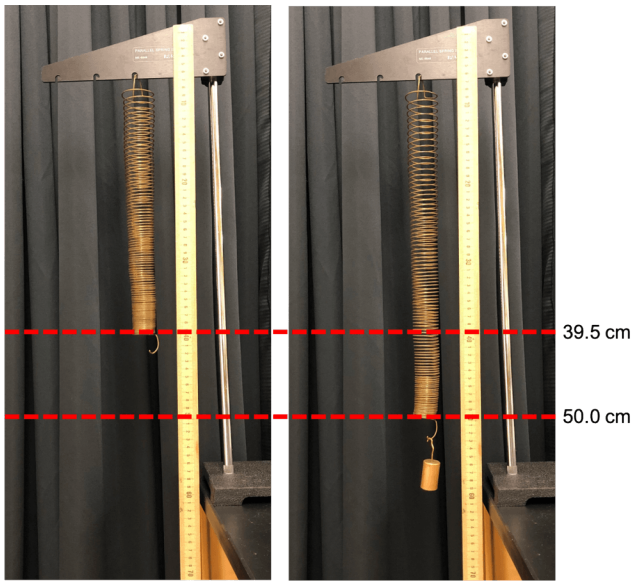


Figure 12.7 – A single spring has an equilibrium position at 39.5 cm on the meter stick (left). When a 100 g mass is placed on the spring, it stretches to 50.0 cm on the meter stick (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

When springs are connected together, the overall spring constant may change. Springs can be connected in series such that the springs join together like one long line, each spring connected together end to end (see Figure 12.8, left). When a mass is placed on series springs, each spring will stretch out based on its spring constant, as both of the springs feel the full force of that mass.

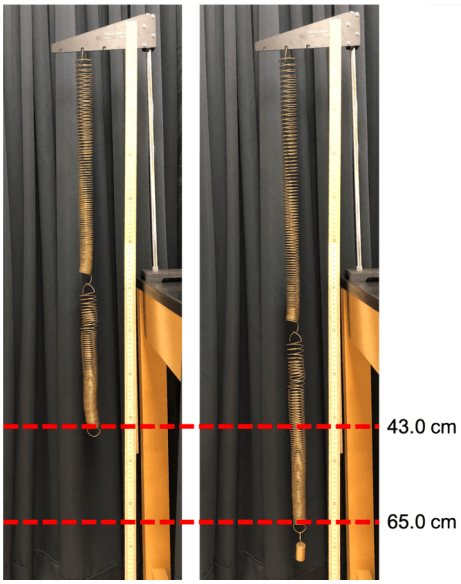


Figure 12.8 – Two series springs have an equilibrium position at 43.0 cm on the meter stick (left). When a 100 g mass is placed on the series springs, they stretch to 65.0 cm on the meter stick (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The series springs in Figure 12.8 deflect 0.22 m (22 cm) when a mass of 0.1 kg (100 g) is placed on them. The weight of the mass is 1 N. The spring constant of the series arrangement of springs is therefore 4.55 N/m. This is about one half the value for a single spring; the stiffness is reduced.

The effective spring constant of springs in series is smaller than the individual spring constants: the stiffness is lower because the spring combination stretches more than any one individual spring would stretch.

Springs can also be connected in parallel such that the springs join together side-by-side (see Figure 12.9, left). When a mass is placed on parallel springs, the force from that mass is distributed between both springs. Therefore they will stretch out less than they would do if they were individual or connected in series.

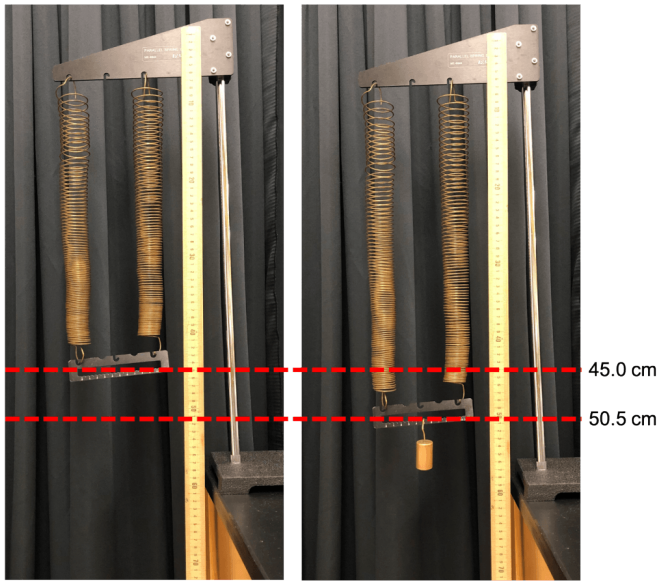


Figure 12.9 – Two parallel springs have an equilibrium position at 45.0 cm on the meter stick (left). When a 100 g mass is placed on the parallel springs, they stretch to 50.5 cm on the meter stick (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The parallel springs in Figure 12.9 deflect 0.055 m (5.5 cm) when a mass of 0.1 kg (100 g) is placed on them. The weight of the mass is 1 N. The spring constant of the parallel arrangement of springs is therefore 18.18 N/m. This is about double the value for a single spring; the stiffness is increased.

The effective spring constant of springs in parallel is larger than the individual spring constants: the stiffness is higher because the spring combination stretches less than any one individual spring would stretch.

Forces in solids

Solids are present in all aspects of our lives. As you read this book, you may be in a house or a building of some sort. That structure is built out of solid objects that are engineered to create a durable, robust building that does not move, even when people walk around in it, work in it, or when weather interacts with the outside of the building. How does that work?

All solid objects deflect to some degree when a force is placed on them. For example, the video below shows a meter stick. As Dr. Pasquale places masses on the meter stick, the meter stick deflects. When a lot of mass is placed on top, the meter stick bends noticeably.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When you stand on the floor, or when you sit in a chair, or when forces act on any solid object, those solids will deflect. The materials used in houses, chairs, tables, and other things that don't visibly bend when we use them have been chosen because they don't deflect much under the influence of the types of forces they will be subjected to. But just because we can't see it, doesn't mean it isn't happening.

In the demo shown in the video below, a laser has been set up to reflect off of a mirror on the door frame and return to a solar panel. That solar panel is then connected to a speaker. The speaker emits noises any time the distance between the laser and the door frame changes, using a principle called interference that will be discussed later in this textbook. This interference principle is capable of detecting really small changes in distance.

When Dr. Fazzini pushes on the door frame, the speaker emits noises, indicating that the distance has changed. When he pushes on the door frame, the metal deflects. Dr. Fazzini has literally bent the wall by pushing on it! Thankfully, the deflection is small and the building stays upright.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The two major forces that act on solid objects are tension and compression. Tension is the force that exists when something is pulled or stretched apart. Compression is the opposite: it is the force that exists when something is pushed or squeezed together.

Going back to the meter stick example, when forces are placed on the meter stick and it deflects, the force causes the bottom of the meter stick to stretch out: that is a tension force. But the top of the meter stick is squeezed together: that is a compression force. This is shown in Figure 12.10. What happens in between the top and bottom of the meter stick? The region along the middle of the stick between the top and bottom of the stick is known as the neutral layer. There are no forces acting in this part of the meter stick.

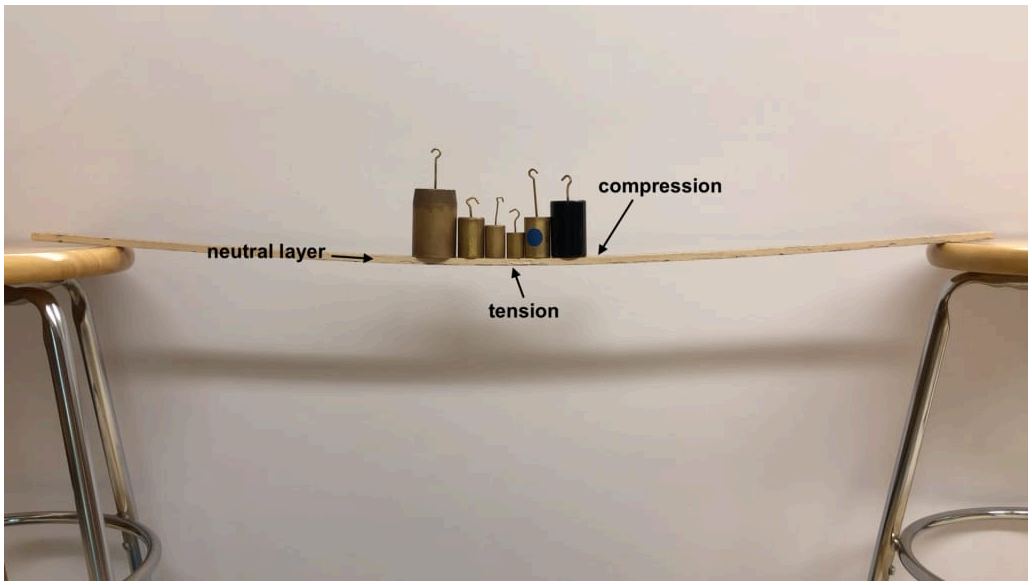


Figure 12.10 – A meter stick deflected by masses experiences compression on top, tension on the bottom, and a neutral layer in between. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The presence of a neutral layer explains the design of I-beams. An I-beam (shown in Figure 12.11) is used in the construction of buildings. If an I-beam is bolted together at both ends, the forces of tension and compression will exist on the top and bottom if the beam is deflected (similar to the meter stick shown in Figure 12.10). Because the tension and compression forces don't exist at the center of the beam, engineers can cut down on the construction materials used in this area. Notice that the top and bottom of the I-beam, which will be subjected to the most forces, have the most material present. In the center, where there are no forces, less material is needed, so the center is carved out. This reduces the cost and weight of the building materials. (The actual forces acting on an I-beam may be more complicated than this, but that discussion is beyond the scope of this textbook.)

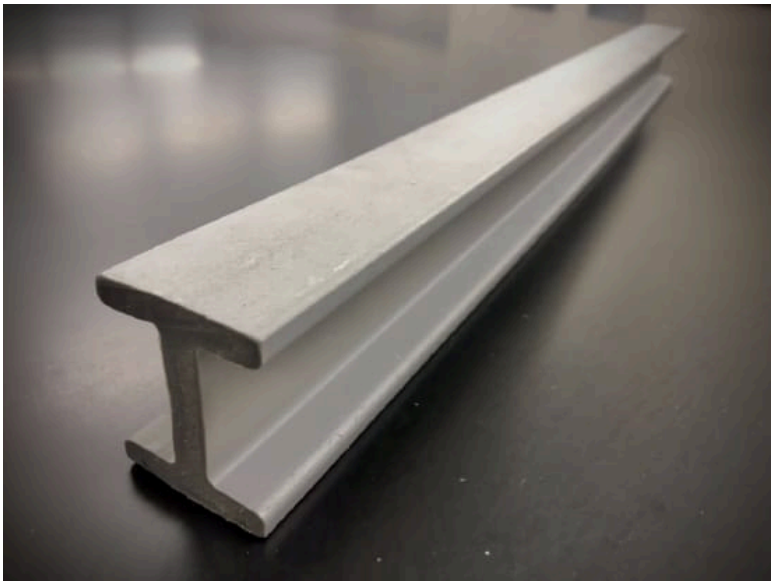


Figure 12.11 – A plastic I-beam. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Another place you may notice compression and tension forces is if you ever see an empty flatbed truck. Notice that the bed of the flatbed truck is curved upward (see Figure 12.12 for an example). When a heavy load of objects is placed on the flatbed, it deflects and becomes flat.



Figure 12.12 – An empty flatbed truck has an upward curve to it. Flatbed by raymondclarkeimages, is licensed under CC BY-NC 2.0.

Scaling

Scaling has to do with the properties of a solid object and how they change when we make an object larger or smaller. This topic has a lot of important ramifications in chemistry, biology, and physics. The properties we are interested in are the surface area of the object, the volume of the object, and the weight of the object. We'll measure these properties for a few cubes to come up with scaling parameters.

Cross-sectional area gives an indication of how strong an object is. If you've ever played sports or done weight training, you may have noticed your muscles increase in cross-sectional area as they become stronger. We can expect that an object with a larger cross-sectional area will be stronger than an object with a smaller cross-sectional area if all other things are kept equal.

Volume and density tell us how massive an object is, or how much it weighs when it is subject to the force of gravity.

Cross-sectional area divided by volume tells us how strong something is compared to its weight. This is an important concept from a structural standpoint. An object that is very strong but weighs a lot will be less structurally sound than something that has the same strength and weighs less. An ant is very strong: the cross-sectional area of an ant's legs is small, but they have very little volume, so the cross-sectional area to volume ratio is rather high. When we look at larger insects and animals, we see that the larger an animal is, the larger the cross-sectional area of its supporting limbs to support their weight. An elephant has very thick legs to keep the cross-sectional area to volume ratio high enough for the elephant to be able to walk around.

Simply increasing the volume of an object (such as an ant) will cause the volume of the ant to increase much more rapidly than the cross-sectional area. From a structural standpoint, this means that we cannot simply make things larger and expect them to be as structurally sound. An ant, blown up to the size of an elephant, would not be able to support its own weight. Sci-fi movies featuring giant spiders destroying cities are getting physics wrong: those giant insects would have a very small cross-sectional area to volume ratio, making those giant insects completely unable to move!

Another important factor in scaling is the total surface area to volume ratio. This has many important applications in heating, chemical reactions, and biology. As an object is scaled up to become larger and larger, the surface area divided by volume ratio will change.

A cube that measures 1 cm by 1 cm by 1 cm (Figure 12.13) has six faces each with a face area of 1 cm^2. The total surface area is therefore 6 cm^2. The volume of the cube is length times width times height: 1 cm^3. The surface area to volume ratio is 6 cm^{-1} (inverse centimeters).



Figure 12.13 – A single cube with edge length of 1 cm. A ruler is shown for reference. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Suppose this cube goes into a magical machine that maintains its properties (such as density) but simply scales it up to have a size of 2 cm by 2 cm by 2 cm (Figure 12.14). Each face has an area of 4 cm^2, so this new cube has a total surface area of 24 cm^2. That's much bigger than before! But now the volume has increased to 8 cm^3, making the surface area to volume ratio a mere 3 cm^{-1}, which is smaller than the previous case.

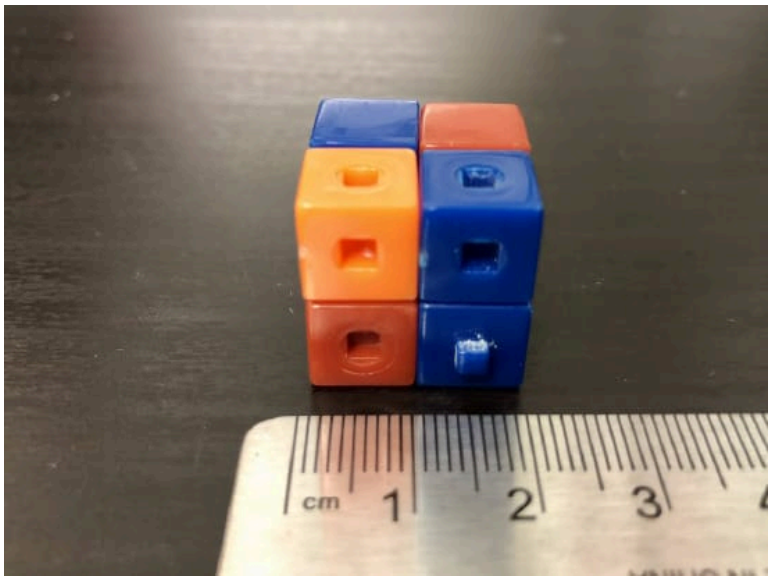


Figure 12.14 – A cube with edge length of 2 cm. A ruler is shown for reference. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Suppose the cube becomes even larger: 3 cm by 3 cm by 3 cm (Figure 12.15). The area of each face is 9 cm^2, and the total surface area is 54 cm^2. The volume is 27 cm^3, so the surface area to volume ratio is 2 cm^{-1}.

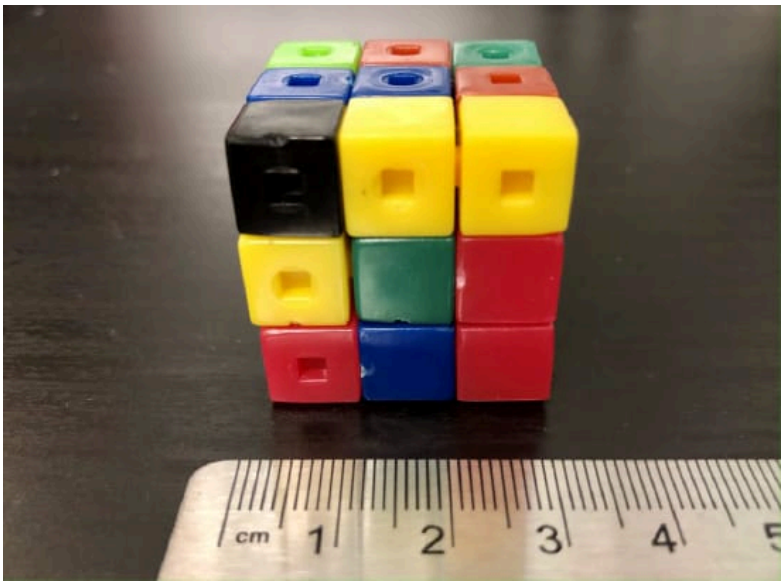


Figure 12.15 – A cube with edge length of 3 cm. A ruler is shown for reference. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

As the cube continues to become larger and larger, the surface area will continue to increase, but it does not increase as fast as the volume. A cube with an edge length of x centimeters (where x can be any number we choose) will have a surface area of $6x^2$, a volume of x^3, and a surface area to volume ratio of $6/x$. The larger x becomes, the smaller the surface area to volume ratio. This data is outlined in the table below for cubes of various edge lengths.

1	1	6	1	6
2	4	24	8	3
3	9	54	27	2
4	16	96	64	1.5
5	25	150	125	1.2
x	x^2	$6x^2$	x^3	6/x

Surface area matters in chemistry, biology, and other related fields. Consider a biological cell. As a cell becomes larger in volume, it requires more nutrients. But nutrients can only be absorbed through the cell wall, which is only as large as the surface area. At a certain point, if a cell becomes too big, it cannot obtain enough nutrients or other chemicals to carry out its functionality. Instead

of becoming larger and larger, cells divide. This keeps the surface area to volume ratio sufficient for cells to continue to survive.

Surface area also plays a large role in heat transfer. Many computer processors have a heatsink placed over the processor (see Figure 12.16). As you browse the internet, play games on your computer, or write up an essay for your English class, the computations that your computer performs cause the processor to heat up. That heat is then dissipated through the heatsink, which turns on a fan to blow cooling air over the system and keep things from melting.



Figure 12.16 – A computer processor contains a metal heatsink with a large surface area. The fan on top blows air over the heatsink to cool the processor. AMD Heatsink and fan, by fir0002, is licensed under GFDL 1.2.

A heatsink that uses airflow to dissipate heat usually has a large surface area. The more surface area there is, the easier it is for heat to get transported away. You also see this in air-cooled cylinders in airplane engines or motorcycles, among other things. An air-cooled airplane engine cylinder shown in Figure 12.17 has the characteristic cooling fins of a heatsink.

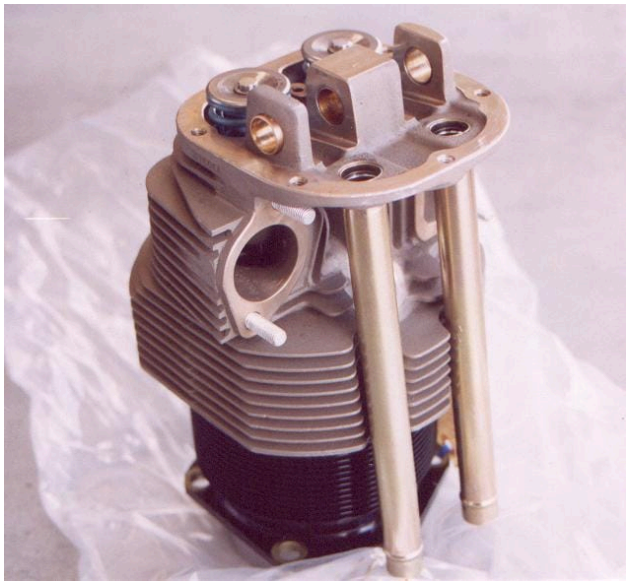


Figure 12.17 – An air-cooled engine cylinder has characteristic fins that increase surface area. Newjug1, by Ali0th, is licensed under CC BY-SA 3.0.

As a final example, consider the video shown below. In this demo, Dr. Pasquale has a tin plate covered with lycopodium powder. This is a dry powder that is also very flammable. When she tries to ignite the lycopodium powder, she has a hard time doing so. Only a little bit of the powder burns before the fire goes out. This is because there is very little surface area available to combust with the surrounding fire and oxygen.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When Dr. Pasquale instead blows individual lycopodium powder into an open flame, the result is that the lycopodium powder explodes. By exposing more of the surface area, the chemical reaction is more readily enabled, and we can see that very dramatically in this example. The two outcomes are shown in Figure 12.18.



Figure 12.18 – Dr. Pasquale is unsuccessful at lighting lycopodium powder on fire when it is placed in a pie plate with minimal surface area (left). When the surface area is increased, Dr. Pasquale is able to create an explosion (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Chemical reactions will be more vigorous when the reactants have more surface area than when they have less. This is true for experiments in test tubes as well as in fires. Grain elevators are at risk of explosion after grains have been newly introduced, stirring up lots of particles and exposing a lot of surface area. If any sparks occur during these times, grain elevators can explode. A stereograph of a grain elevator explosion that occurred in 1878 is shown in Figure 12.19.

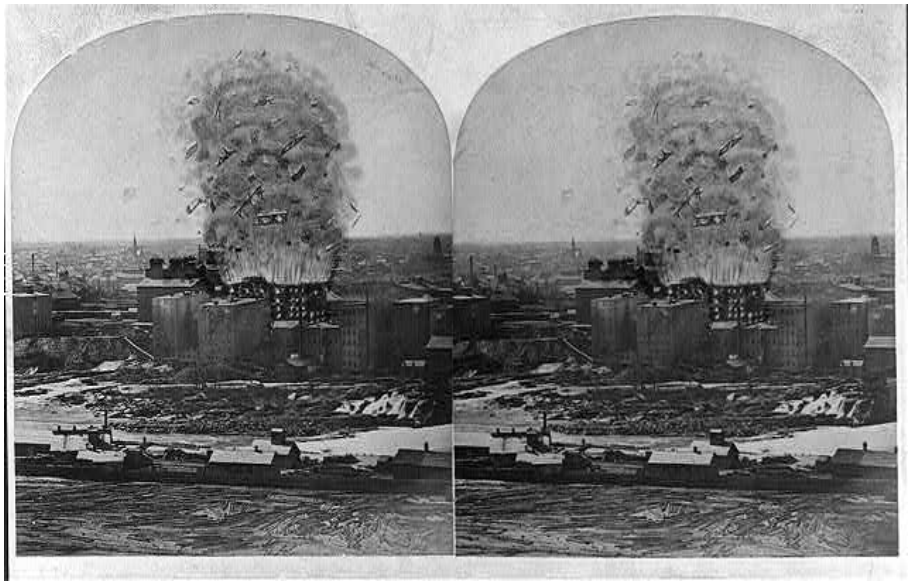


Figure 12.19 – A stereograph of the Washburn Mill Disaster (grain elevator explosion) of 1878. Stereograph of the Washburn A Mill explosion, by Library of Congress , is licensed under Public Domain.

Further reading

- PhET Hooke's law simulation – Play with this simulation to explore Hooke's law for single springs, series springs, and parallel springs.

Practice questions

Conceptual comprehension

1. What would happen if the bed of a flatbed truck started out flat, and then deflected downward when a heavy load was placed on top?

Numerical analysis

1. An object has a mass of 50 g (0.05 kg) and measures 0.02 m^3 (20,000 cm^3). Calculate its density. Use units of g/cm^3 and then re-calculate with units of kg/m^3.
2. An irregularly shaped stone has a mass of 750 g (0.75 kg) and displaces 0.0004 m^3 (400 cm^3) of water when submerged. Calculate its density. Use units of g/cm^3 and then re-calculate with units of kg/m^3.
3. A wooden cube has a mass of 20 g (0.02 kg) and each edge of the cube measures 3.3 cm. First, calculate the volume of the cube. Then, calculate its density. Use units of g/cm^3 and then re-calculate with units of kg/m^3.
4. A spring has a spring constant of 200 N/m. If a force of 50 N is applied to the spring, how much will it stretch?
5. A spring with a spring constant of 300 N/m is stretched out a distance of 0.1 meters. Calculate the force required to stretch it.
6. A spring is stretched by 0.05 meters when a force of 25 N is applied. Calculate the spring constant.
7. Suppose you have two identical springs in series. Each spring has a spring constant of 30 N/m. If a force of 5 N is applied to the combination of these two springs, what will be the total distance that the springs will stretch?
8. Suppose you have two identical springs in parallel. Each spring has a spring constant of 30 N/m. If a force of 5 N is applied to the combination of these two springs, what will be the total distance that the springs will stretch?
9. A small cube has sides measuring 4 centimeters each. If you scale it up proportionally to create a larger cube (with equal density), and the sides of the larger cube are now 12 centimeters each, calculate the...
    1. ...total surface area of the new cube.
    2. ...volume of the new cube.

3. ...surface area to volume ratio of the new cube.

Hands-on experiments

1. Find a few irregularly shaped objects (like stones or pebbles) and a measuring cup. Fill the measuring cup with water and record the initial water level. Carefully add the objects one by one, recording the new water level each time. Calculate the volume of each object by subtracting the initial water level from the final water level. If you have a scale to measure mass, calculate the density using the mass and volume of each object.

This page titled Chapter 12: Solids is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 13: Liquids

Summary

A liquid is a state of matter in which molecules or atoms are free to move about much more easily than in a solid; however, the molecules are still packed rather closely. A liquid will change its shape, but cannot be easily compressed to change its volume. Liquids play a huge and important role in our lives. This chapter will explain many important properties about liquids, including why things float and the basic idea of how water systems work.

Pressure

Pressure is a concept that is important in solids, liquids, and gases. Even the pressure of the atmosphere plays an important role in our lives.

The symbol for pressure is the capital letter P. The SI unit of pressure is the pascal (Pa), which can be broken down into N/m^2. Pressure caused by a solid object is equal to force divided by area, or

$$P = \frac{F}{A}.$$

In other words, this is the amount of force divided by the amount of area over which the force is applied.

When Dr. Pasquale participated in the "bed of nails" demo, which was explained in chapter seven, the reason the nails don't hurt her too much when she lies down on them is because the weight of her body, the force involved in the equation, is distributed over a large number of nails. While each nail has a relatively small surface area, the combined surface area of all of the nails causes the pressure to be less than it would be on any given nail. Dr. Pasquale would not have done this demo if there had only been one nail on the board!

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Pressure is also a concept that is important in fluids: liquids and gases. In a liquid, the pressure is equal to the density of the liquid times the gravitational acceleration times the depth, in other words,

$$P = (\frac{m}{V})(g)(depth).$$

If you've ever gone swimming or scuba diving and had your ears pop, that's because the pressure in water increases as you swim deeper and deeper. Note that the equation for pressure does not have to do with how large an area the liquid is distributed over: it only has to do with the density, gravitational acceleration, and depth. If you dive 5 meters under the surface of Lake Michigan, you would experience the same pressure that you would 5 meters under a freshwater pond, even though the two bodies of water are different sizes.

One of the consequences of liquid pressure is that water always "seeks its own level" in a container. If you've ever taken a container of liquid and tilted it, perhaps you've noticed that the water level will always even out. A video of this is shown below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Even in the strangely shaped container shown in the video (and also shown in Figure 13.1), water seeks its own level as a result of keeping all forces due to water pressure in the fluid in equilibrium.

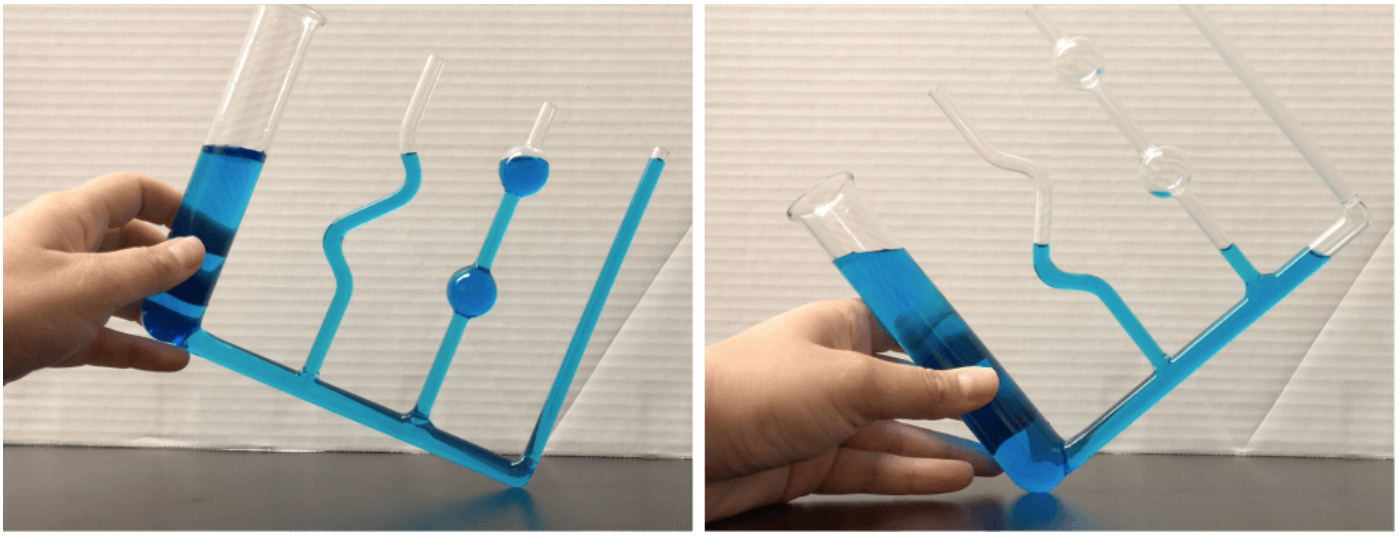


Figure 13.1 – Water always seeks its own level, even in a strangely shaped container. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Consider an example of water that is not equalized after rotating a container, shown in Figure 13.2. The depth of the water is unequal from one end to the other. In the example shown in Figure 13.2, the left-hand side has a higher height above the bottom of the container, and the right-hand side has a lower height above the bottom. This causes more pressure to be present on the left. If there's more pressure, then there is a net force between the left and right side. Newton's laws state that when there's a net force, there will be acceleration. The water will move until the height is equal on both sides, causing an equal amount of pressure on each side resulting in no net force.

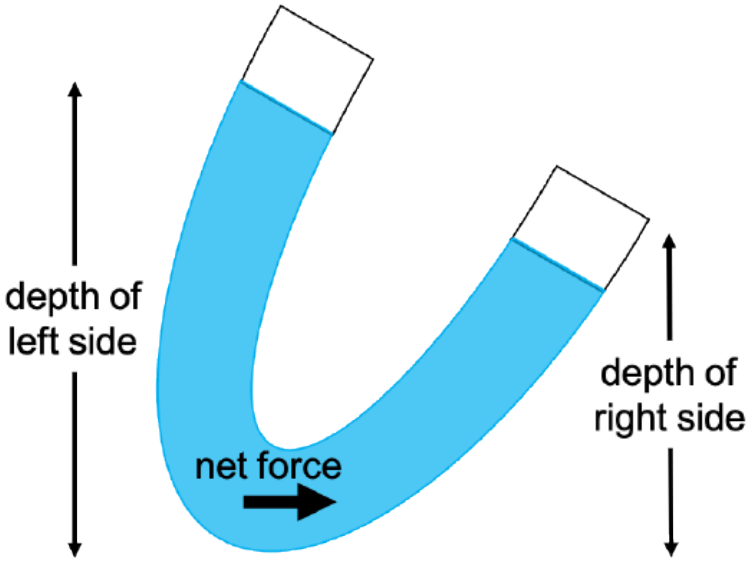


Figure 13.2 – Water that does not seek its own level will experience a net force. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Archimedes' principle

If you've gone swimming, you may have noticed that you feel like you weigh less in water than you do on land. It's true! Liquids exert an upward buoyant force on submerged objects, which means we exert less of a support force if we're standing in a pool or other shallow body of water than we would if we were on dry land. If the density of a fluid is high enough, such as in salt water, the buoyant force may actually be large enough to reduce the support force to zero and cause us to float.

The buoyant force acting on an object placed in a liquid has to do with the differences in the fluid pressure between the top and bottom surfaces of the immersed object. For example, a cube placed in water is going to have pressure exerted on each of the faces of that cube, creating a force on each cube face (as depicted in Figure 13.3). Because pressure increases with depth, the lowest part of the cube has more force than the highest part of the cube because the greater pressure on the same sized area produces a greater force. All of the horizontal forces on the cube cancel out. The difference between the force pushing down on the top surface and the force pushing up on the bottom surface produces a net force pointing up. This difference is the buoyant force.

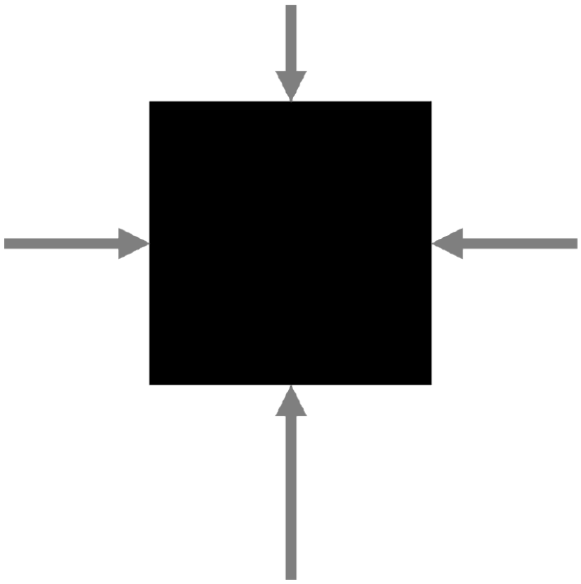


Figure 13.3 – An object that is submerged in a fluid will experience forces on each face due to the liquid pressure. This results in an upward (buoyant) force. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

All objects placed in a liquid have a buoyant force pointing upward. If you place an object that sinks at the bottom of a container filled with water, that buoyant force will act on the object. When the object sinks, it stays put at the bottom of the container. There's no net force. This means the gravitational force acting on the object (mass times gravitational acceleration) is canceled out by the buoyant force and support force. The forces acting on a lead weight that sinks in water are shown in Figure 13.4. The buoyant force causes the support force to be smaller than it would be on land. This is why we feel lighter in a swimming pool.

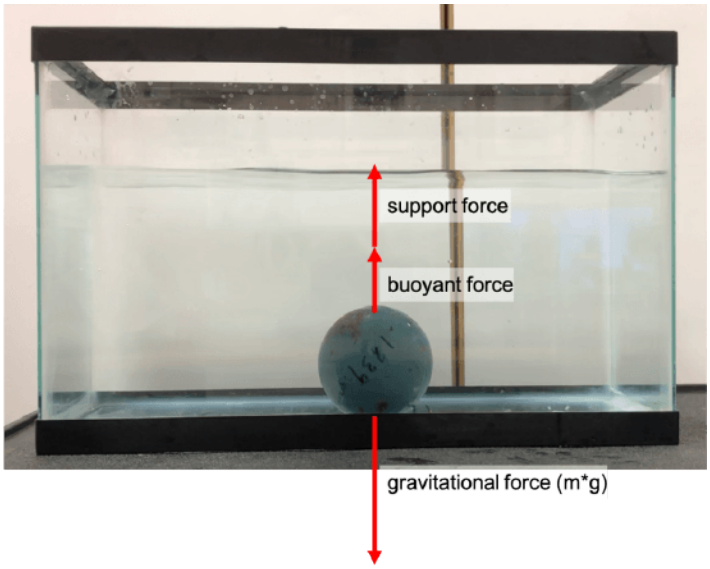


Figure 13.4 – An object that sinks in water experiences a smaller normal force than it would on land. The upward buoyant force and support force together equal the force due to gravity acting on the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

An object that floats will have a buoyant force greater than or equal to the gravitational force acting on the object. If an object that floats is fully submerged in water, the buoyant force, being greater than gravitational force, leads to a net force, causing the object to accelerate upward (see Figure 13.5, left). At some point the object will reach equilibrium as it rests on the surface of the water. The buoyant force decreases as the object displaces less liquid as it moves upward out of the water. When the gravitational and buoyant forces are equal, the object will stop moving and simply float (see Figure 13.5, right).

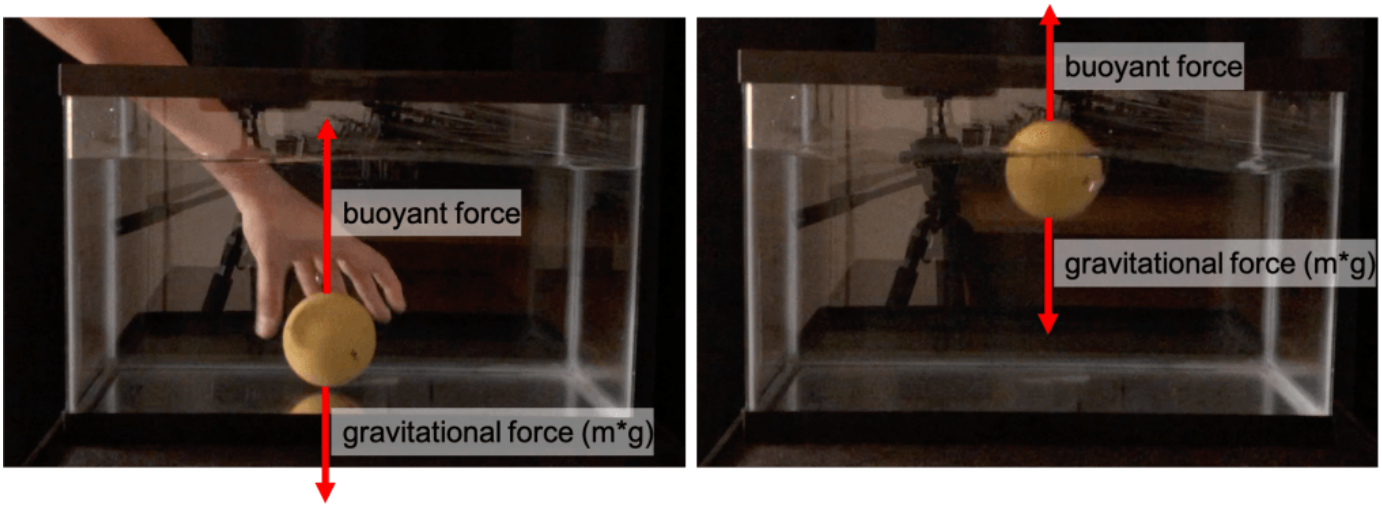


Figure 13.5 – A wooden object that floats in water is released at the bottom of a fish tank. At this point, the buoyant force is greater than the gravitational force, causing a net upward force (left). After moving to the top of the tank, the buoyant force becomes equal to the gravitational force, putting the wooden sphere into equilibrium (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The wooden sphere from Figure 13.5, placed into the bottom of a tank of water, will experience a net force, causing the wooden sphere to accelerate upward until the forces are in equilibrium. This is shown in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Archimedes' principle states that the buoyant force is equal to the weight of the fluid that is displaced when a solid object is placed into a fluid. This means that the buoyant force is equal to the mass of the displaced fluid times the gravitational acceleration. Since density is mass divided by volume, we can write the mass of the displaced fluid as equal to the density of the fluid times the volume of the displaced fluid. Therefore, the magnitude of the upward buoyant force can be written as the density of the fluid times the volume that is displaced times the gravitational acceleration. In other words,

$$F_b = D_{fluid} V_{displaced} g.$$

Buoyant force, being a force, is measured with newtons (N). Recall that newtons are equal to kgm/s^2. In order for units to work correctly in the buoyant force equation, the density of the fluid must be expressed in units of kg/m^3 (not cm/g^3). In addition, volume must be expressed in m^3, not cm^3. If you have different units for density or volume, then you must first do a unit conversion before determining the force in newtons.

The equation for buoyant force implies that for two objects in the same liquid, whichever object displaces the most liquid will experience the larger buoyant force. If both objects displace the same amount of liquid, then they will experience the same magnitude of buoyant force.

A second equation that can be used to quantify buoyant force is that the buoyant force on an object is equal to the weight of the object in air minus the weight of the object in the fluid. In other words,

$$F_b = W_{air} - W_{fluid}.$$

Shown in the video below is another demonstration of Archimedes' principle. Dr. Pasquale has a heavy brass ball sitting on some dried beans. When she shakes the container, the brass ball moves downward to the bottom of the container, and out pops a ping-pong ball! The ping-pong ball has a lower density than the dried beans, and therefore experiences a buoyant force. The brass ball has a high density and sinks. This demonstration only works when Dr. Pasquale shakes the container, creating a type of fluid flow.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Flotation

The principle of flotation states that any object that floats displaces a weight of fluid equal to its own weight. For example, a boat that has a weight of 10,000 N will float if it displaces 10,000 N of water.

A solid piece of iron has a density approximately 8 times greater than that of water. This means that if we were to place a piece of solid iron into a container of water, it would only displace one eighth of its weight in water. This is not enough to make it float.

If iron sinks in water, how is it that boats and submarines can be made out of iron and other dense metals? If you think about the

shape of a boat in a simplistic manner, it is carved out so that it is able to displace its own weight in water. Effectively, the boat has sections of air that cause the overall density of the submerged portion of the boat including the air to be less than that of the fluid on which it is floating.

An object that is less dense than a fluid will float in that fluid. An object that is more dense than a fluid will sink in that fluid.

Pascal's principle

Pascal's principle states that a pressure change everywhere in a closed system of a liquid is equal at all points in the system.

In equation form, Pascal's principle states that

$$\begin{aligned} \Delta P_1 &= \Delta P_2 \\ \frac{\Delta F_1}{A_1} &= \frac{\Delta F_2}{A_2}, \end{aligned}$$

where P_1 and P_2 are pressure changes at two points in a closed system. If a force is applied to point 1, F_1 is the force applied at that point and A_1 is the area over which the force is applied. At point 2, having an area of A_2, F_2 is the output force. In other words: the change in force divided by the area at one point in a system equals the change in force divided by area somewhere else in the system.

This equation is depicted in Figure 13.6. A closed system of fluid containing two pistons (one small, left, and one large, right) shows the relative forces that would be present on each piston to abide by Pascal's principle. Because the piston on the left is small, a small force applied at that point would yield a large force on the large piston (on the right).

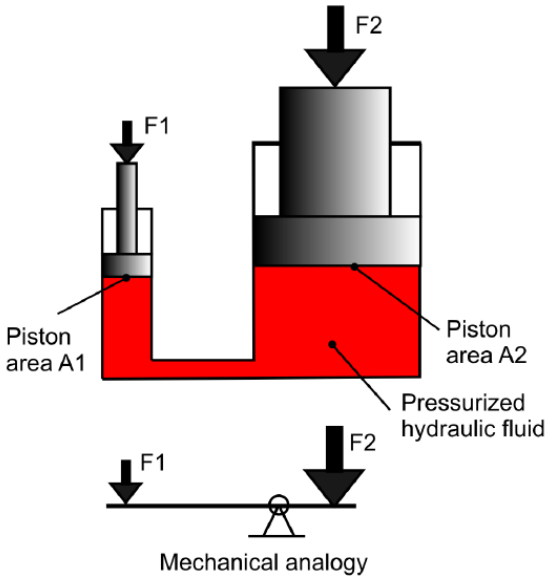


Figure 13.6 – A diagram of Pascal's principle. A closed system with two pistons (one with a small area, one with a large area) shows the sizes of forces that would result due to Pascal's principle. A mechanical analogy using a lever is also depicted. Hydraulic Force Torque, by Lidingo, is licensed under CC BY-SA 3.0.

In Figure 13.6, a mechanical analogy is also shown, as this is similar to the mechanical advantage that can be obtained with a lever (which is a type of simple machine). This is because the volume of fluid pushed out of the small piston is equal to the volume of fluid pushed into the large piston (due to conservation of mass). Because volume equals cross-sectional area times height, the small piston moves a small force through a long distance (height) while the large piston moves a large force through a short distance.

In the video below, Dr. Pasquale has a piston system similar to that shown in Figure 13.6. When she places a mass of 200 g on the smaller piston (on the right), that generates a force of 2 N due to gravity. That force is exerted over a small area, causing a pressure change. As a result, an equal pressure change will occur over the large piston. This results in a mass of 500 g (with a weight of 5 N) lifting up. The force that Dr. Pasquale placed on the small piston was multiplied by 2.5 due to the mechanical advantage provided by Pascal's principle.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This mechanical advantage forms the basis of hydraulic systems. A small force can be exerted on a small area to achieve a large force output on a large area somewhere else. If the mechanical advantage is large enough, this could be used to lift something really large, like a car. Hydraulic lifters use this principle to lift heavy objects. Hydraulics are also used in automobile braking systems. Exerting a small force on a pedal causes our cars to slow down. Hydraulics are also used in large airplanes. Pilots can exert a small force on a control yoke or rudder pedals and a hydraulic system creates a deflection in the airplane flight controls that would otherwise be too heavy for humans to move on their own.

Pascal's principle also forms the foundation for modern water supplies and distribution systems. In the Midwest where it's relatively flat, we have water towers to create water pressure. A photograph of a water tower in Naperville, IL is shown in Figure 13.7. When you open the tap on a sink, there will be water pressure because the building is connected to a water main that's pressurized.



Figure 13.7 – A water tower in Naperville, IL. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The bad news is that if there is a break in the water main at any point, water pressure will be lost everywhere in the water system. When author Dr. Pasquale was living in Boston in 2010, there was a water main break that caused about two million people, including her, to lose water pressure in their homes and apartments. Because the pressure change is transmitted throughout a system, a water break in one spot will transmit everywhere!

Surface tension

Surface tension is a force that acts parallel to a surface of liquid, such as water. This is depicted in Figure 13.8. As a result of surface tension, liquids such as water will tend to minimize their total surface area to minimize the forces acting on each molecule. In short, surface tension describes the resistance of the surface of a liquid to being stretched.

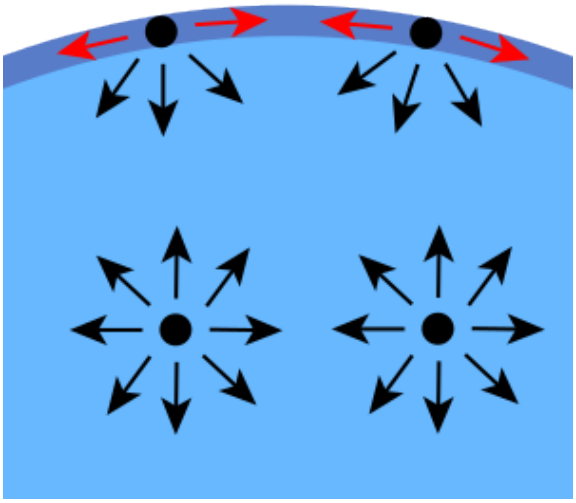


Figure 13.8 – The force of surface tension acts parallel to the surface. Diagram of the cohesive forces on molecules of a liquid, by Booyabazooka, is licensed under Public Domain.

In most liquids, including water, surface tension causes water to form drops. The spherical shape of the drop minimizes the amount of force present on the surface. In a way, you can think of the spherical shape as the most comfortable arrangement for the molecules. A photograph of a water droplet falling out of a pipette is shown in Figure 13.9 (in fact, even the water stuck to the sides of the pipette is spherical or dome-shaped).

In a volume of water, bonds between molecules will cause the molecules in the center of the water to have no net force. Molecules at the surface don't have forces acting on them in all directions, causing a force of surface tension.



Figure 13.9 – A water droplet falling out of a pipette is a spherical shape to minimize the forces of surface tension. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Surface tension explains how objects with enough surface area can rest on the top of water without getting submerged. In Figure 13.10, a paper clip is shown resting at the top of a beaker of water. The paper clip is not floating. Aluminum is more dense than water, so we would expect it to sink. However, the paper clip is sitting on top of the water because the surface tension of the water is greater than the weight of the paper clip. Much like the surface of a trampoline, the surface of the water stretches and sags slightly due to the weight of the paper clip resting on top of it.

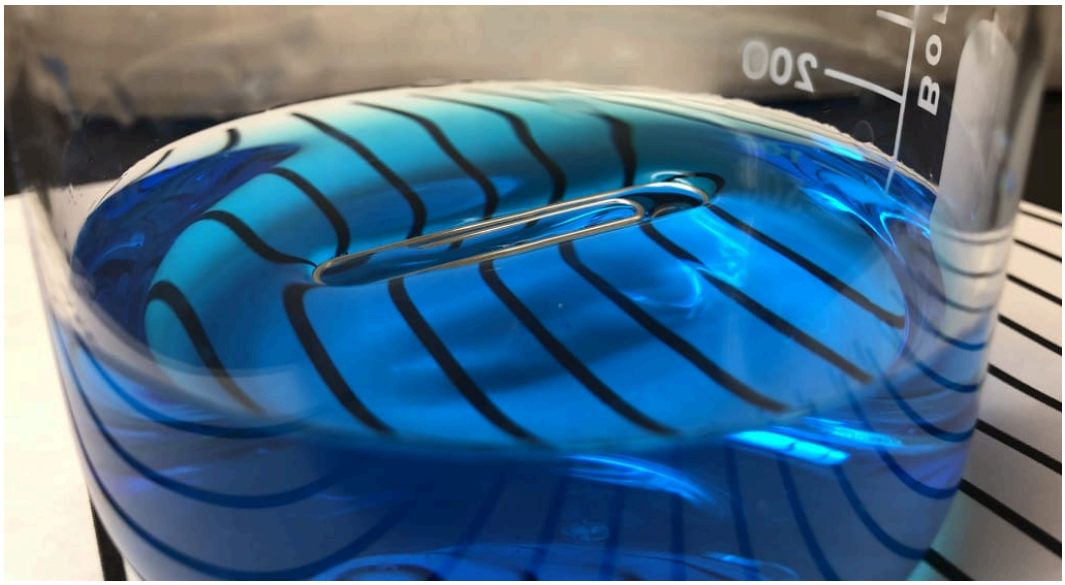


Figure 13.10 – A paper clip appears to float on the surface of water. The force of surface tension holds it up. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale places the paper clip into the beaker of water. When she pushes down on the paper clip, the downward force is large enough to break through the surface tension. The object “falls through” the surface. (In our trampoline analogy, the trampoline’s surface has been torn.) The paper clip then sinks as expected, since the aluminum is more dense than water.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=756#oembed-6

Similar to the paper clip resting on top of water, insects like water striders can walk across water. Surface tension also explains why it hurts to do belly flops into a swimming pool: when your body must break through the force of surface tension, it can be painful!

Capillarity

Capillarity is related to surface tension. When liquids are in a narrow space, they can cause the surface of a liquid to form a curve called a meniscus. If in a narrow enough space, the liquid can rise up the walls of a container. The curvature of the meniscus depends on the competing effects of cohesion versus adhesion. Atoms and molecules attract to other atoms and molecules through electrical interactions of their electrons. The attractive forces between “like” atoms and molecules (such as water molecules to other water molecules) is referred to as cohesion. The attractive forces between “unlike” atoms and molecules (such as between mercury and glass) is referred to as adhesion. Referring to Figure 13.11 (right), the adhesive force between water and glass is greater than water’s cohesive force. In other words, water would rather stick to the glass than next to other water molecules. This arrangement creates a concave meniscus shape as water molecules are drawn up along the glass. This minimizes the forces in the water. In the case of mercury in glass (shown in Figure 13.11, left), the mercury atoms would rather stick to themselves than stick to the glass. This results in a convex meniscus that is opposite to the water-in-glass meniscus as the mercury atoms clump together as much as possible.



Figure 13.11 – Mercury creates a convex meniscus (left), while water creates a concave meniscus (right). Meniscus in mercury and water by karabekirus, is licensed under CC BY-SA 2.0.

Capillarity explains how paper towels absorb moisture, how trees get water from their roots to their branches, and how small amounts of blood can be drawn using a finger stick.

Practice questions

Conceptual comprehension

1. Which of the two spheres shown in Figure 13.12 experiences the larger buoyant force, the wooden sphere (floating, left) or the iron sphere (sinking, right)? Both spheres have identical volumes.

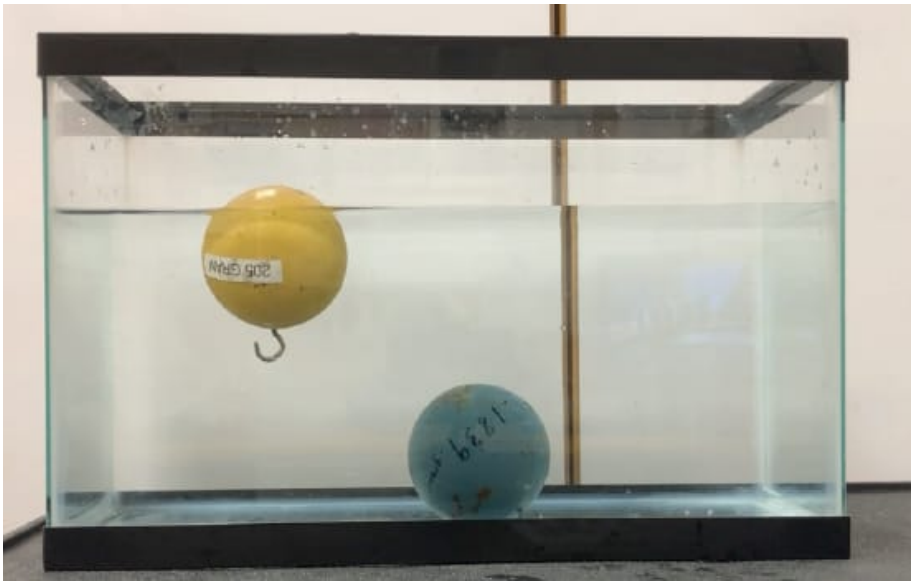


Figure 13.12 – A wooden sphere floats in water (left) while an iron sphere sinks (right). Both spheres otherwise have identical volumes. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Numerical analysis

1. A heavy box with a mass of 30 kilograms is placed on the floor. If the box has a base with dimensions of 0.5 meters by 0.5

meters, calculate the pressure that the box exerts on the floor.

2. The book shown in Figure 13.13 has a thickness of 4.5 cm, a depth of 20.5 cm, and a height of 26 cm. It weighs 22 N. Calculate the pressure exerted by the book on the table when it is...
    1. ...upright (Figure 13.13, left).
    2. ...laying down (Figure 13.13, right).

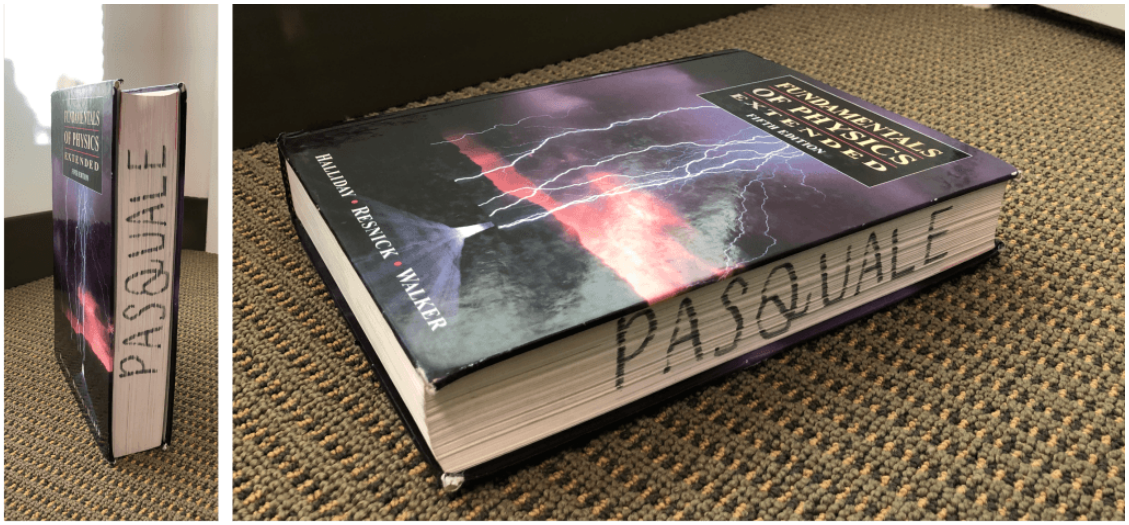


Figure 13.13 – A book positioned upright on a table (left) and on its side (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

3. A person weighing 550 N has a foot area of 160 cm^2 each. Sometimes this person enjoys wearing high heels, which have an area of 1 cm^2 each. Calculate the pressure the person exerts on the floor when they are...
    1. ...barefoot.
    2. ...wearing high heels.
4. A car with a mass of 1,200 kg has tires with an area of contact with the road of 0.02 m^2 each. Calculate the total pressure exerted by all four tires on the road when the car is stationary.
5. A freshwater swimming pool is 2 meters deep. Calculate the pressure at the bottom of the pool due to the weight of the water.
6. A submarine is 30 meters below the surface of the ocean. Calculate the pressure exerted on the submarine due to the weight of the water above it. (Assume the density of seawater is 1,025 kg/m^3.)
7. A solid metal cube with a volume of 0.02 m^3 is completely submerged in freshwater. If the density of the metal is 8,000 kg/m^3, what is the buoyant force acting on the cube?
8. A wooden block with a volume of 0.05 m^3 and a density of 600 kg/m^3 is placed in a container of oil with a density of 900 kg/m^3. Calculate the buoyant force on the wooden block when it is fully submerged in the oil.
9. A ship has a mass of 400,000 kg. Calculate...
    1. ...the weight of the ship.
    2. ...the weight of liquid that must be displaced to cause the ship to float.
    3. ...the volume of freshwater that must be displaced to cause the ship to float.
    4. ...the volume of seawater (1,020 kg/m^3) that must be displaced to cause the ship to float.
10. If a force of 200 N is applied to a piston with an area of 0.02 m^3 in a hydraulic system, calculate the pressure that is exerted on the fluid inside the system.
11. In a hydraulic jack, the small piston has an area of 0.005 m^3, and the large piston has an area of 0.05 m^3. If a force of 500 N is applied to the small piston, calculate the force that will be exerted by the large piston.
12. In a hydraulic lift used to raise a car, the small piston has an area of 0.02 m^3. If a force of 10,000 N is applied to the small piston, calculate the force that will be exerted on the car resting on the larger piston (with an area of 0.1 m^3).

Hands-on experiments

1. You can demonstrate density by making a simple density tower using liquids of different densities, such as water, vegetable

oil, and syrup. Pour these liquids carefully, one on top of the other, in a clear container. Objects (like raisins or small plastic toys) can be added to see if they float or sink, depending on their density relative to the surrounding liquid.

This page titled Chapter 13: Liquids is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 14: Gases

Summary

Gas is a state of matter where molecules are very free to move about and generally do not interact with each other except during collisions. This means that the shape and volume of a gas is free to change. A gas will assume the same shape as the container that it's in, and it can be compressed to take up less space, or expanded to take up more space.

The Earth's atmosphere

The Earth's atmosphere is a layer of gases that surrounds the planet. The atmosphere doesn't just allow us to breathe, it also protects us from UV radiation, traps moisture and heat, and exerts a force on every surface it touches.

The composition of the Earth's atmosphere is approximately 78% nitrogen, 21% oxygen, and 1% argon. There are also trace amounts of some other gases in our atmosphere: carbon dioxide, water vapor, and others.

Because gas molecules are not in contact with each other (except during very brief collisions), gases can be compressed. The atmosphere will be denser at sea level than at higher elevations and altitudes due to the weight of the molecules above pushing down on the molecules below. This is analogous to a human pyramid: in a human pyramid, the people at the bottom are compressed from the forces from all the people above (see Figure 14.1). In our atmosphere the molecules at the bottom are compressed from all the molecules above.



Figure 14.1 – A human pyramid is created by people at the base holding up levels of people above them. This is analogous to the compression of our atmosphere. Escala de 9 Falcons Vilafranca, by Pere López, is licensed under CC BY-SA 3.0.

If we were to take a ride from sea level up and up and up, we would notice the density of the atmosphere decreasing with our altitude. At a certain point, we would require an oxygen tank to allow us to breathe. Though the composition of the atmosphere remains constant for about 100 kilometers, meaning that we still have 20% oxygen, the total number of molecules decreases when density decreases. This is why commercial airplanes have pressurized cabins as they fly through the stratosphere, and why many mountain climbers summiting Mount Everest will take oxygen with them.

The atmosphere of our planet has no real height to it. There is no point in our journey up and up to higher altitudes where the atmosphere stops. Instead, it is more of a gradient. There are a lot of atmospheric molecules close to the surface of the Earth,

and fewer and fewer as altitude increases. Most of the mass of the Earth's atmosphere exists below 100 kilometers, known as the Kármán line. This altitude is used as the designation of where "space" begins and the people who go beyond that altitude are considered to be astronauts. However, atmospheric phenomena such as auroras can still occur above this point.

Atmospheric pressure

At sea level, the pressure we experience due to the atmosphere is approximately 100,000 Pa. That means over a one square meter area, the force exerted by the atmosphere is equal to 100,000 newtons! This value of 100,000 Pa is also known as 1 atmosphere (atm) of pressure.

A device called a barometer is used to measure atmospheric pressure. Old barometers were made of tubes filled with mercury. As atmospheric pressure increased, more force was exerted on the mercury, which would rise up an evacuated tube and indicate higher pressure. A drawing of a mercury barometer is shown in Figure 14.2.If you've ever heard the units "millimeters of mercury" or "inches of mercury," this is where those units of atmospheric pressure come from. Standard sea level pressure is 29.92 inches of mercury. (These units are really a misnomer. Millimeters and inches are units of height and not units of pressure. These units are a shorthand way of describing the pressure caused by a column of mercury that had the specified height.)

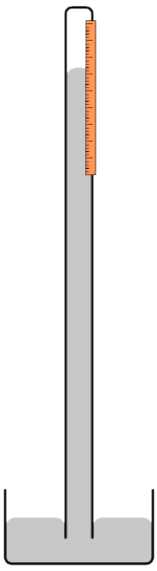


Figure 14.2 – A mercury barometer is a sealed column filled with mercury that empties into a reservoir. When atmospheric pressure presses down on the reservoir of fluid, the height increases (hence the units "inches of mercury"). Mercury barometer, by Danomagnum, is licensed under Public Domain.

Aneroid barometers use compressible cells connected to a pointer to indicate pressure. As pressure increases, the cells are squished and the needle moves to higher values. As air pressure decreases, the cells expand and the needle indicates lower pressure values. An aneroid barometer is shown in Figure 14.3.

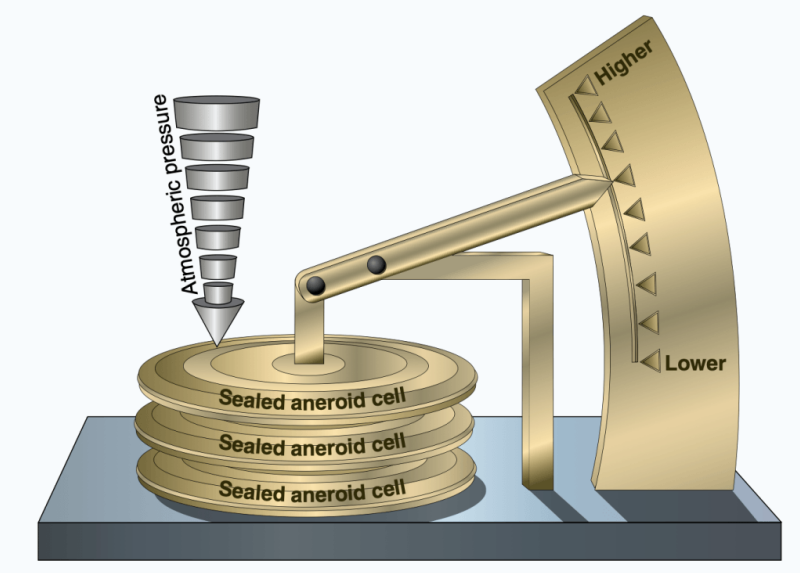


Figure 14.3 – An aneroid barometer is a series of stacked wafers. When they are compressed by atmospheric pressure, a needle rises, indicating a higher value. Aneroid barometer, by Federal Aviation Administration, is licensed under Public Domain.

Today, many devices such as smartphones and watches use tiny barometers made of micrometer-sized electronics. Barometers are extremely useful devices in forecasting weather, and are an essential instrument in airplane control panels. An airplane instrument panel is shown in Figure 14.4. Three key instruments use atmospheric pressure to tell the pilot about the status of the airplane: the altimeter, the airspeed indicator, and the vertical speed indicator. While the instruments shown in Figure 14.4 are analog, even modern aircraft use digital displays of instruments that use the same physical concepts (essentially measuring air pressure).



Figure 14.4 – At least three important airplane instruments use atmospheric pressure. Even commercial airplanes use digital variations of these same instruments. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Our atmosphere is capable of amazing things. Because we are constantly experiencing our atmosphere, we may not realize just how powerful atmospheric pressure is. Suction cups are capable of holding things up. Where does the counterforce to gravity come from? When air molecules are forced out from under the cup, there are many molecules pushing on the suction cup from the side, but very few molecules inside to push back. A really good suction cup can hold up quite a bit of weight. In Figure 14.5, a suction cup placed on a whiteboard holds up a 1 kg mass.



Figure 14.5 – A suction cup placed onto a whiteboard holds up a 1 kg mass. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale places larger and larger masses on the suction cup that's placed on the whiteboard. At the conclusion of the video, she removes the suction cup from the board. When pulling directly outward on the suction cup, it's nearly impossible to remove the suction cup, due to the large force created by air pressure pushing on the cup. In order to remove it, it is necessary to break the seal. Either pulling sideways on the suction cup, or pulling up on the edge, will cause air molecules to rush into the space between the suction cup and the whiteboard and break the seal.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In the video below, Dr. Pasquale takes a metal can and removes the lid. She pours some water inside and places it onto a hot plate. The water heats up and begins to boil. Due to buoyancy, the hot water vapor rises and displaces the air, forcing most of the air molecules out of the can and into the surrounding room. At this point, Dr. Pasquale screws the cap back on tightly, and removes the can from the hot plate. Over time, the water vapor molecules inside the can cool down and convert back to their liquid form. When all of the water has condensed, there are very few molecules left inside the can, so very little pressure is being exerted outward on the can, but a lot of pressure is being exerted inward on the can. This causes it to collapse very dramatically!

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The sequence of events that occurs in the can crush video is depicted in Figure 14.6.

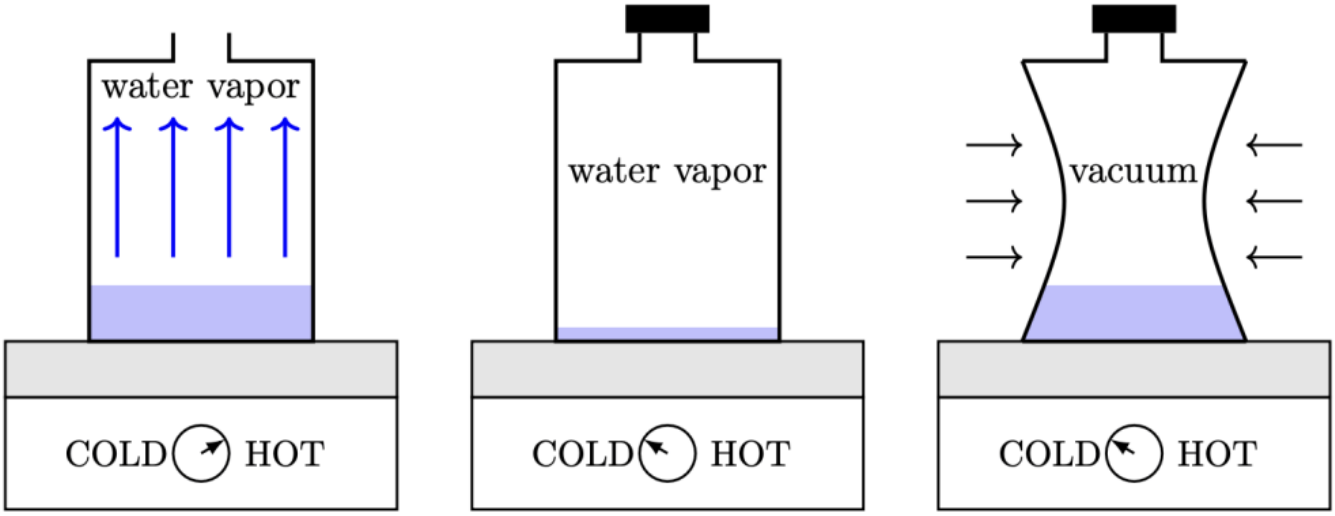


Figure 14.6 – The can crush works by displacing the air with water vapor, created by boiling water (left). When the air has been displaced, the cap is placed on the container and the heat is turned off (middle). After the water vapor has condensed back to liquid form, there is a vacuum left behind, causing the air pressure to collapse the can (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Boyle's law

Boyle's law describes the relationship between pressure and volume of a fixed amount of gas when the temperature and number of molecules in a closed system are held constant. It states that the pressure times the volume is constant. In a closed system, if the volume decreases, it will cause the pressure to increase. If the volume increases, it will cause the pressure to decrease. We can state that in equation form as saying

$$P_1V_1 = P_2V_2,$$

where P_1 and V_1 are the pressure and volume at time one, and P_2 and V_2 are the pressure and volume at time two. In other words, the product of pressure and volume before the change is equal to the product of pressure and volume after the change as long as the temperature and amount of gas remain constant.

In the video below, Dr. Pasquale uses a cylinder and piston apparatus. The cylinder is closed, keeping the number of molecules of gas inside constant. A temperature gauge indicates a mostly constant temperature throughout the experiment. The volume of gas inside the cylinder is controlled by moving the piston up and down. A pressure gauge shows the pressure inside the cylinder, subtracting out the background atmospheric pressure. This is known as gauge pressure. Whereas absolute pressure is measured relative to vacuum, gauge pressure is measured relative to ambient atmospheric pressure. This can be related with the equation

$$P_{gauge} = P_{absolute} - P_{atmospheric}.$$

As Dr. Pasquale decreases the volume of the cylinder, the pressure increases. It actually becomes hard to decrease the volume of the cylinder once it starts getting very small, due to the large amount of pressure that has built up in the cylinder.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A graph of the pressure and volume relationship is shown in Figure 14.7. (Download this data [XLSX, 9 kB]) The pressure (converted to units of atm) decreases as the volume increases, showing a decreasing trend on the graph. The product of pressure and volume, however, remains constant with volume, as predicted by Boyle's law.

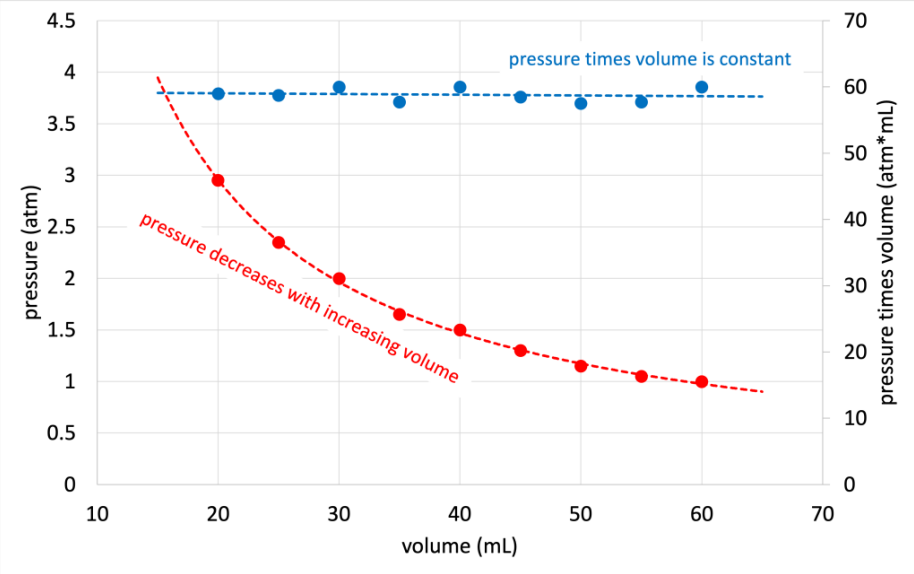


Figure 14.7 – A graph of pressure versus volume shows a decreasing trend as volume increases. The product of pressure and volume remains constant as volume increases. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video, once Dr. Pasquale compresses the cylinder all the way down, she releases the air. This causes the pressure gauge to indicate zero again. (The background atmospheric pressure has been subtracted out, as gauge pressure, not absolute pressure, is being measured. This is similar to a car tire pressure gauge: even if it indicates a pressure of zero for a flat tire, there is still one atmosphere of background atmospheric pressure inside the tire.) Now as she increases the volume of the cylinder, the pressure decreases, indicating negative on the pressure gauge. This relationship continues until she increases the volume back up to maximum again.

A more general version of Boyle's law is the ideal gas law, which also considers temperature and the number of molecules involved. We will not go into details of this law other than to say that the ideal gas law tells us that the product of pressure times volume is proportional to temperature. In other words, as the temperature decreases, the product of pressure times volume decreases as well. If you have a car with an automatic tire pressure sensor, you may notice that it often triggers a low pressure alert in the winter due to the cold weather causing tire pressure to decrease.

Archimedes' principle

As was discussed in the chapter on liquids, Archimedes' principle discusses the buoyant force that exists when a fluid is displaced by another object. The buoyant force is equal to the weight of the displaced fluid. In liquids, this makes you feel lighter when you go swimming, for example. Both Archimedes' principle and the principle of flotation exist in gases, too.

Any gas that is more dense than air will sink; any gas that is less dense than air will rise. Because the density of air decreases as the height above the surface of the Earth increases, any object that experiences a buoyant force will rise up until its density matches the surrounding density of the atmosphere.

In the video below, Dr. Pasquale has lit three candles in a large beaker. The beaker is initially just filled with air. She places a piece of solid carbon dioxide inside the beaker. The solid carbon dioxide sublimes, or turns directly into a gas (which is a process that will be discussed in chapter 17 of this textbook). The gaseous form of carbon dioxide is more dense than air. As the carbon dioxide builds up, it sinks and displaces the air in the beaker. Eventually the level of carbon dioxide rises up in the beaker and causes the candles to snuff out. Combustion requires oxygen to burn, and when the carbon dioxide displaces it, there is no more oxygen for the candles to remain lit.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Hot air balloons (shown in Figure 14.8) use buoyancy to fly. Hot air is less dense than cold air. Therefore, by heating up an envelope with hot air, we can achieve flight. Airships such as blimps achieve flight by using a gas that's less dense than air, usually helium, inside their envelope. (Historically, a different gas that is less dense than air, hydrogen, was used in blimps. However, hydrogen is extremely flammable, and this led to the Hindenburg disaster. For this reason, hydrogen is no longer used as a buoyant gas in airships.)



Figure 14.8 – A hot air balloon obtains lift because hot air is less dense than cold air. New 2017 Cameron hot air balloon in flight, by EliavCohen303, is licensed under CC BY-SA 4.0.

If all it takes to rise up in our atmosphere is a less dense gas than air, why not use balloons to go to outer space? Consider a balloon filled up with helium gas. At sea level, helium is much less dense than the atmosphere. If you've ever let go of a helium-filled balloon, you've probably watched as it flew up into the sky. As the balloon rises, two things happen. The surrounding atmospheric pressure decreases with altitude. So the balloon will expand, and the difference in density between the balloon and its surroundings will decrease.

In the video below, Dr. Pasquale has a jar with an uninflated balloon. In the beginning, the surrounding pressure is relatively high and the volume of the balloon is low. It is safe to assume that the balloon is a closed system and that no molecules can either enter or exit through the balloon in any way. As Dr. Pasquale pumps air out of the jar, the volume of the balloon increases. The pressure in the balloon causes it to inflate because there are fewer molecules pushing in on it from outside.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

If we picture this balloon on a journey upward through our atmosphere, one of two things will limit the balloon's flight. Either the balloon will rise up to the point where its density is equal to that of the surrounding atmosphere. Or the balloon will expand to the point where the envelope of the balloon bursts. Balloons are certainly used for a lot of high-altitude research and weather measurements, but they are not a feasible method to achieve space flight. For that, we require rocketry.

Bernoulli's principle

We just learned about two methods to fly around in our atmosphere: hot air balloons and blimps. What about airplanes and gliders? How do they work?

Bernoulli's principle states that as the speed of a fluid increases, the pressure decreases. Another related principle is the Venturi effect, which states that as the area of fluid flow decreases, speed must increase. Both of these properties relate to airplane flight.

The venturi effect comes about due to conservation of mass. If there is a fluid flowing through some pipes, if a section of the pipe has a smaller cross-sectional area, then in order to get the same amount of mass flowing through, the fluid has to speed up in the smaller area. This is shown in an animation in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A carburetor (common in automobiles before fuel injection) uses the venturi effect to mix together gasoline and air to fuel a gas-powered internal combustion engine. While they are not common in automobiles anymore, they are still frequently encountered in the engines that power lawn mowers, chain saws, and small airplanes.

Bernoulli's principle comes about due to conservation of energy. That fluid that speeds up now has more kinetic energy than the surrounding fluid. That additional energy has to come from somewhere since energy cannot be created or destroyed. As a result, the pressure of the fluid decreases in order to compensate. Whenever there is a difference in pressure on one side of an object compared to the other, there will be a net force.

In the video below, Dr. Fazzini demonstrates Bernoulli's principle with two cans that are allowed to freely slide side-to-side by being placed on drinking straws. When he blows air in between the two cans, the increase in speed causes a decrease in pressure between the two cans, leading the two cans to bang into each other due to the greater pressure of the stagnant air on the outer sides of the two cans.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

That video demonstrates a phenomenon you may have noticed if passing a large truck at high speed on the highway. As the airflow between two vehicles (particularly a large one such as a truck) becomes faster than the surrounding airflow, Bernoulli's principle causes them to experience a force that tends to bring them together.

In this next demonstration of Bernoulli's principle, Dr. Fazzini is able to deflect a candle flame (and blow the candle out) by deflecting airflow in different ways.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

At first, Dr. Fazzini has a candle placed on the opposite side of a flat barrier from himself. When he blows air onto the barrier, it creates fast moving air that moves along the sides of the barrier, creating low pressure on the sides of the barrier. Air from immediately adjacent the opposite side of the barrier moves outward to fill into this low pressure area, and the result is that the candle flame deflects toward the barrier. This is depicted schematically in Figure 14.9.

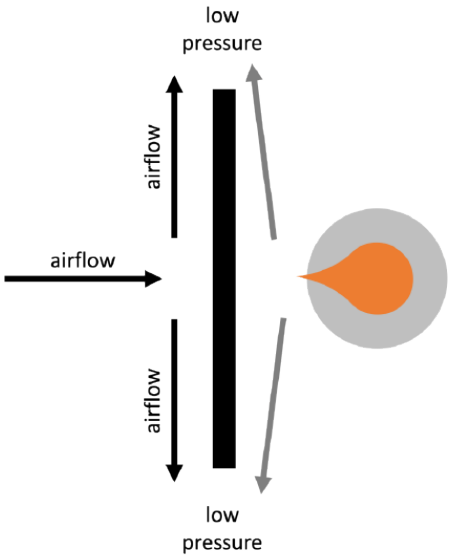


Figure 14.9 – A candle flame deflects toward a rectangular barrier due to the existence of low pressure due to fast airflow around the barrier. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Next, Dr. Fazzini uses a circular barrier between himself and the candle. Now when he blows gently on the candle, the airflow streamlines along the circular barrier and causes the flame to deflect away from the barrier. When he blows hard enough, the candle is snuffed out.

As an additional demonstration of Bernoulli's principle, in the video below, Dr. Fazzini attempts to fill two identical long plastic bags with air. Dr. Fazzini blows into the first bag five times. The volume of air is limited by his lung capacity. After five breaths, the air in the bag constitutes less than half of the bag's total volume.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Next, Dr. Fazzini blows into the second bag. This time, he blows from a position slightly removed from the entrance to the bag. This causes an increase in airspeed, and a decrease in air pressure, causing air from around the room to fill in to that space and enter into the bag. In this case, much more of the bag is filled with air, even though only "one breath" was used to initiate this process.

In the video below, Dr. Fazzini again demonstrates Bernoulli's principle with a ping-pong ball and a funnel. When he blows into the funnel, low pressure is created between the funnel and the ball, causing the ball to get "sucked in" to the funnel. He can hold the funnel upside down, and as long as he continues to blow into the funnel, the ping-pong ball will stay put in the funnel. However, as soon as he blows air above the ball, the ping-pong ball leaves the funnel. This is because the air blown above the ball causes low pressure above the ball and high pressure underneath. The net force is upward.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This next demonstration is similar to the ping-pong ball in the funnel demonstration, and gives further evidence to Bernoulli's principle's importance in flight. In the video below, Dr. Pasquale has a hair dryer and a foam ball. At first, she blows the air directly upward on the foam ball, pushing it up. As she tilts the hair dryer, she redirects the airflow, causing the air to flow faster above and slower below. Dr. Pasquale uses a string to show the airflow, which is moving very fast above the ball, and pretty much not at all below. This creates a pressure differential. The air pressure is lower on the top side of the ball, and higher on the bottom side of the ball. This causes a force of lift that counteracts gravity to keep the ball flying in the air. This is exactly the same physics that allows airplanes to fly!

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In an airplane, the wings are angled so that the air moves faster over the top of the wing compared to the bottom of the wing. Bernoulli's principle states that the fast air on the top of the wing has less pressure than the slow air on the bottom of the wing. This pressure difference leads to a force, which we call lift. As long as lift is initially greater than the weight of the airplane, the

airplane will accelerate up into the air and fly! When the airplane is level in the air, the lift and weight vectors are equal and the airplane will maintain steady flight.

It should be pointed out that the physics that keeps airplanes in the air is more complicated than what is presented in this discussion. For example, aircraft can create lift while upside down. This is because other factors (not just the venturi and Bernoulli effects) are involved in the creation of lift. These factors include Newton's third law, angle of attack (the angle that the chord of an airplane wing makes with respect to the relative wind), and camber (the shape of a wing). When the angle of attack is greater than zero (but less than a critical angle of attack at which point the airflow becomes detached from the wing), the airflow on the wing is deflected slightly downward. By Newton's third law, the reaction force is that the wind pushes up on the wing. Both Bernoulli and Newton contribute to the overall explanation of how airplanes generate lift.

Plasma

The fourth of the four fundamental states of matter is plasma. Plasma is ionized gas. As previously discussed in this textbook, ions are atoms with an unequal number of protons and electrons. They contain a charge. A plasma is a gas where so much energy is present that electrons are removed from the atom. The resultant atoms of the plasma become positive ions floating in a sea of electrons. This separation of electric charge makes plasma capable of conducting electricity.

Plasma is the most abundant state of matter in the universe, as stars are made of plasma. Lightning, gas discharge lamps (such as neon lights), and the aurora lights are all examples of plasma that we may experience here on Earth.

Plasma is particularly valuable as a method of creating light (which will be discussed in much greater detail later in this textbook). For now, it's enough to know that plasma is how many different types of lighting work: fluorescent light, neon light, and sodium street lights. A plasma lamp, as shown in Figure 14.10, generates very colorful plasma emissions as a result of electric discharge.

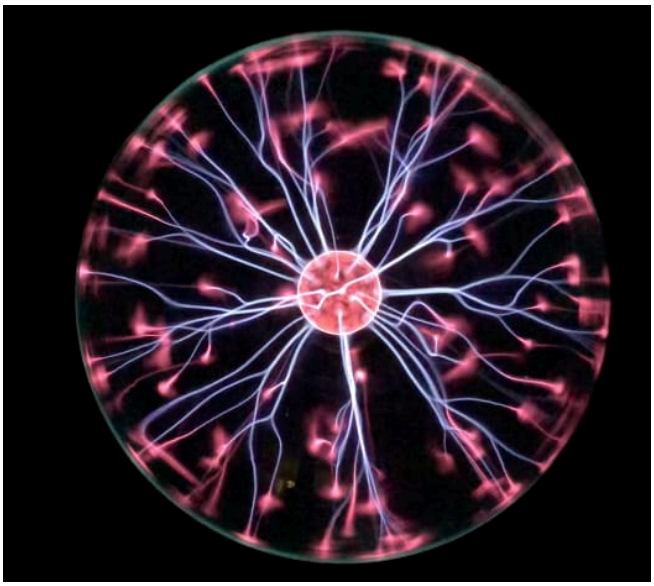


Figure 14.10 – A plasma lamp generates colorful plasma from electrical discharge. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Further reading

- Why don't I feel the miles of air above me that are crushing me down? – This article explains why the force of atmospheric pressure (100,000 N per square meter!) doesn't crush us.

Practice questions

Numerical analysis

1. A gas in a closed container has an initial volume of 4 L and a pressure of 3 atm. If the pressure is reduced to 1.5 atm while keeping the temperature constant, calculate the new volume of the gas.
2. A sample of gas occupies a volume of 10 L at a pressure of 200 kPa. If the pressure is increased to 400 kPa while keeping the number of gas molecules (and temperature) constant, calculate the new volume of the gas.
3. If a gas occupies a volume of 8 L at a pressure of 4 atm, calculate the pressure of the gas if the volume is reduced to 2 L while keeping the temperature and number of molecules constant.
4. A weather balloon is released into the atmosphere with a volume of 10 m^3 at a pressure of 100 kPa. As the balloon ascends, the pressure decreases to 30 kPa. Calculate the final volume of the balloon when it reaches this lower pressure region (you can assume the temperature remains constant, as does the number of gas molecules inside the balloon).
5. A helium-filled balloon has a volume of 5 m^3. If the density of helium is 0.18 kg/m^3 and the density of air is approximately 1.2 kg/m^3, calculate the buoyant force on the balloon when released into the atmosphere.

Hands-on experiments

1. If you have a drinking straw, place it into a cup or glass that's filled with water. Cover the top of the straw with your finger, then lift it out of the glass of water. What happens to the water that's in the straw? How does this relate to the topics discussed in this chapter?

This page titled Chapter 14: Gases is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

IV: THERMODYNAMICS

Chapter 15: Temperature and Heat
Chapter 16: Heat transfer
Chapter 17: Phase change
Chapter 18: Thermodynamics

This page titled IV: Thermodynamics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 15: Temperature and Heat

Summary

Temperature and heat are important phenomena that start to unlock the ideas of thermodynamics. Temperature is something that impacts our lives greatly as we move throughout the seasons. Heat and specific heat capacity are concepts that allow us to understand climate and cooling systems, among other things. Finally, thermal expansion has far-ranging implications from the existence of aquatic life on this planet, to understanding how objects change their sizes from month to month as temperatures change.

Temperature

Temperature quantitatively describes the average translational kinetic energy of the atoms and molecules in a substance. It quantifies the “hotness” or “coldness” of something. The symbol for temperature is the uppercase letter T.

Kinetic energy describes the energy of motion. Therefore, it would be reasonable to expect that a substance at a higher temperature will have more motion than a substance at a lower temperature.

In the video below, there are two beakers filled with water. The beaker on the left is filled with cold water, and the beaker on the right is filled with hot water. (The beaker on the left appears translucent due to the fact that water vapor from the air has condensed on it.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A drop of blue food dye is dropped into each beaker. After watching for several moments, you can see the food dye swirl around and distribute itself much more rapidly in the beaker of hot water, demonstrating that its motion is much greater than that of the cold water. A photograph of each beaker 30 seconds after the introduction of the food dye is shown in Figure 15.1. The dye in the cold beaker (left) has not distributed nearly as much as the dye in the hot beaker (right).

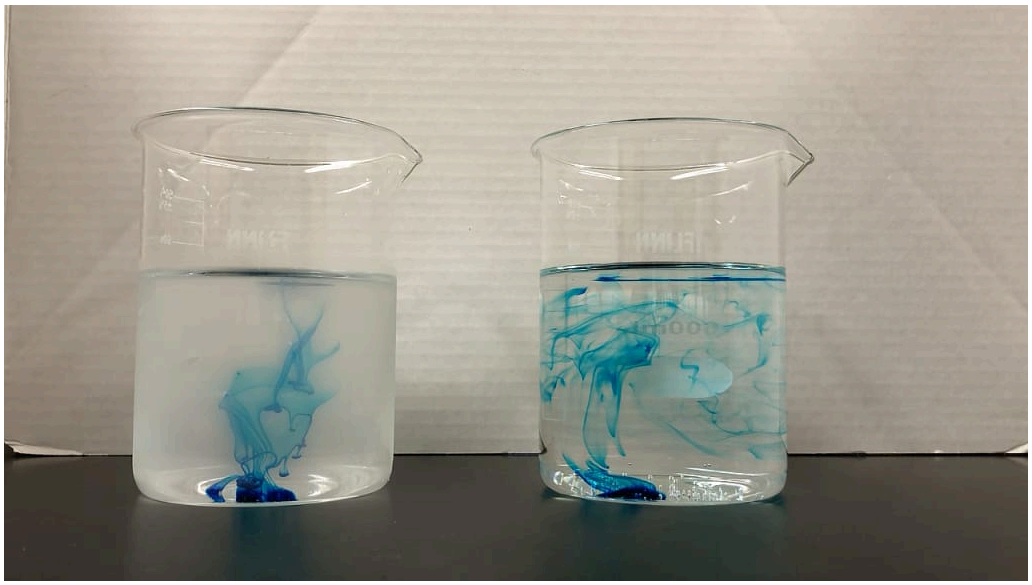


Figure 15.1 – A beaker of cold water (left) and hot water (right) are viewed 30 seconds after placing a drop of blue food coloring in each beaker. The dye in the hot water has distributed much more than the dye in the cold water. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

From a human perspective, temperature is extremely important in helping us determine what to wear when we go outside, and in deciding on what types of indoor and outdoor activities we want to do. Having an objective temperature scale is important in that it cuts out the subjective interpretation of each person's idea of what may be "hot" or "cold." From a physics perspective, an objective temperature scale enables scientists to perform experiments on thermodynamics, climate change, and other important phenomena.

Empirical temperature scales

Now that we understand what temperature is, we need a way to measure it. There are two types of temperature scales that can be used: empirical scales and absolute scales.

An empirical scale means that experiments were performed to derive the upper and lower limits of the scale. The Fahrenheit scale was proposed in the 1700s. The limit for the cold end of the temperature scale was a frozen salt water solution. That temperature was denoted as zero. The limit for the hot end of the temperature scale was the temperature of a human body. This temperature was denoted at 96. (This value was chosen because it is easily divisible by many numbers; it has 12 divisors.) The total number of degrees on this scale was therefore 96, creating the original Fahrenheit temperature scale. This rudimentary scale is shown in Figure 15.2.

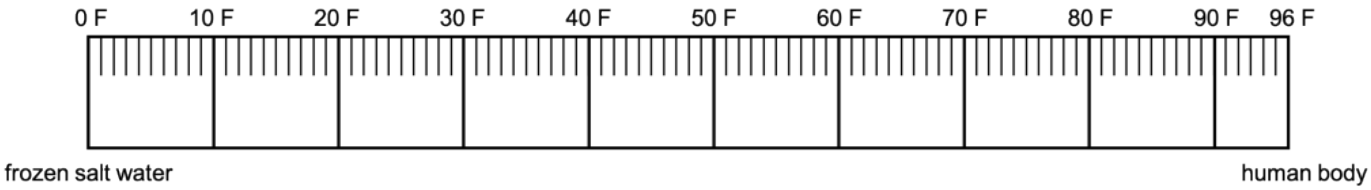


Figure 15.2 – The original Fahrenheit scale was developed by setting 0 as the temperature of a frozen salt water solution and setting 96 as the temperature of the human body. This created a total number of 96 degrees between hot and cold. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The basis of the Fahrenheit scale isn't terribly scientific because it's not repeatable. Depending on how much salt is placed in the freezing salt water solution, and whether or not the person whose body temperature is measured is sick with a fever, there will be different definitions of 0 and 96 Fahrenheit. Even if we standardize how much salt to place in the water, we are still limited by the variations in human body temperature.

Since the development of the Fahrenheit scale, it has been standardized such that temperature measurements using this scale are now repeatable and scientific. This allows meteorologists (and others) in the United States (one of the only countries in the

world that continues to use this scale) to record temperatures in degrees Fahrenheit that are comparable from day to day and over time. While the Fahrenheit scale is admittedly useful as a temperature scale, as we tend to experience the full range of temperatures from zero to one hundred (particularly in varied climates), it was not created with a scientifically sound basis.

The Celsius scale was also developed in the 1700s. The cold end of the temperature scale is established by the freezing point of water at 1 atm of pressure, and is called zero Celsius. (A photograph of a Celsius thermometer reading zero degrees in a bath of ice water is shown in Figure 15.3.) The hot end of the temperature scale is established by the boiling point of water at 1 atm of pressure, and is called one hundred Celsius. This means that there are 100 degrees between the cold and hot ends of the Celsius scale. (Note that the original Celsius scale was actually inverted such that 0 degrees indicated the boiling point of water, and 100 degrees indicated the freezing point of water.)

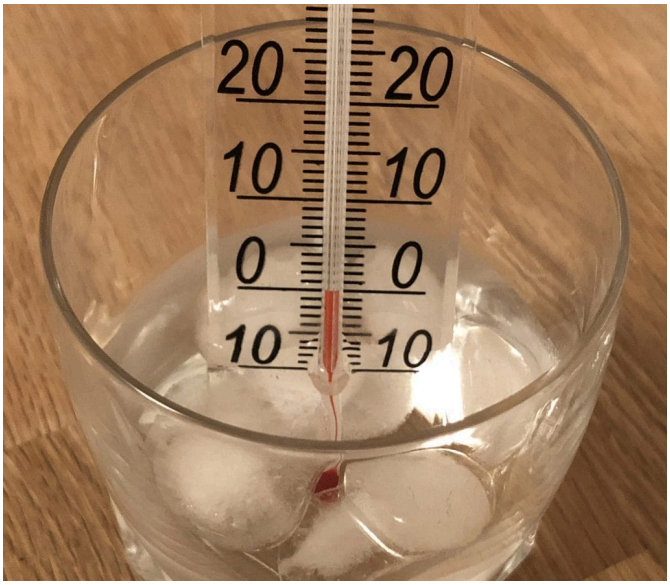


Figure 15.3 – A Celsius thermometer reads zero degrees as it is submerged in a container of ice water. Melting ice thermometer, by Ulflund, is licensed under CC0.

The selection of the cold and hot ends of the thermometer establishes the temperature scale and allows us to measure temperatures in between those two extremes. (We can also extrapolate to measure temperatures that are hotter or colder.) This is shown in Figure 15.4.

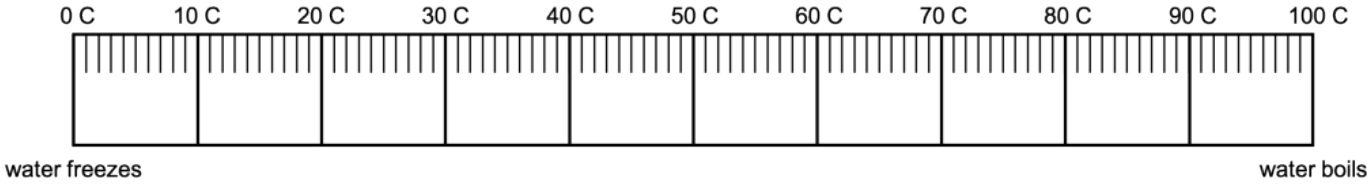


Figure 15.4 – The Celsius scale was developed by setting 0 as the freezing point of water at 1 atm and setting 100 as the boiling point of water at 1 atm. This created a total number of 100 degrees between hot and cold. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Celsius is a scientific scale because it's repeatable. As long as two (or more) scientists use pure water, and freeze and melt that water at one atmosphere of pressure, they will develop identical temperature scales. Because of this, Celsius is the scale that we will use, almost exclusively, to discuss temperature in this textbook.

Although we will not be using the Fahrenheit scale in this textbook, it is still useful to understand how to convert between the Fahrenheit and Celsius scales. To convert between scales, we need to know where the two scales overlap with each other, and how much of a change in one scale relates to a change in the other scale. On the first point: 0 ^{o}C is equal to 32 ^{o}F. On the second point: a change in 1 C^{o} is equal to a change in 1.8 F^{o}. (Note the use of the symbol C^{o} to denote a Celsius degree, which is used to distinguish a change in temperature from a temperature value.) Therefore, to convert from degrees Celsius to degrees Fahrenheit, take the temperature in Celsius, multiply by 1.8, and add 32:

$$T_F = 1.8T_C + 32.$$

To convert from degrees Fahrenheit to degrees Celsius, take the temperature in Fahrenheit and subtract 32 from it. Multiply the result of that subtraction by five-ninths. The result is the temperature in degrees Celsius:

$$T_C = (T_F - 32) * \frac{5}{9}.$$

Absolute temperature scales

The other type of temperature scale that can be used is called an absolute scale. Absolute scales are designed by establishing the lower limit based on a physical limitation. As we will discuss later in this textbook, there is a lowest possible temperature that can be achieved. That limit is known as absolute zero. As far as we know, nothing can be colder than absolute zero. This establishes the lower limit of our absolute scale.

The absolute scale that's used in physics is known as the Kelvin scale. The zero end of the Kelvin scale is absolute zero. Because there's no upper limit to the temperature of an object, there is no upper limit to the scale. (In fact, the upper limit is bounded by the total amount of energy that exists in the universe.) How do we establish the gradations of each unit of kelvin? The Kelvin scale was developed so that a change in one kelvin is equal to a change in one Celsius degree.

Absolute zero is equal to -273 ^{o}C. Therefore, to convert from degrees Celsius to kelvin, simply add 273:

$$T_K = T_C + 273.$$

To convert from kelvin to degrees Celsius, subtract 273:

$$T_C = T_K - 273.$$

Just a note to check your work: there is no such thing as a negative temperature in the Kelvin scale, so if you get a negative answer when converting to kelvin, check your math!

Thermometers

To measure temperature, a thermometer is used. How do thermometers work, and how are they designed?

In general, a thermometer exploits a thermometric property of a substance. A thermometric property is simply a property of the substance that varies with temperature. To achieve the best sensitivity for a thermometer, you want to choose a property that has a very large variation for small changes in temperature. Most analog thermometers (see, for example, Figure 15.5) contain a substance that will remain a liquid at all temperatures that the thermometer can reasonably be expected to experience. For example, in geographic regions with hot summers and cold winters, a thermometer would need to remain liquid between temperatures of roughly -20 ^{o}C and +40 ^{o}C. Modern thermometers are generally filled with dyed ethyl alcohol, which remains liquid between temperatures of -114 ^{o}C and +78 ^{o}C. (Previously, thermometers were filled with mercury, which remains liquid to much higher temperatures, but which is a hazardous toxic element.)

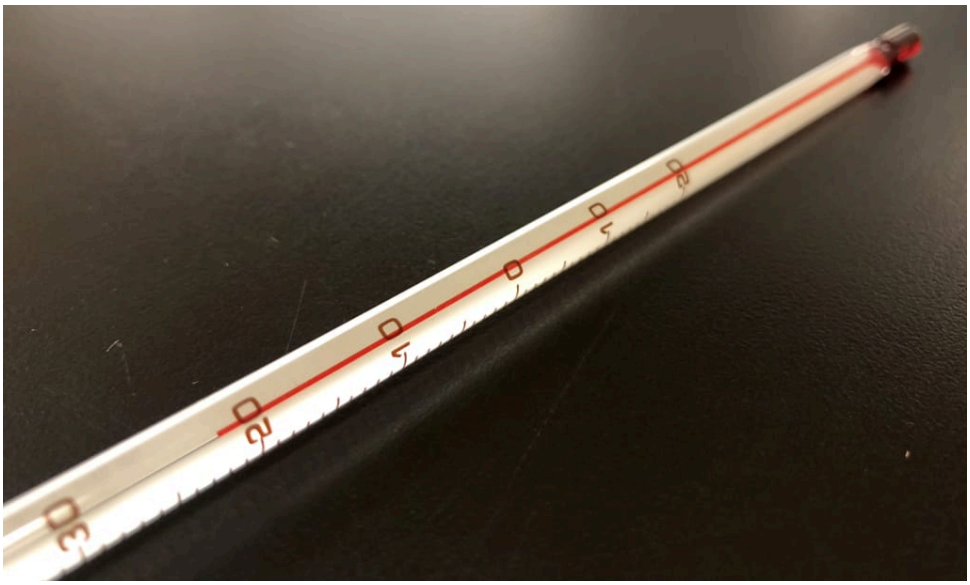


Figure 15.5 – An analog thermometer uses a liquid that will remain in liquid form throughout a wide temperature range. This thermometer uses red dyed alcohol. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Materials such as liquids will generally expand as they heat and contract as they cool. That is, the thermometric property used is the volume of the substance. This means that as the thermometer is placed in cooler temperatures, the level of liquid will fall. As the thermometer is placed in hotter temperatures, the level of liquid will rise. To calibrate the thermometer, it can be placed in freezing water to establish the 0 ^{o}C point, placed in boiling water to determine the 100 ^{o}C point, and then 99 gradations placed in between. To determine the temperature at any given time, the level of liquid inside the thermometer can be read, as after a time it will have risen or fallen to the level corresponding to the current temperature.

A digital thermometer (shown in Figure 15.6, for example) has a device inside whose electrical properties are changed as a function of temperature. The thermometric property in this case of the resistivity of a semiconductor. The relationship between temperature and resistivity is well known. Therefore, by measuring the resistivity, the temperature can be determined. A microcontroller then converts that temperature into numbers and displays them on an LCD screen.



Figure 15.6 – A digital thermometer uses the relationship between temperature and electrical properties to determine the temperature. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Heat

Heat is energy that is transferred from one object to another in response to a difference in temperature. The symbol for heat is the uppercase letter Q. Calories (cal) are the unit for heat. Because heat is a form of energy, joules is another unit that can be used to express heat.

Heat only spontaneously flows from hot to cold, and never spontaneously flows in the opposite direction. To relate this to something more familiar, consider it to be similar to gravity. If a ball is placed at the top of a ramp, gravity will pull the ball downward, causing it to roll down the ramp. Once the ball reaches the bottom of the ramp, gravity will not pull it back up to the top again. The only way for the ball to reach the top of the ramp is for a person, animal, or machine to expend energy in physically moving it back to the top again. Going against gravity requires work.

In much the same way, working against the established flow of heat takes work. Maybe you use a refrigerator or air conditioner. These seem to work against the flow of heat from hot to cold. And they do, but it requires energy to do this. If you've paid an electrical bill during a hot summer month when running the air conditioner, you know it's not cheap to keep things cool. In other words, your refrigerator and air conditioner will not cool the regions you want unless you plug them in and turn them on!

The mechanisms through which heat can flow from one place to another will be the subject of the next chapter of this textbook.

Specific heat capacity

Specific heat capacity defines how difficult it is to change the temperature of a substance. (It can be thought of as "thermal inertia.") The symbol for specific heat capacity is the lowercase letter c. The unit is calories per gram degree Celsius (cal/(g^{o}C)). Specific heat capacity describes how many calories of heat are required to increase (or decrease) the temperature of a certain mass of material by a certain amount.

The equation that relates heat, specific heat capacity, mass, and change in temperature is

$$Q = mc\Delta T,$$

where Q is the heat (measured in calories), m is the mass (measured in grams), c is the specific heat capacity, and ΔT is the temperature change (measured in degrees Celsius).

It is important to note that this equation only works when the temperature of a substance changes, and cannot be used when an item changes phase, for example to calculate the heat required to boil water.

The specific heat capacity of water is 1 cal/(gC^{o}). This means that 1 calorie of heat is required to raise the temperature of a single gram of water by one Celsius degree. (The calorie unit was actually defined on this basis; one calorie was equal to the amount of heat required to increase the temperature of one gram of water by one Celsius degree.) This may not seem like a lot, but water actually has one of the highest specific heat capacities of common substances. It makes water a very good candidate for use in cooling systems, because it can absorb a lot of heat without changing its temperature very much.

Large calories and small calories

The unit of calorie may be familiar if you have looked at the nutrition label for food (such as the one shown in Figure 15.7). The unit of calorie used in food products does relate to the amount of heat required to raise the temperature of water by one Celsius degree, however, in this case, the amount of water is one kilogram, not one gram! These calories are known as "large calories," "food calories," or "kilogram calories" and are sometimes specifically spelled out as Calories with a capital C to distinguish them from the so-called small calorie described above.

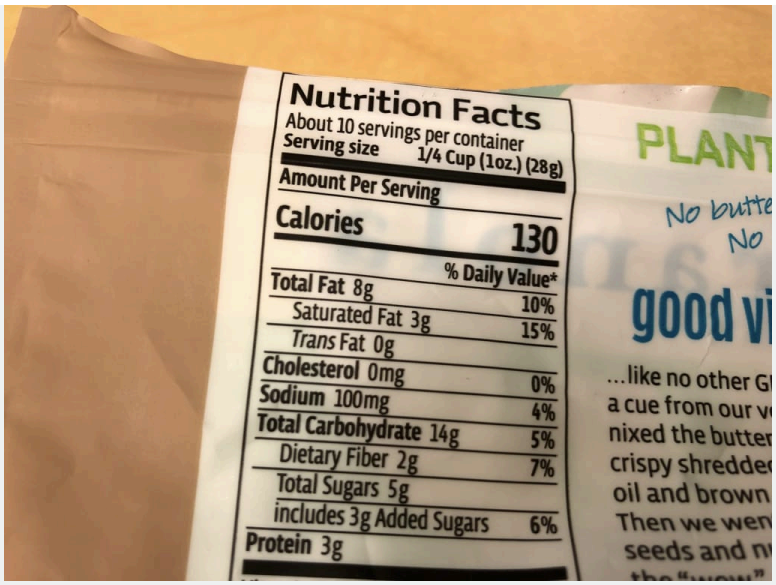


Figure 15.7 – A nutrition label indicates the calorie content of the food, which describes the amount of heat required to raise the temperature of one kilogram of water by one Celsius degree. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This fact also explains why the climate tends to be milder if you live near a large body of water. This could be an ocean, or even a large lake such as Lake Michigan. Because the water in an ocean or lake doesn't change its temperature easily, it means that the water will act as a moderating influence as the seasons change. This also explains why daily temperature fluctuations are much larger in the desert than they are in wetter climates. Humidity in the air acts as a moderating influence to keep temperatures from swinging to extremes between night and day. For example, Figure 15.8 shows the hourly temperature in Phoenix, AZ and Miami, FL on September 14, 2021. (Download this data [XLSX, 9 kB]) In Phoenix, a desert, the temperature demonstrates a much greater variation than in Miami, a much wetter climate.

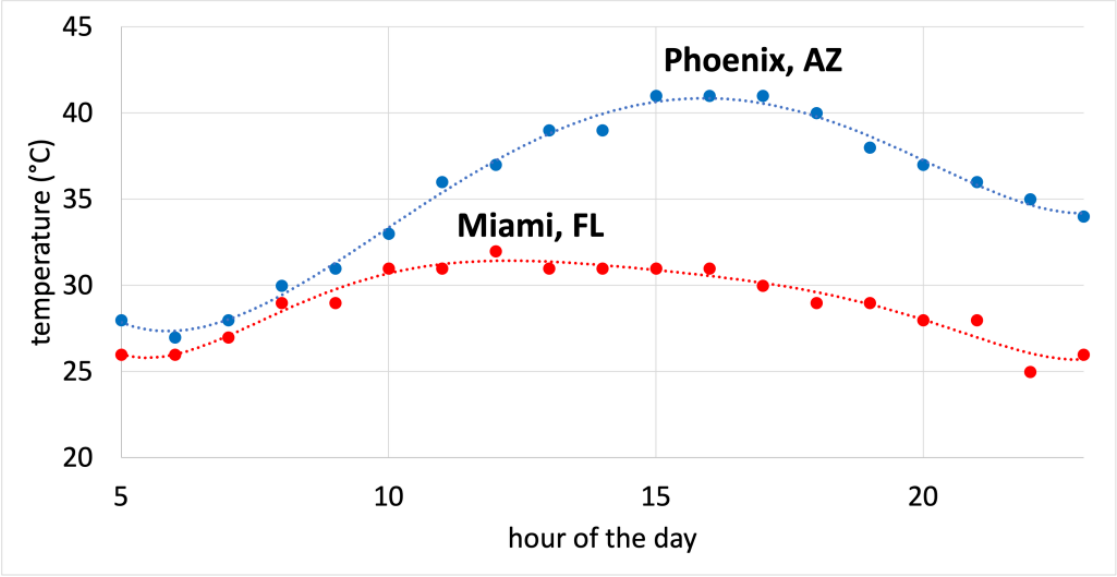


Figure 15.8 – The temperature in Phoenix, AZ (a desert) demonstrates a much wider range of temperatures than the temperature in Miami, FL (a humid climate). This data was taken on September 14, 2021. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Temperature mixtures

When two substances at different initial temperatures are combined together, the hot substance will cool down and the cool substance will heat up until the two substances arrive at thermal equilibrium. At this point the temperature will remain constant.

According to the law of conservation of energy, the energy that is lost by the hot substance will be completely transferred to the cool substance. (This assumes that no heat is lost to the surroundings.) In equation form,

$$\begin{aligned} -Q_{hot} &= Q_{cold} \\ Q_{hot} + Q_{cold} &= 0. \end{aligned}$$

Unless the two substances are composed of the same material with identical specific heat capacities and have the same mass, while both substances will experience a temperature change, the magnitude of change in temperature of each substance may not be the same. (For example, if the hot substance cools down by 10 C^o, the cool object may not warm up by that same amount.) All other things being equal, a substance with a higher specific heat capacity will experience a smaller temperature change than a substance with a lower specific heat capacity.

To determine the final temperature of a mixture of two substances, the conservation of energy equation defined above can be substituted with $Q = mc\Delta T$ to calculate T_{final}. This equation is

$$T_{final} = \frac{m_{hot}c_{hot}T_{i,hot} + m_{cold}c_{cold}T_{i,cold}}{m_{hot}c_{hot} + m_{cold}c_{cold}}.$$

The final temperature should be somewhere in between the two initial starting temperatures. (This is a good “smell test” to use to determine if your final temperature calculation makes sense.)

Thermal expansion

Substances tend to expand when they are heated. In fact, this property was utilized to create analog thermometers, as we saw earlier in this chapter! Because the molecules within the object have more motion as they heat up, it makes sense that most substances will expand when heated.

This is demonstrated in the video below. A metal ball and hoop start out at room temperature. The ball is easily able to pass through the hoop, demonstrating that the ball is smaller than the hoop’s opening. Dr. Pasquale heats up the ball with the blowtorch, which causes it to expand. It has expanded to such a degree that it is no longer able to fit through the hoop. Next, she heats up the hoop with the blowtorch. Because the hoop expands, now the ball is able to pass through again.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

After the ball and hoop cool down again, they return to their original sizes.

Thermal expansion has some very important repercussions in our physical world. Notably, structural objects like bridges and buildings have to be engineered specifically to account for this expansion and contraction in hot and cold weather. Long bridges contain expansion joints (see Figure 15.9). This enables them to stretch out in the heat without buckling or breaking.



Figure 15.9 – An expansion joint in a bridge allows the bridge to expand and contract when the temperature changes without causing damage to the bridge. Expansion joint on a bridge, by Matt H. Wade, is licensed under CC BY-SA 3.0.

Power lines also experience thermal expansion, sometimes in devastating ways. In the summer, power lines expand due to the heat. Not only that, but the load on the electrical lines also increases on hot days due to people using their air conditioning units. The excess load causes even more heat in the power lines, causing them to expand even more. If the expansion causes the power line to come into contact with tree limbs or a bush, then a short circuit can occur. In extreme cases, this may cause the power grid to go offline. In fact, this was one of the contributing causes to the Northeast Blackout of 2003, which caused Dr. Pasquale's family's home to be out of power for a week!

You may notice that tree limbs are trimmed around power lines (see Figure 15.10). One reason for this is to prevent possible short circuits from occurring when the power lines sag.

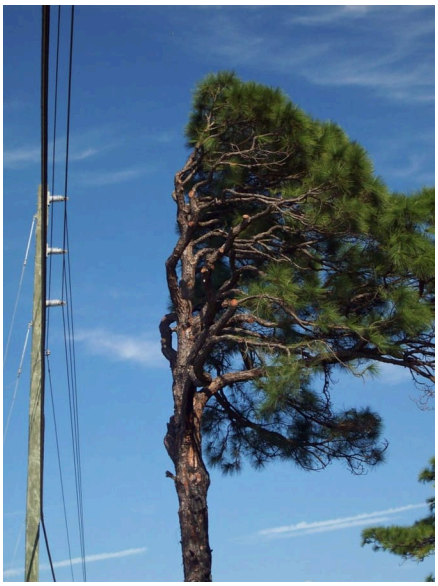


Figure 15.10 – A pine tree has its branches completely pruned from one side to avoid contact with power lines. Drooping power lines can cause electrical shorts if uninsulated wires come into contact with tree branches. Pruned pine by power lines, by dno1967, is licensed under CC BY 2.0.

It's possible to engineer the thermal expansion of a material to be very similar to that of another substance. This is helpful for fillings we get in our teeth because of cavities. When we eat hot food, we want our teeth and the fillings inside of them to expand at the same rate. Composite fillings have been engineered to do just this.

What happens if two substances come together that have different rates of thermal expansion? The video below shows a bimetallic strip. That is, it's made out of two different metals that are connected together. When Dr. Pasquale heats the bimetallic strip with a blowtorch, each metal expands at a different rate, causing it to curl up. The metal with the smaller thermal expansion expands less, which forces it to conform to the inner portion of the curve generated by the metal that expands more.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The bimetallic strip has found a practical application as an element in analog thermostats. A bimetal strip, when heated to a warm room temperature, will be coiled up. As the room cools down to a setpoint, the coil unfurls and eventually comes in contact with some type of sensor that sends a signal to the furnace to turn on and heat the room. Once the room heats up again, the strip coils up and deactivates the furnace. A thermostat using a bimetal strip is shown in Figure 15.11. A second bimetallic strip at the top of the device causes the pointer to indicate current room temperature. Similar bimetal coils are or have been used in refrigerators, computer cooling, and air conditioning systems.



Figure 15.11 – An old home thermostat contains two bimetallic strips. One (at top, connected to the pointers) that causes the temperature pointer to indicate current room temperature. One (at the bottom) that causes a tilt-switch to activate when it is too cold, causing the furnace to turn on. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Negative thermal expansion

While most substances expand when heated, not all of them do. A few substances will contract instead. This is known as negative thermal expansion. Most notably, water is densest at 4 ^{o}C. Below that temperature, water will expand. This is why ice cubes float on liquid water (see Figure 15.12). It's also why aquatic life is able to exist on our planet. Because ice floats, as long as a body of water is deep enough, the water at the bottom will be four degrees Celsius at the coldest time of year, enabling aquatic life to remain alive instead of freezing every winter. The ice also acts as an insulating layer, protecting the water below from even colder air temperatures above the surface of the water.



Figure 15.12 – Ice floats on liquid water. This is an example of negative thermal expansion. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Thermal stress

Whether objects expand or contract when heated, the molecules within the object need a sufficient amount of time to respond to these changes in size. Otherwise, it's possible that the molecules inside an object will experience enough stress to cause them to break.

In the video below, Dr. Pasquale has a glass. There's nothing wrong with it at the start, no scratches or defects at all. She heats the glass up with a blowtorch. Note that the glass stays intact, it just heats up. When Dr. Pasquale plunges the glass into a container filled with ice-cold water, the glass breaks. This is because the expansion and contraction isn't given enough time to occur without causing huge amounts of stress in the glass.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Thermal stress means that we have to be careful with the types of materials we use in and on our stoves and ovens. Specially tempered glass like Pyrex™ is used when wide swings of temperature are expected. (Note that Pyrex™ cannot necessarily be placed in an oven at any temperature!)

Further reading

- SI Units – Temperature – This website from NIST explains the Kelvin and Celsius temperature scales and provides a reference to Fahrenheit as well.

Practice questions

Numerical analysis

1. If the temperature is 20 ^{o}C, calculate the temperature in...
    1. ...Fahrenheit.
    2. ...kelvin.
2. If the temperature is 75 ^{o}F, calculate the temperature in...
    1. ...Celsius.
    2. ...kelvin.
3. If the temperature is 318 K, calculate the temperature in...
    1. ...Celsius.
    2. ...Fahrenheit.
4. Calculate the amount of heat required to change the temperature of 500 g of water from 25 ^{o}C to 45 ^{o}C.
5. Calculate the amount of heat required to change the temperature of 250 g of aluminum from 20 ^{o}C to 80 ^{o}C. The specific heat capacity of aluminum is 0.215 cal/g^{o}C.
6. Calculate the amount of heat required to change the temperature of 2 kg of lead from 100 ^{o}C to 200 ^{o}C. The specific heat capacity of lead is 0.031 cal/g^{o}C.
7. Calculate the amount of heat required to change the temperature of 800 g of water inside a 500 g copper kettle from 20 ^{o}C to 100 ^{o}C. (No phase change occurs in this process.) The specific heat capacity of copper is 0.092 cal/g^{o}C.
8. Calculate the final temperature of a mixture of equal amounts of water at 20 ^{o}C and water at 80 ^{o}C.
9. Calculate the final temperature of a mixture of 10 g of water at 30 ^{o}C and 40 g of water at 90 ^{o}C.
10. If 200 grams of water at 20 ^{o}C are mixed with 300 grams of ethanol at 30 ^{o}C, calculate the final temperature of the mixture. The specific heat capacity of ethanol is 0.584 cal/g^{o}C.
11. You mix 400 g of iron at 80 ^{o}C with 600 g of water at 25 ^{o}C. Calculate the final temperature of the mixture. The specific heat capacity of iron is 0.11 cal/g^{o}C.

This page titled Chapter 15: Temperature and Heat is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 16: Heat transfer

Summary

Heat transfer describes how heat moves from one place to another. It will help you understand everything from how a stove cooks food to how our climate works. Understanding how heat moves around also enables us to create technology that can keep things hot or cold.

Conduction

Conduction is heat transfer that occurs due to physical contact between objects. Conduction also explains how heat is able to transfer within an object.

As discussed in the previous chapter, temperature relates to the average energy of molecules. If molecules are able to easily bump into each other, causing collisions that transfer energy from one place to the next, then those types of substances will be good conductors of heat. Good heat conductors tend to be solids, where molecules are in close physical proximity. In liquids (where the molecules are not rigidly bonded to each other) and gases (where molecules are spread farther apart), it's harder for heat to conduct due to these atomic collisions.

In fact, the solid objects that make the best heat conductors tend to be those that also make the best electrical conductors: metals. In metals, electrons are capable of moving around readily, which provides for a very efficient transfer of thermal energy from one part of a metal to another.

In the video below, there are two pretty much identical ice cubes that have been placed onto different black plates. One of the plates is metal and the other one is foam. Metal is a good heat conductor. Foam is not a good heat conductor. Therefore, the metal is able to conduct heat from the surrounding air into the metal plate, and thus into the ice cube, very efficiently. That ice cube melts very rapidly.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

On the other hand, the foam plate is a poor heat conductor (alternatively: it is a good thermal insulator). That means that it keeps the heat in the atmosphere away from the ice cube, and that ice cube barely melts at all. Note that both ice cubes are surrounded on five sides by air. Air, being a gas, is a rather poor heat conductor. When the ice cube on the metal plate melts, it melts predominantly from the bottom, the side that is in contact with the good heat conductor.

This demonstration also explains why metal objects tend to feel cold to the touch even when they are at the same temperature as their surroundings. When you touch something metal, that metal is able to conduct the heat from your hand away very readily. Similarly, objects made of bad heat conductors, things like plastic, wood, paper, cork, and other insulators, are not going to feel cold to the touch, unless they themselves are cold.

Even among metals, thermal conductivity can vary. In the video below, there is an apparatus with four different metals that connect to a central plate. Dr. Pasquale places candles into each of the metals then lights a Bunsen burner under the central plate. As that central plate heats up, it conducts heat to each of the four metals. The metal that heats up the fastest has the highest heat

conduction, and will melt the candles fastest.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

First, you can see that the heat is going to travel outward. That is to say, the candles closest to the central plate melt first. Second, you can see that the best heat conductor melts the candles the fastest.

That best heat conductor is copper, which contains all of the green candles. The next best heat conductor is aluminum, which contains all of the yellow candles. Then comes brass, with the blue candles. Finally, the worst of the heat conductors of these four metals is iron, which has the red candles.

Remember, heat conduction requires physical contact between materials, or within a material. When you put food onto a frying pan to cook it, that food will cook due to that physical contact: heat conduction. This is why frying pans are made out of metals: they are good heat conductors! The part of the frying pan we have to touch is frequently covered with plastic or is very long and far away from the center of the frying pan, keeping our hands safe from the heat of the stove. The frying pan in Figure 16.1 has a metal base and a plastic handle.



Figure 16.1 – This frying pan has a metal base (good heat conductor) and a plastic handle (poor heat conductor). Frying pan, by JPC24M, is licensed under CC BY-SA 2.0.

One last thing about heat conduction: don't confuse heat conduction with specific heat capacity! While they seem similar, they are actually two completely different things. Heat conduction quantifies the ability of a material to allow thermal energy to move through it from one place to another. Specific heat capacity quantifies how much heat is required to change the temperature of a certain amount of material.

Convection

Convection is heat transfer that comes about due to the motion of molecules themselves. The types of materials that are good at convection are the types of materials where molecules can move around very freely: liquids and gases.

If you've ever held your hands above a toaster while heating something up, you've experienced convection. Air in close proximity to the toaster coils heats up. As that air heats up, it becomes less dense than the surrounding air, causing it to rise. When you

hold your hand above the toaster, you are experiencing that warm air rising up.

In the video below, there is a large beaker filled with water that's a little bit colder than room temperature. Dr. Pasquale places inside of that beaker a smaller flask filled with hot dyed water. The dye is just there to help visualize the motion of the molecules in that flask. What happens is that the hot water, being less dense than the surrounding cool water, is going to experience a buoyant force that causes it to rise up. You can see this from the dyed water moving upward in the beaker.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Eventually, as the hot water moves upward, it exchanges heat energy with the cool water, causing the cool water to increase its temperature and the hot water to decrease its temperature. This process repeats until the entire mixture reaches thermal equilibrium.

Convection occurs in our atmosphere all the time. On hot summer days, the ground is heated up by the Sun (in a process that will be discussed in the next section of this textbook). The air directly above the ground heats up due to conduction, and then the convection process causes that heat to rise up into the atmosphere. This causes a lot of vertical motion in the air.

This is why, in the summer, we tend to see a lot of cumulus clouds (Figure 16.2), which have a lot of vertical height to them. They occur when there is a lot of convection in the atmosphere. Atmospheric convection also causes turbulence in airplanes.



Figure 16.2 – Cumulus clouds are an indication of convection in the Earth's atmosphere. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Convection doesn't happen as much in the winter, because the ground doesn't heat up as much when it's covered in snow. Winter days tend to cause stratus clouds (Figure 16.3), which are very flat, indicating that there isn't a lot of convection, or vertical motion, in the atmosphere.



Figure 16.3 – Stratus clouds are an indication of the absence of convection in the Earth's atmosphere. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Electromagnetic radiation

Electromagnetic radiation is heat that is transferred from one place to another through light waves. Heat from the Sun travels nearly 150 million km through the vacuum of space to our planet as light waves. When those light waves reach the surface of the Earth, objects on the Earth's surface can absorb this energy and will heat up.

The amount that objects heat up has to do with how much of the light they absorb. Black objects absorb (almost) all of the visible wavelengths of light, which is why they appear to be black. This absorption makes black objects heat up much more readily than objects of a different color.

This may be something you experience if you have a dark-colored car. The Sun's light is absorbed by the car, heating it up. This is very nice in cold weather, but can make it hard to cool the car down in hot weather. If you drive a white or silver car, it's very likely you have the opposite experience. Your car remains cooler in hot weather, but does not heat up as much from the Sun's rays in the cold.

This is demonstrated in the video below. There are temperature probes placed in otherwise identical cans. The only difference between the two cans is that one is silver, and the other has been painted black. Dr. Pasquale connects the temperature probes to Logger Pro and turns on a heat lamp. The heat lamp is placed so that it was equidistant from both cans. Right away, the black can starts recording higher temperatures, indicating that it heats up more readily from electromagnetic radiation due to its color. The silver can heats up, but not as quickly.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The data from this demonstration (while the cans were heating up) is shown in Figure 16.4. (Download this data [XLSX, 47 kB])

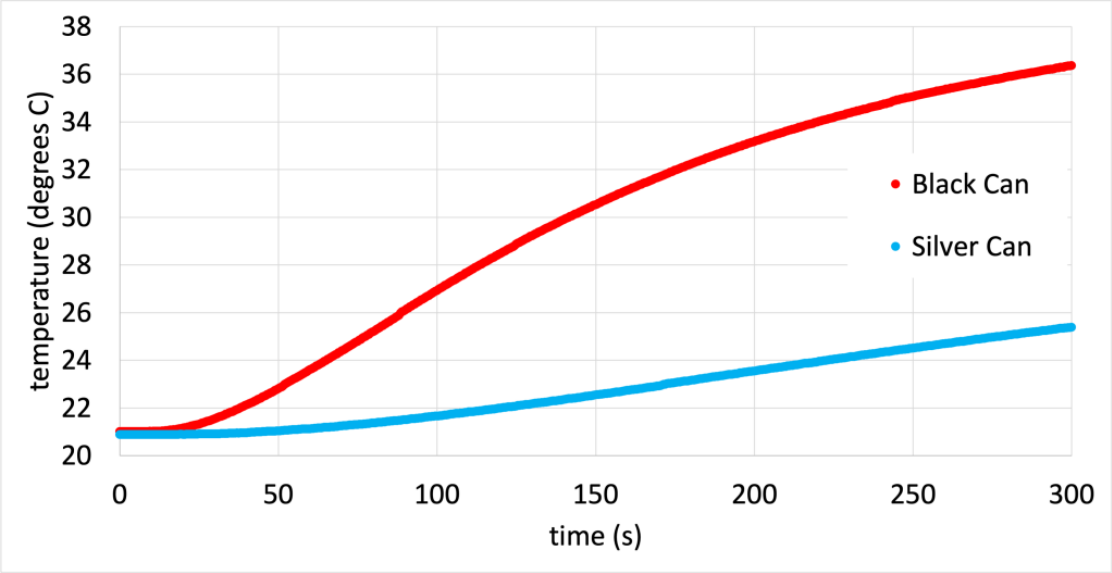


Figure 16.4 – A graph of the data from heating the black and silver cans. The black can heats at a much faster rate than the silver can. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

If a black object heats up faster, what can we say about how quickly it cools down? Dr. Pasquale removed the heat lamp and recorded the temperature of both cans as they cooled off. The black can not only heats up faster, but also cools down faster as well. The data from the experiment (while the cans were cooling down) is shown in Figure 16.5. (Download this data [XLSX, 31 kB])

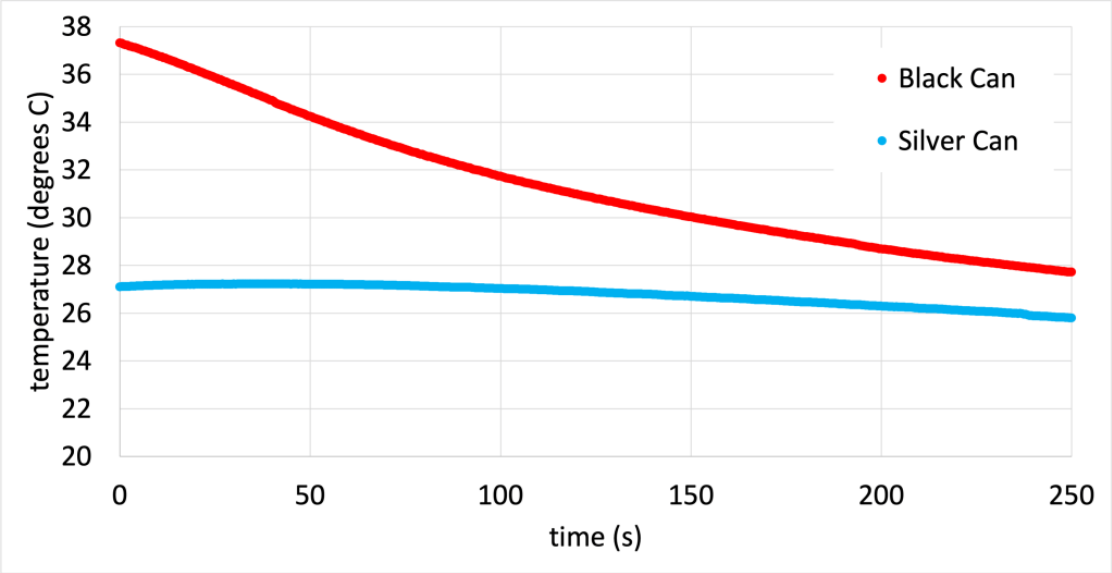


Figure 16.5 – A graph of the data from cooling the black and silver cans. The black can cools down at a much faster rate than the silver can. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The Sun is not the only object that emits electromagnetic radiation. In fact, when the cans from the video above cool down, they are re-emitting that electromagnetic radiation into the air and surrounding objects. Humans also absorb and emit electromagnetic radiation. The reason we don't see it is that the wavelengths of light we emit are in the infrared. Infrared light is not something that the human eye is able to see.

If we use an infrared camera, we can collect infrared light that is emitted from different objects. Figure 16.6 shows an infrared photo of one of the College of DuPage physics classes from a few years ago. Notice that people, who are warm compared to the air, emit a good amount of infrared light, which is then collected by the camera.



Figure 16.6 – An infrared photo of students in the physics classroom at COD. This figure, created by David Fazzini, Ph.D., is licensed under CC BY-NC-SA 4.0.

The students who wear glasses look like they have dark spots around their eyes. This is because glass does not allow infrared light to pass through it. It's not that the students have cold eyes, it's that the infrared radiation emitted by the student's eyes is blocked from reaching the camera.

The hotter an object is, the higher the energy of the light that they will emit. Infrared has relatively low energy and is generally emitted from objects that are relatively cool, such as "room temperature" objects or humans. Things that are very hot, such as stars, are much hotter and emit higher energy light. Light with more energy than infrared is red light, then orange light, all the way through the visible spectrum to blue light and then ultraviolet. A blue star would therefore be much hotter than a red star. We'll revisit these topics when we talk about electromagnetic waves and light emission later in this textbook.

In general, if an object absorbs more energy than it emits, its temperature will increase. If an object emits more energy than it absorbs, its temperature will go down.

The greenhouse effect

The greenhouse effect is a gradual warming process as heat is trapped by gases in the Earth's atmosphere. In short, these gases are transparent to the higher energy light coming from the (hot) Sun. These same gases are opaque to the lower energy radiation emitted from the (cold) Earth. This is analogous to an actual greenhouse (Figure 16.7), which is a beneficial structure used to keep plants from dying off in cold weather. In the case of our planet acting as a greenhouse, warming up the atmosphere and contributing to global climate change, the greenhouse effect is decidedly less positive.

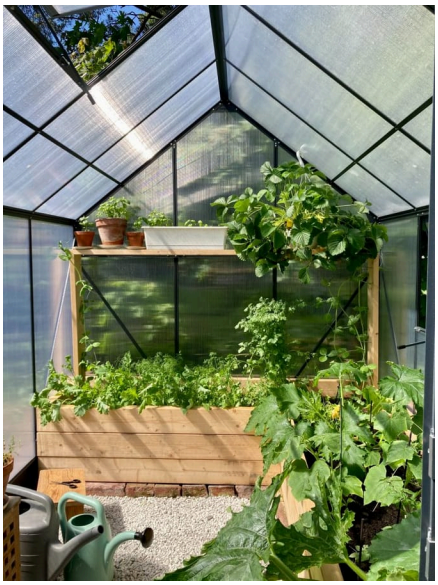


Figure 16.7 – A greenhouse is a building or room with glass walls used to keep plants warm year-round. Private greenhouse, by kallerna, is licensed under CC BY-SA 4.0.

A greenhouse is generally a glass building filled with plants. Sunlight is able to enter into the greenhouse through the windows, and the heat from that light is absorbed by the plants. The plants will re-emit light. Because plants are relatively cool, they emit heat as infrared light. As seen in the infrared photo in Figure 16.6, infrared cannot transmit through glass. Instead, it gets reflected off of glass surfaces. This means that all of the heat from the plants is trapped inside the greenhouse, causing the greenhouse to stay warm even if the outside is cold.

In the video below, two otherwise identical transparent plastic bottles have been placed at equal distances from a heat lamp. One of the plastic bottles has been filled with carbon dioxide, a greenhouse gas. As the heat lamp shines on both of the bottles, the temperature recorded by the computer shows that the bottle filled with carbon dioxide warms up much faster.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The temperature data recorded during this experiment is graphed in Figure 16.8. (Download this data [XLSX, 39 kB])

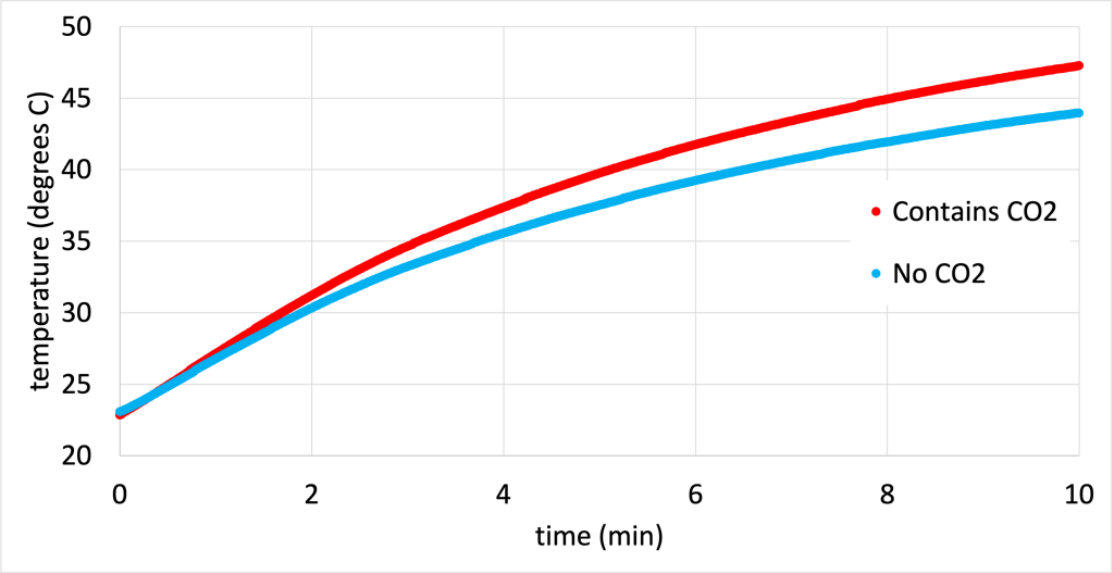


Figure 16.8 – A graph of the data from the greenhouse gas experiment. The bottle containing CO2 warms at a greater rate than the bottle without CO2. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

On our planet, the presence of greenhouse gases causes the re-radiation of electromagnetic radiation that would otherwise be lost if there were no greenhouse gases. This is shown in Figure 16.9.

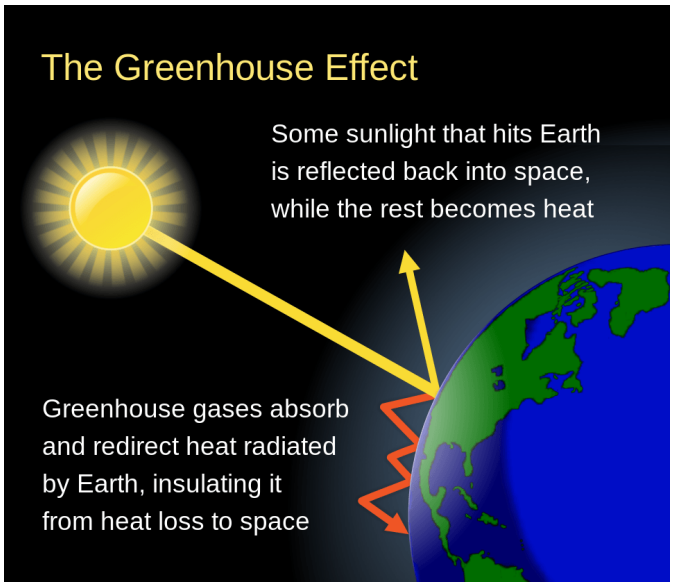


Figure 16.9 – A schematic of how the greenhouse effect contributes to warming of the Earth's atmosphere. Climate change schematic, by Efbrazil, is licensed under CC BY-SA 4.0.

Globally, the greenhouse effect is an issue for our planet as our atmosphere contains a lot of carbon dioxide and other greenhouse gases that have accumulated in large quantities due to human activities. As our planet continues to warm up, this becomes a bigger and bigger problem.

To a certain extent, the fact that our planet is capable of trapping heat makes it possible for our planet to sustain life. However, when the atmosphere heats up to dangerous levels, we see catastrophic weather events. This is an area where technology and policy need to come together to create lasting change that will protect our planet and its inhabitants.

Engineering heat transfer

An insulated beverage container is a great example of how we can use the knowledge of heat transfer to engineer a container

that can keep hot things hot and cold things cold. Insulated beverage containers are made from a double-walled material. Inside the two walls is a vacuum. The inside of the beverage container is silver in color. To keep things hot, or cold, we put a lid on top. (This is depicted in Figure 16.10.) How do all of those features keep our hot things hot and our cold things cold?

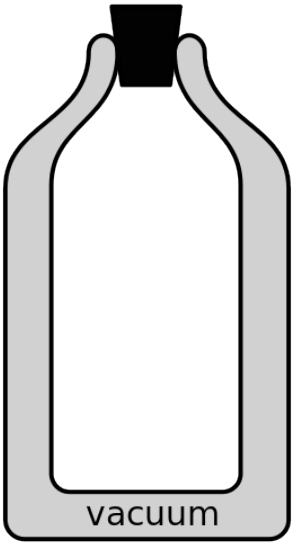


Figure 16.10 – A schematic of a vacuum flask (insulated beverage container). Vacuum dewar flask, by Acdx, is licensed under CC BY-SA 3.0.

First, let's answer the question of "how does an insulated beverage container know if something is hot, and should keep it hot, or if something is cold, and should keep it cold?" The answer is that those two concepts are identical. To keep something hot, we want to prevent heat from leaving the object. To keep something cold, we want to prevent heat from entering the object. Either way, we're eliminating or reducing all opportunities for heat transfer to occur.

To eliminate or reduce heat conduction, the vacuum in the beverage container's walls prevents heat from entering or leaving through the walls of the container. Gas in general is a bad heat conductor, and vacuum is incapable of conducting heat, as there are no molecules to collide in order to move heat around. The container lid is usually made of a heat insulator such as plastic or cork.

To eliminate or reduce convection, the vacuum between the beverage container's walls is also incapable of convecting. Because there are no molecules in a vacuum, they cannot distribute heat by moving around. By placing a lid on the container, we are eliminating convection in the vertical direction by trapping heat inside the container.

Finally, the silver color of the interior of the beverage container greatly reduces heat transfer by electromagnetic radiation. Any heat inside the beverage container is going to reflect off the walls and stay inside. Any heat outside of the container is going to reflect off the walls and stay outside.

In the video below, Dr. Pasquale has placed hot water (90 degrees Celsius) into two different containers: a double-walled calorimeter with a lid, and a metal can (the same kind used on the inside of the calorimeter) without a lid. The calorimeter acts the most like an insulated beverage container. There are two silver cups surrounded by air (not a good heat conductor), and the plastic and cork lid prevents convection.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The data from this experiment is graphed in Figure 16.11. (Download this data [XLSX, 37 kB])

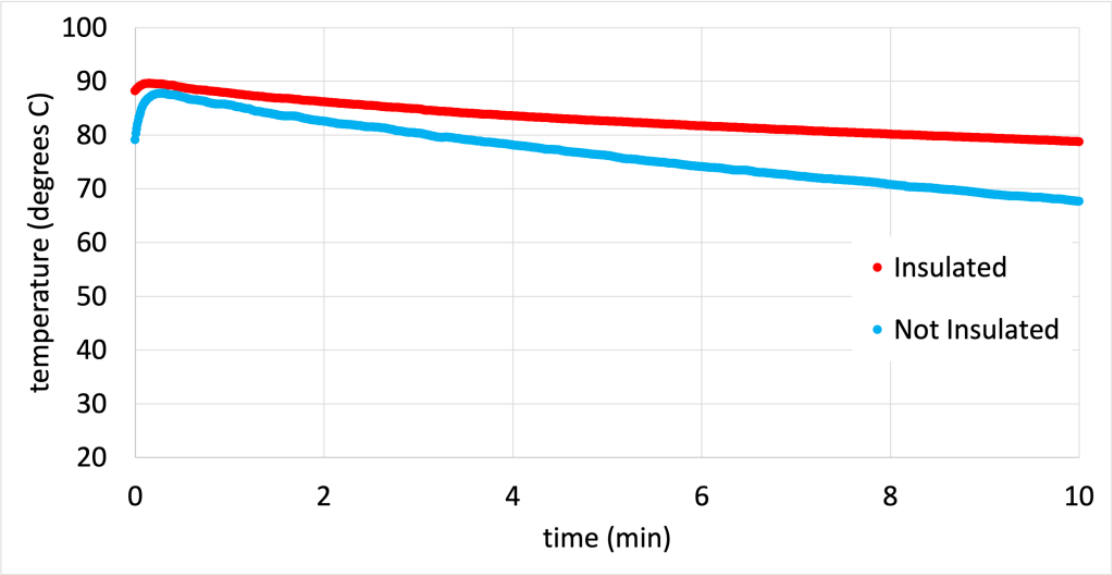


Figure 16.11 – A graph of the data from the experiment demonstrating the cooling of water in a calorimeter vs. in a non-engineered container. The calorimeter retains heat much better than the non-engineered container. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

After placing the hot water into both containers, Dr. Pasquale recorded the temperature changes as they cooled down. Over a span of ten minutes, the double-walled calorimeter held heat much better than the open metal can did. Much of the heat in the metal can was lost due to convection. Even placing a lid on a pot, pan, or your favorite coffee mug will help keep your food and beverages warmer for a longer time.

Solar power

The Earth is constantly receiving energy from the Sun. Solar power can be used to generate electricity for use in our homes using a photovoltaic panel, known as a solar panel (Figure 16.12).



Figure 16.12 – Solar panels arrayed on a rooftop. Solar array mounted on a rooftop, by AleSpa, is licensed under CC BY-SA 3.0.

In the video below, a solar panel is connected to a light-emitting diode (LED). When no light is shining onto the solar panel, there is no electricity generated, and the LED remains off. Once Dr. Pasquale shines a flashlight onto the solar panel, electricity is

generated and the LED turns on. This is a similar process to what happens when sunlight shines onto solar panels connected to a home or to the power grid.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The rate at which we obtain that energy is known as power: solar power. On average, the Sun transmits 1,400 J of energy every second to every square meter of the Earth that it hits. In other words, we receive 1,400 W/m^2. This value is known as the solar constant.

If we consume a certain amount of power, we can use that information to determine how large of a solar panel we would need to generate that amount of electricity. To obtain enough power from solar panels, our power needs cannot exceed the power we receive from the Sun. In equation form, this can be stated as

$$A_{\text{minimum}} = \frac{P_{\text{home}}}{1,400 \text{ W/m}^2},$$

where P_{home} is the average power consumption of our home and $A_{minimum}$ is the minimum area of the solar panels required to obtain that power from the Sun.

The amount of power we need depends on how much we run our appliances. On average, a household in the United States consumes 1,250 W of power in one day. If there are solar panels capable of taking the entire 1,400 W/m^2 of power from the Sun to convert to usable electricity, then solar panels that are

$$\begin{aligned} A &= \frac{1250 \text{ W}}{1400 \text{ W/m}^2} \\ &= 0.89 \text{ m}^2 \end{aligned}$$

would be sufficient to power that household. That size is not very big!

However, solar panels aren't necessarily 100% efficient. A good solar panel is at best 15% efficient as of the time this textbook was written. How does this affect the size of the solar panels needed to power the average household? The solar constant of 1,400 W/m^2 is now only converted at 15%, meaning that the solar panels will only create 210 W/m^2 of power. In this case, the size of the solar panels required will be

$$\begin{aligned} A &= \frac{1250 \text{ W}}{210 \text{ W/m}^2} \\ &= 5.95 \text{ m}^2 \end{aligned}$$

This is a big difference from perfectly efficient solar panels! Making solar panels more efficient is a big topic in engineering and physics research. More efficient solar panels will make renewable energy much more competitive with fossil fuels, and more convenient.

Another factor we must consider, if we do not live at the equator, is that the amount of energy we receive from the Sun is going to change throughout the seasons. This is because the Sun does not necessarily hit all parts of the surface of the Earth at a direct 90 degree angle. The equator may receive that 90 degree light, which means that area will receive more solar power than areas of the Earth that are tilted away from the Sun.

In the northern hemisphere in summer, the Earth's axis is tilted toward the Sun, providing us with more direct light. We therefore obtain more energy from the Sun, and the northern hemisphere climate warms up, giving us hot summers. In the winter, the Earth's axis is tilted away from the Sun, providing us with less direct light. We obtain less energy from the Sun per square meter,

and the northern hemisphere climate cools down, giving us cold winters.

A graphic depicting the seasonal variations in the northern and southern hemispheres due to the Earth's tilt relative to its orbit around the Sun is shown in Figure 16.13.

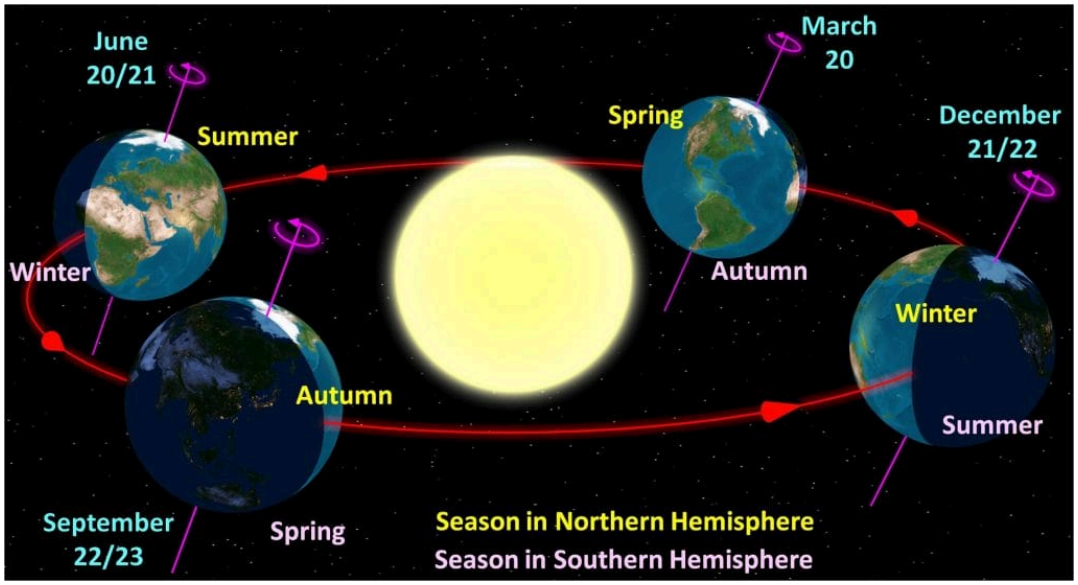


Figure 16.13 – An illustration of the Earth's axial tilt and how this contributes to seasons in the northern and southern hemispheres. Axial parallelism, by Tauʻolunga, is licensed under Public Domain.

Newton's law of cooling

Newton's law of cooling defines the rate at which an object loses heat, or cools down. It states that the change in temperature over time, ΔT (capital T for temperature) divided by Δt (lowercase t for time), is proportional to the temperature of an object minus the temperature of the surroundings. In other words:

$$\frac{\Delta T}{\Delta t} \propto T_{object} - T_{surroundings}.$$

That means when there's a big temperature difference between an object and its surroundings, we will expect it to cool down quickly. But as it cools down, the temperature difference between the object and the surroundings decreases, causing the rate of cooling to decrease as well. This continues to decrease until the temperature of the object asymptotically reaches the temperature of the surroundings.

This type of relationship is known as a differential equation and is described by a function known as an exponential function. The mathematics behind this function is beyond the scope of this textbook. However, we will use the basics of Newton's law of cooling to understand and interpret a graph of the cooling of an object.

In an experimental demonstration, Dr. Pasquale took water at an initial temperature of 94 ^{o}C and recorded the temperature change over time as it cooled down. The temperature of the room was also recorded at a constant value of 22.5 ^{o}C. The temperature recordings were measured for ten minutes.

After the experiment was completed, Dr. Pasquale used Logger Pro software to calculate the exponential equation that represented the data. This was used to extrapolate the data to see what trend would occur if the experiment had been performed for an entire hour. The experimental data and exponential extrapolation are shown in a graph in Figure 16.14. (Download this data [XLSX, 43 kB])

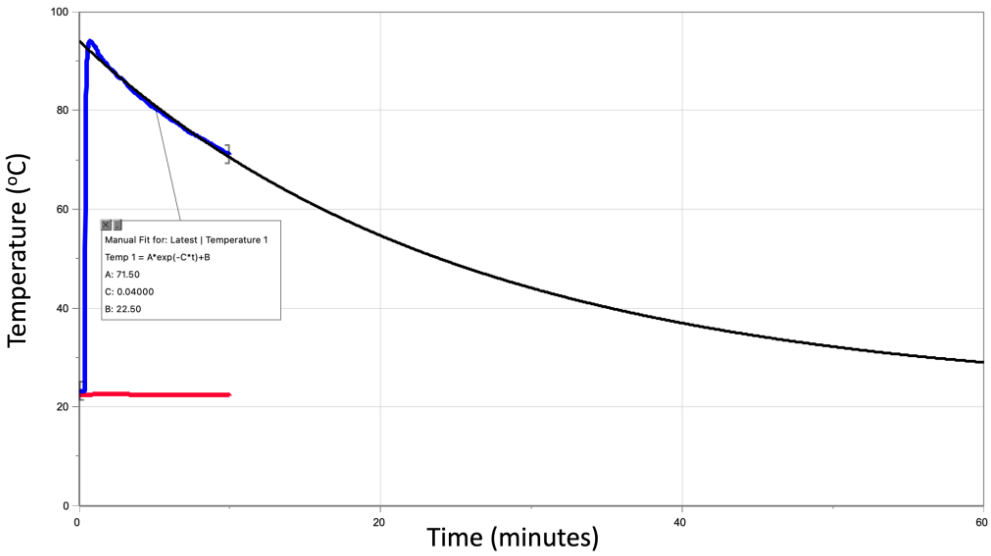


Figure 16.14 – A graph of the data collected during the Newton's law of cooling experiment. The blue data points represent the water, the red data points represent the room temperature, and the black line is the extrapolation of the water temperature over time. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

At the very beginning of the experiment, the temperature of the water was 94 degrees and the temperature of the room was 22.5 ^{o}C. The difference in temperature was 71.5 ^{o}C. At that time, the rate of cooling was 2.9 ^{o}C/min. That is, the slope of the line at that instant in time says that the water will cool off at that rate. Because the function isn't a straight line, the value of the slope is constantly changing, but the value of 2.9 °C/min is valid at this particular data point.

At 7.2 minutes into the experiment, the water had cooled down to 76.1 ^{o}C. At that point the temperature difference is 53.6 ^{o}C, three quarters of the initial value. What we should expect is that the slope of the line at this point will also be three quarters of 2.9 ^{o}C/min. According to the extrapolation, it is 2.1 ^{o}C/min, which is in very close agreement with that expectation.

We can look at a few more points and see that this trend is satisfied for all of the data calculated in this experiment. 17.3 minutes in, we would expect the temperature of the water to be 58.3 ^{o}C and the difference in temperature to be 35.8 ^{o}C, one half the original value. Now the slope of the line at this point would be about one half the original value, 1.4 ^{o}C/min.

Finally, 34.6 minutes into the experiment, the temperature of the water should be 40.4 ^{o}C and the temperature difference should be 17.9 ^{o}C, one quarter of the initial value. We would expect the slope to be one quarter the initial slope: 0.7 ^{o}C/min.

The slope continues to decrease over time. That is to say, as the water gets closer to room temperature, it cools down slower and slower, until eventually it asymptotically reaches the temperature of the room and is in thermal equilibrium.

Further reading

- Solar power – This page from the International Energy Agency gives information about solar power and its importance in reducing the effects of global climate change.

Practice questions

Numerical analysis

1. How large would perfectly efficient solar panels need to be to power a business using 8,500 W of electrical power?
2. If solar panels are only 15% efficient, calculate the size of the solar panels required for the business using 8,500 W of electrical power.
3. A household has 1 m^2 of solar panels, generating 150 W of electrical power. How efficient are the solar panels?

4. Hot, freshly baked cookies are removed from the oven at a temperature of 200 ^{o}C into a room that is 20 ^{o}C. At that moment, the cookies cool at a rate of 4 ^{o}C/min. How quickly do the cookies cool when the temperature of the cookies is...
    1. ...110 ^{o}C?
    2. ...65 ^{o}C?
    3. ...42.5 ^{o}C?

This page titled Chapter 16: Heat transfer is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 17: Phase change

Summary

This chapter expands on the topics of heat and thermodynamics by discussing the states of matter and how substances change their phase. We'll learn about the energy required to create phase changes, and analyze the heat, temperature, and phase diagrams of different substances.

Phases of matter

In our everyday lives, we experience four states of matter: solid, liquid, gas, and plasma.

A solid is a material where the molecules or atoms are very tightly bound together. This gives a solid a very rigid volume and shape. A solid can therefore be defined by the fact that a solid will not change either its volume or its shape.

A liquid is a material where the molecules or atoms are free to move about much more easily than in a solid, however, the molecules are still packed rather closely together. A liquid will change its shape, but cannot be easily compressed to change its volume. When pouring liquid from one container into another, while the shape of that liquid will change, the volume of the liquid will not change.

In a gas, the molecules are very free to move about, and they are not arranged tightly at all. This means that the shape and volume of a gas is free to change. A gas will assume the same shape as the container that it's in, and can be compressed to take up less space, or expanded to take up more space. If you've ever used an air compressor to fill up your car or bicycle tires, then you've changed the volume of air.

A plasma is an ionized gas. That is to say, it's a gas with so much energy that electrons have been ripped away from the atoms. While this might seem like an exotic state of matter, it's something we experience in our daily lives. Stars, including our Sun, are made of ionized hydrogen and helium: plasma. Lightning is a plasma. Even fluorescent bulbs use a plasma to create light.

In general, we can characterize each state of matter by the amount of energy it has. For any given substance, a solid has the least amount of energy, followed by a liquid, followed by a gas, and then plasma, which has the most energy.

While there are four common states of matter, there are in fact a lot of other states of matter. However, they tend to be very exotic. That means they need to be under very special situations to occur. Some states of matter require temperatures very close to absolute zero, or require extraordinarily high pressure to exist. We are very unlikely to ever deal with these states of matter in our day-to-day lives.

Exotic states of matter

States of matter that exist at extremes of temperature and pressure have been discovered throughout the years.

Superfluids, which are created when certain liquids are cooled to extremely low temperatures, experience zero viscosity. This enables them to flow through pores in otherwise solid objects (such as ceramic) and climb the walls of containers and flow to the lowest point.

Bose-Einstein condensate is a state of matter that is also achieved at extremely low temperatures. In this type of matter, quantum physics properties that are usually only observed on extremely small scales become apparent on much larger scales.

There are many other types of states of matter, both experimentally validated and theoretically proposed by scientists. These exotic states, such as those found in neutron stars or in the earliest seconds of the universe, provide fascinating research opportunities for physicists interested in matter.

Vaporization and condensation

Vaporization is the phase change that occurs when a substance converts from a liquid to a gas. Because a gas has more energy than a liquid, that means that energy must be provided to a substance to get it to change from a liquid to a gas.

In the vaporization process, molecules in their liquid state absorb energy. That energy loosens the bonds between the molecules, freeing them to move about as a gas. Heat is required to convert a liquid to a gas. That heat must come from somewhere. When vaporization occurs, the surrounding environment cools down after it provides that heat. It is important to note that because the surrounding environment is typically extremely large with respect to the substance that is vaporizing, the temperature change of the environment is typically not noticeable. (This is similar to the example earlier in this textbook when we discussed how dropping a smartphone will cause the Earth to accelerate upward, but to such a small degree that it can be essentially ignored.)

There are two types of vaporization: boiling and evaporation. Boiling will be discussed in the next section of this chapter. Evaporation is a subset of vaporization that occurs on the surface of a liquid. That is, even if we have a large volume of liquid, evaporation only occurs at the interface between the liquid and surrounding fluid, such as air. To demonstrate this, in the video below, Dr. Pasquale sprays a small amount of alcohol onto a lab bench and places a fan nearby. The alcohol evaporates by changing phase to a gas at the surface. The pool of liquid becomes shallower and shallower until all of the alcohol is converted into a gas.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Evaporative cooling is something that we as humans experience frequently. Our bodies sweat when it's warm. Our skin provides the heat required for that sweat to evaporate, causing our skin to cool down. Evaporative cooling can also be used in homes in areas with a dry climate. In the Chicagoland area, evaporative cooling is not as effective as air conditioning due to our humid summer climate.

To demonstrate that evaporation is a cooling process, in the video below, Dr. Pasquale dips a thermometer in a bottle of alcohol and then waves it around in the air to get the alcohol to evaporate. Heat from the thermometer changes the phase of the alcohol from a liquid to a gas, and as a result the temperature recorded by the thermometer decreases.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The drinking bird (shown in the video below) is another example of the cooling properties of evaporation. The bird's head is dipped into water. As the water evaporates from the bird's head, the cooler temperature causes the pressure inside the head to decrease. This pressure differential causes the fluid inside the bird's body to be drawn up into the head, causing a net torque. A net torque will create rotational acceleration, causing the bird's head to rotate downward. The bird dips its beak and the fluid in the head is able to travel back into the bottom of the bird. This process repeats until the bird's head dries off and evaporation stops.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Condensation is the opposite of vaporization. It is the phase change from a gas to a liquid. As the molecules lose energy in the condensation process, they start to assemble in a tighter form, which gives liquids a more well-defined arrangement than gases. The energy lost by the substance is released into the surroundings, causing them to warm up. (Again, as the surrounding environment is typically very large compared to the condensing substance, this does not usually cause a noticeable temperature change.)

Condensation is something you may have noticed if you have a cold glass of water, or can of soda, sitting out on a warm or hot day. The water vapor in the air surrounding the container will condense into a liquid form, causing beads of liquid on the container (Figure 17.1).



Figure 17.1 – Beads of condensation from water vapor form on a can of soda when the contents of the can are colder than the temperature of the surrounding air. Diet Coke, by Carissa Rogers , is licensed under CC BY 2.0.

It is important to note that, for a given substance, vaporization and condensation occur at the same temperature. Pure water at 1 atm of pressure will vaporize at 100 ^{o}C, and it will also condense at 100 ^{o}C. Whether it converts from a liquid to a gas, or from a gas to a liquid, depends on whether it is gaining or losing energy at this temperature.

Boiling

Boiling is a type of vaporization (change of state from liquid to gas) that occurs throughout the volume of a liquid. Boiling occurs when the average kinetic energy of the molecules in the liquid becomes large enough for the molecules to start to overcome the cohesive force that keeps them attracted to each other. When this occurs, molecules in the liquid separate from each other and convert into the gaseous form as they remain situated throughout the liquid. Eventually, the buoyant force or kinetic energy of the gas bubbles will cause them to rise up and out of the liquid. That is, bubbles of gas are able to form without being squashed out of existence by atmospheric pressure. In the video below, Dr. Pasquale uses a hotplate to boil water. You can see that the water turns into a gas within the liquid. Those gas bubbles move upward due to a buoyant force and then leave the beaker.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Boiling has to do with atmospheric pressure. At standard atmospheric pressure and temperatures, a beaker of water is not going to boil. This is because any gas bubbles that try to form will collapse back into the liquid state due to the pressure of the atmosphere pushing down from above. As the liquid heats up, the energy of the bubbles increases, until a certain point where the bubbles overcome atmospheric pressure. This is known as the boiling point.

At a higher elevations, atmospheric pressure is lower. This means that the temperature at which water will freely exist in the gas phase is lower. In fact, if you have ever cooked cake or brownies from a mix, you may have seen a high altitude note on the box. This is because water boils at a lower temperature at high altitudes, and things may not cook as thoroughly as they would at sea level.

It is possible to boil water at room temperature by reducing atmospheric pressure. Dr. Pasquale demonstrated this in the video below using a vacuum chamber. By pumping out most of the surrounding air, there is no longer enough pressure to collapse any bubbles that form in the water. This beaker of water is boiling at room temperature.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This demonstration emphasizes why humans need to wear pressure suits at extremely high altitudes in flight, or in the vacuum of space. Above 19,000 meters, the fluid in our throat and lungs will boil away. A pressure suit compensates for this by compressing the human body, preventing this boiling process.

Latent heat of vaporization

Latent heat of vaporization describes the amount of heat required to convert one gram of a substance from a liquid form to a gaseous form. Alternatively, it defines the amount of heat that must be removed from one gram of a substance to convert it from a gas to a liquid.

During the vaporization or condensation process, it is important to note that the temperature of the substance remains constant. Any heat entering the liquid does not increase the temperature: it changes the molecular arrangement of the atoms. Any heat leaving the substance does not decrease the temperature: it changes the arrangement of the atoms.

Because the vaporization and condensation temperatures are the same, and the amount of latent heat is the same for both processes, we can only tell if a material is vaporizing or condensing by determining if the object is gaining heat or losing heat. If the substance is gaining heat, then the molecules are being reconfigured into a gas and vaporization is occurring. If the substance is losing heat, then the molecules are being reconfigured into a liquid and condensation is occurring.

Water has a relatively high latent heat of vaporization, especially compared to other common substances. The latent heat of vaporization of water is 540 cal/g. This means that 540 calories of heat must be added to one gram of water at 100 ^{o}C to convert it into one gram of steam at 100 °C.

The equation we can use for calculating the amount of heat required to change the phase of a substance from liquid to gas is

$$Q = mLH_{vaporization}.$$

Heat equals the mass of the substance times its latent heat of vaporization.

Melting and freezing

Melting is the phase change that occurs when a substance converts from a solid to a liquid. Because a liquid has more energy than a solid, that means that energy must be provided to a substance to get it to change its phase.

In the melting process, molecules in their solid state absorb energy. That energy loosens the tight bonds between the molecules, allowing them to move about more freely. Heat is required to convert a solid to a liquid. That heat must come from somewhere. When melting occurs, the surrounding environment cools down after it provides that heat. (Similar to what was discussed above, the surrounding environment is typically large enough not to have its temperature noticeably affected by this process.)

Freezing is the opposite of melting. It is the phase change from a liquid to a solid. As the molecules lose energy in the freezing process, they start to form tight bonds, which gives solids a very well-defined shape. The energy lost by the substance is released into the surroundings, causing them to warm up. Again, the temperature change of the surroundings, which are extremely large with respect to the freezing substance, can typically be ignored. (An example of this would be adding a ladle of liquified molten iron into Lake Michigan. The heat released into Lake Michigan from freezing the molten iron back to its solid form, and then cooling it to the same temperature of the lake, would not be sufficiently high to cause a noticeable change in temperature of the lake.)

It is important to note that, for a given substance, melting and freezing occur at the same temperature. Water will freeze at 0 ^{o}C, and it will also melt at 0 ^{o}C. Whether it converts from a solid to a liquid, or from a liquid to a solid, depends on whether it is gaining energy from the environment or losing energy to the environment at this temperature.

In the winter, salt is frequently placed on roadways and sidewalks to cause ice to melt. Salt will block the bonds between water molecules, making it much harder for liquid water to freeze into a solid. This lowers the freezing point; salt water has a lower freezing point than fresh water. If the freezing point is lower, the melting point is also lower. Frozen salt water will melt at a lower temperature than fresh water.

Latent heat of fusion

Latent heat of fusion describes the amount of heat required to convert one gram of a substance from a solid form to a liquid form. Alternatively, it defines the amount of heat that must be removed from one gram of a substance to convert it from a liquid to a solid.

During the melting or freezing process, it is important to note that the temperature of the substance remains constant. Any heat entering the solid does not increase the temperature: it changes the molecular arrangement of the atoms. Any heat leaving the liquid does not decrease the temperature: it changes the molecular arrangement of the atoms.

Because the melting and freezing temperatures are the same, and the amount of latent heat is the same for both processes, we can only tell if a material is melting or freezing by determining if the object is gaining heat or losing heat. If the substance is gaining heat, then the molecules are being reconfigured into a liquid and melting is occurring. If the substance is losing heat, then the molecules are being reconfigured into a solid and freezing is occurring.

The latent heat of fusion of water is 80 cal/g. This means that 80 calories of heat must be added to one gram of ice at 0 ^{o}C to convert it into one gram of water at 0 °C.

The equation we can use for calculating the amount of heat required to change the phase of a substance from solid to liquid, or from liquid to solid, is

$$Q = mLH_{fusion}.$$

This is similar to an equation discussed earlier in this chapter, but now instead of using the latent heat of vaporization, we use

the latent heat of fusion.

Triple point

The triple point is a particular combination of temperature and pressure at which a substance will exist in liquid, solid, and gaseous forms simultaneously. This might seem outrageous, but it is simply an incredibly interesting property of matter. When held at the triple point, both the freezing/melting and condensation/vaporization processes can be seen happening at the same time, with all three phases (solid, liquid, and gas) existing in equilibrium with each other.

The boiling point (and freezing point) of a given substance generally varies with pressure. Recall that the temperature at which water boils decreases when atmospheric pressure is greatly reduced. The triple point is a unique point that exists at a particular temperature and pressure for a substance.

The video below shows a solvent (cyclohexane) at the triple point. As you watch the video, notice how the processes of freezing and boiling occur simultaneously. The solvent exists in three states of matter (solid, liquid, and gas) at the same time.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Sublimation and deposition

Sublimation is the process by which a substance goes directly from a solid to a gas without going through the liquid phase. Some substances, such as carbon dioxide, cannot exist as a liquid at standard atmospheric pressure.

The video below shows a piece of dry ice as it was recorded over a span of half an hour. While you cannot actually see the carbon dioxide when it turns into a gas, it is clear that the piece of dry ice becomes physically smaller over time, and no liquid is generated. The carbon dioxide goes directly from solid to gaseous form.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Water vapor from the surrounding air creates frost on the surface of the dry ice over time. The temperature at which carbon dioxide will sublimate is just about -80 ^{o}C, much colder than the freezing point of water!

Therefore, we are not only witnessing the sublimation of carbon dioxide, but the deposition of water. Deposition is the opposite process to sublimation: it is when a gas turns directly into a solid without passing through the liquid form. This occurs all the time in winter when water vapor in the air turns into frost.

Deposition of substances such as chromium, gold, platinum, copper, and other metals is frequently used in the semiconductor industry to create electric contacts, sensors, and other micro-scale or nano-scale electronics and optical devices.

Ionization and recombination

Ionization is the process by which a gas is converted into a plasma. As heat is added to a gas, the energy of the gas increases to such an extent that electrons orbiting the atom are ripped away, causing the gas molecules to have an imbalance of electric charge. Recall from earlier in this textbook that an atom with different numbers of electrons and protons is known as an ion.

Recombination is the reverse process. It occurs when a plasma loses energy and gains its electrons back, converting back to a regular gas.

Phase change diagrams

A phase change diagram shows the relationship between temperature and heat in a substance. The x-axis of the graph shows how much heat is added to a substance. In this textbook, that heat will be measured in calories. Along the y-axis is the temperature of the substance measured in degrees Celsius. From these graphs, we can learn a lot about a substance. We can determine the melting/freezing point, boiling/condensation point, latent heat of fusion, latent heat of vaporization, and specific heat capacity of the solid, liquid, and gaseous forms of the substance.

Figure 17.2 shows an example of a phase change diagram for one gram of a random substance. This particular phase change diagram starts in the lower left corner with a temperature of -200 ^{o}C and no heat added. As heat is added, the temperature increases. This increase continues until 50 cal of heat have been added, at which point the temperature hits a plateau at -100 ^{o}C. Then the temperature remains constant at -100 ^{o}C until an additional 150 cal of heat are added. After that, the temperature starts to increase with more added heat. Adding an additional 150 cal of heat causes the temperature to rise to +100 ^{o}C, at which point another plateau is reached. The temperature remains constant through the addition of another 200 cal of heat. Finally, adding 150 cal of heat causes the temperature of the substance to rise from 100 ^{o}C to 500 ^{o}C.

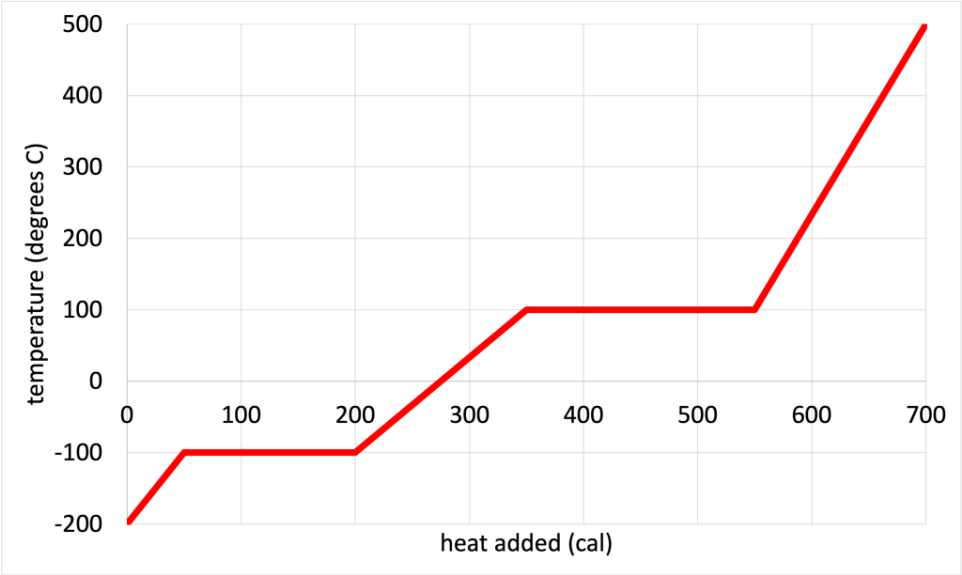


Figure 17.2 – An example phase change diagram for a random substance. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This phase change diagram contains information about the phases and phase changes of the substance. This is shown in Figure 17.3, which has the same phase change diagram as in Figure 17.2, with all of the phases and phase changes labeled. The colder temperatures of the graph represent the substance in its solid form. The middle region represents the substance in liquid form. The hotter region represents the substance in its gaseous form. The melting/freezing point is the colder plateau temperature, and the boiling/condensation point is the hotter plateau temperature. These are phase change temperatures because adding heat doesn't increase the temperature but rearranges the bonds between the atoms in the substance to change phase.

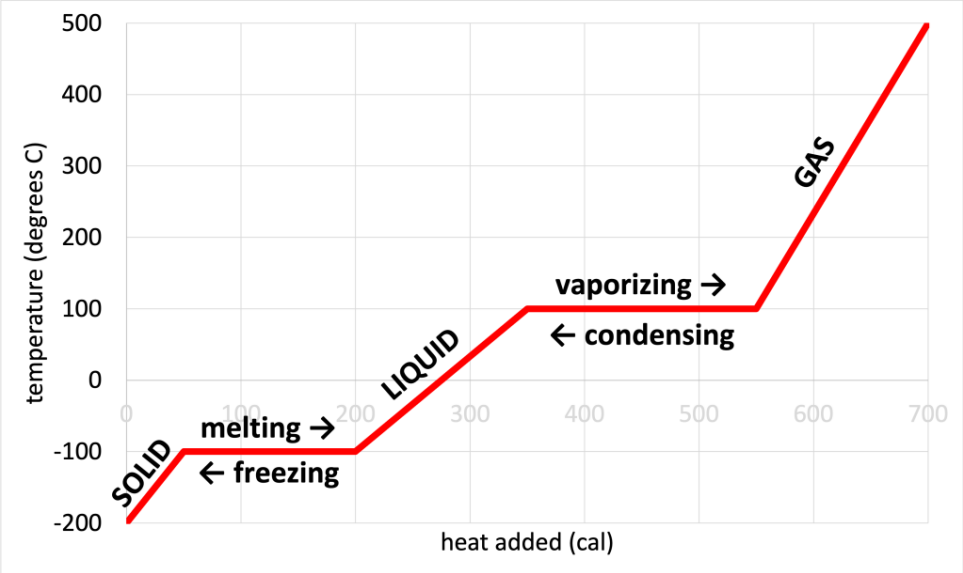


Figure 17.3 – The states of matter and phase transitions for a random substance. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The plateaus also provide information about the latent heats of fusion and vaporization. From this graph, we see that it requires 100 cal of heat to convert the substance from a solid to a liquid, while remaining at a temperature of -100 ^{o}C. Therefore, the latent heat of fusion is 100 cal/g. It requires 200 cal of heat to convert the substance from a liquid to a gas at a constant temperature of 100 ^{o}C. Therefore, the latent heat of vaporization is 200 cal/g.

The rising areas of the graph give information about how much heat is required to change the temperature of the substance while it is in a single phase – this is related to the specific heat capacity. The equation for specific heat capacity is

$$c = \frac{Q}{m\Delta T}.$$

Therefore, to calculate the specific heat capacity of the substance in any form, divide the heat added to the substance by the mass of the substance times the temperature change of the substance.

For the phase change diagram shown in Figure 17.2, in the solid phase, 50 cal of heat increases the temperature from -200 ^{o}C to -100 ^{o}C, a temperature change of 100 C^{o}. The specific heat capacity of the solid form of this substance is 0.5 cal/g C^{o}.

In the liquid phase, 150 cal of heat changes the temperature of the substance from -100 ^{o}C to 100 ^{o}C, a temperature change of 200 C^{o}. The specific heat capacity of the liquid form of this substance is 0.75 cal/g C^{o}.

In the gaseous phase, 150 cal of heat changes the temperature from 100 ^{o}C to 500 ^{o}C, a temperature change of 400 C^{o}. The specific heat capacity of the gaseous form of this substance is 0.375cal/gC^{o}.

The phase change diagram can also be used to answer other questions about temperature change and heat added. For example: “how much heat is required to convert 1 gram of this substance from a solid at -100 ^{o}C to a gas at 100 ^{o}C?” First, identify the two points on the graph. The point where the substance is completely solid at -100 °C is at the start of the plateau representing the latent heat of fusion, before the phase changes to liquid. The point where the substance is completely gaseous at 100 °C is at the end of the plateau representing the latent heat of vaporization, after the phase completely changes to gas. Reading from the x-axis, it requires 450 calories of heat to go from one to the other point.

This process can be repeated to answer other questions. Identify the two points on the graph. Then, use the change in y-axis value to determine the temperature change or the change in x-axis value to determine the required heat added.

Further reading

- The engineering of the drinking bird – This video gives a great explanation of the drinking bird toy. The topics include evaporation, thermodynamics, and rotational motion.

Practice questions

Numerical analysis

1. If 100 g of ice at -10 ^{o}C is heated until it turns into water at 0 ^{o}C, calculate the amount of heat energy that will be absorbed by the ice during this process. (The specific heat capacity of ice is 0.5 cal/g^{o}C.)
2. A pot of water on the stove is boiling vigorously at 100 ^{o}C. If you continue to apply heat, calculate how much energy is required to convert 500 g of boiling water into steam at 100 ^{o}C.
3. A scientist wants to melt a 50 g block of lead at its melting point. Calculate the amount of heat energy required to turn the solid lead at 327 ^{o}C into liquid lead at 327 ^{o}C. (The latent heat of fusion of lead is 5.5 cal/g.)
4. For the phase change diagram shown in Figure 17.4 (which depicts the temperature change vs. added heat for one gram of a mystery substance), calculate the...
    1. ...freezing/melting point of the substance.
    2. ...condensation/vaporization point of the substance.
    3. ...latent heat of fusion of the substance.
    4. ...latent heat of vaporization of the substance.
    5. ...specific heat capacity of the solid form of the substance.
    6. ...specific heat capacity of the liquid form of the substance.
    7. ...specific heat capacity of the gaseous form of the substance.

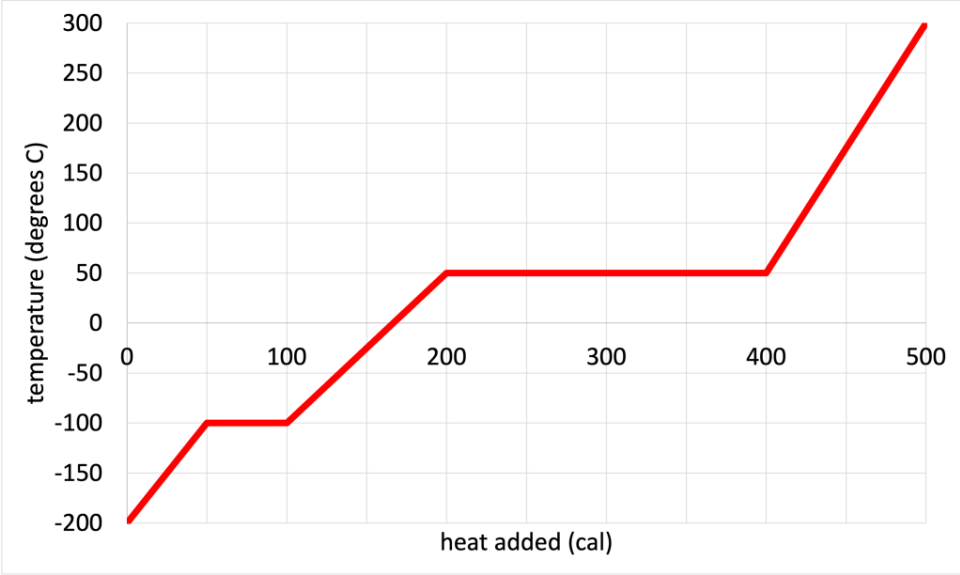


Figure 17.4 – A phase change diagram depicting the change in temperature vs. heat added for one gram of a mystery substance. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This page titled Chapter 17: Phase change is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 18: Thermodynamics

Summary

Thermodynamics is the study of heat flow and its relationship to temperature, work, energy, and entropy. Thermodynamics has huge relevance in our lives, as it explains the workings of and constraints of all heat engines, from power plants to internal combustion engines to refrigerators and HVAC systems.

The laws of thermodynamics play a large role in how scientists and engineers develop efficient power generators and motors. The concept of entropy is always present in our lives, as we find that it's easier to create disorder than order.

Absolute temperature scales

In thermodynamics, absolute scales are important. This is because an absolute scale is established based on a physical limit to how cold things can get. This physical limit establishes the zero end of the scale. The Kelvin scale is an absolute temperature scale. Zero kelvin is equal to absolute zero: the coldest possible temperature.

The Kelvin temperature scale was developed so that a change in one kelvin is equal to a change in one Celsius degree. This makes conversion between Celsius and Kelvin scales a matter of simply adding and subtracting. Zero K is equal to -273 ^{o}C. Therefore, to convert from Celsius to kelvin, add 273. To convert from kelvin to Celsius, subtract 273.

What evidence do we have that there is an absolute lower limit to temperature? How do we know that absolute zero exists?

Experiments establishing the relationship between the temperature of a substance and the volume of that substance were performed in the 1700s. As the temperature increased, volume increased. By extrapolating this data, it is possible to see a temperature at which the volume of the substance would be equal to zero. This is shown in a graph in Figure 18.1.

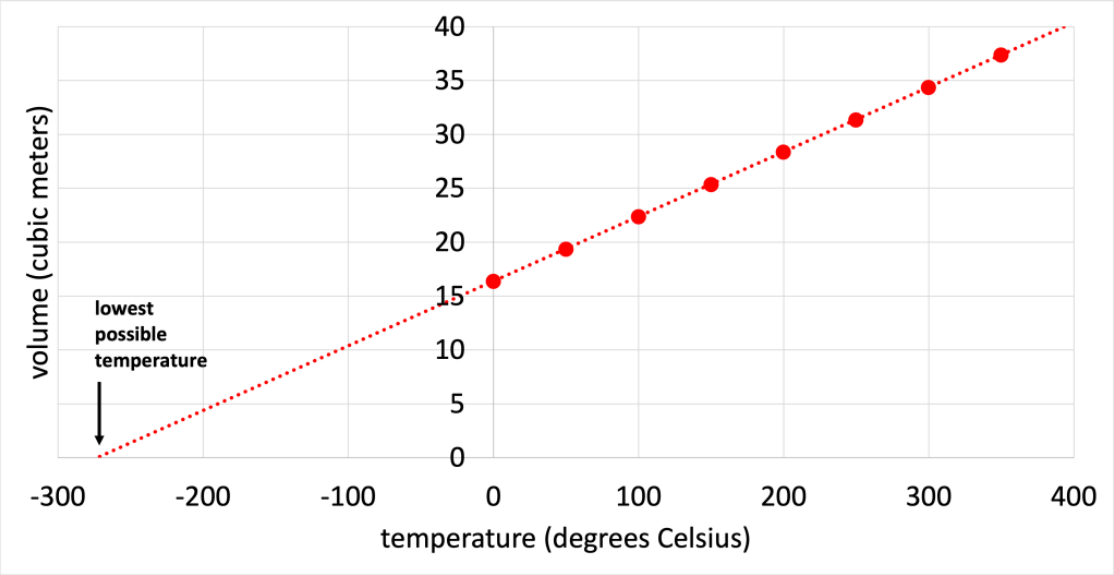


Figure 18.1 – A graph depicting the relationship between temperature and volume of a hypothetical substance. The point at which the volume would be zero is the lowest possible limit to temperature (absolute zero). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This establishes a physical limit to the system: it is not possible to have zero volume, and certainly not possible to have a negative volume. Experiments on other substances found a similar relationship. That is, different gases showed a linear volume vs.

temperature graph that, when extrapolated to zero volume, led to the same temperature. This temperature is known as absolute zero.

Scientists have been able to achieve extremely low temperatures, to within less than a billionth of a degree of absolute zero. To our current level of knowledge, actually achieving absolute zero is not possible. This is because it would take an infinite amount of work to remove all of the heat from a substance. Absolute zero is more of an asymptotic limit than an actually attainable temperature.

The first law of thermodynamics

The first law of thermodynamics is nothing more than conservation of energy with the inclusion of thermal energy. In a closed system, the first law of thermodynamics can be stated in equation form:

$$\Delta E = Q - W.$$

ΔE is the change in internal energy of the system, expressed in units of energy (joules or calories). Q is the heat added to the system, also expressed in either joules or calories. W is the work done by the system, also expressed in joules or calories. Emphasis should be placed on the signs of Q and W when performing calculations. Since Q is the heat **added** to the system, Q will be a negative number if heat is removed from the system. Since W is the work done **by** the system on its surroundings (such as during an expansion), W will be a negative number if the surroundings do work on the system (such as during a contraction).

Internal energy is the sum total of all of the kinetic and potential energies of the molecules in a system. This is usually related to temperature and phase. If the internal energy of a system is increased by adding heat, and that system does no work, then the molecules in that system will increase their kinetic energy or change their phase.

The first law of thermodynamics tells us that we cannot get energy for free. There is no such thing as a machine that can be made to do work with no energy input. It would violate conservation of energy. Simply put: “we cannot get something for nothing.”

Adiabatic processes

An adiabatic process is a specific type of thermodynamic process in which the heat added to or removed from a system is zero. The system is either very well insulated or the process occurs so quickly that there is not enough time for heat to be added or removed from the system. If work is done ON the system, W will be negative, and the internal energy of the system will increase.

This is demonstrated in the video below. This syringe contains the head of a match. As you can see, the initial state is that the match has not been ignited. First, Dr. Fazzini pushes down slowly on the syringe. This process is not adiabatic; work is done on the gas inside of the syringe, but that heat has enough time to leak through the walls of the syringe, causing the temperature inside to remain relatively constant. The second time, Dr. Fazzini pushes quickly and creates an adiabatic process. When the fire syringe is compressed very quickly with enough force, there is not enough time for heat to leak out of the system. The work done on the system causes the internal energy of the match head to increase to such an extent that the match head ignites and briefly sparks a fire.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This process is used in diesel engines. Internal combustion engines require a mixture of fuel and air to be ignited to create mechanical work. In a gasoline engine, spark plugs are used to ignite that mixture. In a diesel engine, there are no spark plugs. Instead, the work done during the adiabatic compression of the piston in the engine provides the increase in internal energy that causes the ignition of the fuel and air mixture, creating mechanical work output.

The second law of thermodynamics

Before discussing the second law of thermodynamics, it is important to consider the direction that heat will spontaneously flow. Heat will only spontaneously flow from hot to cold. It will never spontaneously flow from cold to hot. This is similar to gravity, which will cause a ball to roll down a hill, but never up. To get something back to the top of the hill, we have to exert energy to do so. Similarly, while heat will not spontaneously flow from a region of lower temperature to a region of higher temperature, that does not mean we cannot make heat flow "uphill." To get heat to flow from cold to hot, as an air conditioning unit or refrigerator does, external energy input is required. That fact that nature has a preference for the direction in which heat will flow is governed by the second law of thermodynamics.

Consider a metal rod that is hot on one end and cold on the other end at some initial point in time. As time passes, the hot end will cool down and the cold end will warm up until the entire rod is at a common intermediate temperature. The thermal energy flows from the hot end to the cold end. However, the reverse process will never occur spontaneously even though it does not violate conservation of energy. (The greater kinetic energy of the atoms and molecules at the hot end could be gained at the expense of decreased kinetic energy of the atoms and molecules at the cold end. Still, this never happens spontaneously.)

A heat engine (which will be discussed in greater detail in the next section of this chapter) is a device that uses a difference in temperature to do useful work while operating in a cycle. The second law of thermodynamics tells us that not all heat input will be converted to useful work in this process. There will always be wasted heat. So, while the first law of thermodynamics tells us that "we can't win," the second law tells us "we can't break even" either.

In the video below, Dr. Pasquale uses hot and cold water to do useful work in a demonstration of a heat engine. This demo is a very simple model of a factory where widgets come in on one level on a conveyor belt, get raised up to a higher level, and then get moved away to another area of the factory on a second conveyor belt higher up. The heat engine raises the mass (so that it can move along the higher conveyor belt), and then lowers down so a new mass can be placed into the system. By placing the bulb of this heat engine into hot water, the gases inside the bulb warm up and expand, causing the mass to rise up to the second level. While in this demo there is no conveyor belt, much less a whole factory, Dr. Pasquale can use her hand to move the mass elsewhere.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

After the first mass has moved along in the model factory, the system is ready for a second mass. How can the system reset back to the lower level? The bulb needs to cool off somehow. It is currently warm, and now Dr. Pasquale must throw away that heat to reset the process. Whether Dr. Pasquale waits for the bulb to cool off in the atmosphere, or cools the bulb by plunging the bulb into cold water, either way, that heat energy does not do any useful work. Instead, it slightly raises the temperature of the surroundings, which does not equate to any useful mechanical process.

Dr. Pasquale repeats this process several times in the demo. Every time she needs to reset the system, heat energy is wasted. This is an inevitable result of physics that is formalized by the second law of thermodynamics.

Heat engines

A heat engine is a device that uses a difference in temperature to do useful work while operating in a cycle. Heat engines are used in many types of power plants: nuclear, coal, and gas turbines heat water into steam to spin a coil of wire around a magnet, generating electricity. Internal combustion engines that use fossil fuels to do mechanical work, especially in automobiles, are heat engines. On a more basic level, the refrigerator in your home is a heat engine operating in reverse.

The drinking bird, shown in the video below, uses a heat difference due to the evaporation of water from the bird's head. This causes a fluid to flow up into the bird's head and create a torque that tips the bird. When the bird's head tips over, fluid is released

back into the base of the bird and it tips back upright again. This process will repeat until all of the fluid evaporates.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

While the drinking bird is arguably not a terribly useful heat engine, and is certainly a very inefficient heat engine, it produces mechanical work from a temperature difference.

The Stirling engine, shown in the video below, uses a heat source (boiling water underneath the engine) and a cool source (room temperature air above the engine) to expand and compress gas that causes the motion of pistons. Without a temperature difference between the hot piston and the cold piston, gas would not expand or contract, and there would be no mechanical work output.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The video below demonstrates a heat engine known as a Hero's engine. There is a large flask filled with water placed over an open flame. That flame causes the water to boil, creating steam. The steam releases through openings in the piping. This steam generates a torque that causes the engine to rotate. This rotation is mechanical work.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Note that this demo caused water to splatter out of the openings in the piping, demonstrating really well the concepts of centripetal force and tangential velocity!

Engine efficiency

The measure of how well a heat engine converts thermal energy into useful work is called the efficiency of the engine. To put it simply, the efficiency is just what you "get out" divided by what you "put in." What you get out is the useful work done by the engine. What you put in is the heat that enters the engine from the high temperature region. That is,

$$e = \frac{\text{work output}}{\text{heat input}}.$$

A perfect engine would be one in which all of the entering heat is converted into useful work. Such an engine would be 100% efficient because there would not be any heat expelled into the low temperature region. The second law of thermodynamics tells us that the efficiency of a heat engine can never be 100%. Some of the heat in the system will always be transformed to unusable forms of energy and wasted.

As shown in the video above (demonstrating the simple model of a factory), there must always be heat dumped into the low temperature region in order to reset the engine for the next cycle. This can perhaps be visualized by looking at a block diagram of a heat engine, shown in Figure 18.2. This confirms the conclusion of the second law of thermodynamics, which essentially says that heat input cannot be converted completely into useful work.

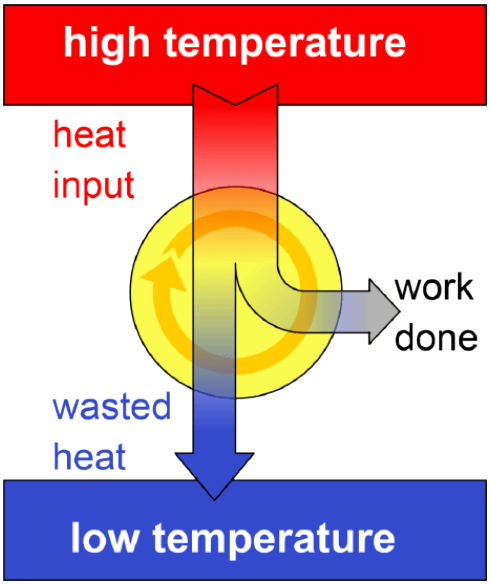


Figure 18.2 – In a heat engine, useful work can be done due to the difference in temperature between a hot and cold source. However, some heat will be wasted and lost in the process. Heat engine by Alyssa J. Pasquale, Ph.D., (adapted from original work by MikeRun) is licensed under CC BY-SA 4.0.

If the efficiency cannot be 100%, what is the best efficiency that we can have within the laws of physics? This problem was solved by Sadi Carnot in 1824. While the mathematics involved is beyond the scope of this textbook, the result of Carnot's discovery is that the highest efficiency attainable for any heat engine under ideal conditions (no friction or other losses such as turbulence or unwanted heat transfer) depends only on the absolute temperature of the hot and cold regions. This highest efficiency is called the Carnot efficiency and is equal to the temperature of the hot area minus the temperature of the cold area, all divided by the temperature of the hot area. In other words

$$e_{carnot} = \frac{T_h - T_c}{T_h}.$$

This equation can be restated as

$$e_{carnot} = 1 - \frac{T_c}{T_h}.$$

The temperature scale used in this equation must be an absolute scale: Kelvin.

If we look at the equation for Carnot efficiency, the only way the efficiency of a heat engine can be 100%, representing a perfectly efficient heat engine, is if the temperature of the cold area is zero: absolute zero. That isn't possible, so the efficiency of a heat engine will never be 100%. We can also see from the equation that efficiency can be maximized as much as possible by increasing the temperature difference between the hot and cold areas. If the temperature difference is small, the efficiency will be very low.

The heat engine demonstrating the model factory used hot water at a temperature of 98 ^{o}C (371 K) and cold water at a temperature of 10 ^{o}C (283 K). To calculate the Carnot efficiency of this heat engine, the temperatures must be expressed in units of kelvin. The Carnot efficiency is calculated below, and is rather small due to the very small difference between temperatures of

the hot and cold sources.

$$\begin{aligned} e_{carnot} &= \frac{371K - 283K}{371K} \\ &= 0.24 \\ &= 24\% \end{aligned}$$

In reality, all real heat engines operating between two given temperatures will always have an efficiency that is less than the Carnot efficiency. This is due to physical limitations of any engine or machine that come about due to friction or other losses.

Refrigerators

If you reverse all the arrows in the heat engine shown in Figure 18.2, you would do work on a system in order to draw heat from a low temperature region and then dump all that energy into a high temperature region (as shown in Figure 18.3). In other words, a heat engine operating in reverse is a refrigerator. A refrigerator is a device that uses work input from an external source to remove heat from an already cold region while operating in a cycle. All the energy that enters the system ($W_{input} + Q_{fromcoldregion}$) is then expelled into the high temperature region.

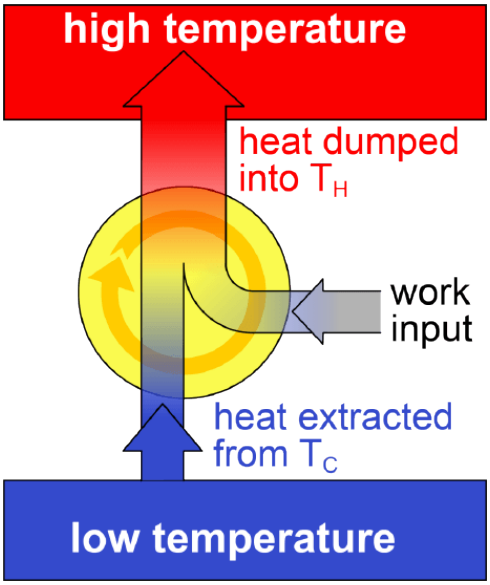


Figure 18.3 – In a refrigerator, work input can cause heat extracted from a cold region to be dumped into a hot region. Refrigerator by Alyssa J. Pasquale, Ph.D., (adapted from original work by MikeRun) is licensed under CC BY-SA 4.0.

A measure of the refrigerator's ability to do this is called the coefficient of performance (COP). The symbol for coefficient of performance is the lowercase Greek letter kappa (κ). Like the efficiency of a heat engine, COP is also defined as what you "get out" divided by what you "put in." In this case, however, what you get from the refrigerator is the heat extracted from the low temperature region. What you put in is the work you do to extract the heat from the low temperature region. This equation is

$$\kappa = \frac{Q_C}{W_{input}}.$$

A perfect refrigerator is one that would require no work to extract heat from the low temperature region. Of course, the net result of such a device would be a spontaneous flow of heat from a cold place to a hot place. As mentioned earlier, the second law of thermodynamics forbids this from happening. There are no perfect refrigerators. Put simply: your refrigerator will not work unless you plug it in!

As with the heat engine, there is an upper limit to how well a refrigerator operating between two temperatures can accomplish its task. The parameter that describes this is the Carnot coefficient of performance (COP). The formula for the Carnot COP is

$$\kappa_{carnot} = \frac{T_c}{T_h - T_c}.$$

All real refrigerators operating between two given temperatures will always have a COP that is less than the Carnot COP. This is

due to the physical limitations of any real machine (friction, drag, and other dissipative forces).

Throughout the ages, humans have been fascinated with the idea of perpetual motion machines. There have always been people who have tried to create a machine that will run forever without any input of energy. The first law of thermodynamics tells us that perpetual motion machines that create energy from nowhere are impossible. The second law of thermodynamics tells us that perpetual motion itself is impossible. For any machine to operate in a closed cyclical loop forever and ever, energy must be continually fed into the system to keep it moving.

Entropy

Entropy describes the amount of disorder in a system. When we consider a system, entropy is related to the number of different states that the system can be in. A simple example of a small system is a single coin. The coin can be either in its heads or tails configuration: two states (Figure 18.4).



Figure 18.4 – A single coin has two possible states: heads and tails. States of one coin by Alyssa J. Pasquale, Ph.D., (adapted from US One Cent Coin, public domain, US Mint) is licensed under CC BY-NC-SA 4.0.

When a second coin is introduced into the system, there are now four different states the system can be in: both tails, both heads, and two states consisting of one heads and one tails (Figure 18.5).

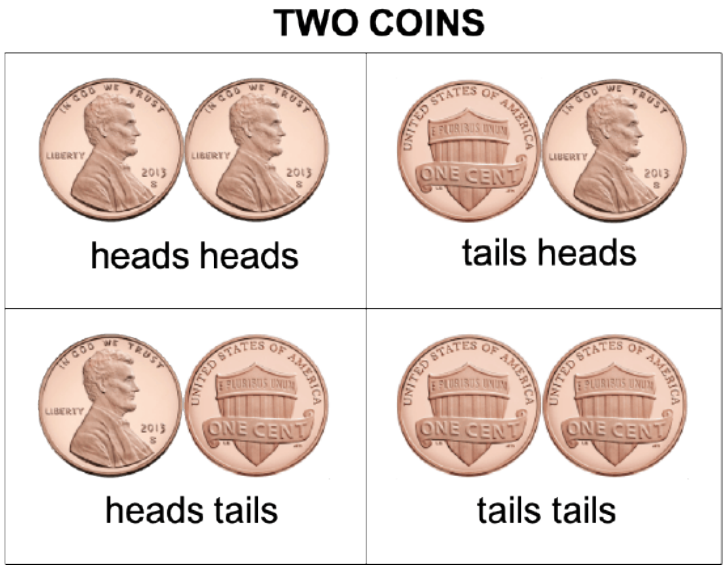


Figure 18.5 – Two coins have four possible states: heads/heads, heads/tails, tails/heads, and tails/tails.States of two coins by Alyssa J. Pasquale, Ph.D., (adapted from US One Cent Coin, public domain, US Mint) is licensed under CC BY-NC-SA 4.0.

The more coins that are introduced into the system, the more possible states there are.

A disordered system of multiple coins would be composed of a random arrangement of heads and tails among all the coins. A highly ordered system would be one where all of the coins are the same: all heads up, or all tails up. When the system consists of a single coin, the probability of getting all heads is 50%. When the system consists of two coins, the odds of having all heads is 25%. The more coins that are introduced into the system, the harder it is to randomly toss all of the coins in the air and have them all come up as heads. The more coins there are, the more disordered states there are compared to highly ordered states. If there are ten coins, for example, there are 1024 different states, but only one corresponds to all heads, and only one corresponds to all tails. More likely, there will be roughly equal numbers of heads and tails: disorder.

Most systems are much more complicated and have many, many more states than a coin. The video below shows a demonstration of different states of a beaker filled with colored cubes (purple and green). Before recording the video, Dr. Pasquale took all of the cubes out of the jar, manually sorted them, and put all of the green cubes on the bottom and all of the purple cubes on top (Figure 18.6, left). This is a highly ordered state of the system.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When Dr. Pasquale shakes the beaker, the cubes become arranged in a random manner and go into one of the many possible disordered states (Figure 18.6, right). Theoretically, it's possible that she could continue to shake this jar for a very long time and get back to the highly ordered arrangement of green on bottom and purple on top. But it's extremely unlikely due to the huge number of possible states. It is easy to increase the entropy of the system, but in order to decrease the entropy, Dr. Pasquale has to manually work to organize the cubes.

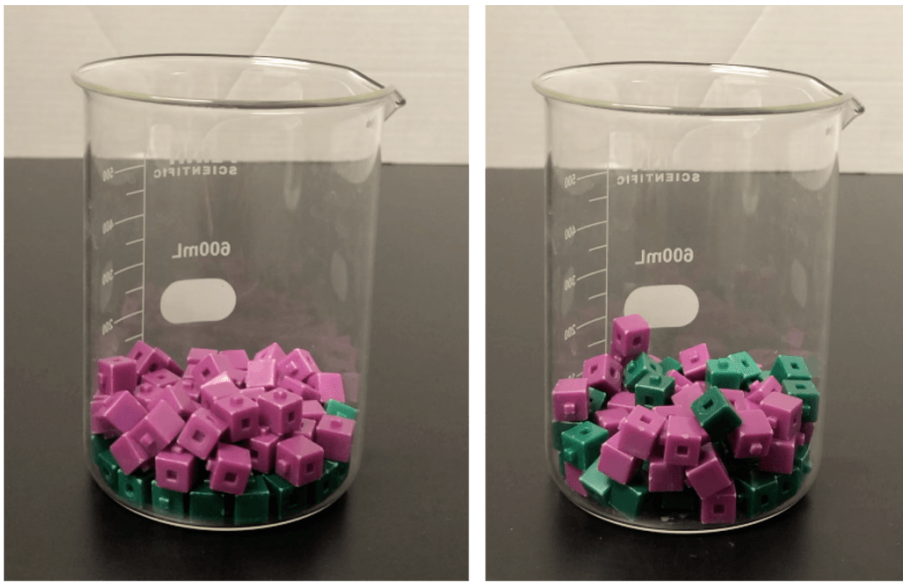


Figure 18.6 – A beaker filled with cubes has been manually arranged into a highly ordered state with all of the cubes sorted by color (left). After shaking the beaker, the cubes are in a highly disordered state (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale has a beaker of water and a container of blue food dye. There are many molecules of water in the beaker. When she places a few drops of food dye into the beaker, the two different substances, water and dye, are still relatively ordered. Most of the dye is in one location, and the water is still clear. Over time, however, the dye distributes throughout the water and the entropy increases. When Dr. Pasquale stirs the water, the dye distributes completely.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

How would it be possible to get the water to become clear again (in other words, how could the entropy of the system be decreased)? How can the dye be separated from the water? This will not happen naturally. No amount of stirring or random motion will separate the blue food dye from the water. To decrease the entropy of the beaker of water, it would be necessary to distill the water or filter it out, manually putting energy and effort into the process.

One thing that the second law of thermodynamics tells us is that entropy increases with time. We saw that with the shaken cubes, and the food dye in the water. If you put effort into decreasing entropy somewhere, for example, by organizing the cubes, you yourself create heat from your exertions and increase the entropy of your surrounding environment. A refrigerator, which acts to decrease the entropy inside the fridge and cool down food, creates heat and actually causes an increase in entropy elsewhere. This is one of the reasons why perpetual motion is impossible. There is no way to decrease entropy in one spot without increasing it more somewhere else. We truly cannot get something for nothing!

By now, you should begin to see why a system left alone tends towards increasing disorder: with any given state being equally as probable as any other state, there are more ways to have disorder than there are to have order. Consider your surrounding environment. Think about how many molecules of air surround you, and how many different states they can be in. Most systems in our lives have many states they can take, and the highly ordered states only represent a very miniscule fraction of those possible states. Entropy is all around us.

We can consider entropy to describe the arrow of time. Some processes in physics are symmetric. That is, you could play a video of Dr. Pasquale walking around a room, and you would not necessarily be able to tell if the video was being played in forward or reverse. Dr. Pasquale can walk backward just as easily as she can walk forward. However, if you were to watch the video below, you would know that it is played backward. Why? Because you know that it is impossible to shake a can and obtain a more ordered state! Entropy describes the arrow of time. While it is possible to decrease entropy in some part of the universe, doing so will increase the entropy in another part of the universe by an amount that is greater than the decrease. Entropy of the universe always increases with time.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Further reading

- Fire syringe explanation – This website goes into a “deep dive” of the physics behind the fire syringe. The mathematics used is outside the scope of this textbook, but as an overview the website gives a lot of information into this particular demo and adiabatic processes.

Practice questions

Conceptual comprehension

1. Why do you think the Celsius scale isn't used to calculate Carnot efficiency?

Numerical analysis

1. A gas in a piston-cylinder system absorbs 200 J of heat while doing 150 J of work on the surroundings. Calculate the change in internal energy of the gas.
2. A chemical reaction releases 5000 J of energy as heat while doing 3000 J of work on the surroundings. Calculate the change in internal energy of the system.
3. A closed container of gas has an initial internal energy of 4000 J. Heat is added to the gas, increasing its internal energy to 6000 J while the gas does 2000 J of work on the surroundings. Calculate how much heat was added to the gas.
4. A heat engine operates between two reservoirs at temperatures of 500 ^{o}C and 100 ^{o}C. Calculate the Carnot efficiency of this engine.
5. A steam engine has a hot reservoir at 300 ^{o}C and a cold reservoir at 30 ^{o}C. Calculate the Carnot efficiency of this engine.
6. A heat pump is used to heat a building. It extracts heat from the outside air at -10 ^{o}C and delivers it to the inside at 40 ^{o}C. Calculate the Carnot efficiency of this heat pump.
7. A hypothetical engine operates between two reservoirs at temperatures of 800 K and 300 K. Calculate the Carnot efficiency of this engine.
8. A single coin has two states, heads and tails. Consider a system consisting of four coins. Calculate...
    1. ...the number of states of that system.
    2. ...the percentage of highly ordered states (consisting of all heads or all tails) in that system.
9. A single die has six states, corresponding to each face of the cube. Consider a system consisting of two dice. Calculate...
    1. ...the number of states of that system.
    2. ...the percentage of highly ordered states (consisting of both dice showing the same number) in that system.
10. Consider a system consisting of three dice. Calculate...
    1. ...the number of states of that system.
    2. ...the percentage of highly ordered states (consisting of all three dice showing the same number) in that system.

Hands-on experiments

1. Take two different liquids with distinct colors, such as water and colored water or milk, and gently pour them into a clear container. Initially, the two liquids will have a clear interface, but over time, they will mix and become one uniform color. Discuss how this demonstrates the increase in entropy as the liquids mix and achieve a more disordered state.

This page titled Chapter 18: Thermodynamics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

V: SOUND AND WAVES

Chapter 19: Vibrations and Waves
Chapter 20: Sound
Chapter 21: Musical sounds

This page titled V: Sound and waves is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 19: Vibrations and Waves

Summary

Many aspects of our lives are dictated by waves. Light waves are used to transmit Wi-Fi and cellular signals, sound waves are used to communicate with others, and even the motion of a pendulum that keeps time in a grandfather clock has wave properties. There are many periodic events in our lives that occur at repeatable time intervals: sunrises, moon phases, seasons, heartbeats, and more.

This chapter will discuss different waves and wave properties, and start a fascinating conversation about waves that will unlock the mysteries of sound and light in chapters to come.

Vibrations and waves

A vibration is a periodic back and forth motion that remains fixed in one location. Examples of vibrations include a swing moving back and forth (like a pendulum) or a mass bobbing up and down on a spring. The video below demonstrates this periodic back and forth motion using the example of a swing set. (While the swing moves back and forth, this periodic motion does not propagate through space. Therefore, it is an example of a vibration.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A wave is a traveling vibration that transfers energy from one place to another. There are many different types of waves: light waves, sound waves, water waves, gravitational waves, seismic waves, and more.

The term periodic oscillation refers to the types of motion that waves make. This is a repeating pattern that causes many wave properties. A wave is periodic motion that gives rise to a variation in some value (intensity, energy, pressure, or some other property) over time as the wave property propagates through space. The video below demonstrates both the time-varying and space-varying properties of a wave as it travels. It shows that, at a single point in space, the wave properties vary with time. It is also possible to see how a wave varies in space at a single snapshot in time.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In the video above, and in many types of periodic motion, waves may look like a “squiggle” that moves up and down in space and time. Formally, this particular “squiggle” shape, shown in Figure 19.1, is known as a sinusoid. This comes from the trigonometric

function “sine.” While we will not make use of any trigonometric functions in this textbook, the shape of a sine (or sinusoidal, meaning “sine-like”) wave appears often enough to informally define it here. The wave shown in Figure 19.1 depicts the value of a wave as it moves through space at a single snapshot in time.

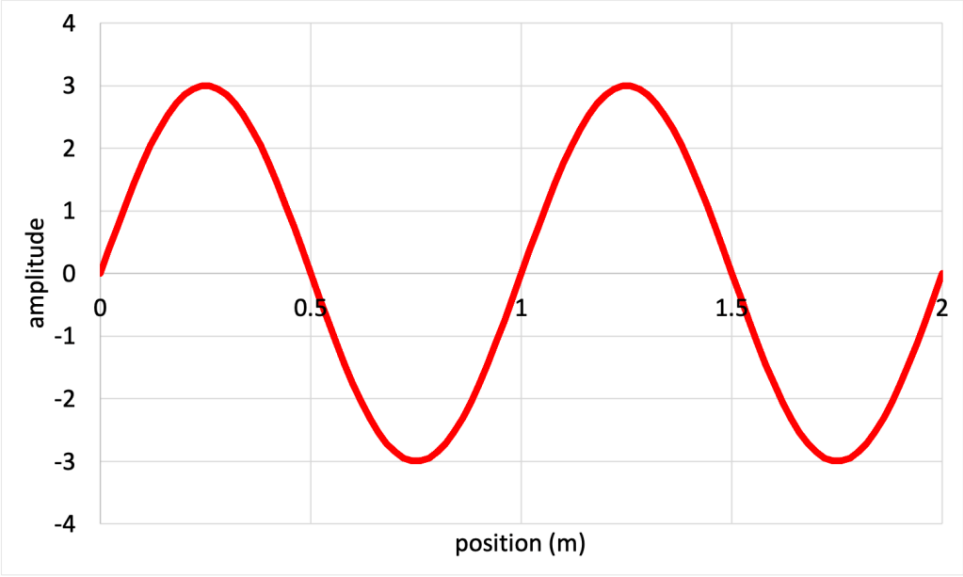


Figure 19.1 – An example of a wave where the amplitude is graphed with relation to position. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Each of the high points on the wave is known as a peak or crest. Each low point of the wave is a valley or trough. The center position is known as the equilibrium value of the medium. This would show us the undisturbed value of the property (energy, intensity, pressure, or another property) of the medium; in other words, the value that a particular property of the medium would have if it was at rest.

Wave properties

In order to characterize a wave, we need to be familiar with the terms used to describe waves. Wave amplitude describes the maximum displacement from the equilibrium position. In other words: how far away from the equilibrium position are the crests and troughs? Depending on the wave, the amplitude could describe energy, intensity, or even just height. (For example, the amplitude of a mass bobbing up and down on a spring could represent length. The amplitude of a swinging pendulum may represent degree from the vertical.) So we will not necessarily define one particular unit for amplitude, as it could be describing one of many physical parameters.

The video below shows an animation of a wave where the amplitude gradually increases over time.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The wavelength of a wave describes the shortest distance between two identical repeating points on a wave. Wavelength is most easily measured from crest to crest (or from trough to trough) on a wave. More precisely, wavelength is the distance between subsequent repeating points on the wave (where the wave has the same value **and** the same slope). The requirement of the same slope is important, otherwise if two adjacent points with just the same value are selected (without paying attention to slope), a value less than the wavelength will be measured. (This is why it is simpler to measure wavelength from crest to crest or from

trough to trough.) Figure 19.2 shows a few possible ways to measure wavelength using this definition.

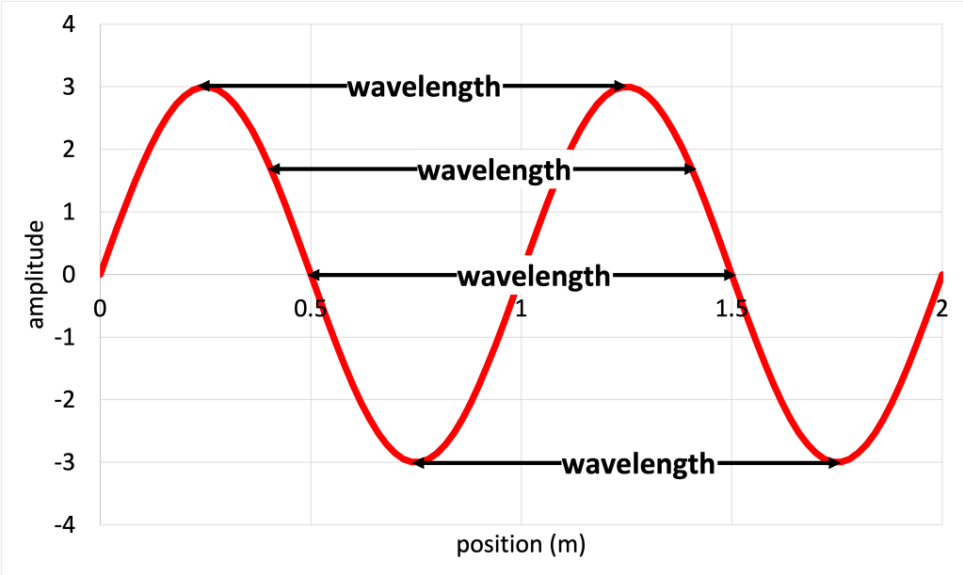


Figure 19.2 – An example of a wave that shows five examples of measuring the wavelength. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The symbol for wavelength is the lowercase Greek letter lambda (λ). It has units of distance: meters or centimeters.

The video below shows an animation of a wave where the distance between adjacent wave crests gradually decreases over time.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Figure 19.1 was annotated to include the crest, trough, equilibrium, amplitude, and wavelength. This is shown in Figure 19.3.

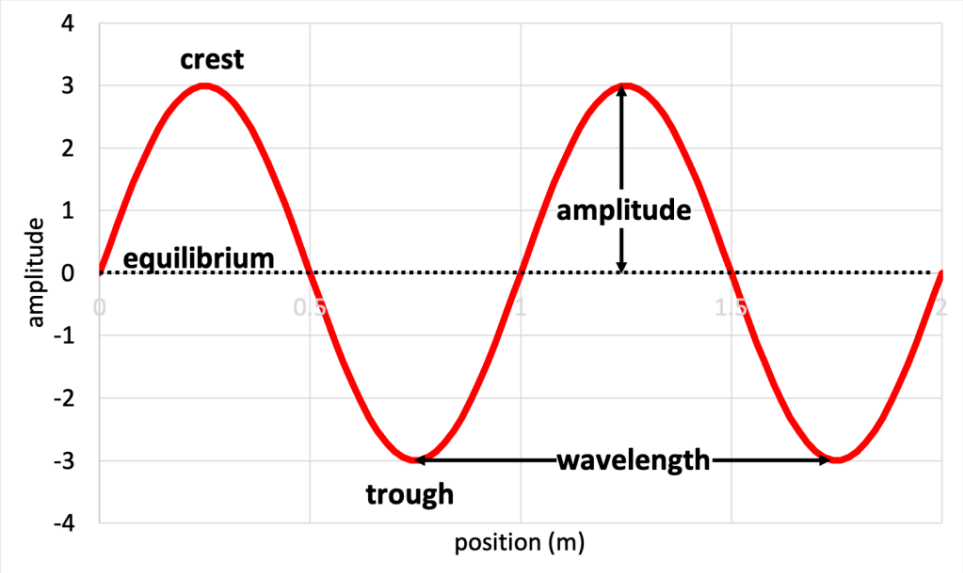


Figure 19.3 – A graph of wave amplitude with respect to position has been marked up to indicate the position of the crest, trough, and equilibrium. The amplitude and wavelength of the wave are also depicted. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The period of a wave describes how long it takes for one complete oscillation (out and back motion) to occur. An oscillation, simply put, is a complete motion from crest to crest or from trough to trough. (More accurately, an oscillation is the time it takes for a particle to return to its initial position and be moving in the same direction. As with wavelength, it is important not to ignore the second requirement in this definition!) In a pendulum, an oscillation relates to the amount of time it takes for the pendulum to swing out and back again. The symbol for period is the capital letter T, and the unit is seconds.

Frequency describes how many wave oscillations occur per second. Frequency is represented by the symbol f and is measured in the unit hertz (Hz). One hertz is equal to one oscillation divided by a second. Frequency and period are related by the equation

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

In other words, frequency is equal to one divided by the period. The higher the frequency of oscillation, the faster the wave squiggles in time, and the shorter each period will be.

Wave speed describes the speed at which a wave moves. We know that speed is equal to distance divided by time. In this case, the distance described by the wave is the wavelength, and the time is described by the period. This means that the wave speed is equal to wavelength divided by the period. In equation form,

$$v = \frac{\lambda}{T}.$$

Or, we can rewrite the equation to say that wave speed is equal to wavelength times frequency,

$$v = \lambda f.$$

If wave speed is constant, then this equation can tell us the relationship between wavelength and frequency.

Types of waves

Waves tend to be characterized by the motion that the wave itself makes compared to how the wave travels through space.

Longitudinal waves

In a longitudinal wave, the oscillations (back and forth motion) of the wave travel in the same direction as the wave itself. The video below shows an animation of many particles oscillating in a longitudinal wave. The motion of the particles is parallel to the

motion of the wave itself.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Sound waves are examples of longitudinal waves. This type of wave will be explored in more detail in the next chapter of this textbook.

Transverse waves

In a transverse wave, the vibrations of the wave move perpendicularly to the direction of motion of the wave. In the video below, a transverse wave is animated as it travels through time and space. A red point on the wave demonstrates the motion of oscillation, which is perpendicular to the direction that the wave propagates.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Water waves

There are also some waves that are neither purely longitudinal nor purely transverse. Water waves in the ocean, for example, are usually a mixture of both longitudinal and transverse. There is both up-and-down and side-to-side motion in ocean waves.

The video below shows an animation of a water wave. Each point in the wave can be seen moving with both parallel and perpendicular motion compared to the motion of the wave itself.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Traffic waves

It might be strange to think of traffic as a wave, but you should think about it next time you're at a traffic light. At a red light, cars stop and line up. When the light turns green, the motion of the cars will be forward, but the direction that the information in the wave moves is backward. The first car in line moves forward, and then the car behind that one moves, and the car behind that one, and so on.

The video below is an animation of a simulation of cars (each one is represented as a square) stopped at a red light. When the light turns green, each car subsequently starts moving. When a car is red, that means it has started moving. The position of the red square, representing the information transmitted by the wave, moves backward, even though the cars move forward.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Torsional waves

Another type of wave is known as a torsional wave. This is a type of twisting back and forth in a periodic manner. The video below shows a mass twisting back and forth in a torsional wave.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A torsional wave is what caused the Tacoma Narrows Bridge to collapse in 1940. Winds blowing across the bridge created and amplified the torsional wave in the bridge, which eventually overcame the strength of the structure, causing its collapse.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Interference

We know that two objects of matter cannot occupy the same space at the same time. Waves are energy, not matter, so they can exist in the same place at the same time. When waves overlap with each other they interfere. That is, the value of one wave at that point adds to the value of the second wave. The total amount of energy present is equal to the sum of those two values.

If two waves interfere where the crest overlaps with crest, and trough overlaps with trough, then the situation is called totally constructive interference. The result is a wave with greater overall amplitude than either one individually. If the two waves have identical amplitudes, the constructive interference will yield a wave with twice the amplitude. The video below shows an animation of two wave pulses interfering constructively.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

If two waves interfere where crest overlaps with trough, and trough overlaps with crest, then the situation is called totally destructive interference. If the two waves have the same amplitude, then if we sum up the contributions of each wave, they cancel out. The result is a wave with zero amplitude. This condition of destruction is temporary: as the waves continue to propagate, as they stop overlapping with each other, the interference will end. The video below shows an animation of two wave pulses interfering destructively.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

It is also possible to have many other types of interference that are neither purely constructive nor purely destructive (the interference may be “partially constructive” or “partially destructive”). The video below shows an animation of the interference of two waves, which is neither necessarily constructive nor destructive.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Interference happens all the time: when pebbles are dropped in a pond, when light waves interact, and when sound waves come together. The interference of light waves was used to prove that the speed of light is constant, which was one of the foundational experiments of the 19th century.

Standing waves

Standing waves are waves that oscillate in time but that do not move at all in space. They occur due to interference between two waves that have equal amplitude and frequency that travel in opposite directions.

Two fixed ends

In the video below, Dr. Pasquale generates standing waves in a coiled rope by shaking one end up and down. Both ends of the coiled rope are held firmly into place. These are known as fixed ends as they are physically unable to move. This creates one wave, which reflects off of the back end of the coiled rope and travels back toward her arm. By shaking at the right frequency, the two waves: the forward wave and the backward wave: will interfere just right to create a standing wave.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Locations on a standing wave that are fixed are known as nodes. Locations where the wave oscillates between minimum and maximum values are called antinodes. These are depicted in Figure 19.4.

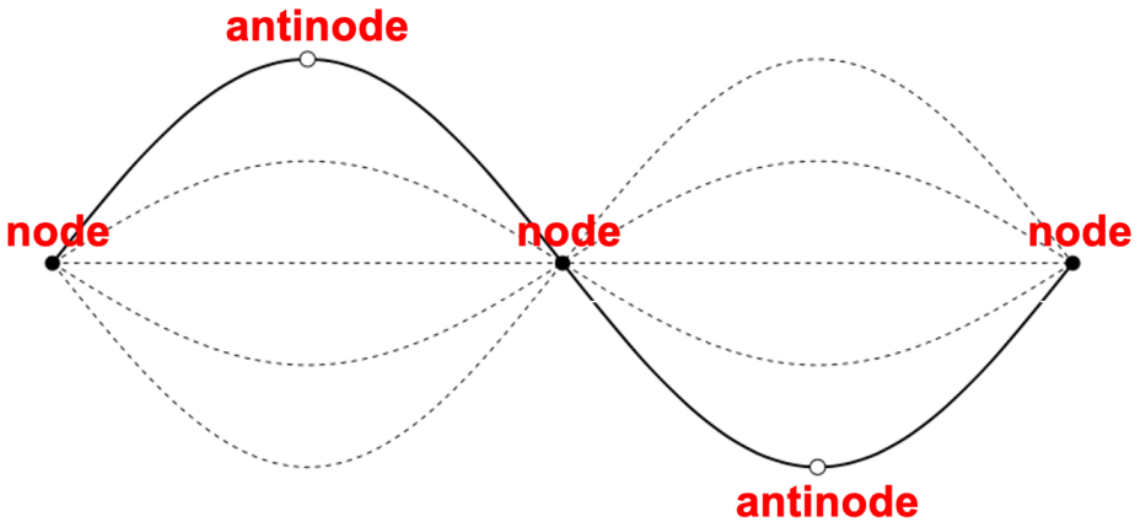


Figure 19.4 – The nodes and antinodes of a standing wave are labeled. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Each standing wave that can be generated is known as a harmonic. When there are two fixed ends, each end of the standing wave is a node. The first harmonic is the wave with one antinode and two nodes. The first harmonic is also sometimes referred to as the fundamental mode. The first harmonic represents one half of a wavelength of the full wave (in other words: the distance between two successive nodes is half of a full wavelength).

The second harmonic has two antinodes and three nodes as shown in Figure 19.4. The second harmonic represents an entire wavelength of the full wave. The third harmonic has three antinodes and four nodes, and represents one and a half wavelengths. The information for the first six harmonics (with two fixed ends) is given in the table below.

1	1	2	0.5
2	2	3	1

3	3	4	1.5
4	4	5	2
5	5	6	2.5
6	6	7	3

The first six harmonics are shown graphically in Figure 19.5.

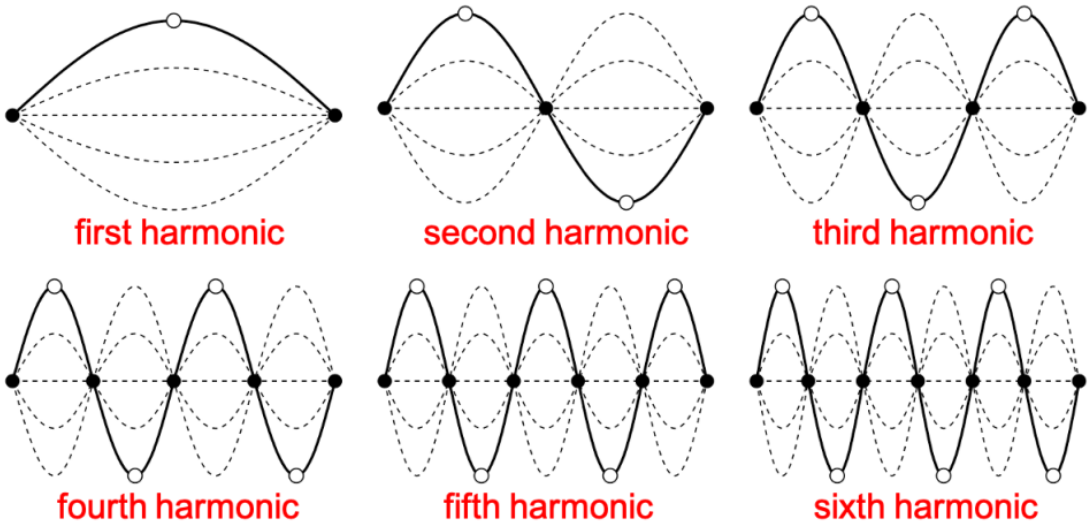


Figure 19.5 – The first six harmonics of a standing wave with two fixed ends. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

For a given medium (for example, the coiled rope used in the video above), the wave speed of a standing wave will be constant. Because the wavelength of each standing wave is related to the harmonic, the frequency will also be related to the harmonic number.

As the harmonic number increases, so does the frequency. In fact, the frequency of a harmonic is related to the number of the harmonic and the frequency of the first harmonic. In equation form: the frequency of the nth harmonic is equal to n, the harmonic number, times the frequency of the first harmonic:

$$f_n = nf_1,$$

where f_1 is the frequency of the first harmonic.

Non-fixed ends

It is possible for a standing wave to have, for example, one fixed end (a node), and one end that is free to move up and down. (It

is assumed that the ring slides up and down on the post with negligible friction.) This is shown in Figure 19.6.

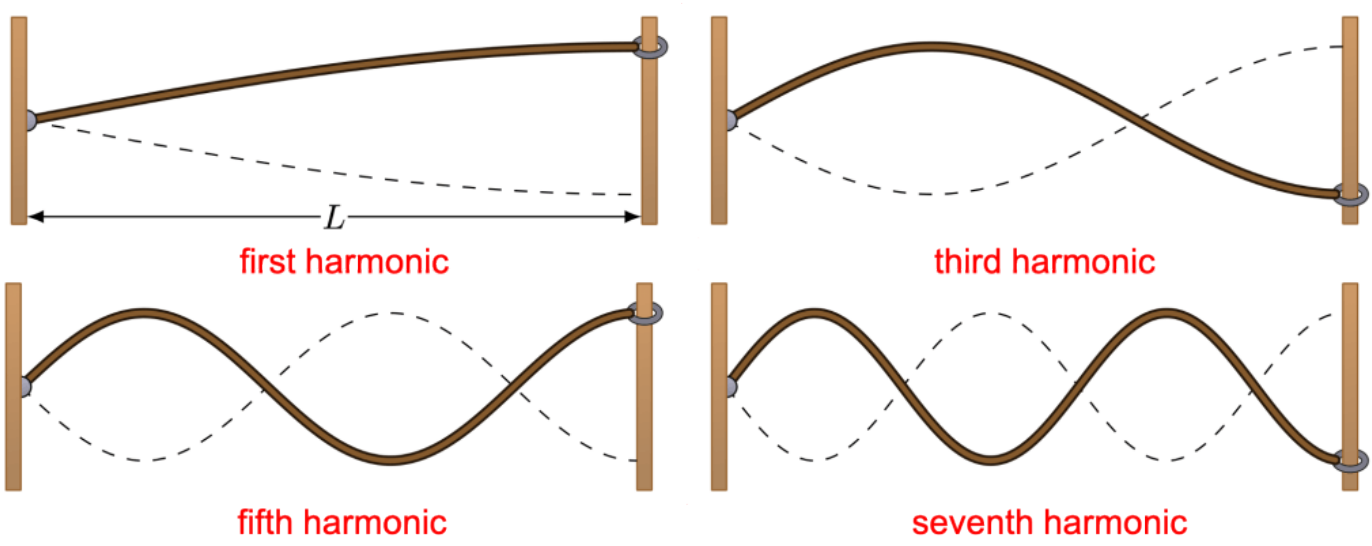


Figure 19.6 – Standing waves with one fixed and one free end. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

In this case, the fixed end of the standing wave is a node, and the free end of the standing wave is an antinode. Whereas standing waves with two fixed ends support multiples of half wavelengths, standing waves with one open end only support multiples of odd fourths of wavelengths. Because of this configuration, only odd harmonics are generated. The information for the first six harmonics (with one fixed end) is given in the table below.

1	1	1	0.25
3	2	2	0.75
5	3	3	1.25
7	4	4	1.75
9	5	5	2.25
11	6	6	2.75

From this table, we can see that in each length L between supports of the rope, there are $n/4$ wavelengths present, where n is the harmonic number. Solving for wavelength, we can see that

$$\lambda_n = \frac{4L}{n}.$$

As discussed with standing waves with two fixed ends, the wave speed of a standing wave will be constant (assuming the medium stays the same). Because the wavelength of each standing wave is related to the harmonic, the frequency will also be related to the harmonic number. In fact, the same relationship between the harmonic frequency and first harmonic exists with standing waves with one fixed end! That is,

$$f_n = nf_1,$$

where f_1 is the frequency of the first harmonic.

Just as it is possible to have transverse standing waves (depicted in Figures 19.4, 19.5, and 19.6), it is also possible to create longitudinal standing waves (for example, from sound waves). As with transverse standing waves, when a longitudinal wave is created in a cavity (such as a pipe or a musical instrument), any time there is a closed end, there will be a node of the longitudinal wave. Any time there is an open end, there will be an antinode of the longitudinal wave. This is how sound waves are generated in many musical instruments.

The first three harmonics of a longitudinal standing wave generated in a pipe with one closed and one open end are shown in Figure 19.7. The pipe is filled with water (depicted in gray in the figure), and the water level creates a closed end. The black dots in this figure represent air molecules. At the locations without the arrows, the molecules essentially do not move; these are the standing wave nodes. The dots shown with the arrows indicate the locations where the molecules vibrate with their greatest amplitude; these are the standing wave antinodes. The solid arrows indicate that air molecules at those points are at a maximum at one snapshot in time (similar to the solid lines shown in Figure 19.5 and Figure 19.6). Note that two successive antinodes will have the particles on opposite sides of the equilibrium position. The dashed arrows show the same thing half a period later (similar to the dashed lines shown in Figure 19.5 and Figure 19.6).

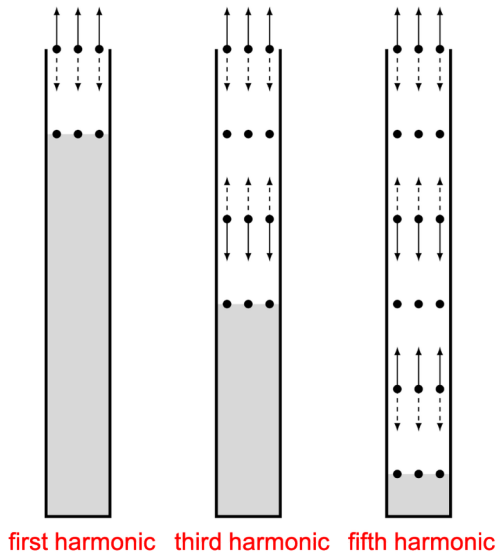


Figure 19.7 – The first three harmonics of a longitudinal standing wave generated in a pipe with one closed and one open end. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In this particular figure (Figure 19.7), the wavelength of each wave remains fixed; the length (distance between the open end of the pipe and water level) is changed to support the standing wave. If the wavelength of the standing wave is known, the length can be calculated as

$$L_n = \frac{n\lambda}{4}.$$

The Doppler effect

Have you ever noticed that when an ambulance or police car drives by, the pitch of the sound of the siren seems to change? This

is due to the Doppler effect. The Doppler effect is the perceived change in frequency of a sound wave as heard by an observer when either the source of the sound or the observer moves through the medium.

When a wave is traveling at a certain speed relative to an observer, the wavefronts become bunched up in the direction of motion. This is because the source moves toward the previously emitted pulse before emitting the next pulse. This shortens the wavelength and (since the wave speed is constant) increases the frequency. Frequency and pitch are related in sound waves, which is why a siren sounds higher in pitch when an ambulance is approaching your location.

In a similar manner, as a wave travels away from the observer, the wavefronts are spread farther apart. This increases the wavelength and decreases the frequency. This means that a siren sounds lower in pitch when it is moving away from your location.

An animation of the Doppler effect is shown in the video below. The wavefronts are compressed in the direction of travel, and spread out in the direction opposite travel.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Figure 19.8 (left) shows the wavefronts of a stationary wave, where all wavelengths are equal in all directions. Figure 19.8 (right) shows the wavefronts of a moving wave, where the Doppler shift is evident.

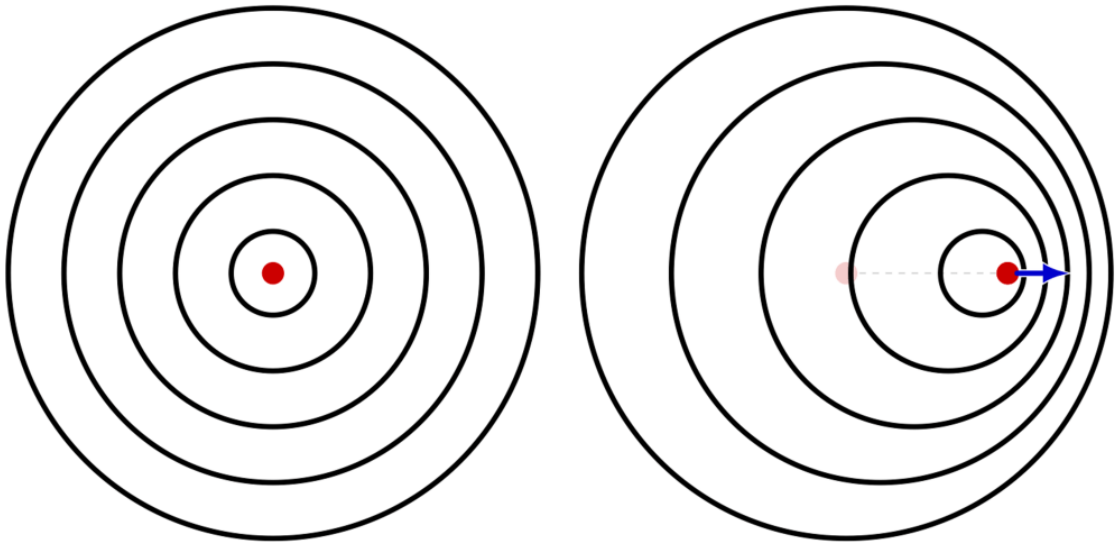


Figure 19.8 – A stationary wave generates wavefronts of equal wavelength and frequency (left). A moving wave has a smaller wavelength (higher frequency) in the direction of motion and longer wavelength (lower frequency) in the direction opposing motion (right). This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

If, instead of the wave (source) moving relative to the observer, the observer moves relative to a stationary source, the results of the Doppler effect are qualitatively the same (a change in perceived pitch of the sound), but are not numerically the same. There is an equation that can be used to determine the exact numerical change in perceived frequency of the sound wave:

$$f_{new} = f_{original}\frac{v \pm v_{observer}}{v \mp v_{source}},$$

where f_{new} is the perceived frequency heard by the observer, $f_{original}$ is the original frequency emitted by the source, v is the speed of sound at that temperature and in that medium (the speed of sound will be formalized in the next chapter of this textbook),

$v_{observer}$ is the speed of the observer with respect to the medium, and v_{source} is the speed of the wave source with respect to the medium.

While this equation may look complicated, it is really just a compact way of dealing with all of the possible ways (in a one-dimensional scenario) that the source and observer can move relative to each other and through the medium. Note that when you are using this equation, if either the source or the observer approaches the other, the frequency f_{new} increases. If the source or observer moves away from each other, f_{new} decreases.

All of the different scenarios including this Doppler equation are outlined in the table below.

observer moves toward a stationary source	$f_{new} = f_{original}\frac{v+v_{observer}}{v}$
observer moves away from a stationary source	$f_{new} = f_{original}\frac{v-v_{observer}}{v}$
source moves toward a stationary observer	$f_{new} = f_{original}\frac{v}{v-v_{source}}$
source moves away from a stationary observer	$f_{new} = f_{original}\frac{v}{v+v_{source}}$
source and observer both move toward each other	$f_{new} = f_{original}\frac{v+v_{observer}}{v-v_{source}}$
source and observer both move away from each other	$f_{new} = f_{original}\frac{v-v_{observer}}{v+v_{source}}$
source chasing the observer	$f_{new} = f_{original}\frac{v-v_{observer}}{v-v_{source}}$
observer chasing the source	$f_{new} = f_{original}\frac{v+v_{observer}}{v+v_{source}}$

Depending on the wave, the speed of motion can have profound applications due to the Doppler effect causing the wavefronts to bunch together. In a boat, water is the physical medium in which waves propagate. If we drop a pebble into a pond, the ripple pattern is even and uniform.

If, however, we continuously drop pebbles while in motion, the Doppler effect will cause the wavefronts to bunch together. At a certain point, when the pebble moves at the same speed as the speed of the water waves, a wake will form. This happens with boats, which generate bow and stern waves when they travel across water faster than the disturbance they create would travel across the water. Figure 19.9 shows V-shaped waves called bow waves generated by a barge and several ducks.

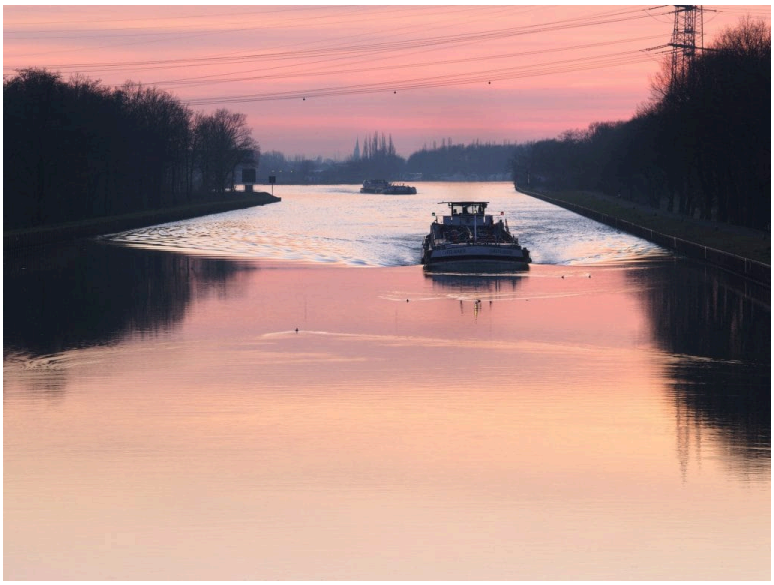


Figure 19.9 – A bow wave generated by a barge (and smaller bow waves generated by ducks). Bow wave of a canal barge, by Arnold Paul, is licensed under CC BY-SA 2.5.

For sound waves, when the speed of an object travels faster than the speed of sound, this disturbance creates a sonic boom. This is a shockwave created due to the overlapping of wavefronts at a speed faster than the air can meaningfully respond to the changes in pressure caused by the wave. This compresses the air and creates the shock wave we call a sonic boom.

A sonic boom travels in a three-dimensional cone shape whose angle corresponds to the exact speed of the supersonic object. The faster the motion of the object, the smaller the angle of the cone becomes. This is demonstrated in Figure 19.10, where an object is depicted traveling at exactly the speed of sound (left), creating a sonic boom normal to the direction of motion. An object moving faster than the speed of sound (middle) has wavefronts that overlap at an angle, and an object moving even faster (right) has wavefronts that overlap at a smaller angle relative to the direction of motion of the object.

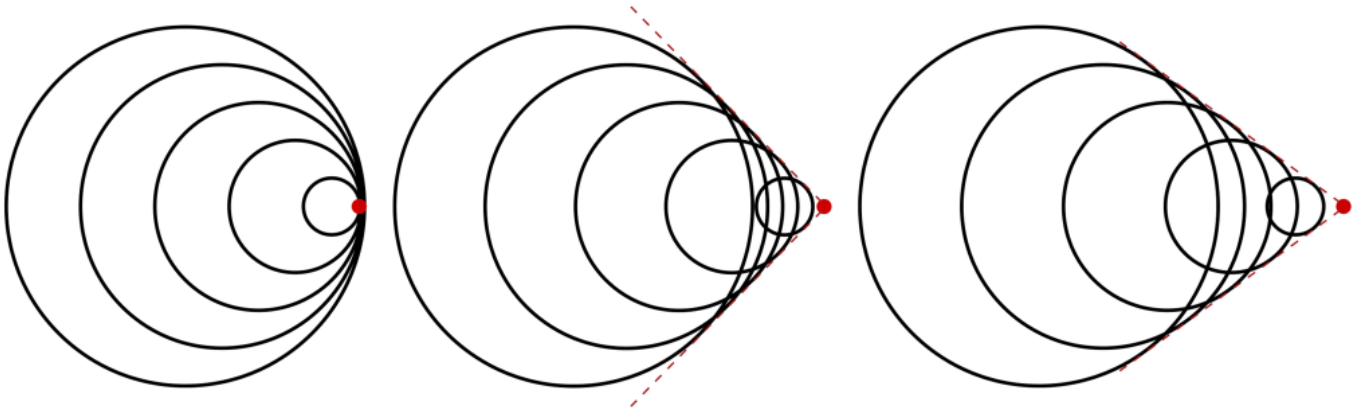


Figure 19.10 – An object moving at exactly the speed of sound creates wavefronts that overlap normal to the direction of motion (left). Objects that move faster than the speed of sound create wavefronts where the angle of the cone generated from the overlapping wavefronts becomes smaller as the speed increases (middle, right). This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

The shockwave generated by a space shuttle orbiter is shown in Figure 19.11.

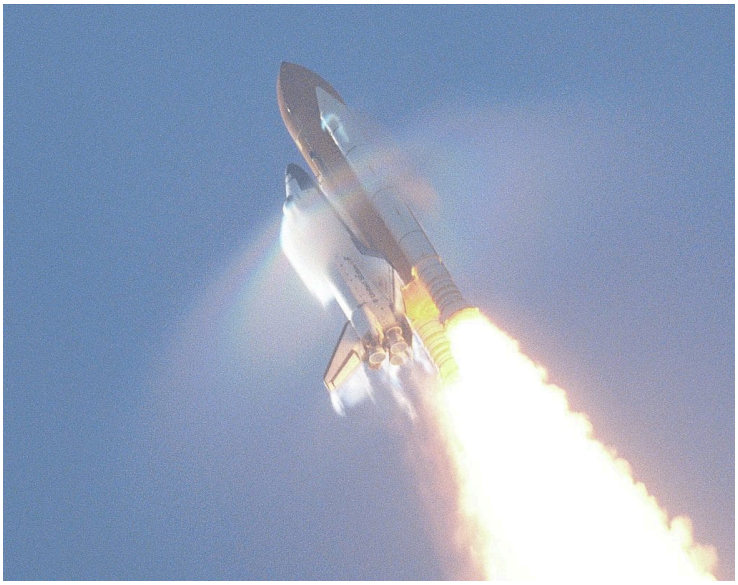


Figure 19.11 – A shockwave formed by the space shuttle orbiter taking off. KSC00pp1416, by NASA Kennedy Space Center, is licensed under Public Domain.

Further reading

- Acoustics and vibrations animations – This website contains animations of many different types of waves (longitudinal, transverse, etc.).
- Circular harmonics – Animations of circular harmonics, which are the waves that are created on a circular membrane, such as a drumhead.

Practice questions

Numerical analysis

1. Use the graph in Figure 19.12 to calculate the...
    1. ...amplitude.
    2. ...wavelength.
    3. ...wave speed, assuming the frequency is 4 Hz.

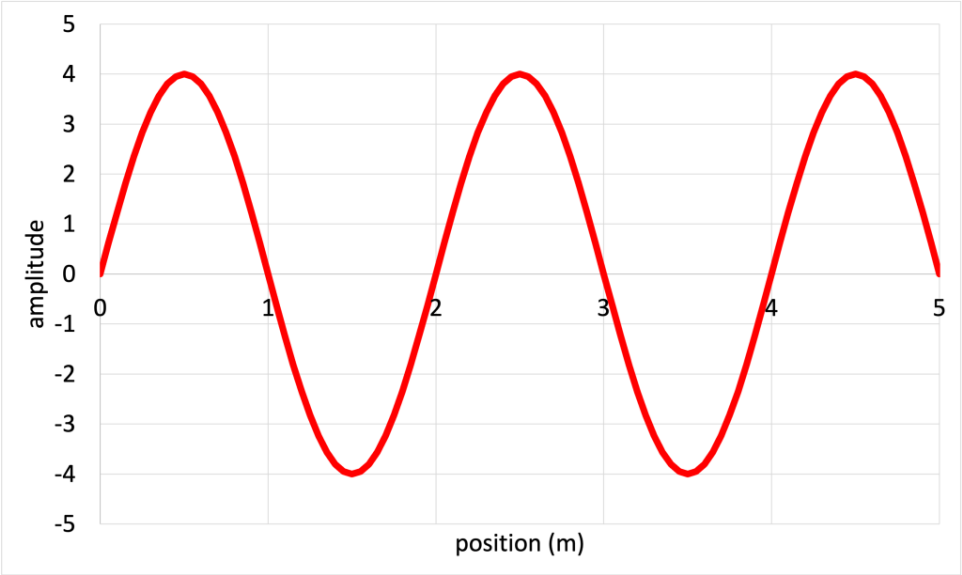


Figure 19.12 – A graph of a wave where the amplitude is graphed with relation to position. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

2. Calculate the frequency of a wave that completes 10 oscillations in 2 seconds.
3. If a wave has a period of 0.5 s, calculate its frequency.
4. A wave with a wavelength of 2 m travels through a medium at a speed of 20 m/s. Calculate the frequency of the wave.
5. If a wave has a frequency of 100 Hz and a wavelength of 0.1 m, calculate its wave speed.
6. A standing wave has a first harmonic frequency of 440 Hz. Calculate the frequency of the...
    1. ...second harmonic.
    2. ...third harmonic.
7. A standing wave has a third harmonic frequency of 900 Hz. Calculate the frequency of the first harmonic.
8. The first harmonic of a standing wave has a wavelength of 2 m. Calculate the length of the string used to generate the standing wave.
9. A standing wave on a guitar string that is 30 cm long has three nodes. Calculate...
    1. ...the harmonic of the standing wave.
    2. ...the wavelength of the wave.
    3. ...the frequency of the wave, assuming the wave speed is 340 m/s.
10. A standing wave on a rope that is 1.2 m long has five antinodes. Calculate...
    1. ...the harmonic of the standing wave.
    2. ...the wavelength of the wave.
    3. ...the frequency of the wave, assuming the wave speed is 8 m/s.

Hands-on experiments

1. If you have a Slinky™, use it to generate a longitudinal wave. Then use it to generate a transverse wave.
2. Fill a shallow tray or a baking dish with water. Drop small objects like coins or pebbles into the water to create ripples. Observe how these waves move, reflect off the edges, and interfere with each other. You can also use a small stick (or your finger) to generate waves by tapping the water's surface.

This page titled Chapter 19: Vibrations and Waves is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 20: Sound

Summary

Sound waves are vibrations of increasing and decreasing pressure that travel though some type of material medium: a solid, liquid, gas, or even plasma. Sound waves are of great interest in audio and music applications. Sound waves can also amplify energy and cause objects to resonate.

Sound waves

Sound waves are longitudinal waves that are caused by variations in air pressure between two points. Because sound waves are caused by pressure disturbances, they require a medium to travel through.

Speakers turn an electrical signal into vibrations. In the video below, Dr. Pasquale placed some small pieces of paper on top of a speaker, whose membrane vibrates back and forth. The frequency of the source controls how quickly these vibrations occur. At low frequencies, it is easy to see the pieces of paper moved around by the vibrations.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The frequency of a sound wave controls the pitch of that wave. In the video below, Dr. Pasquale has a sound source that creates a continuous tone of sound. In increments, she increases the frequency of the sound source. As the frequency increases, the pitch of the tone increases as well. In the demo, Dr. Pasquale used a microphone to capture the sound wave vibrations and show the pressure vs. time graph for the waves. As the frequency of the sound source increases, you can see the oscillations occur more and more frequently as the frequency increases. It is important to note that sound waves are longitudinal. The waves on these graphs appear as sine waves because they show the variation in pressure over time, not the actual vibrations in the air moving back and forth longitudinally.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The pressure vs. time graph for a low frequency (left) and high frequency (right) sound wave from the video are shown in Figure 20.1.

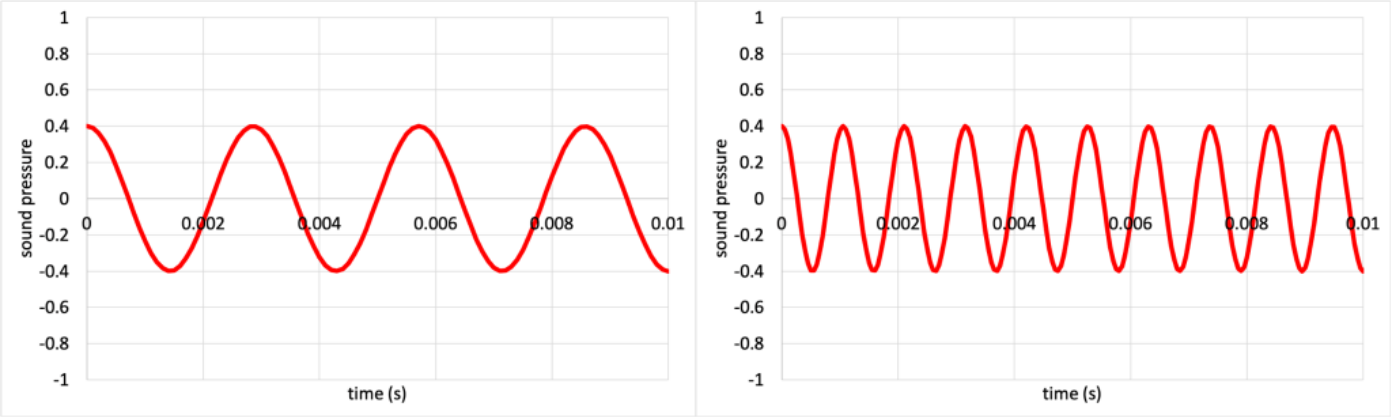


Figure 20.1 – A graph of sound pressure vs. time for a low frequency sound wave (left) and a high frequency sound wave (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The amplitude of a sound wave is proportional to the loudness, or volume, of that wave. In the video below, Dr. Pasquale uses a sound source that creates a continuous tone of sound. The frequency is kept constant, but the amplitude is increased for the first half of the video, then decreased for the second half of the video. As the amplitude increases, volume increases, and vice versa. The pressure vs. time graph shows the amplitude changing over time as the sound source is changed.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The pressure vs. time graph for a small amplitude (left) and large amplitude (right) sound wave from the video are shown in Figure 20.2.

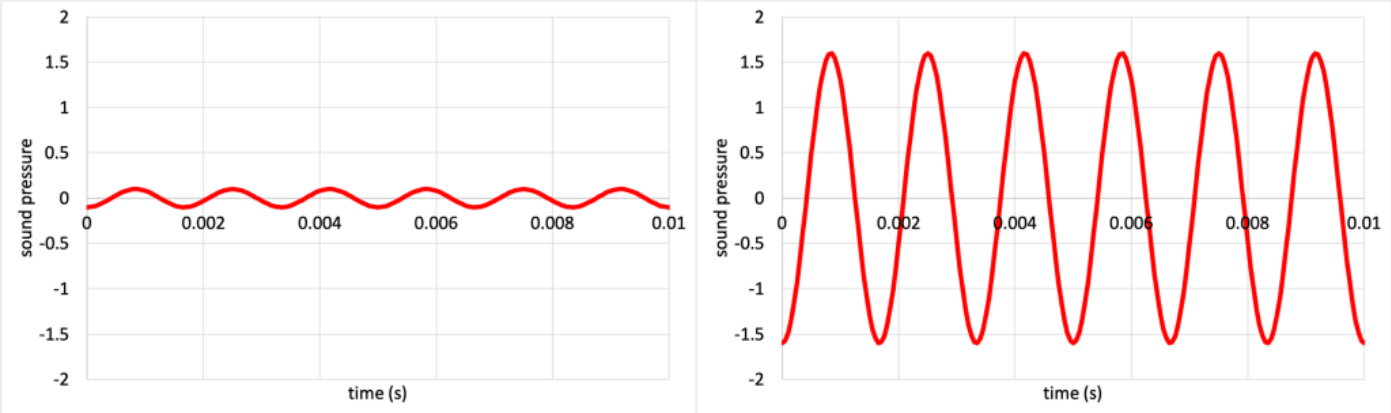


Figure 20.2 – A graph of sound pressure vs. time for a small amplitude sound wave (left) and a large amplitude sound wave (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Musical instruments

Musical instruments work by creating vibrations: sound waves. Different types of instruments create sound waves differently: by creating a vibration in a string, by blowing air across a reed or a mouthpiece, or by creating vibrations on a membrane stretched over a drum. The quality of the sound waves, known as timbre (pronounced TAM-ber), is produced by the materials, design, and quality of the instrument itself.

Large instruments (such as a tuba or a bass violin) are capable of supporting standing waves with a large wavelength. Small instruments (such as a piccolo or kazoo) support standing waves with small wavelengths. Figure 20.3 shows people playing large brass instruments (left) and a piccolo (right) for scale.



People playing a euphonium and tuba, which are large brass instruments (left) and a person playing the piccolo, a small woodwind instrument (right). This figure is adapted from Euphonium and tuba (left) by Elf (Public Domain) and Joueuse de piccolo (right) by Roman Bonnefoy (CC BY-SA 3.0).

Because of the relationship between frequency, wavelength, and wave speed, large instruments will create sound waves with low frequencies, and small instruments will create sound waves with high frequencies. This explains why large instruments create low-pitched sounds and why small instruments create high-pitched sounds.

Range of human hearing

Humans are capable of hearing sounds that range from very low pitch to very high pitch. Our range of hearing is approximately 20 Hz to 20,000 Hz. This of course varies from person to person, and changes as we age. Starting around 30 years old, the higher range of our hearing tends to degrade. This means that the older you are, the lower the frequency of your upper range of hearing will be. That means a high-pitched noise that bothers a young person may not even be audible to an adult.

Sound waves that have frequencies above 20,000 Hz are known as ultrasonic. Ultrasound is used in many applications in our lives. The motion detectors (Figure 20.4) that have been used to create many of the video demonstrations in this textbook use an ultrasonic wave that reflects off the nearest surface and bounces back to the detector. Ultrasound is also used frequently in medical diagnostics and imaging.

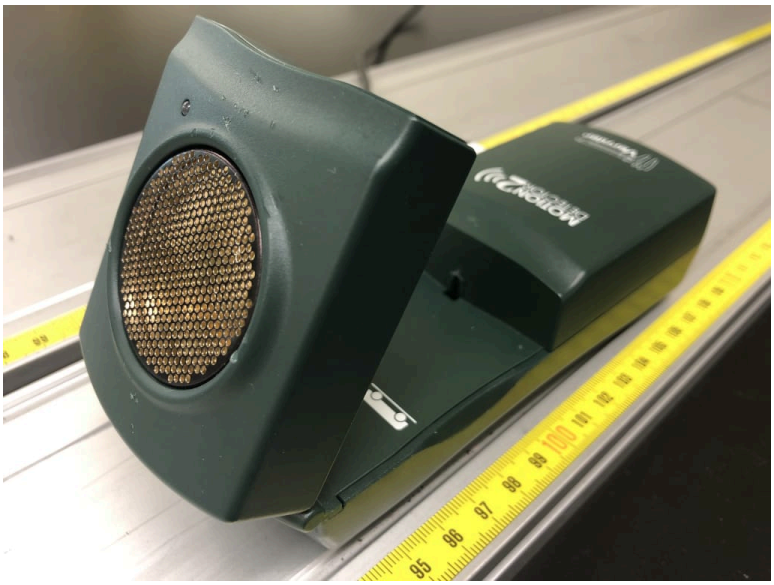


Figure 20.4 – This laboratory motion detector uses ultrasonic waves to measure distance. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Sound waves that have frequencies below 20 Hz are known as infrasonic. Many natural sources such as earthquakes, volcanoes, and waterfalls emit infrasound. Infrasound detectors can therefore be used to monitor these phenomena.

The speed of sound

How fast do sound waves travel? That's a simple question with a somewhat complicated answer. Because sound waves are pressure vibrations that move through a medium, the speed that those vibrations move is dependent on the medium. Lots of things can affect the speed of sound.

First, consider what happens to the speed of sound in different phases of matter. Sound waves travel when molecules bump into each other, transmitting energy from one place to another. If molecules are more rigid and spaced more closely together, then the speed of sound will be faster than if molecules are loosely connected and spaced far apart. This means that sound travels fastest through solids, where molecules are tightly packed, and most slowly through gases, where molecules are very far apart. The speed of sound through a liquid will be somewhere in between. Sound waves will not travel at all through a vacuum, as there are no molecules present to cause pressure vibrations.

Most of the time, humans experience sound waves as they move through air: a gas. So what is the speed of sound in air? There is no one answer to this question. The biggest factor that influences the speed of sound in air is the temperature. To a small degree, things such as humidity and wind speed can affect the speed of sound as well.

It is possible to use an equation to determine the speed of sound in air. This equation was experimentally derived, and air temperature is the independent variable. The equation states that the speed of sound at a given temperature is equal to

$$331\ \mathrm{m/s} + (0.6\ \mathrm{m/\left(s^{\circ}C\right)})\,T,$$

where T is the air temperature in degrees Celsius.

Room temperature is approximately 20 $^{\circ}$C. Using the equation to determine the speed of sound at that temperature:

$$\begin{aligned} v &= 331\ \mathrm{m/s} + (0.6\ \mathrm{m/\left(s^{\circ}C\right)})\,(20^{\circ}\mathrm{C}) \\ &= 331\ \mathrm{m/s} + 12\ \mathrm{m/s} \\ &= 343\ \mathrm{m/s}. \end{aligned}$$

The speed of sound is much, much slower than the speed of light. This is why we may see things that are far away and notice that the sound of the object is out of sync with what we see. For example, somebody playing basketball may dribble the ball, and the sound of the ball bouncing off of the ground will reach our ears after the ball actually bounces. The farther away something is, the more pronounced the effect will be.

Note that the speed of sound does **not** depend on the properties of the wave itself: frequency and amplitude do not affect the speed of sound. If you speak louder or softer, that does not change how long it will take the sound of your voice to reach a point a fixed distance away.

Reflection and refraction

When waves travel, they can change their motion and properties in different ways. One thing that can happen with waves is called reflection. This occurs when waves bounce off of a surface. If you've looked in a mirror, you've experienced the reflection of light waves. Sound waves can reflect as well. When discussing sound waves, a reflection is sometimes called an echo, especially when a sound wave reflects off of a surface that is far away, causing a delay between the initial sound source and the reflected wave.

The reflection of ultrasonic waves is frequently used to determine the distance between two objects.If the temperature is known, the speed of sound can be calculated, thus distance can be determined by using the velocity equals distance divided by time equation. The time will be how long it takes for a wave to leave an ultrasonic sensor, bounce off of an object, and return to the sensor. In fact, this is how the motion sensor shown in Figure 20.4 operates.

It is also possible for sound waves to refract, or bend. This happens when the speed of sound changes, for example, if there is a big temperature difference that a sound wave travels through. We will discuss the concept of refraction in more detail later in this textbook in the context of light waves.

Interference and beats

Like any other waves, sound waves can interfere. This can cause strange things to happen acoustically when sound waves reflect off of surfaces, say, in a concert hall. It's possible that waves can constructively and destructively interfere in ways that cause an orchestra or band to sound less than ideal at certain positions in a concert hall. Acoustic spaces such as concert halls are usually designed with strangely shaped panels designed to randomly reflect sound waves to reduce the bad-sounding consequences of interference so that members of the audience can enjoy the sounds of the music. An example of acoustic tiles in a concert hall are shown in Figure 20.5.



Figure 20.5 – Specially shaped tiles positioned on the ceiling of a concert hall create acoustics that sound good throughout the space. Heichal Hatarbut, by Etan J. Tal, is licensed under CC BY-SA 3.0.

The interference of sound waves leads to another effect: beats. Beats occur when two sound waves of very similar frequency interfere with each other. This interference creates an interesting wave that has an amplitude that increases and decreases, and sounds like a regular increase and decrease in the volume of that sound. As the frequencies of two interfering waves become closer together, our ears eventually begin hearing the two different pitches together as a single musical tone.

In the video below, Dr. Pasquale turns on two sound sources at frequencies of 480 Hz and 489 Hz. Perhaps you can hear the

sound of the beat, which comes across as an increase and decrease in the sound of the tones.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The sound pressure of the sound waves was recorded using a microphone. When both pure tones play at the same time, the graph, as shown in Figure 20.6, shows a very fast oscillation (which has a frequency close to that of the pure tones, represented by a red line in Figure 20.6) superimposed with a slow oscillation (represented by a black dashed line in Figure 20.6). (Download this data [XLSX, 49 kB]) That slow oscillation is the beat oscillation and is caused by the interference of the two waves.

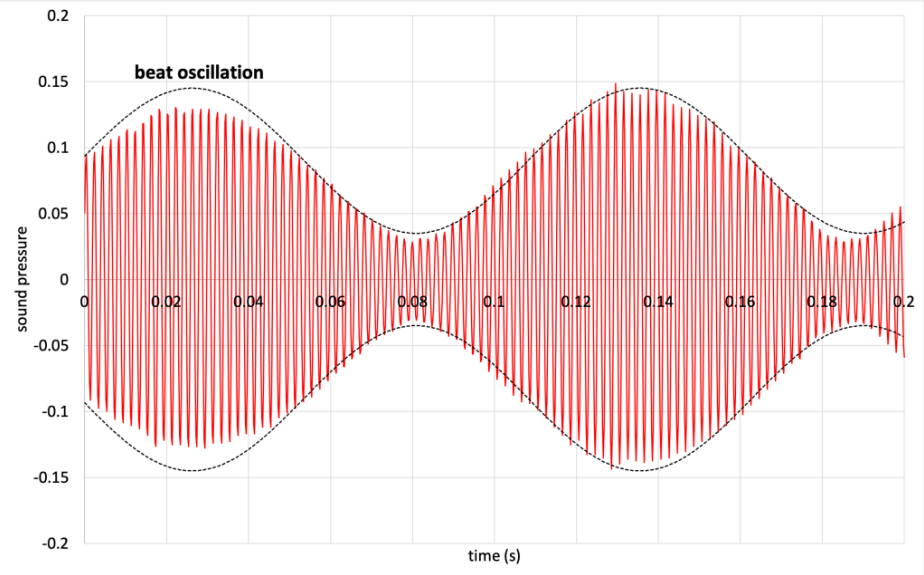


Figure 20.6 – Two pure tones playing simultaneously creates a graph of the interference between the sound waves. The red lines on the graph represent the sound pressure from the two pure tones. The black dashed lines on the graph represent the beat frequency. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The frequency of a beat can be calculated using an equation. The frequency of the beat is equal to the absolute value of one frequency subtracted from the other. In equation form:

$$f_{beat} = |f_1 - f_2|.$$

In the case of the audio example from the video above, the two tones have frequencies of f_1 = 489 Hz and f_2 = 480 Hz. Therefore, the beat frequency is 9 Hz.

Forced vibration and resonance

Striking a tuning fork with a mallet causes it to vibrate. There is energy in these vibrations. In the video below, Dr. Pasquale strikes a tuning fork, then places the vibrating tuning fork into a beaker of water. The water splashes out of the beaker, which is a great visual indication of the energy stored in the tuning fork after it was forced to vibrate.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A forced vibration occurs when one object causes another object to vibrate at a certain frequency. In the video below, Dr. Pasquale strikes a tuning fork to get it to vibrate. Then she places the tuning fork on a table, which causes the table to vibrate at the same frequency as the tuning fork.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Plucking a guitar string causes the body of the guitar to vibrate as well, providing another example of forced vibration. This occurs in many musical instruments.

On the other hand, the natural frequency of an object refers to the frequency of the sound wave that object will tend to produce. A tuning fork is a device built to have a natural frequency that is a pure tone which can be generated by striking the tuning fork with a mallet. This tone comes from the vibrations of the tines of the fork, with smaller tuning fork tines generating higher pitched sounds than larger tuning fork tines. A collection of tuning forks of different sizes is shown in Figure 20.7.



Figure 20.7 – Three tuning forks of different size create different tones when forced to vibrate. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale strikes three different tuning forks, demonstrating that the frequency (pitch) of the tones generated by the tuning forks is related to the length of the tines.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

If an object is forced to vibrate at its natural frequency, that phenomenon is known as resonance. One example of this can be demonstrated by swinging on a swing set. If you pump your legs at the natural frequency, then the amplitude of your swing will increase. Simply pushing the swing at the same frequency of its oscillation will cause the amplitude of the swing to increase (shown in the video below), an example of resonance.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In the video below, Dr. Pasquale demonstrates resonance using an apparatus that contains two sets of three pendulums (each with different lengths) that are connected by a tube. When Dr. Pasquale gets the largest pendulum swinging, it moves back and forth at its natural frequency. That vibration travels through the tube and excites the identical pendulum on the other side, causing it to move back and forth as well. This forced vibration from the left-hand pendulum is equal to the natural frequency of the right-hand pendulum, so resonance occurs and both pendulums continue to move. The two smaller pendulums also experience that forced vibration, but do not start to move because they have different natural frequencies.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Dr. Pasquale then repeats this experiment by moving the middle pendulum, which causes the middle pendulum to oscillate on the other side. The same phenomenon happens with the small pendulum. Resonance occurs when the forced vibration is equal to the natural frequency of the object.

Standing waves also demonstrate the phenomenon of resonance. This can occur in any system that can oscillate: the coiled rope used to show standing waves in chapter 19 of this book, the strings or cavities of a musical instrument, our vocal chords, and even a standing wave excited in a glass bottle, demonstrated in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Resonance is a phenomenon that is also seen in other physical properties, not just sound. Lasers use optical resonance to create strong beams of coherent light. Electric circuits use resonance to tune to particular radio or television stations.

Resonance can be used in many beneficial applications. However, because resonance leads to an amplification of energy, we must be cautious. If energy is left to amplify too much, the energy build-up within the object undergoing resonance may eventually cause it to break apart. If an object such as a bridge or structure is caused to vibrate at its resonant frequency, it may collapse. Engineers take this into consideration when designing buildings, bridges, airplanes, and other structures.

Further reading

- Resonance – This 30-minute video from The Mechanical Universe series demonstrates the concept of resonance using the notable Tacoma Narrows Bridge disaster. (Note: there is some consensus by experts that resonance was not the cause of the Tacoma Narrows Bridge disaster, but instead a concept called aeroelastic flutter, which is too complicated to discuss in this textbook. That said, this video is still very interesting and has lots of good scientific information about resonance.)
- Acoustic levitation – Styrofoam beads are levitated in the interference patterns created in sound waves.
- Hearing ranges – This Wikipedia article discusses the hearing ranges of a wide variety of animals, including humans.

Practice questions

Numerical analysis

1. Calculate the speed of sound in air at...
    1. ...0 ^{o}C.
    2. ...-30 ^{o}C.
    3. ...40 ^{o}C.
2. A student measures the speed of sound in a lab and finds it to be 341 m/s. Calculate the temperature of the lab.
3. A student measures the speed of sound as it travels in an oven and finds it to be 522 m/s. Calculate the temperature of the oven.
4. Two tuning forks produce sound waves with frequencies of 450 Hz and 452 Hz. Calculate the beat frequency when both pure tones are played at the same time.
5. If two speakers emit sound waves at frequencies of 680 Hz and 685 Hz simultaneously, calculate the beat frequency heard by an observer standing nearby.
6. A musician plucks two guitar strings simultaneously. One string has a frequency of 250 Hz, and the beat frequency is 6 Hz. Calculate the two possible frequencies that could be generated by the other guitar string.

Hands-on experiments

1. Create a rubber band guitar by stretching a rubber band over the top of an empty tissue box or a cardboard tube. Pluck the rubber band and observe how the pitch changes when you stretch the rubber band tighter.
2. Fill several glasses with different amounts of water (alternatively, use a single glass and add to or remove water from it). Gently tap the glass(es) with a spoon and observe how the pitch changes with the amount of water. Does the pitch increase or decrease when the water level is closer to the top of the glass?
3. Make a drum. Find a container with an opening on the top, like a plastic container or an aluminum can. Stretch a balloon over the open end and secure it with a rubber band. Tap the balloon and observe the different sounds you can create by varying

the tension on the balloon.

This page titled Chapter 20: Sound is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 21: Musical sounds

Summary

What is it that distinguishes music from noise? Why is it that two musical instruments playing the same note sound different from each other? How can we characterize the loudness of different sound waves? All of these questions, and more, will be answered in this chapter!

Music versus noise

There are a practically infinite number of combinations of different frequencies and amplitudes of sound. Some of these combinations may sound pleasing to us, and others may not. Usually, we categorize the pleasing sounds as musical sounds. Perhaps you consider jazz, classical, rock, or rap to be musical. Other sounds that are not pleasing can be categorized as noise. This might include sounds such as banging together pots and pans, dropping a hammer on the floor, or scratching fingernails on a blackboard.

What is it that distinguishes music from noise? To a certain extent, this is a subjective question that depends on the person listening to the sounds. You may consider certain types of sounds to be musical, whereas someone else might consider them to be more like noise.

From a physics perspective, we can broadly categorize different types of sounds as either music or noise based on the types of frequencies contained in the sound. Most sounds are not pure tones consisting of only a single frequency of sound waves. Instead, they are the combination of a few, or sometimes many, tones combined together.

Using a concept called Fourier analysis, we can look at the different frequencies that make up each sound. The mathematics behind Fourier analysis is very complicated and is outside the scope of this textbook. But the conceptual idea is that Fourier analysis decomposes a wave into each individual frequencies that make it up. A graph of a Fourier analysis therefore plots the amplitude of a wave on the y-axis and the frequency of the wave components on the x-axis. Frequencies that have a large amplitude are more well-represented in the wave than frequencies that have a small amplitude.

A tuning fork is designed to produce a single frequency of sound, although other harmonics can be created in the tuning fork, just as a string can support multiple standing wave harmonics. Figure 21.1 shows both the graph of the sound pressure of the wave generated by the tuning fork as it changes through time (top) as well as the Fourier analysis of the wave (bottom). The Fourier analysis of the tuning fork shows a single predominant peak at 490 Hz, the same frequency that is stamped on the side of the tuning fork. A few other harmonics exist at greatly reduced amplitudes (the second harmonic is a very small bump at 980 Hz, and the third harmonic at 1470 Hz is barely perceptible on the graph). This is consistent with the sound pressure graph that shows a single oscillating wave at a frequency of 490 Hz.

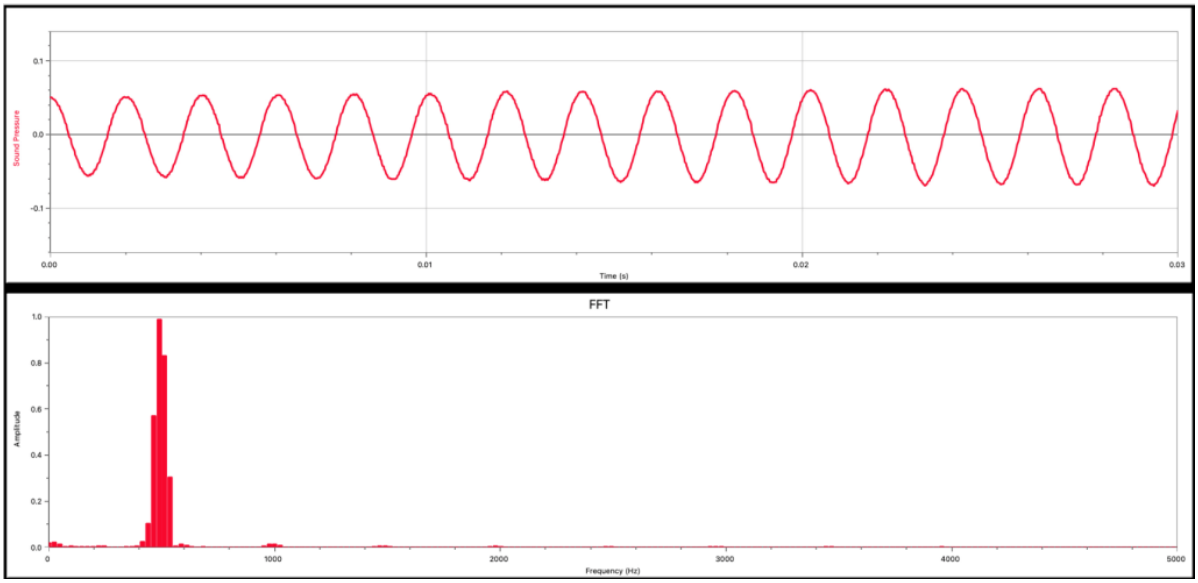


Figure 21.1 – A graph of sound pressure vs. time for a sound wave generated by a tuning fork (top) and the Fourier analysis of that sound wave (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The graphs in Figure 21.2 were generated by a piano playing a single musical note. The sound pressure vs. time graph (top) shows a wave that is more complicated than that of the tuning fork, but it is still possible to see periodic oscillations in the graph. The Fourier analysis graph (bottom) shows that the piano does not produce one single frequency of sound wave. There are more frequencies of waves represented in this sound compared to the tuning fork. However, the frequencies that do exist in the Fourier analysis are periodic. That is, they are spaced out at regular intervals.

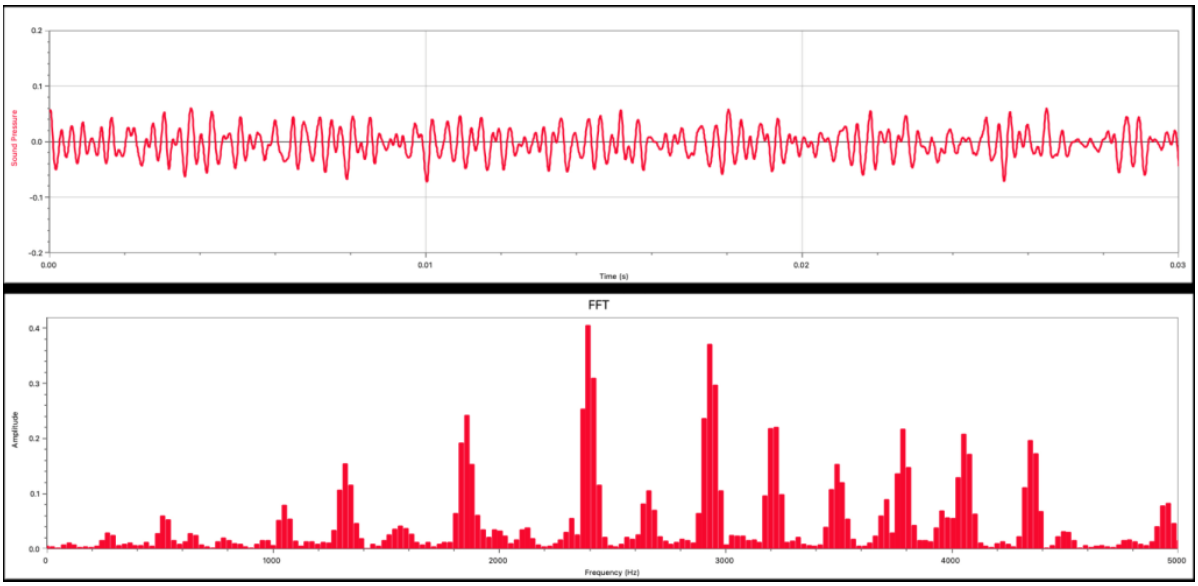


Figure 21.2 – A graph of sound pressure vs. time for a sound wave generated by a piano playing a musical note (top) and the Fourier analysis of that sound wave (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The graphs in Figure 21.3 are for a more “noise”-like sound, created by Dr. Pasquale snapping her fingers. The sound pressure vs. time graph (top) shows a very complicated wave that has many strong oscillations at the start of the sound, followed by many other oscillations. The Fourier analysis (bottom) shows the frequencies of the sound wave generated by the snapping sound. There are many frequencies present in the sound wave, and they have no periodic spacing or clear pattern to them. This Fourier analysis shows a very random spread of frequencies.

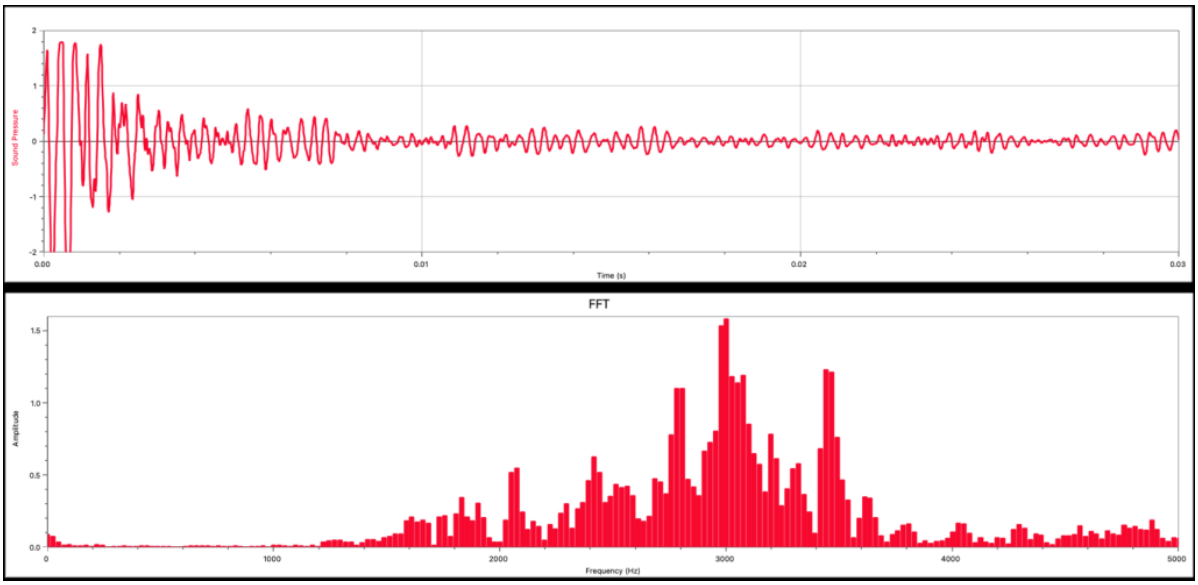


Figure 21.3 – A graph of sound pressure vs. time for a sound wave generated by Dr. Pasquale snapping her fingers (top) and the Fourier analysis of that sound wave (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Finally, the graph in Figure 21.4 shows the sound pressure vs. time graph (top) and Fourier analysis (bottom) for the sound generated by a white noise machine. The white noise machine creates many different frequencies of sound. The sound pressure vs. time graph is very messy looking, and the Fourier analysis shows a very broad spectrum of frequencies that are represented in the sound wave.

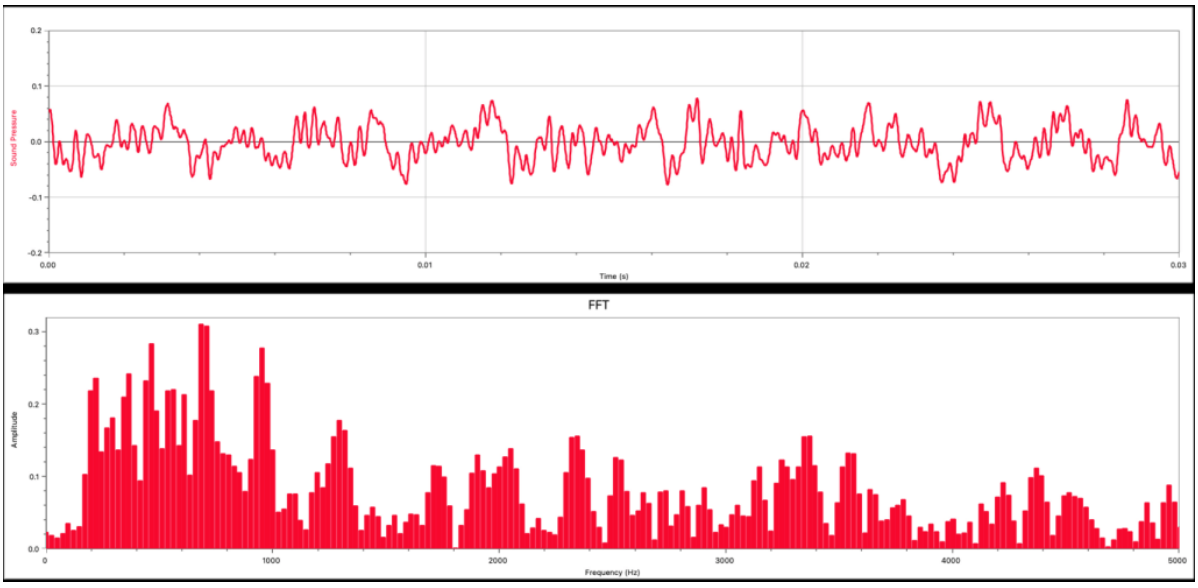


Figure 21.4 – A graph of sound pressure vs. time for a sound wave generated by a white noise machine (top) and the Fourier analysis of that sound wave (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

What conclusions can we make from these graphs? Generally speaking, musical sounds, such as the tuning fork and piano, consist of just a few frequencies of sound waves. We call this type of Fourier analysis discrete. Discrete just means that there are only a few particular frequencies that predominate. Sounds that we consider to be noise have very continuous, broad, or random frequency distributions in their Fourier analysis.

One last comment about musical sounds. Musical instruments are designed to support different standing waves and create sounds at different frequencies. For example, the note “middle C” has a frequency of 256 Hz. If there are five different instruments all playing middle C at the same frequency, what is it that causes each instrument to sound different?

In fact, each instrument will have a slightly different Fourier analysis. The exact frequencies that make up any particular musical

note will be a little different from any other instrument. This characteristic is called timbre. It has to do with the design of the instrument, the materials it is made from, and how many harmonics are created in the instrument.

Loudness versus intensity

Loudness has to do with the subjective quality of the amplitude of a sound wave. Perhaps you like to turn the volume all the way up, and consider that to be the best level of volume for listening to music. Somebody else may consider that same volume to be too loud. For a single individual, loudness is related to the intensity of a sound wave, but among individuals, loudness is not something that can be agreed upon.

On the other hand, intensity describes the amount of energy per unit of area that reaches a certain location every second. In other words, it is equal to the amount of power divided by the area over which that power is distributed. Intensity is an objective, physical quantity that we can measure. The unit for intensity is W/m^2. It is related to the energy that is transmitted by the sound wave. (In fact, intensity can be measured for all waves, not just sound waves.)

The human ear is capable of hearing sounds that are both very quiet and very loud. This means that the values of intensity that we can hear go from extremely small numbers to extremely huge numbers. The lowest intensity sound that an average human ear can hear is 10^{-12} W/m^2. That is: 0.000000000001 W/m^2, an extremely small number. On the other end of our ability to hear, a rocket engine has an intensity of 10^6 W/m^2. That is: 1,000,000 W/m^2, a very large number. Between these two values are a huge range of intensity values that we can hear.

For this reason, using intensity values is not the best way to describe sound intensity. Instead, scientists use a relative scale. In other words, how much more intense is one sound compared to another? To determine the relative intensity of sounds, we divide by a baseline value. For sound waves, the baseline that is used is known as "the threshold of human hearing," which is 10^{-12} W/m^2. Using this value, we can see how much more intense one sound is compared to that threshold. The threshold of human hearing has a relative intensity of one (it is exactly equal to itself). The rocket engine has a relative intensity of

$$\frac{10^6\ \mathrm{W/m^2}}{10^{-12}\ \mathrm{W/m^2}} = 10^{18}.$$

This method (using a relative scale) has eliminated very tiny numbers from being used when determining the loudness of sounds, but it still encompasses an enormous range of values from quiet (1) to loud (10^{18} or higher).

To solve this issue, we use a logarithmic scale to discuss intensity. The logarithm just looks at the exponent of the power of ten, and ignores the base (that number ten that we're taking a power of). In addition, we multiply that logarithm by the number 10. This is known as the decibel scale. The table below shows several different values of intensity, the relative intensity compared to the threshold of human hearing, the logarithm, and the logarithm multiplied by 10.

10^{-12}	1	0	0
10^{-10}	100	2	20
10^{-8}	10,000	4	40

10^{-6}	1,000,000	6	60
10^{-4}	100,000,000	8	80
10^{-2}	10^{10}	10	100
10^{0}	10^{12}	12	120
10^{2}	10^{14}	14	140
10^{4}	10^{16}	16	160
10^{6}	10^{18}	18	180

The logarithm multiplied by ten is known as the decibel scale. Many common sounds can be represented by their decibel (dB) value. This scale solves the problems represented by using intensity and relative intensity by creating a scale that generates values between 0 dB and about 100 dB for common sounds. The table below gives the intensity in both W/m^2 and decibels for several common sounds.

threshold of human hearing	10^{-12}	0
rustling leaves	10^{-11}	10
whispering	10^{-10}	20
“normal” conversation	10^{-8} – 10^{-6}	40 – 60

vacuum cleaner	10^{-5}	70
heavy traffic	10^{-3}	90
riveter	10^{-2}	100
loud rock concert	10^{0}	120
jet airplane	10^{3}	150
rocket engine	10^{6}	180

At a certain level of intensity, around 85 decibels, our ears will become damaged. People working around loud noises for long periods of time need to wear earplugs or headphones to protect their hearing. If we're going to be exposed to relatively low intensity noises for short periods of time, earplugs may be sufficient. For longer exposures, or for more intense noises, such as working around jet engines, earmuffs (Figure 21.5) may be needed.



Figure 21.5 – Earmuffs may be needed to prevent hearing loss for people exposed to loud sounds such as jet engines. Various earmuff hearing protectors, by JiriMatejicek, is licensed under CC BY-SA 4.0.

At intense enough levels, sound waves can cause physical pain, which occurs around approximately 120 decibels. Above this level, it is very important to wear protective equipment on our ears!

Earplugs are used to decrease the intensity of sound. Usually on the label of a package of earplugs is a decibel value that is

included (Figure 21.6). This indicates how much the decibel level will decrease. If you're using a gas-powered leaf blower with an intensity of 90 dB, wearing 30 dB earplugs will reduce the intensity to 60 dB.

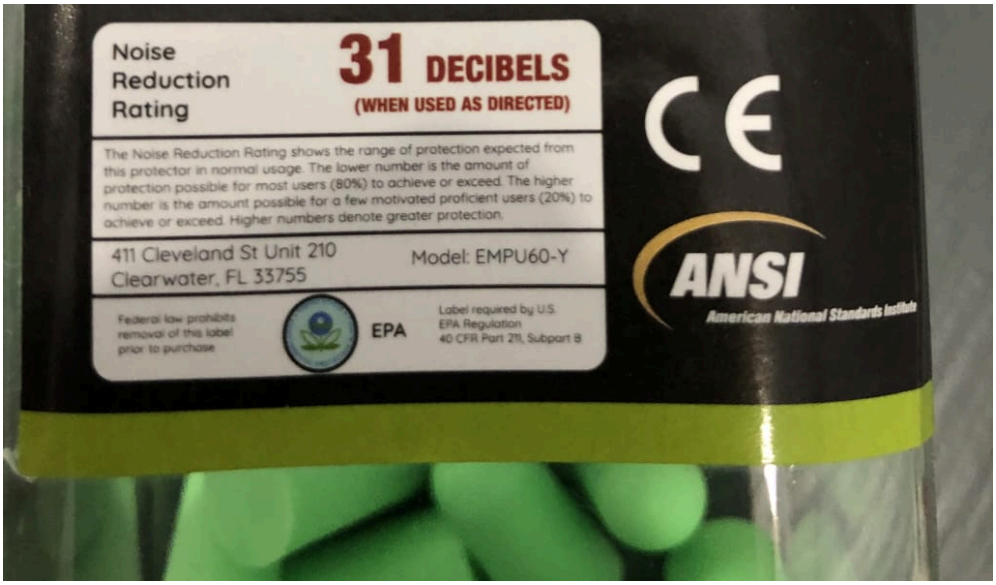


Figure 21.6 – A package of earplugs lists the reduction in intensity as 31 dB. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

How much more intense (louder, if you will) is a rocket engine compared to a jet engine? This question of "how much louder" becomes slightly difficult to answer when using a logarithmic scale. A logarithm tells us the exponent on the number ten. In other words, it tells us how many zeros are after the one in a number. A jet engine is 150 dB. Something with an intensity of 160 dB is 10 times more intense than the jet engine. Something at 170 dB is 100 times more intense. The rocket engine, at 180 dB, is one thousand times more intense than the jet engine.

So when calculating "how much more intense," remember that every time the value 10 is added to the decibel value of intensity, the W/m^2 intensity is multiplied by 10. A difference of 10 dB is 10 times more intense, but a difference of 20 dB is 100 times more intense, a difference of 30 dB is 1,000 times more intense, and so on.

Further reading

- Colors of noise – This chapter discusses white noise; there are many different "colors" of noise pertaining to the frequencies of sound contained within the noise. This Wikipedia page allows you to listen to the different types of noise.

Practice questions

Conceptual comprehension

1. From the Fourier analysis shown in Figure 21.7, was the sound wave most likely a musical sound or a noisy sound? Explain your answer.

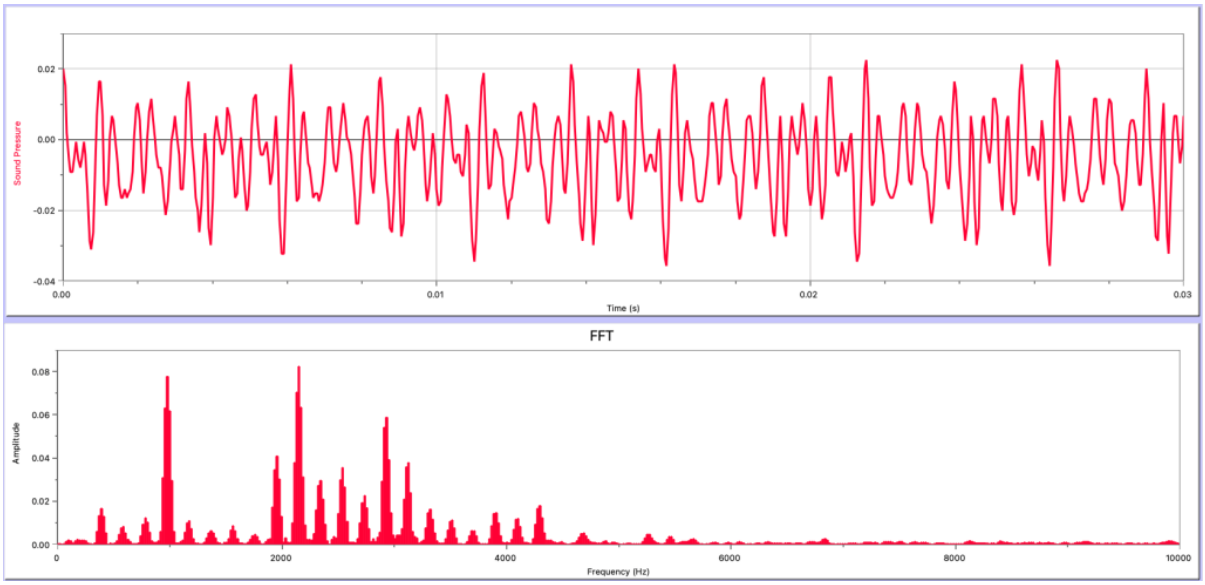


Figure 21.7 – A graph of sound pressure vs. time (top) and the Fourier analysis (bottom) of an unknown sound wave. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

2. From the Fourier analysis shown in Figure 21.8, was the sound wave most likely a musical sound or a noisy sound? Explain your answer.

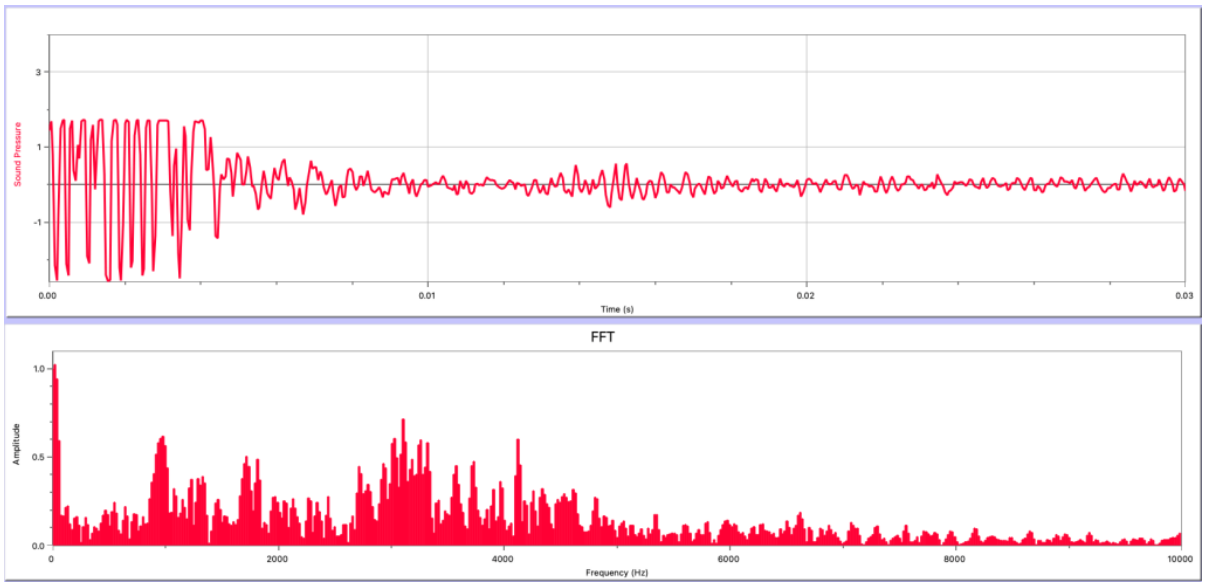


Figure 21.8 – A graph of sound pressure vs. time (top) and the Fourier analysis (bottom) of an unknown sound wave. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Numerical analysis

1. The Fourier analysis shown in Figure 21.9 was generated by Dr. Pasquale blowing into a glass bottle. On average, the crests of the wave shown in the sound pressure vs. time graph (top) are spaced apart by 3.8 ms. Calculate the frequency of the peak value in the Fourier analysis graph (bottom).

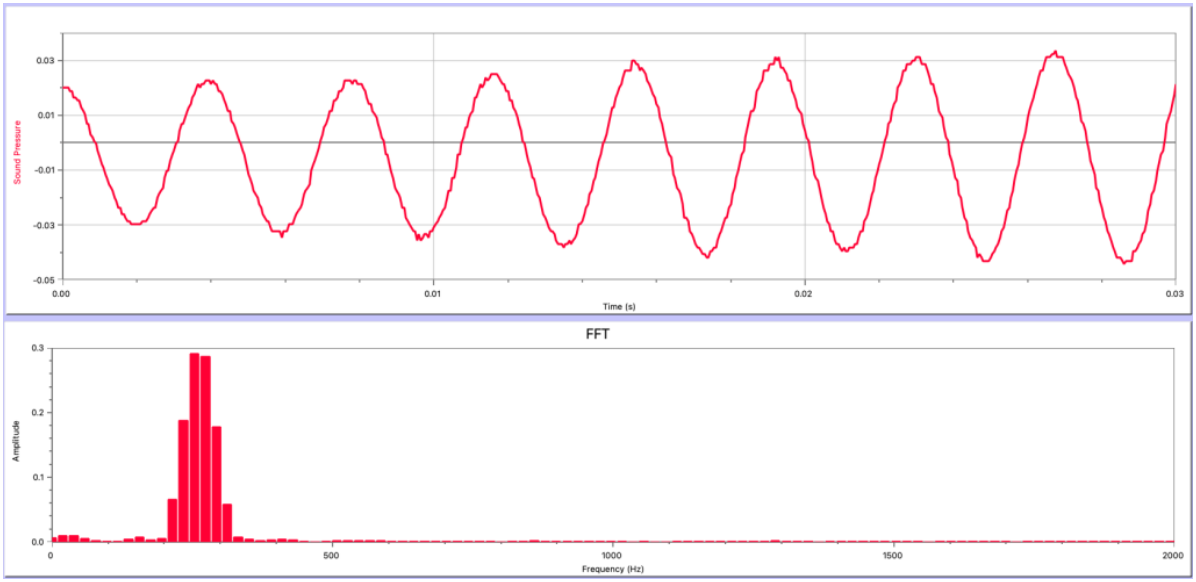


Figure 21.9 – A graph of sound pressure vs. time for a sound wave generated by Dr. Pasquale blowing into a glass bottle (top) and the Fourier analysis of that sound wave (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

2. The Fourier analysis shown in Figure 21.10 was generated by Dr. Pasquale whistling. On average, the crests of the wave shown in the sound pressure vs. time graph (top) are spaced apart by 7.35×10^{-4} s. Calculate the frequency of the peak value in the Fourier analysis graph (bottom).

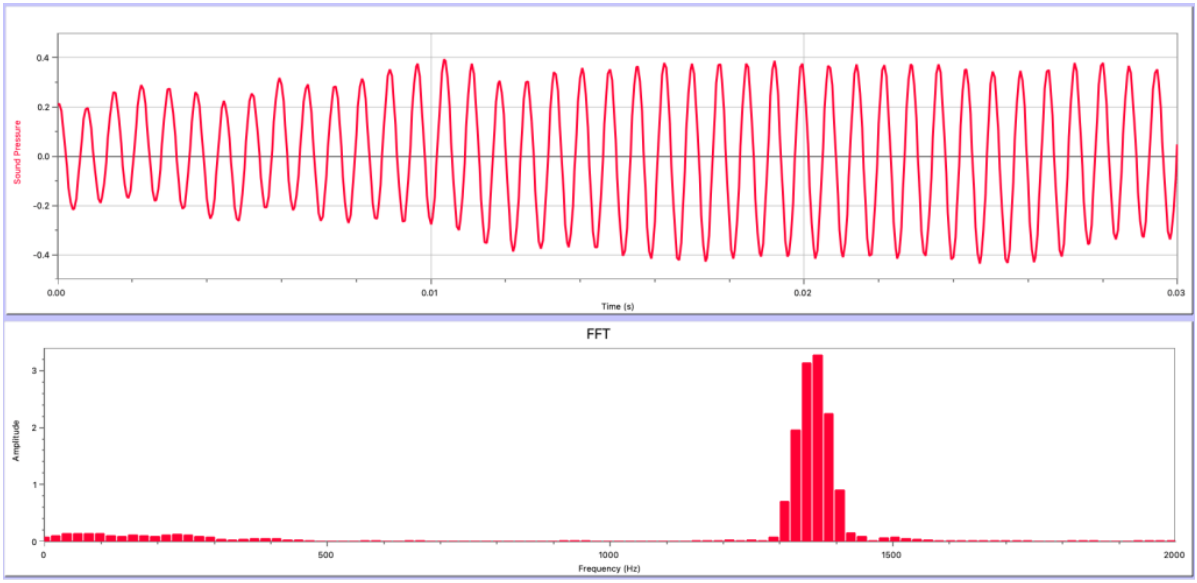


Figure 21.10 – A graph of sound pressure vs. time for a sound wave generated by Dr. Pasquale whistling (top) and the Fourier analysis of that sound wave (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

3. The intensity of a sound is 10^{-5} W/m^2. Calculate the sound intensity in dB.
4. How much more intense is a 50 dB sound compared to a...
    1. ...40 dB sound?
    2. ...30 dB sound?
    3. ...20 dB sound?
    4. ...10 dB sound?
5. An amplifier increases the intensity of a signal by a factor of 1,000. If the original sound level was 80 dB, calculate the new intensity (in dB) after amplification.

Hands-on experiments

1. Use your smartphone and an app such as phyphox. Go to the “audio scope” and press the play button to record several sounds and look at the sound pressure vs. time wave. Then go to the “audio spectrum” and press the play button to see the Fourier analysis of different sounds. Try whistling at several different frequencies (or play different musical notes on an instrument) and see how the Fourier analysis changes.

This page titled Chapter 21: Musical sounds is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

VI: ELECTRICITY AND MAGNETISM

Chapter 22: Electrostatics
Chapter 23: Electric circuits
Chapter 24: Magnetism
Chapter 25: Electromagnetic induction
Chapter 26: Light
Chapter 27: Color
Chapter 28: Reflection and refraction
Chapter 29: Light waves
Chapter 30: Light emission

This page titled VI: Electricity and magnetism is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 22: Electrostatics

Summary

Electrostatics introduces an important new property of matter: electric charge. This property unlocks the concepts of electricity and magnetism, which we'll be exploring in the next several chapters of this textbook. This discussion will explain what electric charges are and how they give rise to forces and electric fields.

Charge

Electrical charge is a fundamental property of matter that causes particles that carry a charge to experience a force when in the presence of an electric field. This is somewhat analogous to mass, which is a fundamental property of matter that causes objects of mass to experience a force when in the presence of a gravitational field. Just as mass is conserved, charge is also a conserved quantity. This means that the total charge in any closed system must remain constant over time. Charge can neither be created nor destroyed.

The term electrostatics, which is the topic of this chapter, has to do with charges that do not move. We will discuss all of the fascinating properties of charges that do move in the next chapter on electric circuits.

Charge has a symbol of the lowercase letter q. (It's possible you may see the uppercase letter Q used for charge in other textbooks.) We use the unit coulomb, which has a symbol of the capital letter C.

There are two types of charges: positive charges and negative charges. Every proton carries a positive charge, and every electron carries a negative charge. The magnitude of the charge of an electron is equal to the magnitude of the charge of a proton. The value of this charge is known as the elementary charge, and has a symbol of the lowercase letter e. The value of the elementary charge is

$$e = 1.6 \times 10^{-19}\ \text{C}.$$

Each proton carries a positive charge of $+e$, and each electron carries a negative charge of $-e$. In a material or object that is neutral, the magnitude of positive charge is exactly equal to the magnitude of negative charge.

In addition to being conserved, charge is also quantized. This means that charge can only come in particular values which are integer multiples of the value e. This is because there can only be whole numbers of protons and electrons. As it is not possible to have a fraction of an electron, or a fraction of a proton, the amount of charge that any object can have must come in integer values of the magnitude of the elementary charge. That is, $|q| = Ne$, where N is an integer.

There are two possible interactions that charges can experience. Two like charges (positive and positive, or negative and negative) will repel. That is, if you could somehow hold two like charges together, they would experience a force that causes them to move apart from each other. Two unlike charges (positive and negative) will attract. Therefore, if you could somehow hold two unlike charges apart from each other, they would experience a force that causes them to come together.

We can visualize the effect of this force by looking at how charges interact on a low-friction turntable. In the video below, two acrylic rods are rubbed with paper. This removes electrons, causing a positive charge in both rods. When one of the rods is free to rotate, and the other rod is held nearby, we can see that the first rod rotates away. Positive charges repel other positive changes.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In the video below, two PVC rods are rubbed with paper. This process adds electrons to the PVC, causing a negative charge. Holding one negatively charged PVC rod near the one on the turntable causes the PVC rod on the turntable to rotate away, demonstrating electrostatic repulsion. Negative charges repel other negative changes.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When the negatively charged PVC and positively charged acrylic are held nearby, now the positive rod rotates toward the negative rod, demonstrating a force of attraction between the unlike charges.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The above video demonstrations show how charged objects behave when held near other charged objects. What about when a charged object is held near a neutral object?

In the video below, Dr. Pasquale places one of the charged rods close to a neutral object (a wooden meter stick resting on a pivot). When the charged rod gets close to the meter stick, it experiences a torque, causing it to rotate. This demonstrates an attractive force with the charged rod. Both positively and negatively charged rods create an attractive force with neutral objects.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

How is it possible to give objects positive or negative charge? We know that atoms have a positively charged nucleus comprised of protons and neutrons, and that electrons orbit the nucleus. It is very complicated to add or remove protons from an atom. It

requires a nuclear process such as radioactive decay, fission, or fusion. Because electrons orbit the nucleus, it is much simpler to add or remove electrons than to add or remove protons. Therefore, to give an object a positive charge, we remove electrons and to give an object a negative charge, we add electrons.

An ion with a charge of negative two has that charge because it has two excess electrons. An ion with a charge of positive three has that charge because it has three fewer electrons than protons.

Coulomb's law

If two like-charged particles are held near each other, they will move away from each other if let go. If they start at rest, then move away from each other, that means the two charges will speed up, indicating that an acceleration is present. From Newton's second law, we know that when there's acceleration, there's a net force. Charged particles exert forces on each other.

Coulomb's law is the equation that quantifies the force that exists between particles. The magnitude of this force is given by the equation

$$|F| = k\frac{|q_1||q_2|}{d^2}.$$

That is, the force equals a constant (k) times the charge of one of the particles, times the charge of the second particle, divided by the distance between the two particles squared. The unit of force, as always, is the newton.

The constant k is known as Coulomb's constant and has a value of $9 \times 10^9\ \mathrm{Nm^2/C^2}$.

Note the similarities with the equation for gravitational force! Both forces have a constant term, a multiplication of the two physical properties (mass for gravity, charge for electrostatic force), and are divided by the distance squared. This means that electrostatic force, just like gravity, obeys the inverse-square law. Halving the distance between two charges will quadruple the force between them, while doubling the distance will cut the force to a fourth of its original value.

Gravitational forces and electrostatic forces also have some significant differences. One of the biggest differences is that the electrostatic force is much stronger than gravity. To attempt to quantify this, we can calculate both the electrostatic force and gravitational force that exist between a single proton and single electron in a hydrogen atom. The ratio of electrostatic force to gravitational force can be calculated as

$$\begin{aligned}\frac{F_{\text{electrostatic}}}{F_{\text{gravitational}}} &= \frac{k\,|q_{\text{proton}}\,|\,|q_{\text{electron}}\,|}{G m_{\text{proton}}\,m_{\text{electron}}} \\ &= \frac{\left(9\times10^{9}\mathrm{Nm^2/C^2}\right)\left(\left|1.6\times10^{-19}\mathrm{C}\right|\right)\left(\left|-1.6\times10^{-19}\mathrm{C}\right|\right)}{\left(6.67\times10^{-11}\mathrm{Nm^2/kg^2}\right)\left(1.67\times10^{-27}\ \mathrm{kg}\right)\left(9.11\times10^{-31}\ \mathrm{kg}\right)} \\ &= \frac{2.3\times10^{-28}\ \mathrm{N}}{1.01\times10^{-67}\ \mathrm{N}} \\ &= 2.27\times10^{39}.\end{aligned} \tag{22.1}$$

That is, for a single proton and a single electron in a hydrogen atom, the electrostatic force is nearly 40 powers of ten stronger than the gravitational force between the two particles! (Note that the radius of the hydrogen atom, being equal for both the electrostatic and gravitational forces, cancels out in the calculation.)

The fact that the electrostatic force is so much stronger than gravity may be something that you've noticed, perhaps if you've ever stuck a balloon to the wall after rubbing it in your hair. Dr. Pasquale demonstrates this in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The force from a few charged particles on the balloon is easily able to overcome gravity. This also happens when small pieces of plastic stick to your hands. (Or when foam peanuts stick to a cat, as seen in Figure 22.1.) These situations are not defying gravity: electrical charge that's built up on an insulator (such as foam or plastic) will cause it to have an electrostatic force with any neutral object nearby.



Figure 22.1 – A cat is covered in foam packing peanuts due to electrostatic attraction between the charged foam and cat's fur. Cooper and styrofoam, by Sean McGrath, is licensed under CC BY 2.0.

The second big difference between the electrostatic force and gravity is that electrostatic forces come in two types: attractive and repulsive. There are two types of charge, and that means the force can go in one of two directions: toward and away. Meanwhile, gravity is only an attractive force. Mass can only be positive, not negative, so we will only ever experience attractive forces with other objects of mass.

When using Coulomb's law, we need to be careful to ensure that the sign we obtain in our force calculation is consistent with the direction of that force. As described in appendix A of this textbook, vectors pointing to the right are positive, and vectors pointing to the left are negative. Vectors pointing up are positive, and vectors pointing down are negative. We will want to ensure that the direction of force corresponds with this convention.

Therefore, when calculating an electrostatic force, first calculate the magnitude of that force by using the equation. Once you've finished calculating the magnitude, look at the types of charges involved to determine the direction.

Consider a charge (charge A) $q_A = +10\ \mu\text{C}$ (that is: 10×10^{-6} C) that's held 0.2 meters away from a second charge (charge B) $q_2 = -5\ \mu\text{C}$ (-5×10^{-6} C). What is the force that charge B causes on charge A?

First, calculate the magnitude by ignoring the signs of the charges and using Coulomb's law.

Next, analyze the situation to determine the direction of the force. Because B is causing the force, A is the charge that is moving

as a result. In which direction will charge A move? Because the force is attractive (there are unlike charges involved), charge A will move toward charge B. If that direction is to the right, then the force is positive. If that direction is to the left, then the force is negative.

What is the force that charge A causes on charge B? We can solve this in two ways. The first is to see that the magnitude using Coulomb's law will be the same, and because now B is in motion, it moves in the opposite direction as charge A. Or we can recall from Newton's third law that these are action-reaction pairs, meaning the forces will be equal and opposite.

Conduction

Materials can be characterized by how easily they allow electrons to move around. We looked at a similar property when we learned about heat transfer. In fact, materials that are good at moving heat around also tend to be good at moving electrons around (and vice versa).

Conductors

We call materials in which electrons can easily move around electrical conductors. If you think about what wires are made of, maybe you can think of the types of materials that make good electrical conductors: metals. In metals, electrons are able to move around easily. This means we use them in all of our electric circuits (for example, the circuit in Figure 22.2) as well as in power lines and the charging cable you use for your electronic devices.

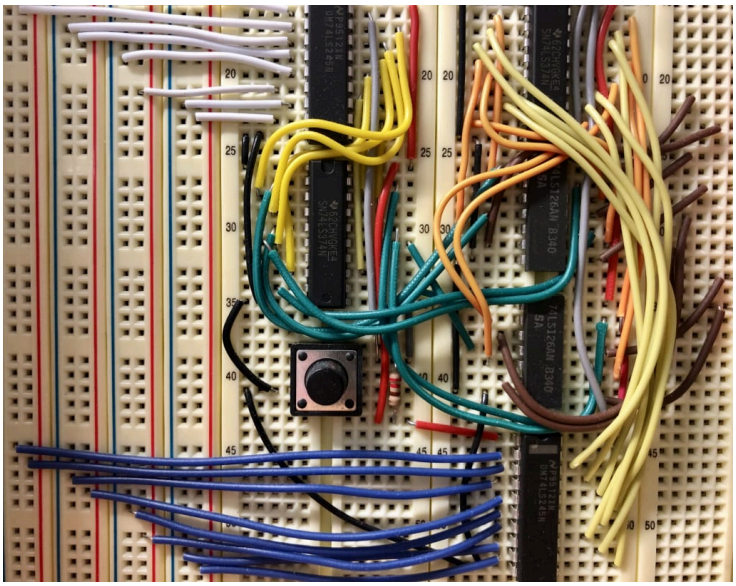


Figure 22.2 – An electric circuit built with integrated circuit chips and many wires. All of the wires are copper conductors coated with plastic insulation. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Insulators

Materials that are not good electrical conductors are called insulators. Plastic, paper, air, wood: these are all good electrical insulators. In these materials, electrons are very tightly bound to the nucleus of their host atoms. However, when water contains impurities such as salt, then it becomes a very good conductor.

In fact, the wires in the electric circuit shown in Figure 22.2 are coated with an insulator: plastic. This is to prevent the conductors inside each wire from touching each other and sending electric charges flowing through the wrong paths in the circuit. The charging cables used for your electronic devices are similarly coated in a plastic insulator. The fact that air is an insulator prevents us from getting shocked when we walk near electrical outlets!

Semiconductors

The electrical properties of some materials are somewhere in between those of a conductor and those of an insulator. Their electrons aren't super tightly bound to the nucleus as with insulators, but they're also not as freely flowing as with conductors. We call these materials semiconductors. Silicon is probably the most widely known elemental semiconductor. It's used in just about every integrated circuit in every computer and electronic device. Semiconductors are useful because they can be turned into an electrical switch that turns on (acts like a conductor) or off (acts like an insulator) by applying something called voltage. These electrical switches are called transistors. If you're reading this as an ebook, the computer or phone you're using likely has many billions of these transistors inside of its processor.

Superconductors

Another fascinating type of material is a superconductor. Superconductors are materials that are capable of allowing electrons to flow with no resistance whatsoever; that is, no energy will be lost to heat when electrons move about in these materials. Superconductors are generally made from certain elements or ceramic materials that are cooled down to just a few kelvin. That is, only a few degrees above absolute zero!

If you've ever had an MRI, then you've benefited from technology that uses superconductors. Superconducting electromagnets are used in MRI machines. They are cooled to a low temperature using liquid helium, causing a rhythmic pumping sound in the background as the MRI machine runs.

Because these materials exhibit zero resistance at very low temperatures, widespread use of superconductors is currently impractical. In the early 1980s, a new class of superconductors whose superconducting transition temperatures were greater than 77 K (-196 °C, or -321 °F) was discovered. This temperature is significant because this meant that these superconductors could be cooled with inexpensive liquid nitrogen. However, these materials tend to be ceramics and have very poor mechanical properties. A great achievement of physics would be the creation of a room temperature (300 K) superconductor that could also be made into wires. It will likely be some time before our infrastructure will involve such materials.

Charge distribution

Charge distribution refers to how charges distribute themselves or move around in different materials. In metals, electrons are able to move about freely. This means that when electrons are introduced into a metal they can move around as needed to reduce the effect of the repulsion between each individual electron. The same situation occurs when electrons are removed from a conductor. Electrons redistribute to cause the least amount of energy due to the repulsive forces between positive charged ions.

We can see this in an electroscope, as shown in the video below. When electrons are introduced into the electroscope, which is made of metal, the free piece of metal in the electroscope moves away. Because the negative charges have distributed themselves throughout the conducting material, we can see the repulsive force causing the free piece of metal to rotate to get as far away from the surrounding negatively charged metal as possible.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The same thing occurs when electrons are removed from the electroscope, shown in the video below. The conducting material of the electroscope becomes positively charged and the repulsive force causes the free piece of metal to rotate away from the rest of the positively charged metal.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Conduction

Conduction occurs whenever two or more conducting materials come into contact with each other. Charge is able to flow freely between the conductors, causing the electrons to move in such a way as to minimize repulsive forces. We will discuss the effects of conduction much more in the next chapter of this textbook when discussing electric circuits. The flow of charge, known as current, causes many interesting electrical properties and plays a large role in our 21st century lives.

Grounding

Grounding is a special type of conduction where charges redistribute themselves to an area that can accept lots of charges without causing a large buildup of electrons or positive ions. The Earth is a great ground source. Because the Earth is so large, it can accept a lot of electrons without causing a buildup of charge. When you plug in your electronics, the ground connection to the earth is provided by either a copper wire connected to plumbing (as in Figure 22.3), or a copper wire buried in the ground.



Figure 22.3 – A copper plumbing pipe connected to the Earth has an electrical ground wire connected to it. Metal water pipe used as grounding electrode, by z22, is licensed under CC BY-SA 3.0.

Sometimes, grounding can occur through our bodies. Humans are relatively good conductors. When Dr. Pasquale touches the electroscope in the video below, charges move around to the ground through her hand and her skin. We can see that she has discharged the electroscope as the freely moving piece has rotated back to its equilibrium position. There is no longer a charge on the metal causing a repulsive force.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

As long as there isn't too much charge flowing through a person when they act as a conduit to ground, this can be safe. This becomes very dangerous only when the flow (movement) of charge becomes too high. The mere presence of charge is not in itself dangerous.

In the video below, Dr. Pasquale holds onto a van de Graff generator. As it runs, electrons build up on her skin and hairs. After a time, we can visually see the effect of this charge buildup as her hairs begin to separate from each other: they are experiencing the force of electrostatic repulsion. The hair on her head and on her skin will stick up.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When Dr. Pasquale sticks her finger out, a corona discharge emanates from her finger. This is caused by the air around her finger becoming ionized. Unfortunately, the sound was not captured well enough on the video above to hear. If you have ever stood in the vicinity of large transmission lines, the corona discharge sounds like a staticky hiss.

Note that a lot of electric charge transfers to Dr. Pasquale's body when she touches the van de Graff generator. As long as she does not let that charge flow rapidly away from her, this is perfectly safe. She stands on a giant plastic block when she is touching the van de Graff generator so that charge cannot flow through her feet into the floor. She is surrounded by insulators. The van de Graff generator would be dangerous if Dr. Pasquale were to touch an object and cause the electrons to flow through her body.

At the conclusion of the video, Dr. Pasquale is grounded when she touches a wire connected to a very large resistor. This electrical component, which will be discussed in the next chapter of this book, ensures that the charge does not flow away from her body too fast. It causes the charge to flow away at a safe rate.

Electrostatic discharge

While the presence of charge in itself is not a danger, the movement of charges, known as current (which will be discussed in the next chapter of this textbook) is where the danger lies. Perhaps you can recall being shocked while touching a doorknob after walking across a carpeted floor in the winter. This shock (known as electrostatic discharge), does not come from the charge buildup itself, but from the movement of that charge when it is able to discharge from your body.

We can see electrostatic discharge using a Wimshurst machine, as shown in the video below. Rotating the machine causes a buildup of opposite charge on the two spheres. As long as the charge remains surrounded by insulators, it will remain on the metal sphere. As soon as a second sphere comes close enough, the charge will discharge, even through the gap of air in between the two spheres! Dr. Pasquale filmed this demo with the lights off to make the sparks easier to see.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In insulating materials, charges cannot move around freely. Charges will tend to stay where they are placed. When rubbing a balloon in your hair, electrons are transferred to the balloon. Because the balloon is made of an insulating material, the electrons are unable to redistribute around the entire balloon. Only the portion of the balloon that was rubbed against your hair is able to stick to the wall. The other side of the balloon does not have a buildup of electrons and remains neutral. This side of the balloon will not stick to the wall.

Induction

Induction is another way that charges can be caused to redistribute themselves in different materials. Unlike conduction, induction does not always require physical contact between objects to cause charges to move.

In the video below, we can see how charges redistribute when Dr. Pasquale holds a negatively charged (gray PVC) rod close to (but not touching) the top of the electroscope. The charged rod is not touching any of the metal contacts, but we can see from the movement of the metal piece in the electroscope that charges are moving around. Electrostatic repulsion causes electrons to move to the bottom of the electroscope, to be as far away from the negatively charged rod as possible. While the net charge on the electroscope is still zero, there is an overall polarization of the metal. The abundance of negative charges in the bottom of the electroscope causes repulsion (see Figure 22.4, left), and the movable piece rotates as a result of that force. When the charged rod is moved away from the electroscope, the effect goes away.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The same effect occurs with a positively charged rod (shown in the video above with the clear acrylic rod). When holding it close to (but not touching) the top of the electroscope, the electrons experience an attractive force, causing them to move to the top of the electroscope. This creates a buildup of positive charge in the bottom of the electroscope (see Figure 22.4, right), causing the moveable piece to rotate as a result of the repulsion between these positive charges.

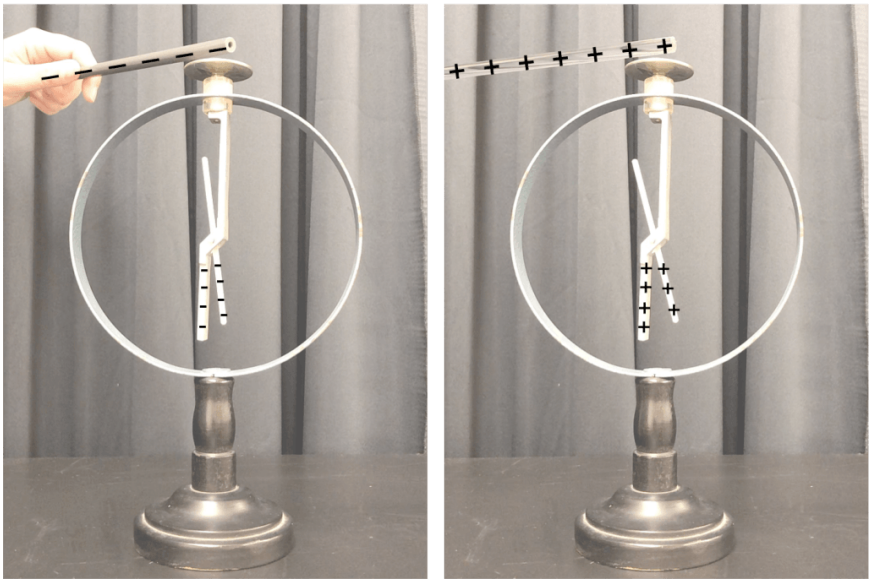


Figure 22.4 – Induction in metals occurs due to charge redistribution in the conducting material from electrostatic repulsion. This can occur due to the presence of a negatively charged object (left) or a positively charged object (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Charging by induction

If we want to use this effect to cause a more permanent charge in metals, we can use two pieces of metal instead of one. Placing a negatively charged object nearby two touching metal objects will cause the charges in the metal to redistribute. Repulsion causes the electrons in the metal to move as far away from the charged object as possible. As a result, the metal closest to the charged object will now contain a positive charge (Figure 22.5, left). If the two metal objects are then separated, and the charged object moves away, we are left with two charged pieces of metal (Figure 22.5, right). Each piece now holds an opposite charge of equal magnitude. This process is known as charging by induction.

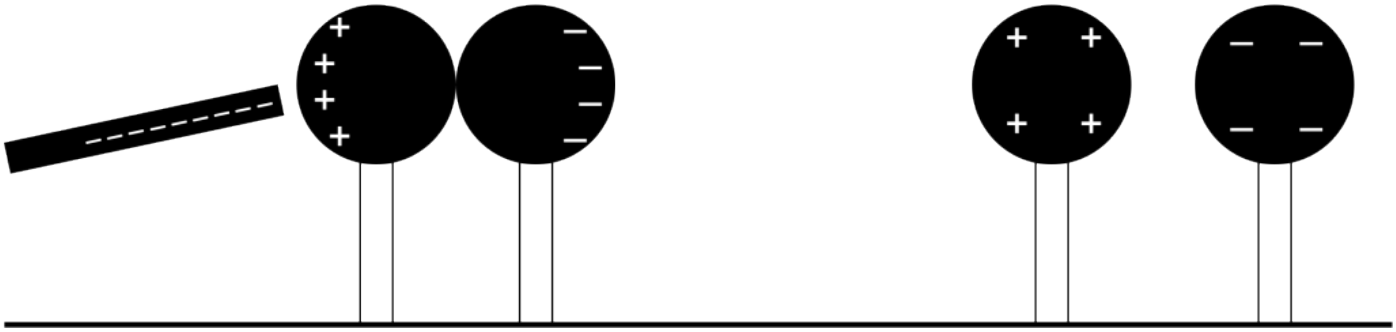


Figure 22.5 – To charge objects by induction, first bring a charged object near two touching conductors (left). Then, physically separate the two conductors and remove the charged object (right). Even after removing the charged object, both conductors are now charged. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Induction in insulators (polarization)

Induction can also occur in insulators, but the effect is slightly different due to the fact that electrons cannot move around freely in insulators. If we could look at the arrangement of the molecules surrounding the balloon sticking to the wall (as in the video earlier in this chapter), we would see that while the electrons cannot move, the molecules will become polarized. In effect, the molecules reorient themselves so that the negatively charged parts of the molecules are as far from the balloon's negative charges as possible. This is shown in Figure 22.6.

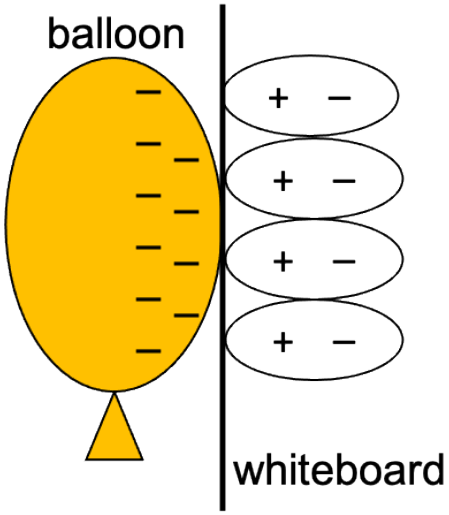


Figure 22.6 – The negative charges in the balloon cause the molecules in the whiteboard to realign themselves to minimize electrostatic repulsion. Note: this diagram is not to scale. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

These explanations of induction in metals and insulators explains a phenomenon we saw earlier in this chapter in the video of the charged rods attracting the meter stick: charged objects are attracted to neutral objects. Note in Figure 22.6 that the positive charges in the wall are closer to the negative charges on the balloon due to electrostatic attraction and that the negative charges in the wall are farther from the negative charges on the balloon due to electrostatic repulsion. Since the electric force between single electric charges varies inversely as the square of the distance between them, the force weakens more quickly than the distance increases. This means that the attraction is greater than the repulsion, so the balloon sticks to the wall. If a positively charged object is brought near the neutral wall, the atoms in the wall will polarize the opposite way. Either way, the attractive force outweighs the repulsion. Again, charged objects attract neutral objects.

Electric fields

Recall from chapter nine that fields are a model that allows us to see how objects in the presence of forces will interact. An electric field shows us how charged objects will interact as a result of electrostatic forces. We can draw an electric field by showing how a positively charged object would move as a result of being in the presence of another charged object.

The electric field of a charge distribution can be mapped by noting the trajectory that a small (so as not to affect the electric field of the charge distribution) positive test charge follows when placed at various points in space around the charge distribution. In the simplest case, we can draw the electric fields corresponding to single point charges. The electric fields shown in Figure 22.7 emanate symmetrically from a single positive charge. Because fields depict how a positive charge would move if placed near this single positive charge, they must depict the force of repulsion and indicate movement away from the positive charge. The arrows depicted on the field point away from the center as a result. Note that there are really an infinite number of electric field lines from any charge distribution and that any diagram will show just a few representative lines.

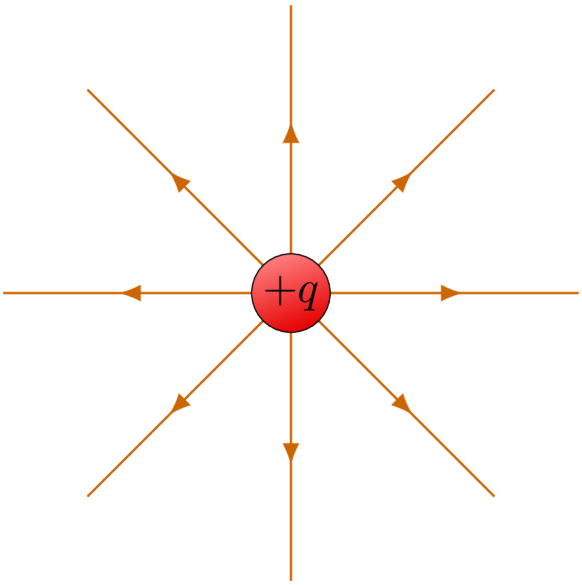


Figure 22.7 – A positive point charge has electric field lines that point radially away. Positive point charge, by Izaak Neutelings, is licensed under CC BY-SA 4.0.

The electric fields shown in Figure 22.8 surround a single negative charge. Because field lines depict how a positive charge should move if placed near this negative charge, they depict the force of attraction and indicate movement toward the negative charge. The arrows depicted on the field point toward the center as a result.

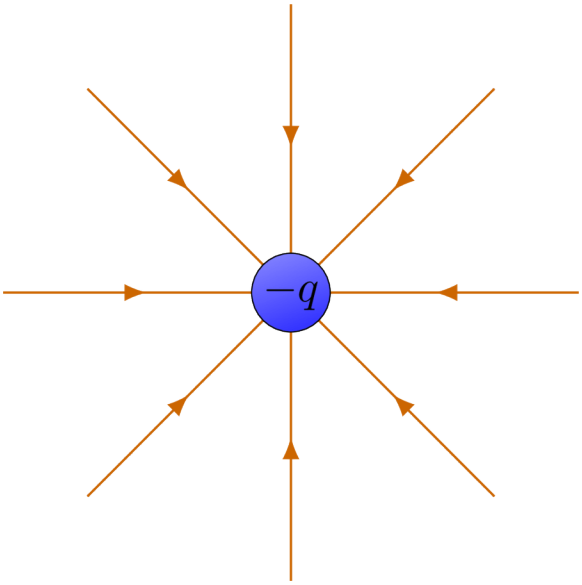


Figure 22.8 – A negative point charge has electric field lines that point radially toward the center. Negative point charge, by Izaak Neutelings, is licensed under CC BY-SA 4.0.

Electric field diagrams can be drawn for more complicated arrangements of charges, and serve as a useful model for us to determine how positive charges would move in these scenarios. (For example, the electric field lines between two attracting point charges are shown in Figure 22.9.) In addition, just as the gravitational forces are strongest when the gravitational field shows the arrows closest together, the electrostatic forces are strongest when the electric field arrows are closest together.

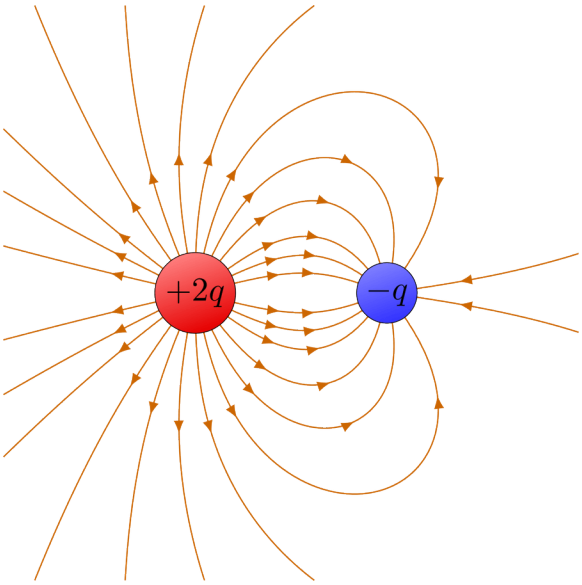


Figure 22.9 – Two attracting point charges have a more complicated electric field diagram. The field lines always point away from positive and toward negative charges. Attracting point charges, by Izaak Neutelings, is licensed under CC BY-SA 4.0.

Electric potential

Just as objects of mass in the presence of a gravitational field have gravitational potential energy, charged objects in the presence of an electric field have electric potential energy. This stored energy can be used to do useful work: charge your cell phone, turn on your lights, or run your refrigerator. We use EPE as the symbol for electric potential energy. Because it is a form of energy, we use joules as the unit. The equation for electric potential energy is

$$EPE = k\frac{q_1 q_2}{d}.$$

This describes the energy stored as the result of holding two charges (q_1 and q_2) a distance d apart from each other. The variable k in the equation is Coulomb's constant, defined earlier in this chapter.

Even if there are not two charges, the presence of a single charge creates electric potential, the possibility to create electrical energy. There doesn't have to be a second, or third, or fourth charge around to interact with the electric field caused by the first one. Electric potential is equal to

$$EP = k\frac{q}{d},$$

which can be seen as equal to EPE divided by charge.

Electric potential can also be defined as describing the amount of work it takes to move a charge through an electric field, divided by the amount of charge. This definition helps to describe the units for electric potential: joules (work) divided by coulombs (charge). A J/C is also equal to a volt (defined by the capital letter V). This concept of potential is very important in electronics. A difference in electric potential between two points is known as voltage.

The quantities of electric potential energy (EPE) and electric potential (EP) are analogous to the quantities of gravitational potential energy (GPE) and gravitational potential (GP). This is described in the table below.

	gravitational	**electric**
potential energy	$GPE = mgh$	$EPE = k\frac{qq_1}{d}$

potential	$GP = \frac{GPE}{m} = gh$	$EP = \frac{EPE}{q_1} = k\frac{q}{d}$

Practice questions

Numerical analysis

1. If charge A of +5 μC is placed 2 meters away from charge B of -3 μC, calculate the magnitude of the electrostatic force between the two charges. If charge A is located to the left of charge B, determine the direction that...
    1. ...charge A would move due to charge B.
    2. ...charge B would move due to charge A.
2. If charge A of +8 μC is placed 5 centimeters away from charge B of +12 μC, calculate the magnitude of the electrostatic force between them. If charge A is located to the left of charge B, determine the direction that...
    1. ...charge A would move due to charge B.
    2. ...charge B would move due to charge A.
3. Two electrons, each with a charge of -1.6×10^{-19} C, are separated by a distance of 1×10^{-10} m. Calculate the magnitude of the electrostatic force between them.
4. Three point charges are placed in a straight line with equal separations of 0.1 meters between each charge. The charges are +6 μC (left), -2 μC (middle), and +8 μC (right). Calculate the net electrostatic force on the...
    1. ...+6 μC charge due to the other two charges.
    2. ...-2 μC charge due to the other two charges.
    3. ...+8 μC charge due to the other two charges.

Hands-on experiments

1. Rub a piece of plastic with a paper towel or piece of cloth and then bring it close to small pieces of paper or a small piece of thread. Observe how the plastic attracts the paper or thread.
2. Charge a plastic comb (or other plastic object such as a balloon) by rubbing it with a cloth and then bring it close to a stream of water without touching it. Observe how the water stream bends towards the comb. This shows the attraction between the charged comb and the water.
3. Try the balloon on the wall experiment: Inflate a balloon and rub it against your hair or clothing to create a static charge. Then, stick the balloon to the wall and observe how it clings to the wall due to electrostatic forces.

This page titled Chapter 22: Electrostatics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 23: Electric circuits

Summary

Electric circuits are a ubiquitous part of our daily lives. This textbook was written using a computer, which was plugged into a wall outlet that provides voltage to the circuits inside the computer. Perhaps you are reading this book on a smartphone, tablet, or computer that uses circuitry to function. The smartphone or tablet is powered with a battery, and the computer plugs into the wall. Learning the fundamentals of electric circuits allows us to better understand the basics of how all of our electronics work.

Voltage

Voltage is a property of electric circuits that's very prevalent in our lives. Voltage has a symbol of the capital letter V. The symbol for voltage is the capital letter V, and the unit is the volt, which also has a symbol of the capital letter V, as described in the previous chapter of this textbook.

Voltage is a measure of electric potential, a property we discussed in the previous chapter. Usually, voltage is measured as a difference in electric potential between two points. This is why you may hear the term "potential difference" used in discussions of voltage.

Voltage can be created in different manners. Chemical means can be used to separate positive and negative charges, for example, in a battery. A battery is a source of constant voltage. AA, AAA, and D-cell batteries are commonly used in household items like flashlights. These output a constant 1.5 volts. A 9-volt battery generates 9 volts of potential difference. Coin cell batteries used in smaller electronics like calculators usually generate either 1.5 volts or 3 volts. A selection of batteries is shown in Figure 23.1.



Figure 23.1 – Batteries use chemical means to create an elecric potential (voltage). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Alternatively, a generator can be used to produce voltage. Power plants generate electricity using a process we'll discuss later in this textbook. The value of this voltage changes over time. In the United States, our power generation creates a voltage that fluctuates at a frequency of 60 hertz and that has an average voltage of 120 volts in our homes. A fluctuating voltage value is shown in Figure 23.2. (Download this data [XLSX, 171 kB]) It has the same general shape as the voltage we receive out of the electrical wall outlets in our homes.

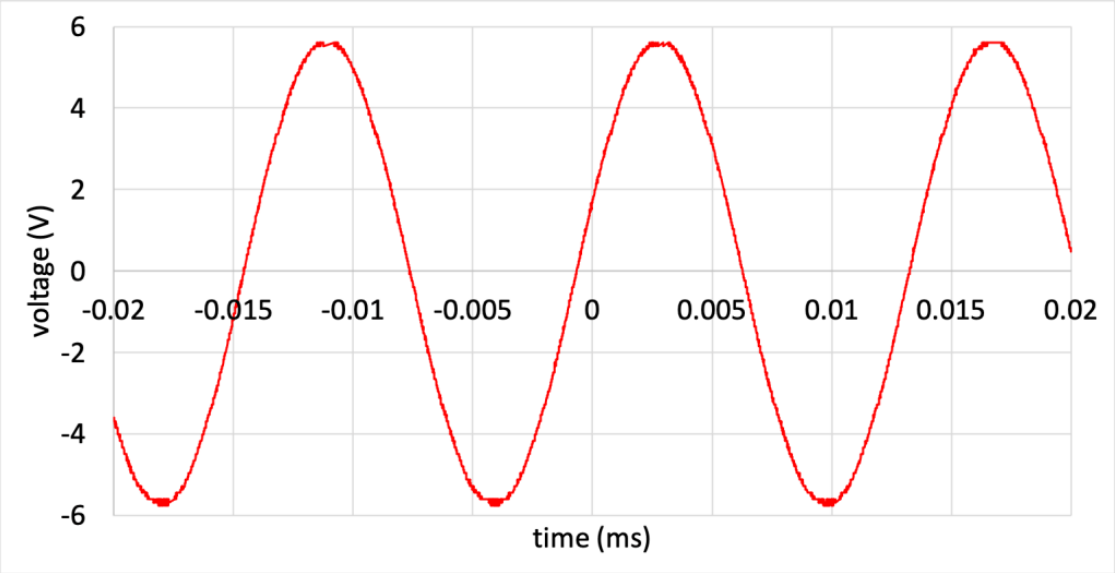


Figure 23.2 – A fluctuating voltage signal that looks similar to the voltage that comes out of the wall outlets in the United States. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Resistance

Resistance is the property of a material that impedes the flow of electrons throughout that material. The symbol for resistance is the capital letter R. The unit is the ohm, which has a symbol of the capital Greek letter Omega (Ω).

The resistance of a material is related to different properties of that material. One of the properties is the resistivity of that material. Resistivity has a symbol of the lowercase Greek letter rho (ρ). The unit used for resistivity is Ωm (an ohm times a meter). For pure substances, resistivity is a property inherent to that material. The smaller the resistivity of a material, the better conductor it is. For example, copper (an excellent conductor) has a resistivity of approximately $1.7 \times 10^{-8}\ \Omega\text{m}$ at room temperature. On the other hand, quartz (which is a very good insulator) has a resistivity of $7.5 \times 10^{17}\ \Omega\text{m}$.

Resistance is equal to the resistivity of an object times its length and divided by its surface area. In other words:

$$R = \frac{\rho L}{A}.$$

We can use this equation to determine the effects of each of these variables on resistance. The larger the resistivity, the larger the resistance. The longer the object, the larger the resistance. This makes sense as there is more material to move across. However, the larger the cross-sectional area, the smaller the resistance. This makes sense because a wider area offers more pathways for the electrons to move.

This equation is used to create electric circuit elements known as resistors. Resistors (see Figure 23.3) are used in electric circuits for many reasons: to create particular circuit properties and to amplify certain frequencies of electronic waves, among many other applications. They are frequently used by electrical and electronics engineers, and can be found in nearly all electric circuits.

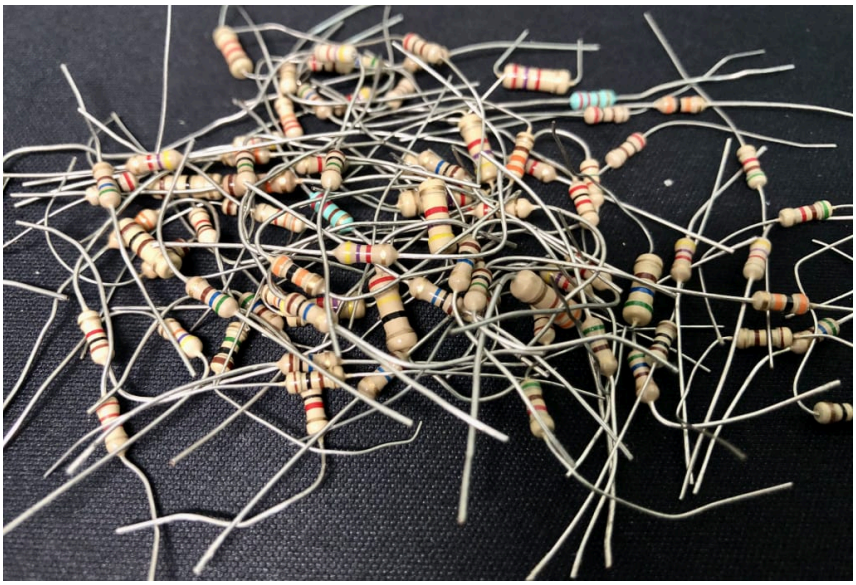


Figure 23.3 – A handful of resistors that all have varying resistance values. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Current and Ohm's law

Current describes the flow of electric charge through a circuit. Current flows in a circuit when there is a complete path including a voltage source. Using a battery, this means there must be a complete path between the positive and negative ends of the battery.

The parts of an atom that are capable of moving through a conductor are the electrons. Electrons will flow in a closed path. Because an electron has a negative charge, electrons will be attracted to the positive end of a battery. Electron flow therefore goes from negative to positive in a circuit.

Current has a symbol of the capital letter I. The units used for current is the amp (A), short for ampere. One amp is equal to one coulomb per second. This means that one amp of current is equal to one coulomb of charge moving past a given point every second.

Note that current is defined by the movement of positive charge in a circuit. We now know that positive charges (protons) do not move freely. While this definition is technically incorrect, it has become a standard for scientists and engineers. Current flows from positive to negative in a circuit. This means that the direction of current is opposite to the direction of electron flow in a circuit.

The relationship between current, voltage, and resistance is given by Ohm's law. Ohm's law states that

$$V = IR.$$

In other words, the voltage of an element (such as a resistor or a light bulb) in a circuit equals the current through that circuit element times the resistance of that circuit element. Ohm's law can be used to analyze an entire circuit or just to analyze particular sections or elements of a circuit. For example: Ohm's law can be used to calculate the voltage over one particular resistor in a circuit that contains more than one resistor. It can also be used to calculate the voltage over multiple resistors in a circuit.

As an example, if a circuit that includes a voltage source of 9 volts contains a resistor of 5,000 ohms, how much current flows through the circuit? First, rearrange the variables in Ohm's law to determine that current equals voltage divided by resistance. Then, solve for current.

$$\begin{aligned} I &= \frac{9\ \text{V}}{5000\ \Omega} \\ &= 0.0018\ \text{A} \\ &= 1.8\ \text{mA} \end{aligned}$$

Power

Power describes the rate at which energy is consumed in a circuit element. This is the same physical property that we discussed earlier in this textbook, which has a symbol of the capital letter P and units of watts. The equation for power in an electric circuit is

$$P = IV.$$

Power equals current times voltage.

Power can be generated in a circuit (batteries, power supplies, and generators can do this), and also consumed in a circuit (by resistors, light bulbs, and so on). Power is conserved, which means that all of the power consumed in a circuit is equal to all of the power generated in a circuit.

From a practical perspective, power is consumed by doing useful things in a circuit. An electric stove consumes electrical power and uses it to generate heat to cook food. A light bulb consumes electrical power and uses it to generate light. The tuning circuits in radios and televisions consume electrical power to allow us to listen to music or watch our favorite TV show. The electronics in your smartphone consume electrical power to operate the device.

Light bulbs are sold with a power rating in watts. That power rating tells us how much energy the light bulb will consume every second when it is in use. When comparing across identical types of light bulbs (for example, comparing between incandescent light bulbs), the power rating is proportional to the brightness of the light that the light bulb will emit. A photograph of multiple boxes of household light bulbs, and their power ratings, is shown in Figure 23.4 (on the left are incandescent light bulbs, on the right are LED light bulbs). Later in this textbook, we will discuss the different types of light emission used to create these light bulbs, and why exactly the incandescent light bulbs have larger power ratings than LED lights.



Figure 23.4 – Incandescent light bulbs with power ratings of 60 W and 100 W (left). LED light bulbs with power ratings of 4.5 W and 8 W. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

For now, it is sufficient to point out that incandescent light bulbs have larger power ratings than LED light bulbs of similar brightness due to the fact that incandescent light bulbs generate a lot of heat. Some of the electrical power going into the incandescent light bulbs will be emitted as light, with the rest being emitted as heat. LED light bulbs are much more efficient at converting electric energy to light. Furthermore, LED light bulbs tend to last longer than incandescent light bulbs. This efficiency is better for the environment and saves us money on our electric bills.

Speaking of electric bills, if you have ever paid one, you have probably noticed that the units used in the bills are kilowatt-hours. This is a unit of energy. Energy equals power times time. Because kilowatt is a unit of power, and hours are a unit of time, a kilowatt-hour is a unit of energy. One kilowatt-hour is equal to consuming one thousand watts of power during the time period of an entire hour.

Our electric bills charge us for the electrical energy we consume every month. For reference, Figure 23.5 shows the electrical

energy used by Dr. Pasquale in one year, split into months. Note that she uses more electrical energy in the summer when running an air conditioner to cool her home, and less electrical energy in the winter, when she uses natural gas (a different source of energy) to heat her home.

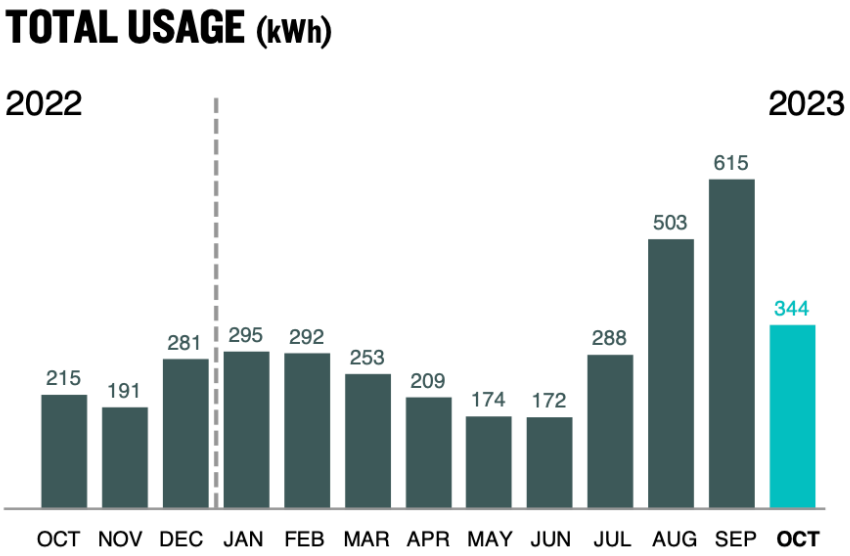


Figure 23.5 – Electrical energy consumed by Dr. Pasquale's household in 2022-2023. Note that this value decreases in winter when natural gas is used to heat her home, and increases in summer when electric-run air conditioning is used to cool her home. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Circuits

A circuit is a closed path where current flows. Circuits can consist of different circuit elements: voltage sources (such as batteries), resistors, light bulbs, motors, and so on. Some circuits have only a few elements. Some circuits have many elements and are much more complicated. All of the circuits described in this textbook can be analyzed using Ohm's law, conservation of energy, and conservation of charge. Electrical engineers take classes in circuit analysis to learn how to design circuits with specific properties.

Circuit diagrams

A circuit can be drawn as a circuit diagram. This uses symbols for each circuit element (shown in Figure 23.6) to represent how the circuit is connected. A circuit diagram acts as a kind of "blueprint" for an engineer or technician to build or fix a circuit, or to learn about how a circuit works.

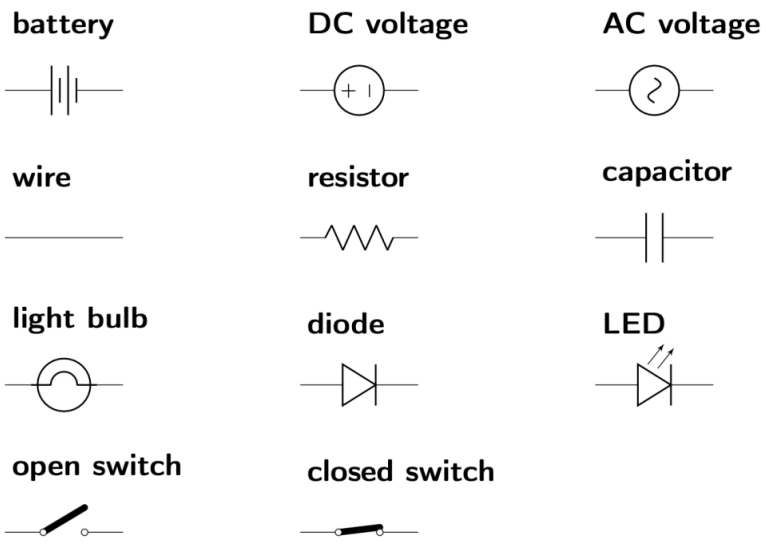


Figure 23.6 – Circuit symbols for common circuit elements. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

For example, a simple circuit can consist of a battery and a resistor, with wires used to connect the battery and resistor in a closed path. The diagram of this circuit is shown in Figure 23.7. If the battery and resistor have particular values, those values would be included on the circuit diagram next to the corresponding circuit element's symbol.

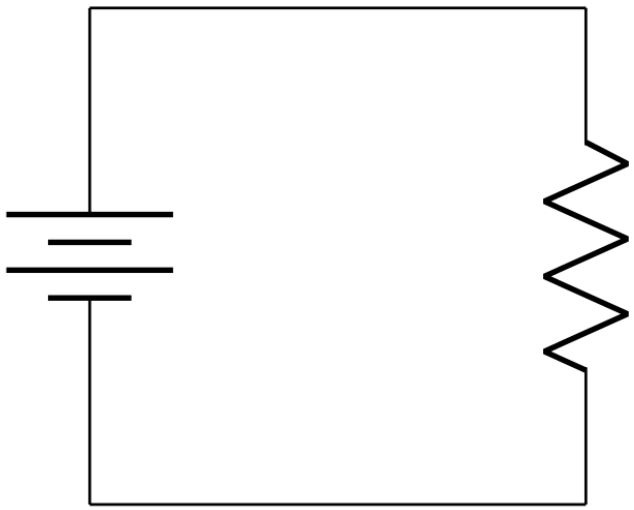


Figure 23.7 – A simple circuit diagram showing a battery connected with wires to a resistor. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Given this circuit diagram (and knowing the value of the battery and resistor), we could then use Ohm's law to calculate the value of the current through the resistor and the power equation to determine how much power is output from the battery and therefore consumed by the resistor.

While there are many different types of circuits, for the purposes of this textbook, the two primary configurations of circuit elements are series and parallel.

Series circuits

A series circuit is one where all of the circuit elements are contained in a single path. (It is possible to have a more complicated

arrangement of circuit elements, but in this case, we will only consider single elements that are all arranged in series.) Because there is only one path, current must flow through all of the elements contained in that path. Due to conservation of charge, every element in that path will experience the same amount of current because any charge that flows into a circuit element must flow out.

Because there is only one path for current to flow through, when two or more circuit elements (such as the light bulbs) are connected in series, if one of them becomes broken or disconnected and disrupts the flow of current, the other circuit elements will also become disconnected.

Consider the series circuit consisting of a battery and two light bulbs, as shown in Figure 23.8.

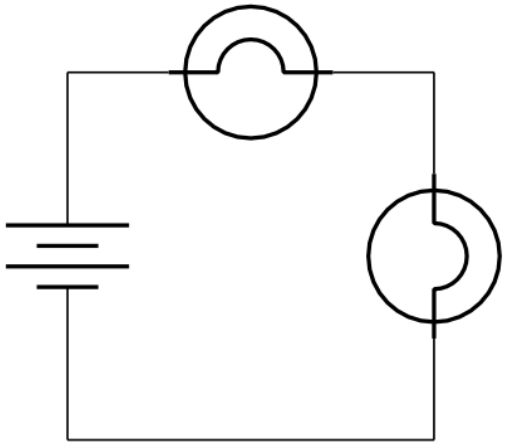


Figure 23.8 – Two light bulbs connected to a battery in a series arrangement. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale demonstrates this. When she removes one of the light bulbs from the circuit, the other one turns off. This is true regardless of which light bulb is disconnected from the circuit. As soon as the light bulb is connected again, the light bulbs turn on again. (Note that the battery is not shown in the video, but is connected to the light bulbs by the thick red and black wires.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Many cheap holiday lights are like this. If one of the light bulbs breaks, the rest of the light bulbs turn off as well. Dr. Pasquale demonstrates this in the video below by removing one of the light bulbs in a strand of holiday lights by taking one of the light bulbs out. This turns off all of the other light bulbs in the strand of lights.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In a series circuit consisting of two or more resistors, because current will be forced to flow through all of the resistors in order to travel through the circuit, the total resistance of the circuit will be greater than the resistance of any of the individual resistors. In a series circuit, the total resistance of the circuit is equal to the sum of each individual resistor in the circuit. That is,

$$R_{total,series} = R_1 + R_2 + R_3 + ... + R_n.$$

Due to conservation of energy, the voltage experienced by each resistor in a series circuit (containing two or more resistors) will be smaller than the voltage of the battery. If two series resistors have the same value, then the voltage drop over each resistor will be one half the battery voltage. If three series resistors have the same value, the voltage drop over each one will be one third the battery voltage.

In a series circuit with unequal resistor values, the resistor with the largest resistance value will have the largest voltage drop. This is because the current in a series circuit is equal throughout the circuit (as mentioned above), and Ohm's law states that voltage is proportional to resistance, all other things being equal. Each individual voltage drop will be equal to the current through the series circuit multiplied by the resistance of the individual resistor, as described by Ohm's law.

Regardless of the values of the number of resistors or resistance values of each resistor, the individual voltage drops over each resistor will be equal to the voltage provided by the source. That is,

$$V_{source} = V_1 + V_2 + V_3 + ... + V_n.$$

Consider a series circuit consisting of a 6 V battery and three identical resistors, shown in Figure 23.9.

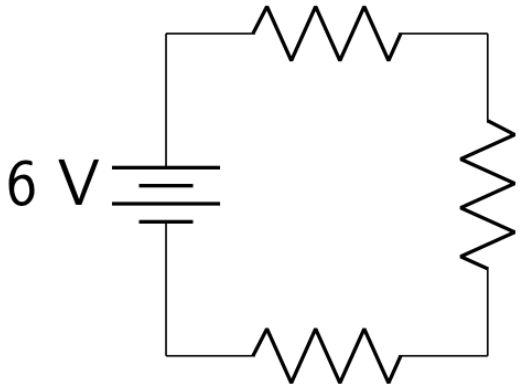


Figure 23.9 – Three identical resistors connected to a battery in a series arrangement. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale has built this circuit. She uses a device called a multimeter to measure the voltage that the battery produces (6 V), and the voltage over each individual resistor. Each resistor has a voltage drop of 2 V, shown by the multimeter, verifying that the voltage is evenly split between the identical resistors.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

It is possible to calculate how much power is consumed by each resistor in a series circuit by using a combination of Ohm's law and the power equation. In the power equation, the individual voltage of each resistor is not necessarily known. However, we know from Ohm's law that the voltage of any individual circuit element is equal to the current through that element times the resistance of that element. This allows us to calculate that

$$\begin{aligned} P &= I(IR) \\ &= I^2R. \end{aligned}$$

That is, in a series circuit, the power consumed by any individual resistor is equal to the current through the circuit, squared, times the resistance of the individual resistor. We can calculate the current through the circuit by using Ohm's law. The current will be equal to the voltage source divided by the sum of all of the resistances in the circuit:

$$I = \frac{V}{R_1 + R_2 + R_3 + ...}.$$

Parallel circuits

A parallel circuit is one where circuit elements are all contained in their own path in the circuit. In other words, they are connected end-to-end, hence the name parallel. Because of conservation of energy, each circuit element that is connected in parallel will experience the same amount of voltage. If they are connected in parallel with a battery or voltage source, then that voltage will be equal to the battery or source voltage.

Because there are multiple paths for current to flow, when two or more circuit elements are connected in parallel, if one of them becomes broken or disconnected, the other circuit elements will continue to function normally.

Consider the parallel circuit consisting of a battery and two light bulbs, as shown in Figure 23.10.

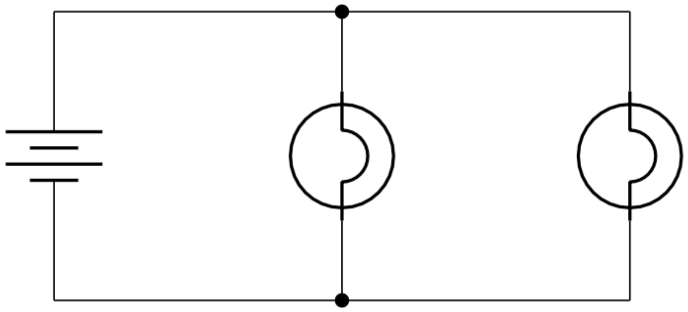


Figure 23.10 – Two light bulbs connected to a battery in a parallel arrangement. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale demonstrates this circuit. When she removes one of the light bulbs from the circuit, the other one remains illuminated. This is true regardless of which light bulb is disconnected from the circuit. (Note that the battery is not shown in the video, but is connected to the light bulbs by the thick red and black wires.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

An example of a parallel configuration of circuit elements would be our homes. They are wired in parallel. If you turn off the kitchen light, your refrigerator is still able to continue working. (Note that some aspects of home wiring are in series, which is why light switches can be used to turn off lights. But just because you turn off the lights with a switch does not mean that power is disconnected to other parts of your home.)

In a parallel circuit consisting of two or more resistors, because current can divide up and flow through multiple paths as it travels through the circuit, the total resistance of the circuit will be less than the resistance of any of the individual resistors. In a parallel circuit, the **reciprocal** of the total resistance of the circuit is equal to the sum of the **reciprocal** of each individual resistor in the circuit. That is,

$$\frac{1}{R_{total,parallel}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + ... + \frac{1}{R_n}.$$

This equation confirms the statement above that the total resistance decreases in a parallel circuit (when compared to each individual resistor value).

Due to conservation of charge, the current through each resistor in a parallel circuit (containing two or more resistors) will be smaller than the current through the battery. If two parallel resistors have the same value, then the current through each resistor will be one half the current through the battery. If three parallel resistors have the same value, then the current through each resistor will be one third of the current through the battery.

In a parallel circuit with unequal resistor values, the resistor with the largest resistance value will have the smallest current. This is because the voltage in a parallel circuit is equal throughout the circuit (as mentioned above), and Ohm's law states that current is inversely proportional to resistance, all other things being equal. Each individual current will be equal to the voltage across the parallel circuit divided by the resistance of the individual resistor, as described by Ohm's law.

Regardless of the values of the number of resistors or resistance values of each resistor, the individual currents through each resistor will be equal to the current provided by the source. That is,

$$I_{source} = I_1 + I_2 + I_3 + ... + I_n.$$

Consider a parallel circuit consisting of a battery and three identical resistors, shown in Figure 23.11. (In this case, it does not necessarily matter what the voltage of the battery is.)

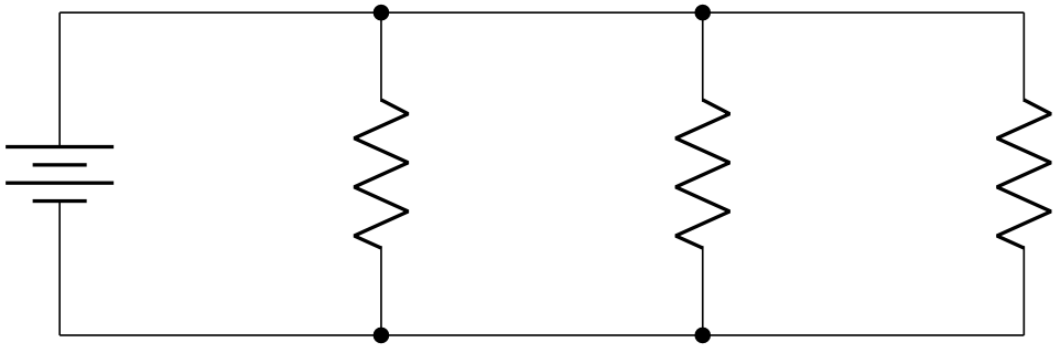


Figure 23.11 – Three identical resistors connected to a battery in a parallel arrangement. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale has built this circuit. She uses a multimeter to measure the current that flows through the battery (1.8 mA), and the current through each individual resistor. Each resistor has a current of 0.6 mA, shown by the multimeter, verifying that the current is evenly split between the identical resistors.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

It is possible to calculate how much power is consumed by each resistor in a parallel circuit by using a combination of Ohm's law and the power equation. In the power equation, the individual current through each resistor is not necessarily known. However, we know from Ohm's law that the current through any individual circuit element is equal to the voltage over that element divided by the resistance of that element. This allows us to calculate that

$$P = V(\frac{V}{R})$$
$$= \frac{V^2}{R}.$$

That is, in a parallel circuit with all circuit elements wired in parallel with the battery or voltage source, the power consumed by any individual resistor is equal to the voltage source of the circuit, squared, divided by the resistance of the individual resistor.

Open circuits

In an open circuit, there is no complete path for current to flow. Electrons do not move, and none of the circuit elements will operate. A switch can cause an open circuit. Sometimes this is useful, for example, when we want to turn the lights off in our bedrooms to go to sleep at night. A switch opens the circuit and prevents current from flowing through the lights.

A circuit consisting of a battery, switch, and a light bulb will be a closed circuit when the switch is closed (Figure 23.12, left). In this configuration, current will flow and the light bulb will light up. When the switch is opened, the circuit is an open circuit, no current will flow, and the light bulb will turn off (Figure 23.12, right).

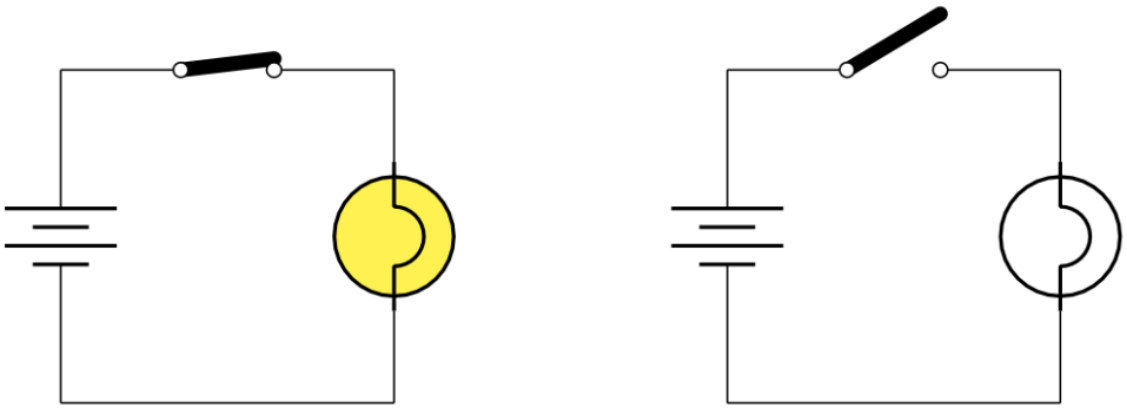


Figure 23.12 – When a switch connecting a battery to a light bulb is closed, current will flow and the light bulb will light up (left). When the switch is open, current will not flow and the light bulb will remain off (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This circuit is demonstrated in the video below. Dr. Pasquale starts with the circuit from Figure 23.12 with the switch closed, and the light bulb is illuminated. When she opens the switch, current no longer flows and the light bulb turns off. (Note that the battery is not shown in the video below, but is connected with the thick red and black wires.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Short circuits

A short circuit is a circuit with a very low resistance path that electrons can take. Electrons will proportionally travel through the path of least resistance. If there is a path of zero resistance, such as with a wire (which does not have exactly zero resistance, but its resistance is extremely small), then all (or most all) current will flow through that path. Because resistance is very low, current will become extremely high. If current becomes very high, so will power. This can cause a lot of heating, and eventually, wires and other circuit elements can melt. Short circuits can be very dangerous, and electrical engineers learn how to spot short circuits and avoid them when building circuits.

Consider the circuit in Figure 23.13 containing a battery, a light bulb, a switch, and wires to connect the circuit elements together. The light bulb and the switch are all connected in parallel with the battery.

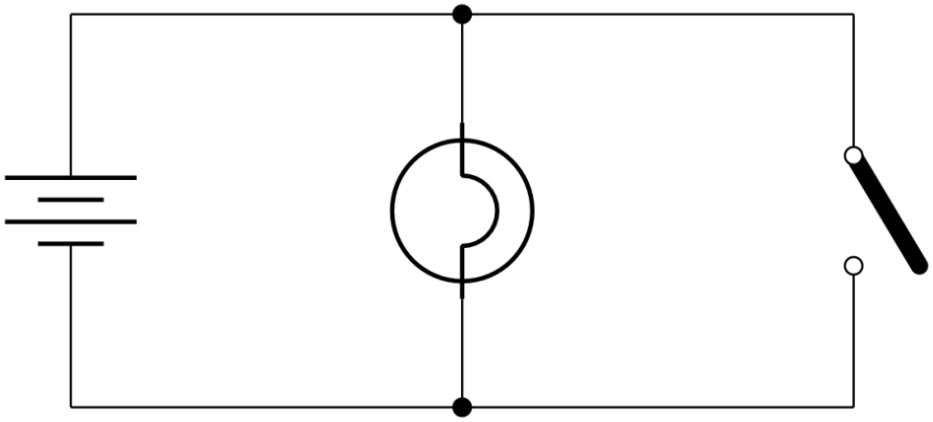


Figure 23.13 – A circuit diagram showing a battery, light bulb, and switch all connected to each other in parallel. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale uses the switch to cause a short circuit in the circuit from Figure 23.13, creating a low-resistance path that bypasses the light bulb. (When a switch is closed, it acts like a wire.) Note that the battery is not shown in the video, but is connected to the circuit through the thick red and black wires. When Dr. Pasquale closes the switch, current flows through the switch instead of through the light bulb, causing the light bulb to turn off. Dr. Pasquale can't keep the switch closed for too long, as the short circuit will drain the battery in a short amount of time. The wires and switch will also heat up, possibly melting the wires if it is left in this condition for too long.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Alternating and direct current

The voltage and current in a circuit can exhibit one of two different properties. One possibility is that the voltage and current stay constant over time. This is known as direct current or DC. In a DC circuit, our measurements of voltage and current will stay the same.

Batteries are capable of producing a DC voltage. AA batteries generate a constant 1.5 V. Most consumer electronics require DC to either function or to charge their batteries. Most of these devices contain a converter that generates the correct DC voltage to carry this out. (The basics of this converter will be explained at the end of this chapter.)

On the other hand, it is possible for voltage and current signals to change over time. Usually they will oscillate in a sinusoidal pattern, like a wave (as shown above in Figure 23.2). This is known as alternating current or AC. A device known as an oscilloscope can be used to record how voltage changes over time. This data was used to create the graph shown in Figure 23.2 which shows the voltage plotted with respect to time.

For reasons we'll discuss in a later chapter of this textbook, power plants generate alternating current. The wall outlets in our home therefore provide AC signals. The frequency of the voltage oscillation is 60 Hz in the United States, and the average value of the voltage is 120 V in the US. When traveling to other countries, it is important to bring an adapter to ensure that our electronic devices continue to function when plugged in to different values or frequencies of voltage.

Diode: one-way valve

A diode is a circuit element that acts like a one-way valve. A diode only allows current to flow through in one direction. (On the circuit diagram, the direction that current can flow is in the same direction that the triangle points.) When connected in a DC circuit, a diode can prevent current from flowing if it is connected backwards. A light-emitting diode, LED, is a diode that also emits light. When connected into a circuit in the forward direction, current flows through and the LED lights up. When connected backwards, current is blocked and the LED remains off. This is shown in the video below. (Note that the voltage source is not shown in the video.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In an AC circuit, a diode allows current to flow in only one direction. It transmits through the positive current and blocks the negative current. We can see this effect in the video below as the LED only lights up when it allows current to flow through, every half cycle.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

A circuit (shown as a circuit diagram in Figure 23.14) built with a regular diode (one that does not emit light) will block negative voltages from the AC signal.

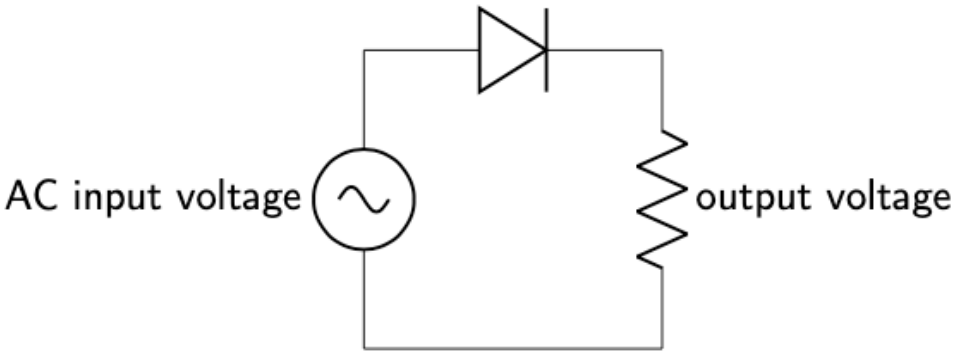


Figure 23.14 – A rectifier contains an AC voltage source and a single diode. The diode will block any negative voltages from passing through the circuit. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This circuit was built by Dr. Pasquale, and the input and output voltages were recorded on an oscilloscope and plotted in Figure 23.15. (Download this data [XLSX, 210 kB]) The AC input signal is red, and the output signal, after negative voltages have been blocked, is shown in blue. This circuit is called a rectifier. A rectifier is a circuit that only allows current to flow in one direction.

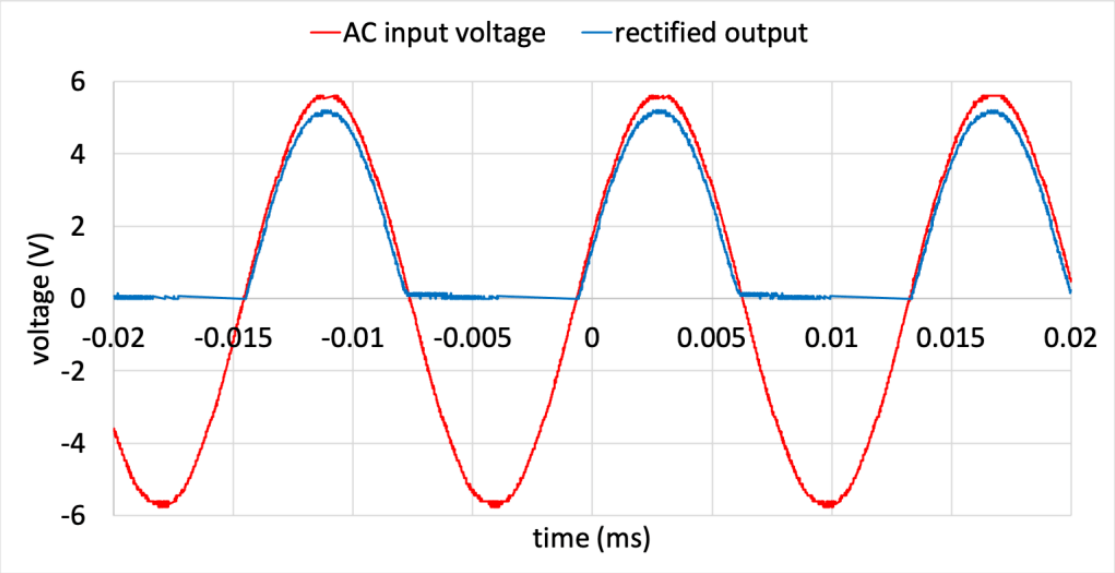


Figure 23.15 – A half-wave rectifier blocks all of the negative voltages from an AC input voltage (red curve) from reaching the output (blue curve). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Full-wave rectifier

Because every half cycle, the diode in Figure 23.14 blocks current, this would not be an efficient circuit to convert AC to DC. For that we can use a circuit called a full-wave rectifier. The configuration of diodes in a full-wave rectifier circuit (shown in Figure 23.16 as a circuit diagram) causes current to always flow through in one direction.

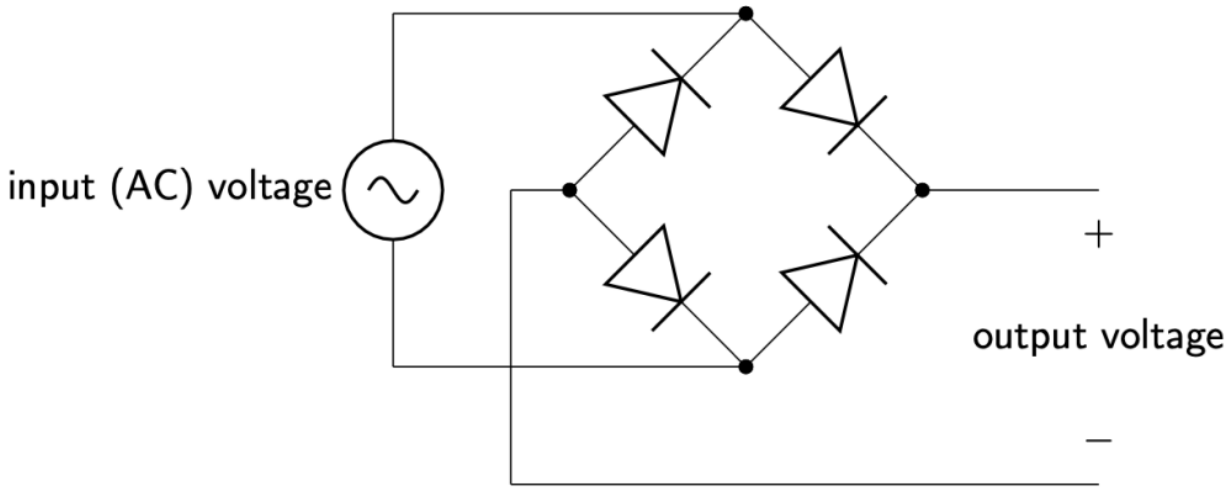


Figure 23.16 – A full-wave rectifier contains an AC voltage source and a configuration of diodes that causes the output voltage to always have a positive sign. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Figure 23.17 shows the AC input voltage (red curve) and the rectified output (blue curve). (Download this data [XLSX, 245 kB]) The output voltage is changing (so it is still AC), but it is now always positive. (The first half of the video below shows this same circuit.)

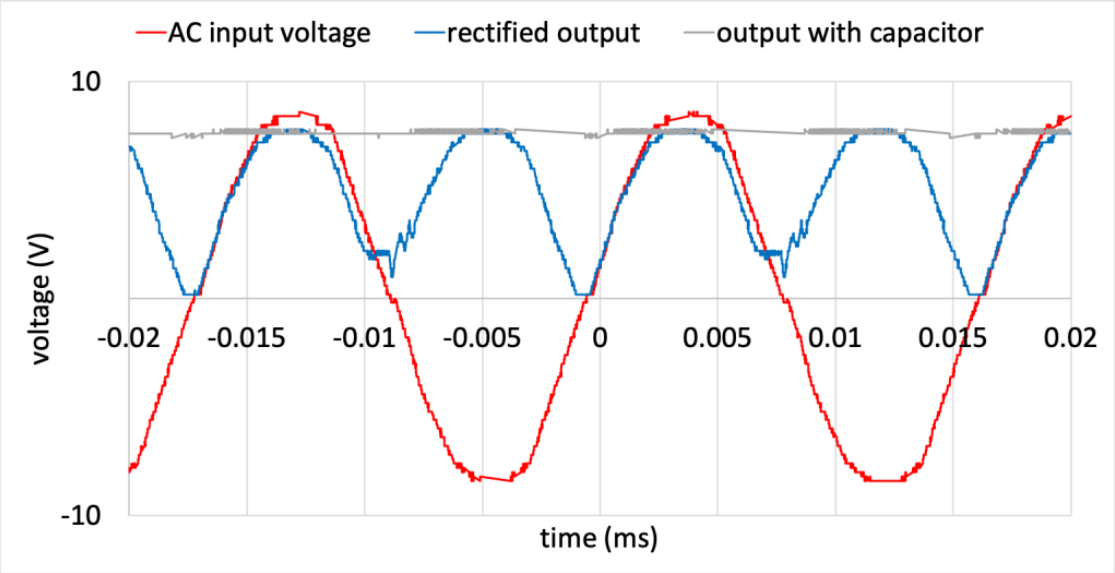


Figure 23.17 – A full-wave rectifier converts an AC input voltage (red curve) to all positive voltage values (blue curve). If a capacitor is used in the circuit, the voltage will become a constant DC value (gray curve). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Capacitor

To smooth out the oscillations in this output voltage and to obtain a nice smooth DC voltage, we use a device called a capacitor. In its simplest form, a capacitor is created from two plates of conducting material that allow equal and opposite charge to build up on them, as shown in Figure 23.18. In between the two plates is an insulator (charge cannot flow between the two plates through this insulator). When connected to a power source, positive charge will build up on one plate and negative charge will build up on the other plate. This leads to a storage of electric potential energy, which can be utilized if the power supply is later removed or disconnected from the circuit.

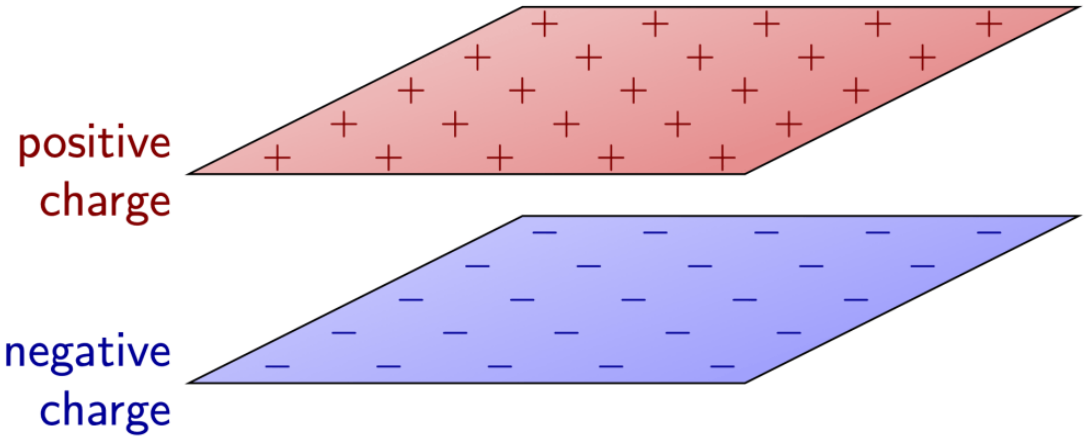


Figure 23.18 – The simplest type of capacitor consists of two plates separated by an insulator. Positive charge will build up on one plate, and negative charge will build up on the other plate, when connected to a power source. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Because a capacitor stores electric potential energy, it is capable of smoothing out changes in voltage in a circuit. This forms the basis of the converters used to charge DC-powered electronics that obtain power from a wall outlet. In the video below, Dr. Pasquale adds a capacitor to the full-wave rectifier, and the oscilloscope output shows the AC signal converting to a constant DC value. This data was measured and recorded in Figure 23.17, above, in the gray curve.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Further reading

- Resistor color codes (4-band) [PDF, 52 kB] – This chart explains how to read the colored stripes on a 4-band resistor to determine its resistance.
- Voltage, current, resistance, and Ohm's law – This website tutorial from SparkFun explains the basics of the concepts that are introduced in this textbook chapter.
- Why is it when one light goes out, they all go out? – So THAT Explains It! – Episode 12 – Dr. Pasquale contributed to this video as part of the "So That Explains It!" video series at the College of DuPage to explain why holiday lights stop working when one of the light bulbs in a strand breaks.

Practice questions

Conceptual comprehension

1. Consider the series circuit containing two light bulbs (call them A and B), shown in Figure 23.8, and the parallel circuit containing two light bulbs (call them C and D), shown in Figure 23.10. If these circuits are both built with an identical voltage source and identical light bulbs, compare the brightness of each of the four light bulbs...
    1. ...will A and B have the same or different brightness? Why?
    2. ...will C and D have the same or different brightness? Why?
    3. ...will the series light bulbs have a different level of brightness than the parallel light bulbs? Why or why not?
2. In a series circuit with three light bulbs, if one light bulb burns out, what happens to the brightness of the other light bulbs? How does this compare to a similar situation in a parallel circuit?
3. When you add more light bulbs to a series circuit, how does it affect the overall brightness of the light bulbs in the circuit compared to adding more light bulbs to a parallel circuit with the same power supply?

Numerical analysis

1. A resistor with a resistance of 100 ohms is connected to a 3 V battery. Calculate the current through the resistor.
2. If a 12 V battery is connected to a circuit, and the current in the circuit is 0.5 A, calculate the resistance of the circuit elements.
3. A light bulb has a resistance of 60 ohms. Calculate the voltage that is required to make it draw a current of 0.12 A.
4. If a circuit element is connected to a 120 V power source and draws a current of 5 A, calculate the power consumed by the device.
5. A 60 W light bulb is connected to a 240-volt power source. Calculate the current through the light bulb.
6. In a circuit, a resistor with a resistance of 40 ohms is connected to a 6 V battery. Calculate the power consumed by the resistor.
7. A series circuit consists of two resistors with values of 800 ohms and 1200 ohms. They are connected to a 10 V battery. Calculate...
    1. ...the total resistance of the circuit.
    2. ...the power consumed by the 800 ohm resistor.
    3. ...the power consumed by the 1200 ohm resistor.
8. In a series circuit, three resistors with values of 10 ohms, 15 ohms, and 20 ohms are connected in series with a 5 V battery. Calculate...
    1. ...the total resistance of the circuit.
    2. ...the power consumed by the 10 ohm resistor.

3. ...the power consumed by the 15 ohm resistor.
4. ...the power consumed by the 20 ohm resistor.

9. In a parallel circuit, two identical 100 ohm resistors are connected to a 5 V power source, creating a total current of 0.1 A flowing through the source. Calculate...
   1. ...the current through each resistor.
   2. ...the power consumed by each resistor.

Hands-on experiments

1. If you have one or two AA or AAA batteries, a small light bulb, and some wire, connect the wires and the light bulb in series with the battery. The light bulb should illuminate. If you have two batteries, connect them together in series with the light bulb and see how that affects the brightness of the light bulb. If you have different bulb sizes, try them out and see how it affects the circuit. (Note: this experiment will not work with household light bulbs, which generally require 120 V to work.)
2. If you have two AA or AAA batteries, an LED, and some wire, connect the circuit elements together in series and see how the LED must be connected to allow current to flow. Flip the LED backwards and see what happens.

This page titled Chapter 23: Electric circuits is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 24: Magnetism

Summary

Magnetism describes the properties of magnetic objects, which are capable of attracting and repelling other magnets. From using a magnet to post artwork and photos on a refrigerator, to navigating with a compass, to the protection that the Earth's magnetic field provides for us, magnetism plays a large role in our lives.

Magnetism

As we've learned, atoms contain a nucleus filled with protons and neutrons. Orbiting around the nucleus are electrons. The motion of these charged particles gives rise to magnetism. Most substances are not magnetic, which is because the magnetic effects of each atom cancel out, leading to an overall effect of no magnetism. When the magnetic effects of each atom don't cancel out, the effect is a permanent magnet.

In some materials, the magnetic effects of individual atoms couple to those of neighboring atoms causing regions of atoms to align their internal magnets. Such regions are called magnetic domains. These microscopic regions may contain many billions of atoms. Each domain can then be thought of as a tiny permanent magnet. In an ordinary bar of iron, the individual domains are randomly oriented, so that ordinary bar of iron has no overall magnetic behavior. However, by placing the iron in a strong magnetic field or by stroking it in one direction with a magnet, the directions of each magnetic domain can be rotated into alignment. If the domains are somewhat aligned, the bar will become weakly magnetic. If all the domains are aligned in one direction, the iron bar will become strongly magnetic with opposite poles on each end. It is also possible to go the other way. If a magnet is heated, or damaged by getting repeatedly dropped (for example), it is possible for these domains to be knocked out of their alignment, and the magnet can become demagnetized.

For example, gadolinium is a metal that acts as a permanent magnet at temperatures less than about 20 ^{o}C. At room temperature, gadolinium will stick to magnetic surfaces, attract, and repel other magnets, as we would expect from a permanent magnet. This is demonstrated by Dr. Pasquale in the video below. At first, the piece of gadolinium is capable of attracting to another magnet. After heating the gadolinium with a heat gun, the magnetic domains become randomized, and the gadolinium is no longer magnetic. The gadolinium loses its magnetic attraction and gravity causes it to fall down. (The magnetic domains will become realigned after it cools back down again.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Analogous to electrical interactions arising due to electric charges, magnetic interactions arise due to magnetic poles. The term poles was coined by Petrus Peregrinus after experimenting with magnets in the 13th century. In 1269, he carved a spherical magnet from an iron ore known as magnetite. Peregrinus noted that small pieces of iron placed on the sphere would align themselves along lines that pointed toward diametrically opposite sides of the sphere in much the same way that lines of longitude on the Earth meet at the geographic poles of the Earth. A magnetic pole is a region of a magnet where the forces involved in magnetic interactions are the strongest. Just as there are two types of charged particles (positive and negative), there are two types of magnetic poles: north and south. However, unlike electric charges, it is not possible to isolate an individual magnetic pole. Magnetic poles always come in opposite pairs. That is, every magnet contains both a north and a south pole.

Magnetic fields

The magnetic field describes the influence that a magnet has on other magnets and charged particles. We've described similar fields for gravity and electric charges. One of the interesting things about magnetic fields is that they can be easily visualized by using iron filings or a compass.

Iron filings, because they are magnetic, will align themselves along the magnetic field lines of a magnet. In Figure 24.1, iron filings are shown arranged around a bar magnet. (A bar magnet is a long rectangular type of magnet with opposite poles at each end; different types of magnets will be explained below.)

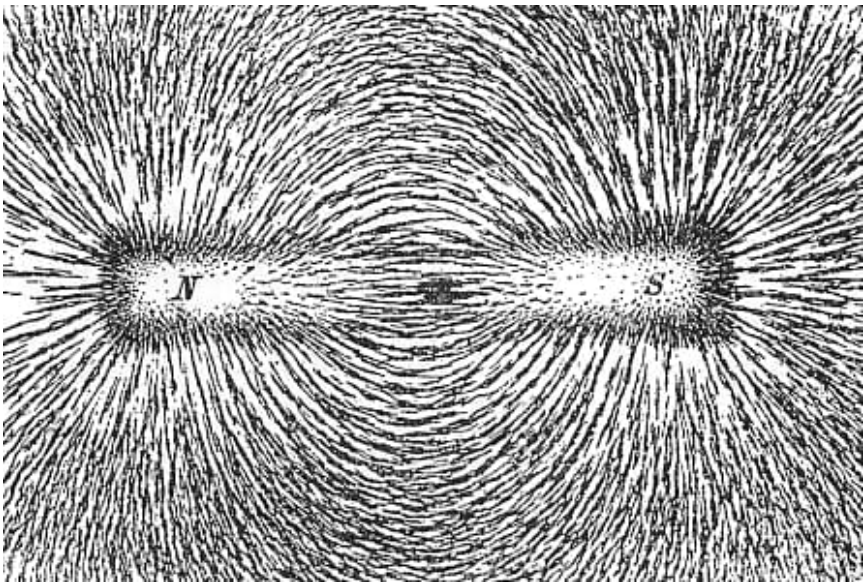


Figure 24.1 – Iron filings around a bar magnet will arrange themselves in line with the magnetic field of the magnet. In this case, they align from pole to pole (end to end) of the bar magnet. Magnet, by Newton Henry Black, is licensed under Public Domain.

For a single bar magnet, the magnetic field loops around the magnet. The convention of a magnetic field is that, outside of the magnet, the field points from north to south pole. Inside the magnet, the field points from south to north.

A compass is a magnet balanced on its center of mass in a fluid or on a low friction pivot. That magnet is free to rotate, and will align itself with the strongest magnetic field in the vicinity. Compasses are used to navigate, as the Earth has a magnetic field (which will be discussed at the conclusion of this chapter). The painted end of a compass (which corresponds to the north pole of the compass) orients itself to point in a northerly direction on the surface of the Earth. This implies that the Earth itself behaves like a giant bar magnet.

In the video below, when a compass is held close to a magnet, the magnetic field of the magnet is strong enough to overcome the Earth's magnetic field, and the painted end of the compass will point to the south pole of the magnet and away from the north pole of the magnet. We can see how the alignment of the compass changes as it is moved around a bar magnet.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Similarly to the attractive and repulsive forces involved with different types of charges (positive and negative), magnetic poles experience two types of interactions. Two like poles (north and north, or south and south) will repel. That is, if two like poles are held together, they will experience a force that causes them to move apart from each other. Two unlike poles (north and south)

will attract. Therefore, if two unlike poles are held apart from each other, they will experience a force that causes them to come together. This is demonstrated in the video below, where Dr. Pasquale uses a low-friction turntable to allow magnets to freely rotate based on this force of magnetic attraction or repulsion. The north pole of a compass needle aligns itself in a northerly direction. If the north pole of the compass is attracted to the northern hemisphere of the Earth, then the pole in the northern hemisphere must actually be the south pole of Earth's magnet.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Figure 24.2 depicts the magnetic field lines around attracting poles: a north and south pole. The magnetic field lines flow from one attracting pole to the next.

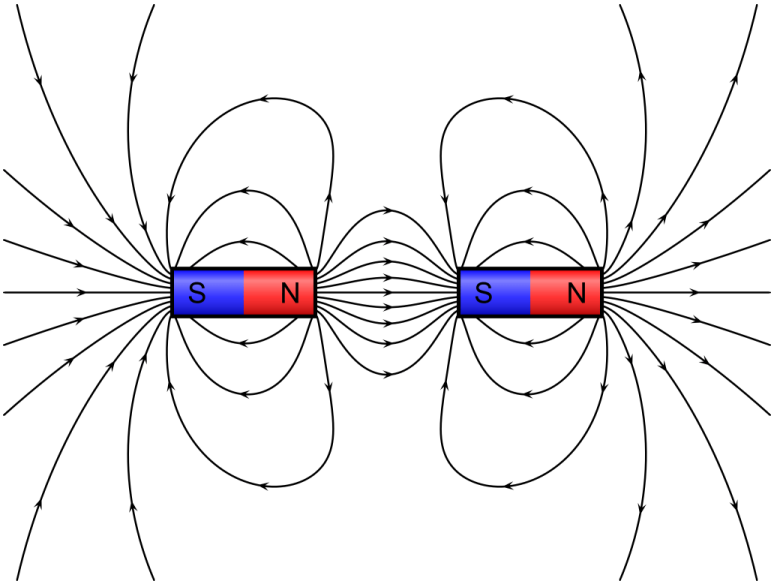


Figure 24.2 – Two attracting bar magnets have field lines that flow from the north pole of one magnet to the south pole of the other magnet. VFPt cylindrical magnets attracting, by Geek3 and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Figure 24.3 depicts the magnetic field lines around repelling poles: two north poles. The magnetic field lines flow away from the repelling pole.

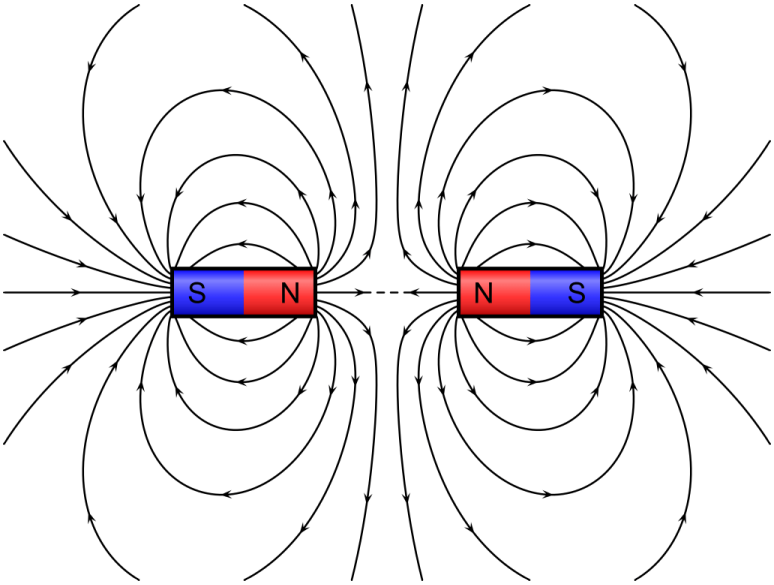


Figure 24.3 – Two repelling bar magnets have field lines that flow away from the repelling poles. VFPt cylindrical magnets repelling, by Geek3 and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

All magnets have both a north and a south pole. As far as physicists are currently aware, there is no way for a magnet to have only a north pole, or only a south pole. Magnetic poles always seem to exist in pairs. That is, the dipole (two equal and opposite poles) is the simplest magnetic structure. This leads to an interesting question: what happens if a magnet is broken in half?

In the video below, Dr. Pasquale has a bar magnet that contains a north and a south pole. When she splits it apart into two pieces, those two pieces each have their own north and south poles. Dr. Pasquale can prove that by using a compass to visualize the poles. This can additionally be demonstrated by showing the attraction and repulsion that exists between both magnet halves.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Types of magnets

Permanent magnets are made of different types of materials. Many low-cost magnets are made of iron oxide ceramics. Rare-earth metals are used to create very strong and small permanent magnets. Magnets also come in many different shapes and sizes. The videos used to demonstrate magnets in this chapter so far have used bar magnets. These are long rectangles with the north pole at one end, and the south pole at the other end.

A horseshoe magnet is essentially a bar magnet that has been bent into a U-shape. The north and south poles are still at opposite ends of the bar, but due to the bent shape, they are now located much closer together than they would be in a bar magnet. The resulting magnetic field is more concentrated in the region between the poles. The magnetic fields around a horseshoe magnet are depicted in Figure 24.4.

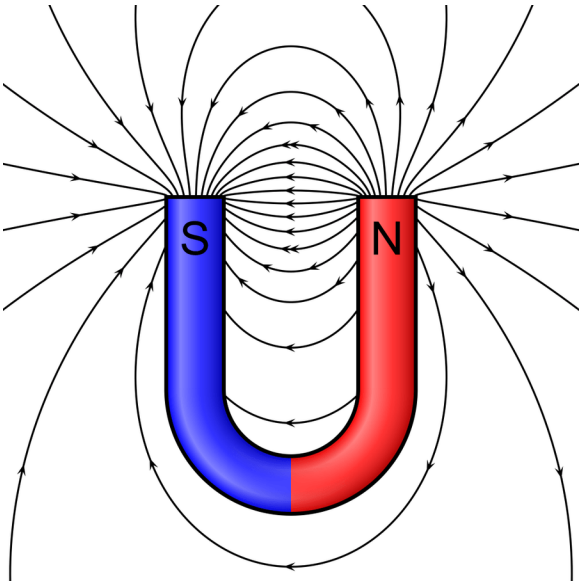


Figure 24.4 – A horseshoe magnet is simply a bar magnet bent into a U shape. This causes the two poles to be closer than they would be in a bar magnet. VFPt horseshoe-magnet, by Geek3 and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

A disk magnet is a short cylindrical magnet. The poles of the disk magnet are located at the top and bottom surfaces of the disk. The magnetic fields around a disk magnet are depicted in Figure 24.5.

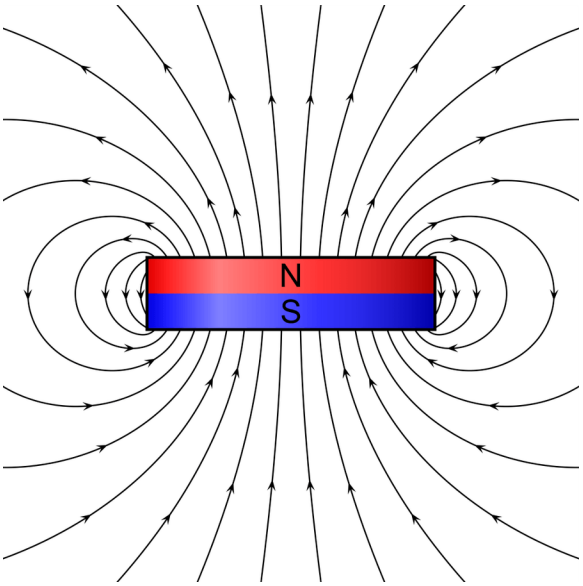


Figure 24.5 – A disk magnet is a short cylinder with a north pole at the top and south pole at the bottom. VFPt flat cylinder magnet potential, by Geek3 and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

The poles of a disk magnet can be visualized by using a compass, shown in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Many rare-earth magnets, so named because they are made of rare-earth elements, are shaped into disk magnets. These rare-earth magnets are extremely strong, and care should be taken when handling them.

A ferrofluid is made up of liquid that can be attracted to a magnet. This is typically accomplished by suspending tiny coated magnetic particles in a liquid. Dr. Pasquale demonstrates the magnetic properties of a ferrofluid in the video below by holding a disk magnet near the ferrofluid.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Electromagnets

Permanent magnets have the advantage of being magnetic all the time, hence the name permanent magnet. However, sometimes there is a benefit in being able to control the magnetism of a material.

In 1820, the Danish physicist Hans Christian Oersted discovered (during a lecture demonstration, no less!) that current through a wire created a magnetic field. In the video below, Dr. Pasquale connected a wire to a battery through a switch. When the switch is open, no current flows through the wire. The compass points in its natural orientation. When Dr. Pasquale places the compass above the wire and closes the switch, the flow of current creates a magnetic field, causing the north pole of the compass to deflect. The direction of the north pole of the compass indicates the direction of that field. When she places the compass below the wire and closes the switch, we can see that the north pole of the compass points in the opposite direction, indicating that the magnetic field points in the opposite direction below the wire.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The magnetic field generated by a straight wire takes the shape of concentric circles around the wire. This means the field points in different directions above and below the wire. Figure 24.6 depicts this field for a wire viewed “head-on,” that is, with the wire depicted as perpendicular to the page (or screen, if you are reading this book on an electronic device). The current in this particular scenario is depicted as directed into the page. The direction of the magnetic field line curl can be determined by the “right-hand rule.” Using your right hand, point your thumb in the same direction as the current. Your fingers will then curl in the direction of the magnetic field around the wire. The direction of the magnetic field line curl will be in the opposite direction if the current flows out of the page.

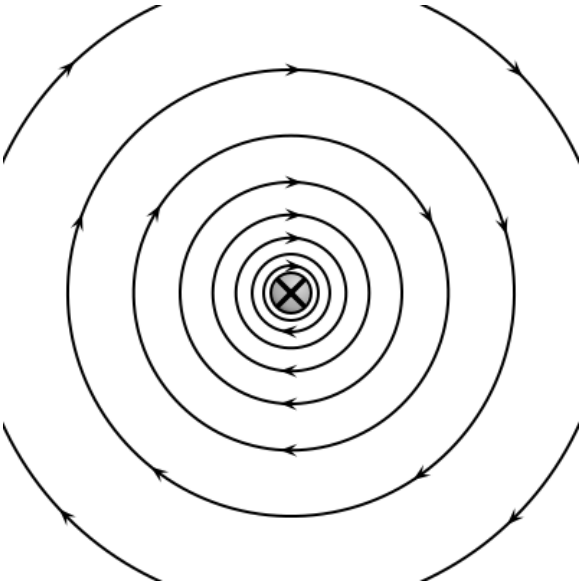


Figure 24.6 – The magnetic field surrounding a wire carrying current (into the page, in this case) is arranged in concentric circles, pointed clockwise. VFPt wire in, by Geek3, is licensed under CC BY-SA 3.0.

When the wire is coiled, it increases the strength of the magnetic field created by the current. This is because the coil consists of contributions of current from many hundreds of loops of wire. Dr. Pasquale demonstrates the magnetic field of a coil of wire in the video below. We can see how this field is stronger than the single straight wire by noticing how much more quickly the compass needle orients itself to the magnetic field for the coil as compared to the single wire.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

To create a strong electromagnet, an iron core can be placed into the center of the coil of wire. The current creates a magnetic field that magnetizes the iron core. The magnetism of the iron core then further increases the magnetic field of the electromagnet. Dr. Pasquale demonstrates this magnetic field in the video below. The compass reacts to the magnetic field even faster after the iron core is placed into the center of the coiled wire.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=999#oembed-9

An electromagnet can both attract and repel a bar magnet. Dr. Pasquale demonstrates this in the video below. This verifies the existence of both a north and a south pole in an electromagnet. We can determine which is which by seeing which end of the electromagnet attracts the north pole of a bar magnet, and which end of the electromagnet repels the north pole.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=999#oembed-10

One of the major advantages of an electromagnet is that it can be turned on and off. Dr. Pasquale demonstrates this in the video below by using an electromagnet to pick up paper clips, move them, and then turn the electromagnet off by disconnecting the power source. Electromagnets built using this principle are used to lift and move around objects such as cars or scrap metal.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=999#oembed-11

We saw in these video demonstrations that electricity can create magnetism. We'll learn in the next chapter of this textbook that magnetism can create electricity!

The deflecting force

An electric charge moving in a magnetic field experiences a deflecting force. This interesting force is even more evidence linking electricity and magnetism.

The moving electric charge, which is simply an electric current, will experience a force whenever the current is not parallel to a magnetic field. This force will act perpendicularly to both the direction of current and the magnetic field. This is why it is called a deflecting force. This force is different from other forces we've learned about so far in this textbook, forces such as gravity and the Coulomb force that cause motion along the same direction as the field lines.

Because this is a perpendicular force, we can use the right-hand rule to determine the direction of magnetic force. First, use your index finger to point in the direction of the electric current. Next, use your middle finger to point in the direction of the magnetic field. (The magnetic field points from north to south outside of the magnet.) Point your thumb now perpendicular to both the direction of the current and the direction of the magnetic field. The direction that your thumb points indicates the direction of the deflecting force. This is shown in Figure 24.7. It should be emphasized that the magnetic field in this situation is from some external source. (It is not the magnetic field created by the current in the wire.)

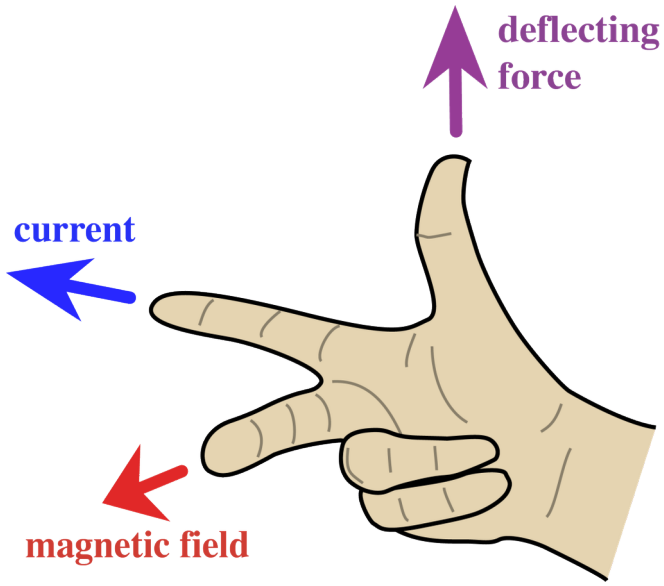


Figure 24.7 – The right-hand rule can be used to determine the direction of the magnetic deflecting force. Right hand rule cross product, by Acdx and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Let's say the current points in the direction pointing out of the page and toward you. If the magnetic field points down, then the deflecting force will point to your right. If the current points out of the screen and toward you, and the magnetic field flips to point upward, then the deflecting force will point to your left. Four different scenarios of the magnetic deflection force are demonstrated in the video below.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=999#oembed-12

The magnetic deflection force has many applications. A basic application is in motors. A direct current (DC) motor contains a battery, a coil of wire, and a magnet. The magnet creates a magnetic field, and the battery produces an electric current in the coil of wire. The deflecting forces produce a torque that causes the coil of wire to rotate, creating mechanical work. A DC motor is demonstrated in the video below.

One or more interactive elements has been excluded from this version of the text. You can view them online here: cod.pressbooks.pub/physics1100/?p=999#oembed-13

The deflecting force is also used in circular particle accelerators. To cause any object to move in a circular path, there must be a centripetal force. That centripetal force can be caused by this magnetic deflecting force acting on charged particles. The Advanced Photon Source (seen in Figure 24.8) at Argonne National Lab in Lemont, Illinois, is a particle accelerator.



Figure 24.8 – The Advanced Photon Source at Argonne National Lab in Lemont, IL, is a large circular building containing a particle accelerator. The magnetic deflecting force keeps particles traveling in a circular path around the perimeter of the building. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In France and Switzerland, the Large Hadron Collider is currently the world's largest particle accelerator, at least at the time of this textbook writing. This particle accelerator is used to simulate conditions that existed just moments after the Big Bang and carry out advanced physics research to help scientists answer questions about how our universe works.

Earth's magnetic field

As previously mentioned, if you wish to navigate without using GPS, a compass is a wonderful tool. The needle in a compass is a magnet, suspended in fluid, that will align itself to the strongest magnetic field in the vicinity. When outdoors, the north pole of a compass will align itself toward the north. Because opposite poles attract, the north pole of a compass attracting to the north indicates that the pole of the Earth in the northern hemisphere is actually the south pole of a magnet.

The Earth's magnetic field is believed to be generated by the currents of molten iron and nickel in the outer core of the planet. Not only does this magnetic field allow us to navigate by use of compass, but it also protects the inhabitants of our planet from solar winds and space weather.

Solar wind consists of charged particles, emitted from the Sun, that would cause the Earth's ozone layer to disintegrate. Without an ozone layer, we would not be protected from the Sun's harmful UV rays, and would become more susceptible to skin cancer.

The Earth's magnetic field deflects the solar wind, and while doing so, generates spectacular aurora (such as the aurora borealis, shown in Figure 24.9) in the process.



Figure 24.9 – The aurora borealis as seen in Iceland at night. Church of light, by AstroAnthony, is licensed under CC BY 4.0.

Further reading

- *Magnetic Storm* (Nova season 31, November 18, 2003) – This Nova episode gives a lot of information about the Earth's magnetic field: why it's important, and how it's changed throughout the millennia.
- Ancient trees show when the Earth's magnetic field last flipped out – This interesting news article from NPR gives some information about the last pole reversal, and how it was detected using trees.
- (Podcast) Magnets: the hidden objects powering your life – This 12-minute podcast gives a helpful introduction to magnets, magnetism, and why magnets are important.

Practice questions

Conceptual comprehension

1. What are the benefits of permanent magnets over electromagnets?
2. What are the benefits of electromagnets over permanent magnets?
3. If the north pole of the magnet inside a compass points to the north pole of the Earth's magnetic field, then what magnetic pole exists in the southern hemisphere of the Earth? Why?

Hands-on experiments

1. Build your own compass. This document [PDF, 108 kB] contains the required materials and instructions.
2. Build your own electromagnet. Using a nail, insulated copper wire, and a battery (such as a AA battery), wrap the copper wire around the nail and leave two loose ends. Connect the loose ends to the terminals of the battery. The nail becomes an electromagnet when the current flows through the wire. Test its magnetic strength by attracting small metal objects like paper clips. (Note: if you keep the wire connected directly to the battery for a long time, the current through the wire will eventually heat it up; to avoid this, disconnect the wire from the battery at intervals.)

This page titled Chapter 24: Magnetism is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 25: Electromagnetic induction

Summary

Electromagnetic induction is the production of an electric field from a changing magnetic field. This property plays a large role in our lives, as it is responsible for the generation of electrical energy and transmission of that energy from power plants to our homes. And as we'll learn in the next chapter of this textbook, this topic opens the door to an understanding of electromagnetism.

Faraday's law

Recall from the previous chapter of this book that it is possible to create magnetism from electricity. Moving charges can create a magnetic field. This is how electromagnets work.

As it turns out, it's also possible to generate electricity from magnetism. In the video below, Dr. Pasquale has connected a coil of wire to a galvanometer, which measures electric current. As she moves a magnet into the coil of wire, the galvanometer indicates that current is flowing through the circuit.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Whenever there is a moving magnetic field in the vicinity of an electric circuit, that changing magnetic field will induce a voltage. When there is a closed circuit, that voltage will create a current, as described by Ohm's law.

Faraday's law of induction can be used to quantify the strength of this induced voltage. Faraday's law is a complicated equation (requiring calculus), so we will not use the exact equation in this textbook. Instead, we will learn the qualitative properties of Faraday's law.

Faraday's law states that the strength of the induced voltage is proportional to the strength and rate of change of the magnetic field, the number of coils in the wire, and the cross-sectional area of the coil of wire.

The stronger the magnetic field of the magnet being used, the larger the induced voltage will be. This effect is demonstrated in the video below where Dr. Pasquale uses small and large magnets to generate induced voltage. The two magnets are otherwise identical except for their size (they have the same area and are made of the same magnetic material), so in this case, the size of the magnet is a good proxy for strength. The galvanometer deflects much less (as there is much less current as a result of the smaller induced voltage) for the small as compared to the large magnet. Dr. Pasquale did her best to control all other variables (the coil is the same for both trials, and she did her best to move both magnets at the same speed).

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The faster the motion of the magnetic field (with respect to the coil of wire), the larger the induced voltage. This relative motion between the magnet and the coil creates a rate of change of the magnetic field. An increasing amount of magnetic field penetrates the coil when the magnet and coil move toward each other and a decreasing amount of magnetic field penetrates the coil when the magnet and coil move away from each other. This is demonstrated in the video below. When Dr. Pasquale moves the magnet into the coil slowly, the galvanometer registers a small amount of current. When she moves the magnet faster, the galvanometer registers a larger amount of current. Note that if the magnetic field does not move, there will be no induced voltage, and therefore no current.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Faraday's law tells us that the magnetic field must be changing within the coil of wire to induce any voltage. If the magnet is held stationary, then the coil of wire can be moved to create this relative motion.

The induced voltage is also proportional to the number of turns and cross-sectional area of the coil of wire. A coil with one or two turns will not experience as much induced voltage as a coil with hundreds or thousands of turns. A coil with a small diameter will have less induced voltage than a coil with a large diameter.

In the video below, this property is demonstrated. As Dr. Pasquale does not have two coils of wire with identical cross-sectional areas and only differing number of turns, or with the same number of turns and different cross-sectional areas, it was not possible to control for either of those two variables. However, she does have a coil of wire with a small area and fewer turns than another coil, which has a large area and many more turns. As Dr. Pasquale moves the magnet, attempting to keep the speed as constant as possible, through both coils, the current measured by the galvanometer is greater in the larger coil with more turns.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The sign of the induced voltage has to do with three things: which pole of the magnet is used, if the magnet is entering or leaving the coil, and which side of the coil (left or right) it is moving toward or away from. By reversing any one of these three things, the sign of the induced voltage will become the opposite of what it was previously.

In the videos below, Dr. Pasquale creates several different scenarios of moving a magnet into and out of a coil of wire (using

different poles, entering and leaving the coil, and entering/leaving from different sides of the coil). As a baseline, Dr. Pasquale **inserts** the **north** pole of the magnet into the **right** side of the coil. This causes the galvanometer to deflect left, which indicates a negative voltage and current.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When Dr. Pasquale then **removes** the **north** pole of the magnet from the **right** side of the coil, the sign changes and the galvanometer deflects right, indicating a positive voltage and current. (One of the three conditions – insertion / removal – was changed from the baseline condition, causing the sign to switch once from negative to positive.)

Dr. Pasquale then changes the magnetic pole and **inserts** the **south** pole of the magnet into the **right** side of the coil. This causes the galvanometer to deflect right, which indicates a positive voltage and current. (Because one condition was changed from the baseline, the sign switches once from negative to positive.)

Then, Dr. Pasquale switches sides of the coil and **inserts** the **north** pole of the magnet into the **left** side of the coil. This causes the galvanometer to deflect right, which indicates a positive voltage and current. (Again, because one condition was changed from the baseline, the sign switches once from negative to positive.)

Note that changing two of these variables simultaneously will cause an effect that cancels itself out (this is similar to multiplying a value by negative one twice, the minus sign cancels out). This is demonstrated at the end of the above video when Dr. Pasquale removes the north pole of the magnet from the left side of the coil. This causes the galvanometer to once again deflect to the left, indicating a negative voltage and current. Changing three of these variables simultaneously will cause the current to change its sign from the baseline scenario.

Lenz's law

When a conductor moves through a magnetic field (or when a magnetic field moves around a conductor), that changing magnetic field creates an induced voltage. A voltage in a conductor will give rise to electric current. In this case, that current is known as an eddy current. Lenz's law defines the direction in which eddy currents will be generated. It states that the current induced in a conductor will itself generate a magnetic field that **opposes** the magnetic field that induced the current.

This effect can be demonstrated by dropping a magnet down a copper pipe. The falling magnet is a moving magnetic field, and the copper pipe is a conductor. The eddy currents generated in the copper pipe will oppose the magnetic field that created them. This causes a repulsion effect between the two magnetic fields, causing the magnet to move much more slowly than we would expect due to the acceleration from gravity. This is demonstrated in the video below. First, a small piece of nonmagnetic metal is dropped down the copper pipe to demonstrate the motion of an object acting only under the influence of gravity. Then, a small magnet is dropped down the copper pipe. Its motion is much slower than that of the nonmagnetic object.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This effect of slowing motion has practical effects. Induction braking (also known as eddy current braking) uses the drag between

an external magnet and eddy currents generated in a moving or rotating conductor to slow something down. Induction braking is used in vehicles such as high-speed trains. Because no parts directly touch during the induction braking process (as would occur during friction braking, where a brake pad creates a friction force to slow down a moving object), induction braking leads to less wear on parts.

The video below demonstrates this induction braking process. A pendulum consisting of an aluminum (conducting) plate is capable of moving back and forth with regular periodic motion. Once a magnet is placed in the path of the pendulum, the currents induced in the conducting plate are directed to oppose the changing magnetic field intercepting the plate. The induction braking process occurs as the eddy currents induced in the plate experience a magnetic force that slows the pendulum and brings it to an abrupt stop.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Generators and motors

The induction of electric potential (as discussed earlier in this textbook, voltage is a difference in electric potential between two points) in a circuit allows us to create generators. Coal, gas, and nuclear power plants work by heating water into steam. That steam spins a turbine that either moves a magnet around a coil of wire, or moves a coil of wire around a magnet. Wind turbines create this relative movement of a magnet and coil of wire by using the power of moving air to cause the blades of the generator to rotate. Hydroelectric power plants harness the power of falling water to create this motion. A schematic of a hydroelectric power plant is shown in Figure 25.1.

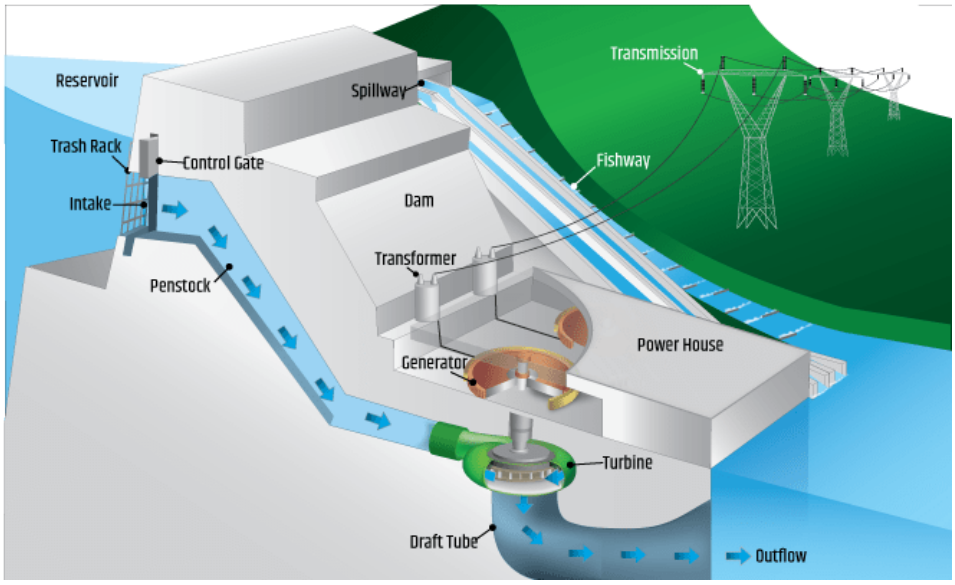


Figure 25.1 – A generator creates electricity by spinning a turbine. The relative motion between a magnetic field and coil of wire produces electricity. Hydroelectric power plant, by United States Department of Energy, is licensed under Public Domain.

In the video below, Dr. Pasquale created a generator out of a salad spinner and two coils of wire. Permanent magnets were placed at regular intervals around the perimeter. When Dr. Pasquale rotates the salad spinner, the magnetic fields rotate. That rotation causes motion, and when the magnets move past the coils of wire, voltage is generated.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When Dr. Pasquale connects the output of the generator to a multimeter, we can see the effect of the generator on the average current that is produced. The more slowly she spins the generator, the smaller the voltage and current that's generated. The faster she spins the generator, the larger the voltage and current that's generated. That makes sense based on Faraday's law of induction.

The voltage created by a generator is AC. As we saw when looking at how the sign of the induced voltage changes in Faraday's law, having a pole move toward or away from a coil of wire will cause a difference in sign. In that particular example, when the north pole moved toward the coil, the sign was positive. When the north pole moved away from that same coil, the sign was negative. In the case of the salad spinner, north and south poles are constantly rotating toward and away from the coils of wire, causing the sign of the induced voltage (and current) to periodically change sign. This is alternating current. This is shown for the salad spinner in the video below where Dr. Pasquale connected the output of the generator to an oscilloscope. (Note that the signal on the oscilloscope is AC but not sinusoidal. This is possibly due to the magnets each having a slightly different strength or not being equally spaced and therefore not generating a "clean" signal.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In summary, a generator is a device that converts mechanical work to electrical energy. The opposite process is also useful. A device that converts electrical energy to mechanical work is known as a motor. (We saw an example of a DC motor in the previous chapter in this textbook.)

The video below demonstrates the fact that a generator and motor are the same concept working in reverse. Dr. Fazzini and Professor Nikolova both hold handheld generators/motors. First, Professor Nikolova uses her handheld device as a generator (using mechanical work to generate electrical energy), which sends electrical energy to Dr. Fazzini's device, causing it to rotate as a motor would (using that electrical energy to generate mechanical work). Then the process occurs in reverse. Dr. Fazzini uses his device as a generator, which causes Professor Nikolova's device to act as a motor.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Transformers

A transformer is an electronic device that is capable of transferring energy from one circuit (known as a primary circuit) to another (known as a secondary circuit) with no physical connection existing between the two circuits. A transformer can also change the value of the voltage induced in the secondary circuit compared to the voltage in the primary circuit, either increasing it or decreasing it. This device brings together what we've learned in this and the previous chapter in this textbook: electric current flowing through a coil of wire generates a magnetic field. When that current is AC, the electric field changes, which generates a changing magnetic field. That changing magnetic field causes an induced voltage in any coils of wire that are nearby.

The symbol for a transformer used in a circuit diagram looks like two coiled objects that are near each other. The circuit diagram in Figure 25.2 depicts a primary coil (on the left) powered by an AC voltage source connected to one coil of a transformer. The secondary coil (on the right) is connected to a galvanometer which can show the output voltage induced in the secondary by the primary.

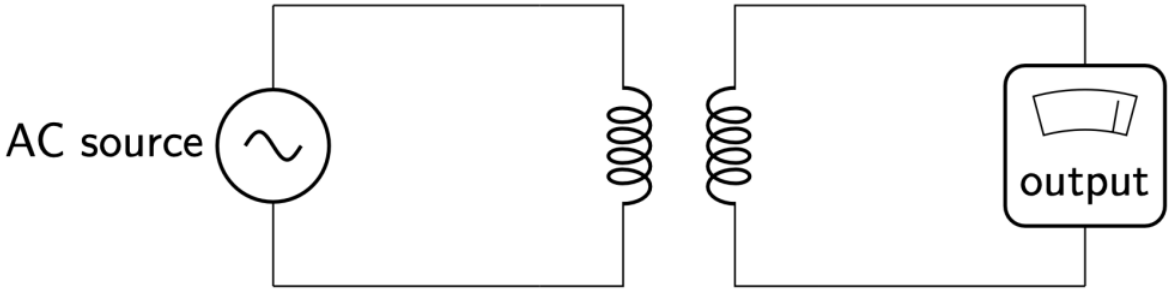


Figure 25.2 – A circuit diagram showing an AC voltage source and primary coil connected to each other in series. To the right of a primary coil is a secondary coil, connected in series with a galvanometer. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale has two coils of wire located near each other. One of the coils, called the primary, is connected to an AC voltage source (this is the large rectangular box located in the bottom right of the video). This generates a changing magnetic field in the primary coil. That changing magnetic field is large enough to overlap with the secondary coil, inducing voltage in the secondary coil. We can see that on the galvanometer connected to the secondary. Because the secondary is connected in a complete circuit, current flows. This current is also AC.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

While the primary and secondary coils are not physically connected, the effectiveness of the transformer is limited by how far away the two coils are positioned from each other. This is shown in the video below. As Dr. Pasquale separates the two coils of wire, the induced voltage decreases, demonstrated by the swings on the galvanometer becoming smaller in magnitude.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

At a certain point, if the secondary coil moves too far from the primary, the changing magnetic field no longer overlaps with the secondary, and no voltage is induced. This is not a long-distance phenomenon.

It's possible to increase the effectiveness of this process by placing an iron core into the two coils of wire. This increases the strength of the induced voltage because the magnetic domains in the iron core will be aligned by the magnetic field generated in the primary. That will cause a further boost to the strength of the magnetic field by increasing how well the primary and secondary devices are coupled together. This is demonstrated in the video below. When Dr. Pasquale places an iron core into both coils of wire, the swings on the galvanometer become much larger, demonstrating a much larger induced voltage.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In the video below, Dr. Pasquale connects the primary to a direct current (DC) source, a battery. In this case, the galvanometer on the secondary no longer deflects. After Dr. Pasquale closes the switch, the primary coil is connected to the source, as indicated by the glowing light-emitting diode inserted into the primary circuit to demonstrate that it is complete. Therefore, the primary circuit is generating a magnetic field. However, that magnetic field does not change over time. It is static. Faraday's law of induction tells us that the only way to generate electricity is to have a **changing** magnetic field. That requirement is not satisfied in this circuit. Therefore, no voltage is induced in the secondary and no current flows in the secondary.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Wireless induction can be extremely useful, and today we benefit from technology that uses short-range induction to wirelessly charge our smartphones and other devices. Dr. Pasquale's phone is seen wirelessly charging in Figure 25.3.

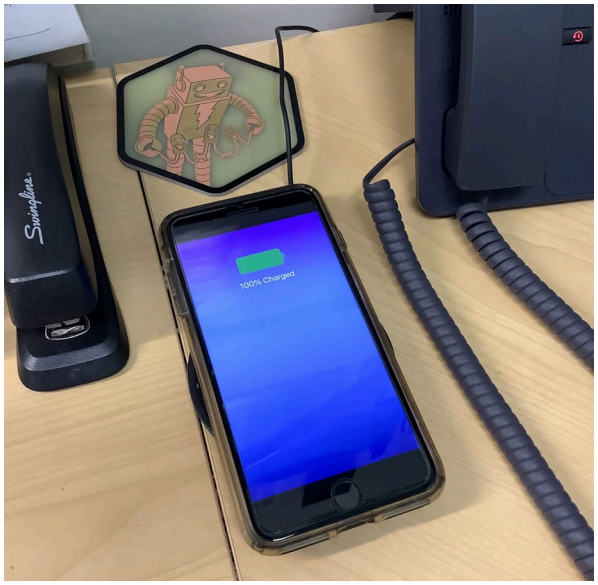


Figure 25.3 – A smartphone is capable of wirelessly charging using electromagnetic induction between two coils of wire. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Transformers aren't just useful for letting us charge smartphones and other devices. They are also an essential circuit component used to change the value of voltage between circuits. It is possible to create a step-up transformer that creates more voltage on the secondary than the primary. It is also possible to create a step-down transformer that creates less voltage on the secondary than the primary.

Whether or not a transformer is a step-up or step-down (or if it does not change the voltage at all) has to do with the number of turns of wire in each transformer. For an ideal transformer (one with no losses and experiencing perfect coupling between both coils of wire), in equation form, this can be written as

$$\frac{V_P}{N_P} = \frac{V_S}{N_S}.$$

The voltage on the primary (V_P) divided by the number of turns in the primary coil (N_P) equals the voltage induced in the secondary (V_S) divided by the number of turns in the secondary coil (N_S).

For example, if a primary coil has 6000 turns and is powered by a 120 V AC source, and the secondary coil has 2000 turns, we can calculate the induced voltage on the secondary coil:

$$\begin{aligned} V_S &= N_S \frac{V_P}{N_P} \\ &= 2000 \text{ turns} \frac{120 \text{ V}}{6000 \text{ turns}} \\ &= 40 \text{ V}. \end{aligned}$$

This is a step-down transformer, because the voltage on the secondary is less than the voltage on the primary.

If, instead, a primary coil has 500 turns and is powered by a 5 V AC source, and the secondary coil has 5000 turns, we can calculate the induced voltage in the secondary.

$$\begin{aligned} V_S &= 5000 \text{ turns} \frac{5 \text{ V}}{500 \text{ turns}} \\ &= 50 \text{ V} \end{aligned}$$

This is a step-up transformer, because the voltage on the secondary is greater than the voltage on the primary.

A step-up transformer sounds pretty great: it generates more voltage in the secondary than existed in the primary! At this point, you're probably very familiar with conservation laws (particularly, that energy cannot be created or destroyed). The voltage induced in the secondary coil does not come for free. We can use the power equation ($P = IV$) to determine what property is therefore

affected in the secondary. Power (the rate of energy consumption in a circuit element) will be equal in both the primary and the secondary under ideal conditions. In other words,

$$P_P = P_S$$
$$V_P I_P = V_S I_S$$

The product of voltage and current in the primary must be equal to the product of voltage and current in the secondary.

A helpful analogy may be to consider a transformer to be the electrical version of a mechanical lever. As previously discussed, a lever allows us to trade force for distance (and vice versa), allowing us to move a heavy output weight with a small input force, at the cost of moving a longer distance on the input than is achieved on the output. A transformer allows us to trade voltage for current (and vice versa).

If the primary device has a 5 V AC source, and contains 10 Ω of resistance, then we can use Ohm's law to calculate the current in the primary to be 0.5 A. The power consumed by the primary is 5 V times 0.5 A: 2.5 W. The power consumed by the secondary must be 2.5 W as well. If the voltage generated on the secondary is 50 V, we can calculate that the current flowing through the secondary will be limited to 0.05 A, or 50 mA.

Cables used to charge household electronics are frequently connected to a transformer. The voltage we have access to in homes in the United States is 120 volts AC. Most modern electrical devices require DC voltage at a much lower value, usually between 5 and 12 volts.

The large blocky part of our charging cables, colloquially referred to as a "wall wart" (four examples of which are shown in Figure 25.4), contains a step-down transformer to convert the 120 volts to a smaller value. Then, a rectifier is used to convert that AC voltage to a DC voltage, using a process we discussed in a previous chapter of this textbook. Finally, there is usually some noise-reducing circuitry inside that "wall wart" to ensure that no voltage spikes get through to harm our electronic devices.



Figure 25.4 – Four "wall warts," which are chunky plastic devices connected on a charging cable that step-down AC, rectify the signal, convert it to DC, and then clean out any noise. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

While we've discussed the transmission process in terms of voltages and coils of wire, in fact, it is the presence of changing electric fields that causes changing magnetic fields, and changing magnetic fields that cause changing electric fields. This induction process does not require the presence of any circuitry to occur. In fact, electromagnetic fields are all around us!

Power distribution and transmission

Now that we've learned about how power plants generate electricity, we can discuss how that electricity gets distributed to businesses, schools, and homes. The overview of this process is shown in Figure 25.5, which highlights the generation, transmission, and distribution process.

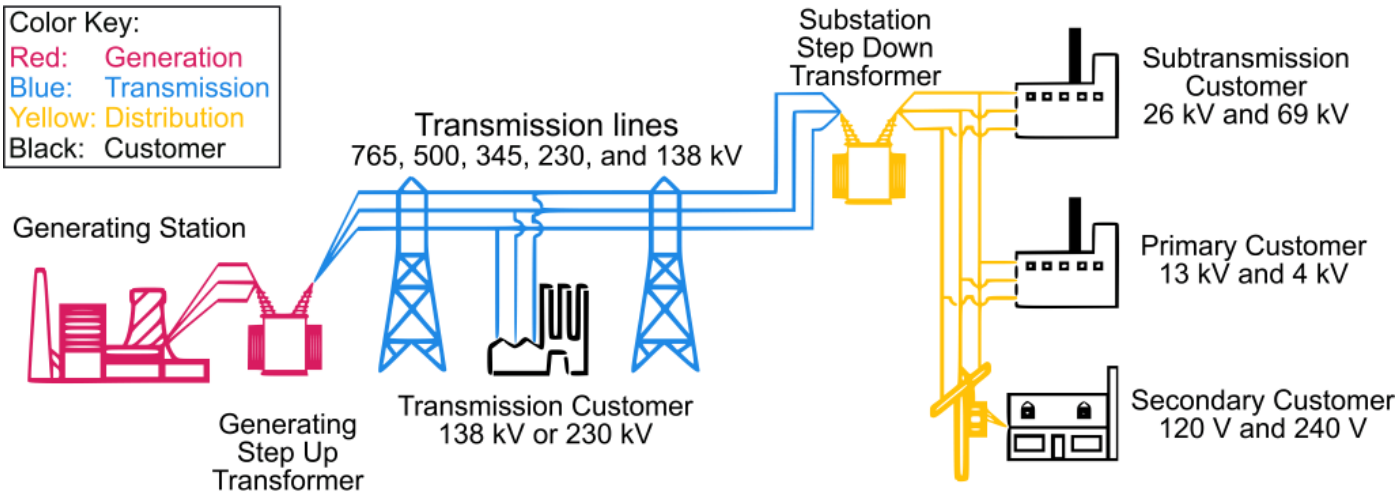


Figure 25.5 – Power is generated at a power plant and undergoes several changes in voltage as it gets sent to businesses and homes. Electricity grid simple North America, by the United States Department of Energy and modified by Alyssa J. Pasquale, Ph.D., is licensed under Public Domain.

The process starts at a power plant, most of which generate power using a turbine. Depending on the type of power plant (nuclear, coal, gas, or wind), the power generated by a power plant will be vastly different. As an example, the Byron Nuclear Generating Station in Ogle County, Illinois (shown in Figure 25.6), has a capacity of approximately two gigawatts (two billion watts of power).



Figure 25.6 – The Byron Nuclear Generating Station has a capacity to produce up to 2 billion watts of power. The big concrete structures in this photograph are cooling towers. Byron nuclear generating station exelon, by Christopher Peterson, is licensed under CC BY 2.5.

The voltage of the electricity generated by a power plant is typically about 25,000 V (25 kV). The frequency of the AC voltage is 60 Hz in the United States. This voltage is stepped up at the power station to a much larger voltage before being sent through transmission lines on the next phase of its journey. These are the extremely huge and tall power lines that do not route electricity to buildings, but instead transmit electricity over long distances. A photograph of transmission lines in Elgin, Illinois, is shown in Figure 25.7.



Figure 25.7 – Transmission lines carry electricity away from power plants, and are connected together via extremely large towers. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The voltage through transmission lines varies, but is usually in the neighborhood of 200 kV to 500 kV. Why is such high voltage used in transmission? Remember that power is conserved in the transmission and distribution process. Using the Byron power plant as an example, if two billion watts of power is generated, and that power is sent through transmission lines at 25,000 V (source [PDF, 18 kB]), then the current will be 80,000 A. The voltage is stepped up at the power plant and sent through the transmission system at 345,000 V. Because power is conserved, the current in the transmission lines will ideally be approximately 5,800 A. In this example, the voltage is stepped up by a factor of 13.8. This requires that the current is decreased by the same factor. Due to the resistance of the transmission lines, the current causes heating and energy losses in the transmission process. The rate of the thermal losses is $P = I^2R$. So, the higher the voltage, the smaller the current, and the more efficient the transmission process will be. In this example, the current being reduced by a factor of 13.8 reduced the power losses by a factor of 190.

At the end of transmission lines is an electrical substation. At this point, the voltage from the transmission lines is stepped down again. The exact voltage level varies depending on the customer(s) ultimately receiving the power. Distribution to neighborhoods will typically be less than 34,000 V (source [PDF, 327 kB]). This power is sent through distribution lines. A suburban substation in DuPage County, Illinois, is shown in Figure 25.8.



Figure 25.8 – An electrical substation steps down high voltage from transmission lines to distribution lines for delivery to our neighborhoods. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Distribution lines are much smaller than transmission lines and send power from the substation to neighborhoods. These are the power lines you will typically see on the sides of the road in the United States (shown in Figure 25.9). Sometimes they are underground, which prevents power outages from heavy winds, fallen tree branches, and winter storms. Because distribution lines are located in neighborhoods, and only transmit electricity over short distances, for safety reasons, the voltage can and should be much lower than what is used in long-distance transmission lines.



Figure 25.9 – Distribution lines carry electricity from a substation to our neighborhoods. These are typically seen along the sides of roads. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Transformers on the power poles holding up distribution lines (the cylindrical shaped object in Figure 25.10 contains a transformer) step down the voltage once more to 240 V for use in residential homes. Some large consumer appliances such as clothes dryers may require this high voltage level. This voltage is then divided into 120 V, which is sent to most of the outlets in a household.



Figure 25.10 – Some distribution lines contain transformers to step down voltage to send to our homes. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

While the voltage of the electricity changes throughout this journey from the power plant, the frequency of the AC voltage remains 60 Hz the entire time, which is why we obtain 60 Hz electricity from wall outlets in the United States.

Further reading

- Power grid? What is that? [PDF, 251 kB] – This article from *IEEE Potentials* magazine talks about the development of the power grid in the United States.
- How it works [PDF, 34 kB] – This informative article from the US Department of Energy has a lot of information about the power transmission and distribution process in the United States.
- The Power Grid – This interactive website lets you determine the effect of power consumption and generation on the power grid.
- Overhead transmission lines – This illustrated glossary of transmission lines from OSHA shows how the voltage of transmission lines can be determined from the design of the towers.
- Electricity 101 – This list of FAQs from the US Department of Energy answers a lot of common questions about power generation, transmission, and distribution in the United States.
- How Solar Storms Could Knock Down Our Power Grid – This video from PBS explains how the electrical grid (particularly in North America) can be susceptible to interruptions from strong solar activity.

Practice questions

Conceptual comprehension

1. A primary coil contains 50 turns and is connected to a secondary coil that contains 100 turns. Is this a step-up or step-down transformer? How do you know?
2. A primary coil contains 50 turns and is connected to a secondary coil that contains 10 turns. Is this a step-up or step-down transformer? How do you know?

Numerical analysis

1. A step-up transformer has 400 turns in the primary coil and increases the voltage from 110 V on the primary to 550 V on the secondary. Calculate the number of turns in the secondary coil.
2. The primary voltage of a step-up transformer is 240 V, with a secondary voltage of 4800 V. If the primary has 200 turns, calculate the number of turns in the secondary coil.
3. The primary voltage in a step-up transformer is 120 V, which generates a secondary voltage of 600 V. If the secondary coil contains 150 turns, calculate the number of turns in the primary coil.
4. The primary voltage of a step-down transformer is 480 V, which induces a secondary voltage of 120 V. If the primary contains 400 turns, calculate the number of turns in the secondary coil.
5. A step-down transformer decreases the voltage from 150 V to 30 V. The secondary contains 50 turns. Calculate the number of turns on the primary coil.
6. For a step-up transformer with a primary voltage of 200 V, a secondary voltage of 1000 V, and a current of 4 A in the secondary, calculate the power consumed by both the primary and secondary coils.
7. A step-up transformer has a primary voltage of 120 V and a secondary voltage of 600 V. If the current in the primary coil is 5 A, calculate the current through the secondary circuit.
8. In a step-down transformer, the primary voltage is 240 V, and the secondary voltage is 120 V. If the current in the secondary coil is 2 A, calculate the current flowing through the primary circuit.

This page titled Chapter 25: Electromagnetic induction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 26: Light

Summary

Light is all around us: some of it we can see, and some of it we cannot. From the Wi-Fi or cellular signal you use to text your friends, to the light waves emanating out of your laptop or smartphone screen, to the infrared radiation you are emitting as heat, the electromagnetic spectrum is possibly one of the most important aspects of our daily lives.

Electromagnetism

As we've learned in the previous two chapters of this textbook, electricity can generate magnetism and magnetism can generate electricity. As fields, a changing magnetic field can generate an electric field, and a changing electric field can generate a magnetic field. This process does not require the presence of a magnet or an electronic circuit. An electromagnetic wave is a transverse wave composed of both an oscillating electric field and an oscillating magnetic field, both at right angles to each other. Those oscillating electromagnetic fields propagate through space in a direction that is perpendicular to both the electric and magnetic fields. This is depicted in Figure 26.1. Note that because the direction of oscillation of the fields is perpendicular to the direction of wave motion, electromagnetic waves are transverse waves.

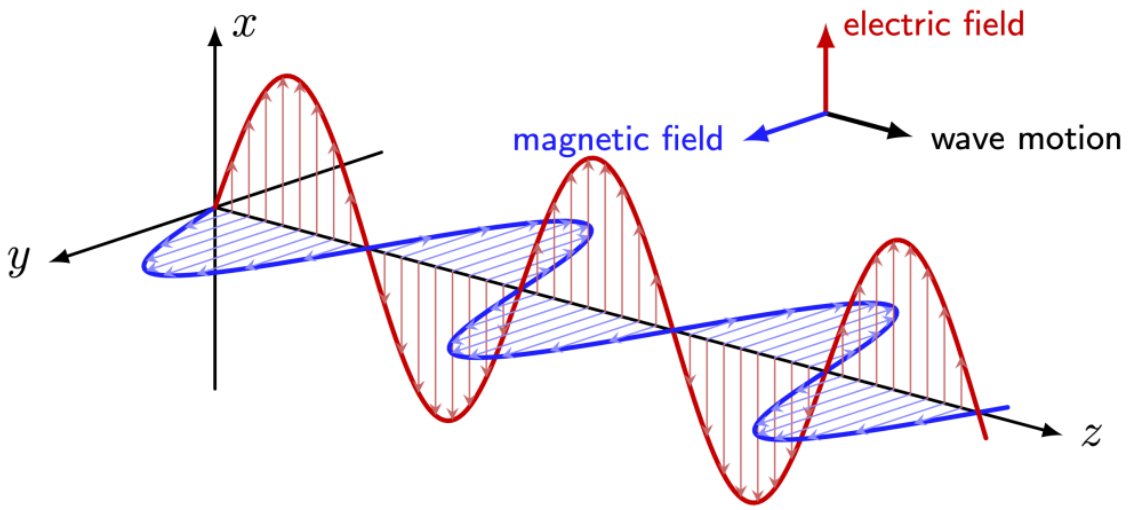


Figure 26.1 – An electromagnetic wave contains electric fields and magnetic fields oscillating at right angles to each other. The direction of motion of the wave is perpendicular to both the electric and magnetic fields. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Electric and magnetic fields have been animated and are shown in the video below. This animation demonstrates the motion of each field, and the motion of the wave itself.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The speed of light, represented by the lowercase letter c, is

$$c = 3 \times 10^8 \text{ m/s}.$$

That is, 300 million meters per second! In the early 1860s, James Clerk Maxwell discovered that the speed of electromagnetic waves is equal to the speed of light, leading him to conclude that light is simply a type of electromagnetic radiation. Maxwell brought together four equations to unify the concepts of classical electricity and magnetism. While these equations (which require vector calculus to solve) are outside the scope of this textbook, it is enough to note that Maxwell's equations are used today by scientists, engineers, and students all around the world to study electromagnetism.

In the early twentieth century, Albert Einstein postulated that the speed of light is constant, and acts as something like a universal speed limit. As far as we know, nothing can travel faster than the speed of light. While electromagnetic waves can travel at the speed of light, this speed serves as an asymptotic upper limit to objects with mass.

Electromagnetic waves are capable of traveling through the vacuum of space. This makes them very different from waves such as sound and water waves that require a medium to travel through. In fact, light travels through the vacuum of space between stars and our planet. Notably, the Sun produces most of the electromagnetic radiation we experience here on Earth.

We can use the speed of light to calculate how long it takes light waves to travel between the Sun and the Earth. Because the velocity is constant throughout the trip, we can use the equation $v = d/t$ to solve for time: $t = d/v$. The distance between the Earth and Sun is approximately 150 billion meters. Therefore, the amount of time it takes light to travel this distance is

$$\begin{aligned} t &= \frac{150 \times 10^9 \text{ m}}{3 \times 10^8 \text{ m/s}} \\ &= 500 \text{ s} \\ &= 8.3 \text{ min}. \end{aligned}$$

If the Sun were to somehow disappear, we wouldn't be aware of that fact until 8.3 minutes later when the light stopped reaching the Earth's surface.

SI prefixes

Our discussion of electromagnetic waves will require us to discuss quantities that are both extremely small and extremely large. Instead of using scientific notation to deal with these quantities, we can instead use prefixes. SI prefixes have been standardized for use with the International System of Units (SI). These prefixes can be used with any type of SI unit: meter, volt, second, gram, and so on. The SI prefixes that are used in this chapter are defined in Appendix B.

The electromagnetic spectrum

There are many categories of waves in the electromagnetic spectrum. Each of these waves can be defined in different categories, and each type of wave has different properties. The electromagnetic spectrum defines each type of wave based on its frequency or wavelength. Because the speed of electromagnetic waves is a constant, wavelength and frequency are inversely related to each other. If frequency is known, the wavelength of an electromagnetic wave is

$$\lambda = \frac{c}{f}.$$

If wavelength is known, the frequency of an electromagnetic wave is

$$f = \frac{c}{\lambda}.$$

While we will discuss the equation used to define the energy of an electromagnetic wave later in this textbook, for now it is sufficient to state that the energy of an electromagnetic wave is proportional to frequency and inversely proportional to wavelength. The different types of electromagnetic waves, arranged from smallest to largest wavelength, are listed in Figure 26.2.

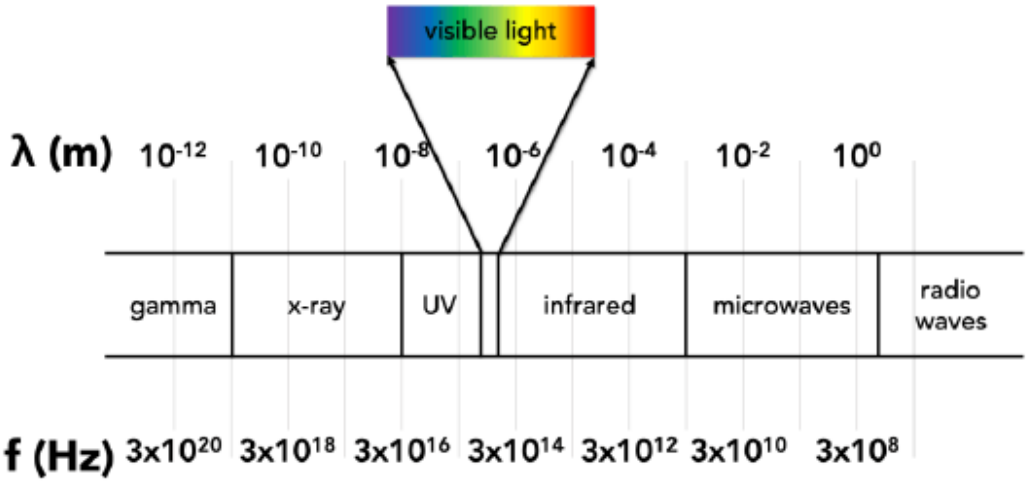


Figure 26.2 – A diagram of different types of electromagnetic waves, arranged from smallest to largest wavelength. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

At the lowest energy, lowest frequency, longest wavelength end of the spectrum are radio waves. These waves have wavelengths greater than 3 m, which correspond to frequencies less than 100 MHz.

Next along the spectrum are microwaves. Microwaves have wavelengths between 3 m and 1 mm, and frequencies ranging from 100 MHz to 300 GHz.

Infrared radiation, or IR, has wavelengths between 1 mm and 700 nm. The frequencies range from 300 GHz to 430 THz.

Visible light is the light we can see with our naked eyes. Visible light is a very tiny portion of the electromagnetic spectrum, encompassing wavelengths from 700 nm to 400 nm. The longest wavelength of visible light corresponds to red, and the shortest wavelength of visible light corresponds to violet. The frequencies of visible light span from 430 THz to 750 THz.

The next three categories of waves along the electromagnetic spectrum are high energy, collectively referred to as ionizing radiation. Ultraviolet, or UV, radiation has wavelengths between 400 nm and 10 nm. These correspond to frequencies between 750 THz and 30 PHz.

Next are x-rays. X-rays have wavelengths between 10 nm and 10 pm, and frequencies between 30 PHz and 30 EHz.

Any electromagnetic radiation with wavelengths less than 10 pm and frequencies above 30 EHz are known as gamma radiation. These are the highest energy waves on the electromagnetic spectrum.

Radio waves

Radio waves, as mentioned, comprise the lowest-energy region of the electromagnetic spectrum. Contrary to what the name implies, the waves we tune into on a radio to listen to music or talk radio are not the only kinds of waves that fall under the category of radio waves.

Radio waves are used predominantly for communications applications. It may be useful to review why radio waves (a type of electromagnetic wave) are used to transmit information instead of sound waves. As previously discussed in this textbook, sound waves are longitudinal pressure waves. They require some type of medium through which to travel (and cannot travel through space). This means that sound waves cannot be used for any kind of communication that routes through satellites (for example), which would require travel through the vacuum of space. In addition, sound waves don't travel very far before they become inaudible. To hear the sound of somebody's voice over long distances, another method for transmitting that sound is required.

To that end, a sound wave can be encoded onto a radio wave and transmitted by means of some electric circuitry and an antenna. Radio waves are part of the electromagnetic spectrum, travel at the speed of light, and can easily transmit through the Earth's atmosphere over much longer distances than sound waves. These radio waves are received by something like a radio or walkie-talkie and decoded back into vibrations through a speaker into sound waves.

The radio wave portion of the electromagnetic spectrum is used for many applications. This is because long-wavelength light can travel over very long distances. Some radio waves can reflect off of the Earth's ionosphere and travel nearly around the globe. Many radio waves can travel through clouds and storms, which means that their transmission will not be affected by weather. The longest wavelengths of radio waves are only limited by the difficulty in generating these waves, which would require extremely large antennas.

The applications of radio waves include telephone signals; military, aircraft, and maritime communication; aircraft navigation; WWVB and other similar time signals; CB and amateur radio; TV broadcasts; and AM/FM radio.

AM and FM Radio

AM and FM refer to the two methods of sending radio waves to a radio receiver. These are amplitude modulation (AM) and frequency modulation (FM). These two modulation methods describe how a sound wave can be converted into a radio wave.

In amplitude modulation (AM), the amplitude of the radio wave (known as the carrier wave) is modified to reflect changes in the sound wave (message signal). This is depicted in Figure 26.3. (Download this data [XLSX, 32 kB]) The amplitude of the carrier wave is multiplied by the amplitude of the message signal, creating an amplitude modulated wave.

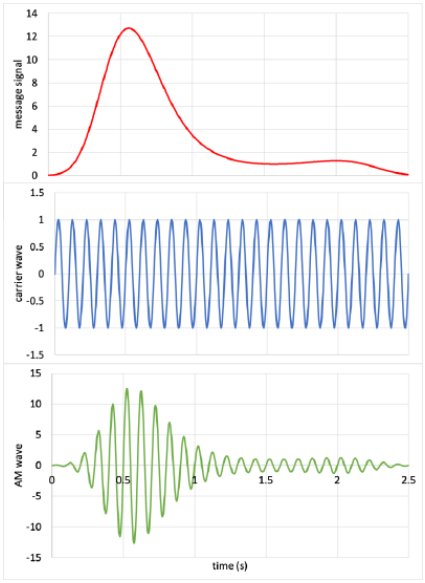


Figure 26.3 – An amplitude modulated (AM) radio wave is generated by altering the amplitude of a carrier wave based on the properties of the message signal. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The frequencies of AM radio stations in the US vary from 540 kHz to 1700 kHz. AM radio, while popular in the early twentieth century, is now used very infrequently, and may be considered by some to be obsolete. Many AM radio stations consist of talk or sports radio. Because of the long wavelengths of AM radio, the signals travel very far. However, because the wavelengths are physically large, they do not transmit well through small spaces like tunnels. AM radio reception may therefore become blocked when driving through tunnels, underpasses, or possibly even parking garages.

In frequency modulation (FM), the frequency of the radio wave (known as the carrier wave) is modified to reflect changes in the sound wave (message signal). This is depicted in Figure 26.4. (Download this data [XLSX, 33 kB]) The frequency of the carrier wave is modified by the amplitude of the message signal, creating a frequency modulated wave.

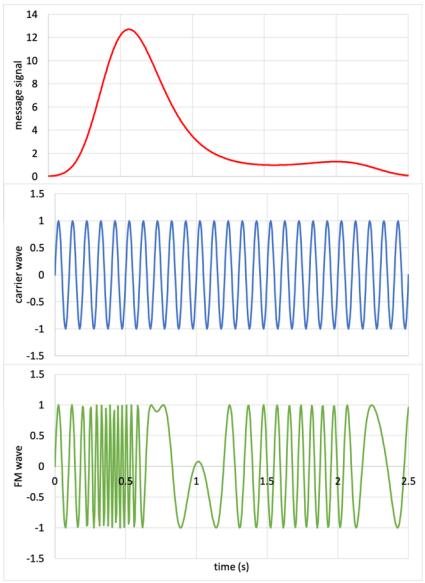


Figure 26.4 – A frequency modulated (FM) radio wave is generated by altering the frequency of a carrier wave based on the properties of the message signal. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The frequencies of FM radio stations in the US vary from 88.1 MHz to 107.9 MHz. FM radio is much more popular than AM radio as it is less susceptible to noise and has much higher sound quality than AM radio. However, due to the shorter wavelengths, FM radio signals cannot travel as far.

Infrared

As we learned in a previous chapter of this textbook, one of the mechanisms of heat transfer is via electromagnetic radiation. Most "cool" objects such as humans, plants, and animals emit heat in the infrared (IR) region of the electromagnetic spectrum. Therefore, the infrared region of the electromagnetic spectrum is extremely important in any area of science and technology that requires us to study heat.

Thermal cameras have a sensor that collects infrared radiation rather than visible light. This allows us to use IR for night vision, thermal imaging, and many other applications. For example, satellite imagery of cloud coverage is used for weather observations that aid in forecasting. After sunset and before sunrise, visible satellite images cannot record any meaningful weather data. However, IR satellites can take images corresponding to the heat reflected from the planet's surface, which correlates to cloud coverage. (As long as there is no extensive snow cover on the ground, the Earth's surface tends to have a much higher temperature than the temperature of clouds. Therefore, the cool areas in an IR photograph typically correspond to areas of cloud coverage.)

Figure 26.5 shows a visual satellite image taken just after sunset in Illinois on October 11, 2018. The entire east coast and most of the Midwest is dark, due to the fact that there was no light present to create an image. While there is evidence of a large storm (Tropical Storm Sergio) near the Baja California Peninsula, and sporadic cloud coverage over the western portion of the United States, no other evidence of storms is present.

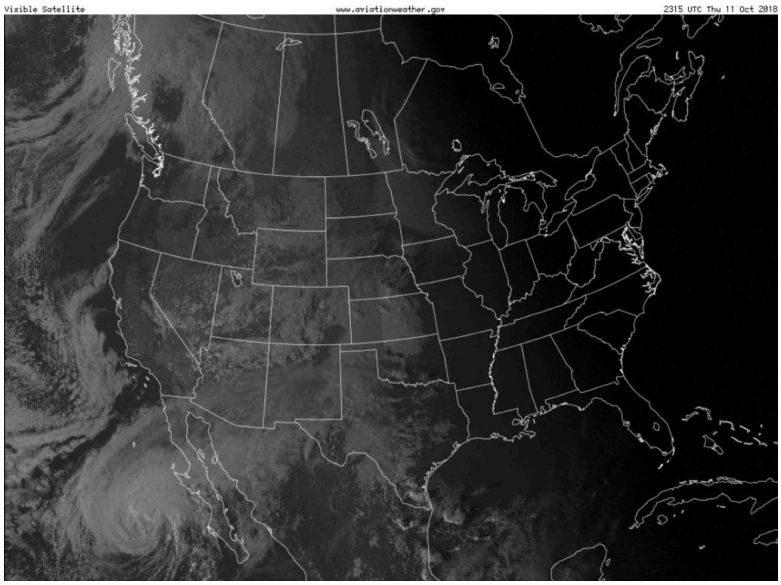


Figure 26.5 – A visible satellite view of the continential United States taken at 18:15 CDT on October 11, 2018. This figure, created by Aviation Weather, is licensed under Public Domain.

An infrared satellite image taken at the same time (shown in Figure 26.6) shows a very different picture. While Tropical Storm Sergio is visible in both images, only the IR satellite image contains evidence of Hurricane Michael that had recently made landfall on the east coast.

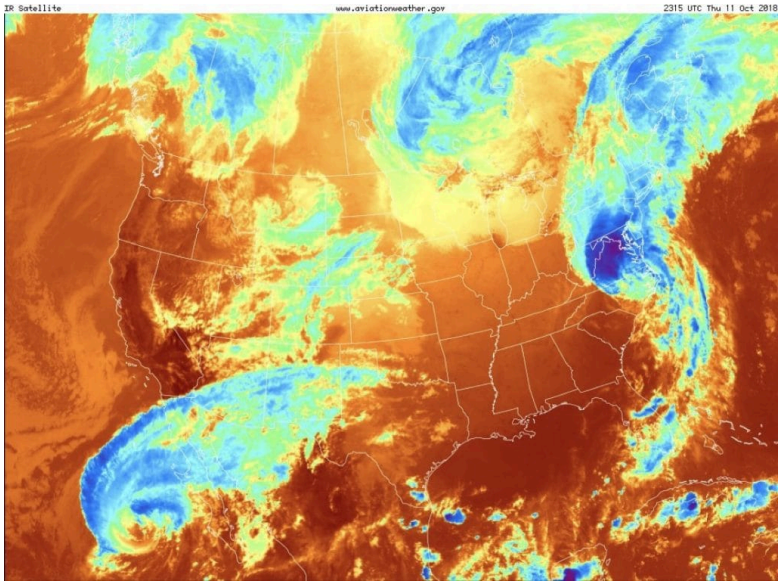


Figure 26.6 – An infrared satellite view of the continental United States taken at 18:15 CDT on October 11, 2018. This figure, created by Aviation Weather, is licensed under Public Domain.

Microwaves

Microwaves travel line-of-sight distances. Unlike radio waves, they cannot bounce off the ionosphere or travel along the curvature of the Earth. For this reason, microwaves are used in applications where signals only need to travel over short distances. Some, but not all, microwaves can transmit through buildings, but again, only over short distances.

Much of our modern life is transmitted over microwaves: Wi-Fi, Bluetooth, cellular, and GPS and other satellite signals all are part of the microwave region of the electromagnetic spectrum. In addition, things such as garage door openers, collision detectors in cars, and radar use microwaves.

Perhaps when you heard the word microwave, you thought about cooking food. In fact, microwave ovens do use microwaves to cook food. At just the right frequency, microwaves can cause the water and fat molecules in food to heat up. Microwave ovens emit microwave radiation at a 2.45 GHz frequency, usually with power levels between 600 W and 1200 W. Most microwave ovens contain a sticker inside the door that shows the frequency of operation and power level. Dr. Pasquale's microwave oven emits at 2.45 GHz and has a power rating of 950 W, as seen in Figure 26.7.



Figure 26.7 – The sticker on a microwave oven contains information about the frequency and power rating that the appliance uses. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

It is interesting to note that most of our Bluetooth and Wi-Fi communications also occur right around the 2.4-2.5 GHz frequency range. Our mobile devices communicate using complex protocols that ensure that one signal does not get confused for another. A laptop's Wi-Fi receiver will not get confused if there is a cell phone nearby receiving a phone call or text message. Also, the reason that a microwave oven is dangerous enough to require metal shielding, and cell phones and wireless routers are not, is due to the power rating. A microwave oven uses microwave radiation at power levels about one thousand times stronger than those used by our modern telecommunication devices.

Ionizing radiation

Electromagnetic radiation with frequencies greater than that of visible light (short-wavelength ultraviolet, x-rays, and gamma rays) is known as ionizing radiation. That's because the energy of these waves is so great that it can ionize atoms. This can be very dangerous for humans, plants, and animals, as it can cause damage to our DNA and other tissues.

Ultraviolet

Ultraviolet (UV) radiation is emitted by the Sun. Some of this UV transmits through the Earth's atmosphere, and is the reason why our skin becomes darker with sun exposure. Sunscreen can help block the damage caused by UV light. Higher frequencies of UV are blocked by the Earth's ozone layer and do not reach the surface of the planet.

While ultraviolet radiation can be damaging to us, that doesn't mean it isn't useful. Mercury vapor lights (such as the one shown in Figure 26.8) generate UV, which then is converted by a phosphor to visible light. This is how fluorescent lamps work, as will be discussed later in this textbook. The integrated circuits comprising your computer processor are fabricated in clean rooms using a photolithography process that uses UV rays to pattern the transistors in the processor hardware. UV is also used in medical imaging, sterilization, bug zappers, and curing lamps.

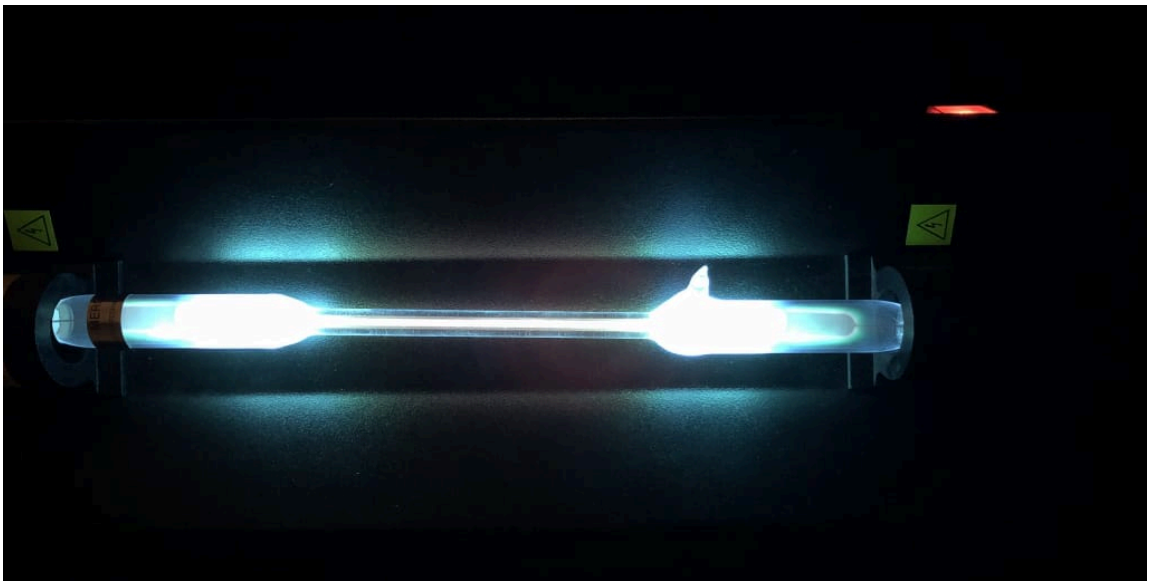


Figure 26.8 – A mercury vapor lamp emits UV light. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

X-rays

X-rays were discovered in 1895 by physicist Wilhelm Röntgen, who discovered rays that had different properties than other known rays at the time. The letter "x" in x-ray signified that the rays were unknown and novel at the time. Röntgen made an image of his wife's hand with x-rays, creating the first radiograph and demonstrating the medical use of these new types of rays.

Unfortunately, the medical hazards of x-rays were not discovered until the late 19th and early 20th centuries; precautions that we use today in medical radiography were not practiced early after the discovery of x-rays, and people suffered burns and even cancer as a result of being exposed to high doses of x-rays for long periods of time.

Since their discovery, x-rays have been used widely in medical imaging. Their ability to permeate through soft tissue and be blocked by our bones makes imaging with x-rays an extremely helpful diagnostic tool. An x-ray image of a broken ankle is shown in Figure 26.9.



Figure 26.9 – An x-ray of a broken ankle. X-rays are useful diagnostic medical tools as they allow us to take images of bones. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Other medical uses of x-rays include computed tomography (CT) scanning (creating multiple x-ray images to obtain cross-

sectional images of anatomy), fluoroscopy (using x-rays to determine the real-time location of anatomy and possibly even surgical implements during a medical procedure), and radiotherapy (purposefully using the ionizing properties of x-rays to treat cancers).

X-rays are also used in airport security to check carry-on bags for the presence of explosives, firearms, or other prohibited items. A radiograph of a backpack is shown in Figure 26.10.



Figure 26.10 – A radiograph of a backpack. The x-rays used to create this image are capable of "seeing into" the bag to detect for firearms, explosives, or other contraband. Xray-verkehrshaus, by IDuke, is licensed under CC BY 2.5.

In addition to the medical uses of x-rays, they are also used in many fields of scientific and engineering research. For example, x-rays can be used to determine the crystalline structure of materials. This was how the double helix shape of DNA was determined in the middle of the 20th century by Rosalind Franklin. The Advanced Photon Source at Argonne National Laboratory creates x-rays for use in scientific research.

Gamma rays

Gamma rays reside at the highest energy end of the electromagnetic spectrum. Gamma rays are produced by radioactive decay, a process we'll discuss later in this textbook. Gamma radiation is extremely hazardous to humans, as it can cause damage to our internal organs and bone marrow. Gamma radiation, while dangerous, does have its applications. Its ability to travel through many objects makes it useful for scanning objects in container trucks at border checkpoints and irradiating objects from harmful bacteria; gamma rays also have limited use in nuclear medicine.

Light waves

Light waves have lots of interesting properties. In fact, we'll learn later in this textbook that light is not just a wave. There is a lot more to it!

There are many things that can happen to electromagnetic waves as they travel between two points. Reflection and refraction have to do with bouncing and bending, which will be discussed in a later chapter of this textbook. If you have ever looked in a mirror, then you have experienced the reflection of light waves.

Light can also be transmitted and absorbed. When you look out a window, visible light transmits through and you can see what's going on outside. If, however, you look at a wall, visible light does not transmit through, and you cannot see through the wall. A combination of reflection and absorption on the other side of the wall prevents the light from getting through to your side. We'll discuss how transmission and absorption play a role in the colors of objects in the next chapter of this textbook.

Whether or not objects allow light waves to pass through gives them different properties. An object is opaque if it does not allow any light to transmit through it. In Figure 26.11 (left), Dr. Pasquale holds a piece of opaque cardboard up between her face and the camera. Because no visible light can transmit through the piece of cardboard, her face is not visible in the photograph. (In

addition, when the photo was taken, she could not see the camera either.)

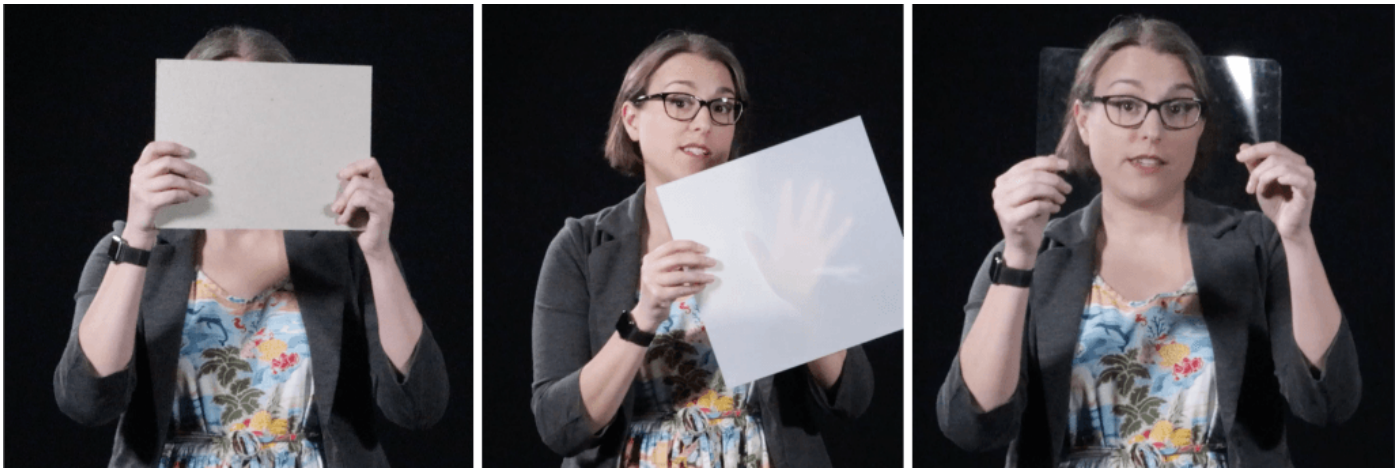


Figure 26.11 – Dr. Pasquale demonstrates the optical properties of something that is opaque (left), translucent (middle), and transparent (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

An object that allows light to pass through, but scatters the light in the process, is known as translucent. This means the image becomes blurry or otherwise hard to make out entirely. In Figure 26.11 (middle), Dr. Pasquale holds a sheet of translucent plastic between her hand and the camera. You can vaguely see her hand behind the sheet, but it is blurry and lacks detail. In general, viewing something behind a translucent object means that you typically can see an object on the other side, but may not be able to give any detail about it.

An object that allows visible light to pass through without getting scattered is known as transparent. In Figure 26.11 (right), Dr. Pasquale holds a transparent sheet of plastic up between her face and the camera. Because visible light can travel through, her face is easily seen in the photograph. (While the photo was taken, she was able to see the camera as well.)

One of the fascinating things about light is that something transparent to one type of electromagnetic radiation may be opaque to another! For example, while glass (such as that used in eyeglasses) is transparent to visible light, it is opaque to infrared light. In Figure 26.12 (left), Dr. Pasquale took a selfie without her glasses on. Her entire face is "visible" to the IR camera, as infrared emitted as heat from her face is collected by the camera's sensor. In Figure 26.12 (right), Dr. Pasquale took a similar selfie with her glasses on. Now her eyes appear dark, as the heat from behind the glass is blocked from reaching the camera's sensor.

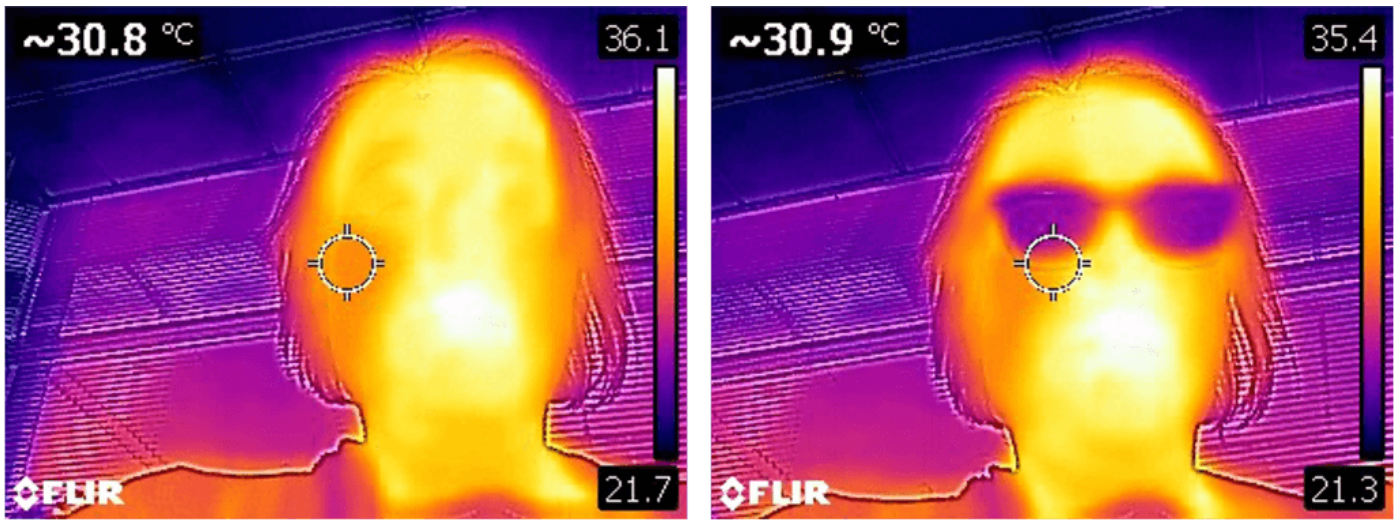


Figure 26.12 – Dr. Pasquale's face without eyeglasses (left) and with eyeglasses (right), demonstrating that glass is opaque to infrared. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Similarly, our atmosphere is transparent to some types of light and opaque to others. This is why some radio waves can travel over long distances: they are reflected by the atmosphere. It's also why satellites using microwaves can send information to our smartphones and GPS devices: the atmosphere is transparent to many microwave frequencies. Figure 26.13 shows the opacity of

the Earth's atmosphere to many different frequencies of electromagnetic radiation.

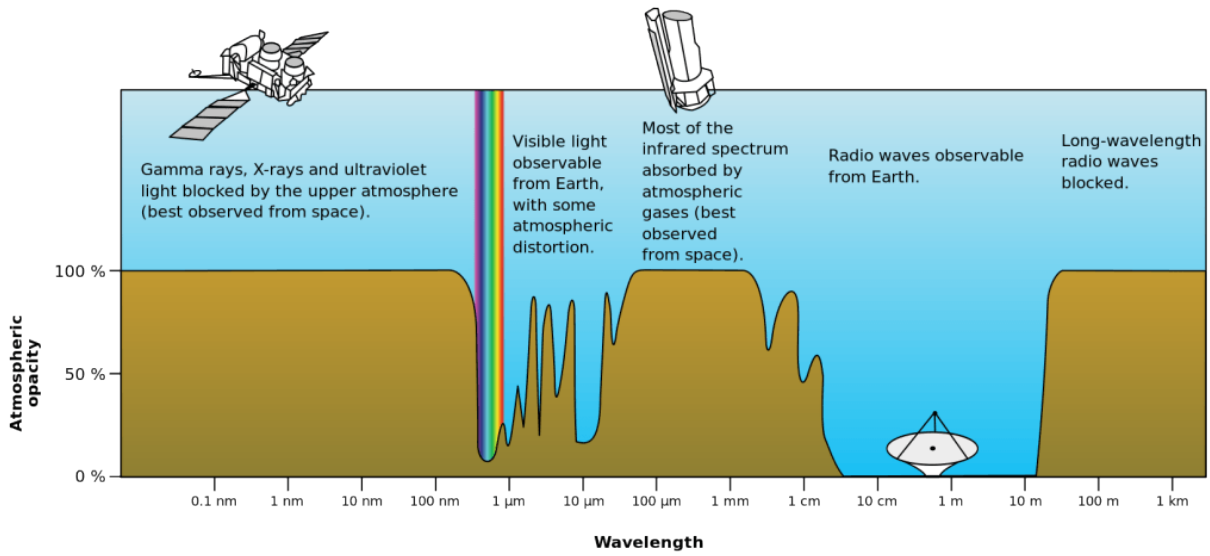


Figure 26.13 – A graphic depicting the opacity of the Earth's atmospheric for different wavelengths of electromagnetic radiation. This figure, created by NASA, is licensed under Public Domain.

Further reading

- Finding the speed of light with peeps – Use your favorite marshmallow candy and a microwave to calculate the speed of light!

Practice questions

Conceptual comprehension

1. What is the evidence that light does not need a material medium through which to travel? In other words, how do we know that light can travel through vacuum?

Numerical analysis

1. Calculate the length of time it takes for an electromagnetic wave to travel from Earth to each of the following...
    1. ...the Moon (which is 3.8×10^8 m, on average, away from the Earth).
    2. ...Mars ($d = 76 \times 10^9$ m when Mars is closest to Earth).
    3. ...Voyager 1 (which was approximately 24×10^{12} m away from Earth as of the time this question was written).
2. Calculate the wavelength of an electromagnetic wave that has a frequency of 500 THz.
3. Calculate the frequency of an electromagnetic wave that has a wavelength of 2 mm.
4. Calculate the wavelength of the radio waves corresponding to the College of DuPage's radio station, WDCB, which broadcasts at a frequency of 90.9 MHz.
5. Calculate the frequency of red light having a wavelength of 633 nm.

This page titled Chapter 26: Light is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 27: Color

Summary

Visible light waves with different frequencies appear as different colors of light. When light waves reflect, transmit, absorb, or scatter selectively based on their frequency, objects take on different colors. This can explain many topics, from how color monitors and phone screens work to why the Earth's sky looks blue during the day and red at sunrise and sunset.

Color

The visible light section of the electromagnetic spectrum encompasses light waves from red to violet. Perhaps you've heard the acronym ROYGBIV which stands for red, orange, yellow, green, blue, indigo, and violet. While visible light is better described as a spectrum rather than as discrete quantities, we consider ROYGBIV to be the major colors of visible light. What distinguishes each color is the wavelength of that light. Red light has long wavelengths, green is in the middle, and violet has the shortest wavelengths of visible light. The table below gives the approximate wavelength ranges for each color of visible light.

color	**wavelength range**
red	700 – 640 nm
orange	640 – 590 nm
yellow	590 – 560 nm
green	560 – 500 nm
blue	500 – 450 nm
violet	450 – 400 nm

Human eyes contain two types of light receptor cells. They are called rods and cones (based on the shapes of these cells). The rods are sensitive to dim light but not to color. The cones are what allow us to perceive colors. There are three types of cone cells: one sensitive to short wavelengths of light, one sensitive to medium wavelengths, and one sensitive to long wavelengths. However, the cones require higher intensity light in order to "fire" and send a signal to the brain. This is why we can see objects in a very dimly lit room, but we cannot distinguish colors very well. Also, when we look at the night sky with the naked eye, almost all the stars appear white. The light intensity of the starlight can trigger the rods, but does not have enough intensity to trigger the cones. When we look at stars through a telescope, the increased light gathering of the instrument will then allow us to see the vast variety of star colors. When light waves of different wavelengths and intensities interact with our rods and cones, we perceive

things as taking on different colors as a result. Physical or inherited issues with cone cells can be a cause of color-blindness.

The three primary colors of light are red, green, and blue. Using color addition, which will be discussed later in this chapter, these three colors alone can be combined in varying intensities to generate all other colors of visible light.

Transmission, absorption, and reflection

When we look at objects, we don't actually see the objects themselves. We see the light that reflects off of them into our eyes. In the case of transparent objects, we see the light waves that transmit through the objects.

Reflection is what occurs when waves bounce off of a surface. When we look at something blue, that means that blue light is reflecting off the object into our eyes. Red and green light waves are absorbed by the object, which is why we don't see those colors. A red object reflects red light and absorbs blue and green light.

Absorption occurs when a material absorbs the energy from a light wave, preventing it from either reflecting or passing through that material.

In general, objects that reflect light into our eyes reflect the colors that we see and absorb the colors that we do not see.

Transmission occurs when light waves pass through an object. When we look through blue transparent objects, those objects transmit blue light and absorb green and red light. When we look through a yellow transparent object, it transmits red and green light and absorbs blue light.

Other objects that we see via transmission of light are light sources themselves. This includes LEDs, street lights, and some light bulbs. Red light sources transmit red light waves to our eyes, green light sources transmit green light waves to our eyes, and so on.

Objects that we see via light transmission transmit the colors we see and absorb the colors that we do not see.

Color addition

Color addition refers to the ability to use light to create secondary colors from primary colors. Mixing various intensities of red, green, and blue light allows us to create any other color of light.

Color addition is related to light we see via transmission. When two primary colors are transmitted together, we see a secondary color. When red and blue of equal intensity are added, a pinkish-purple called magenta is created. When red and green of equal intensity are added, yellow is created. When blue and green of equal intensity are added, cyan is created. Magenta, yellow, and cyan are therefore known as secondary colors. When all three primary colors are added at equal intensities, the result is white light. This is sometimes referred to as the RGB (red-green-blue) color model, and is shown in Figure 27.1. All other colors can be generated with mixtures of red, green, and blue light at different intensity levels.

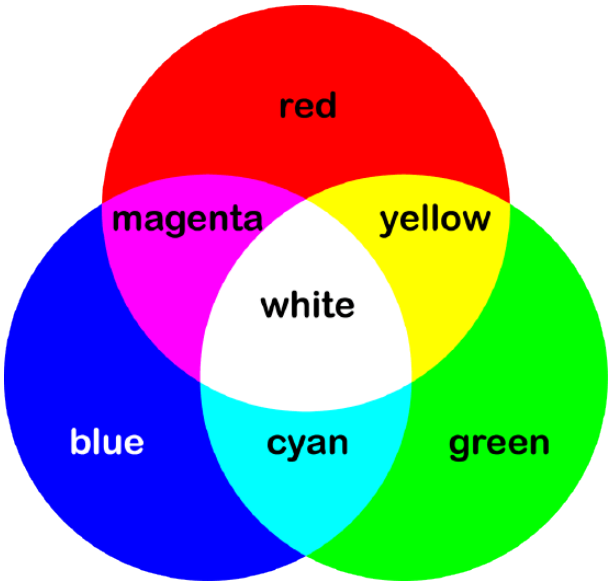


Figure 27.1 – The RGB model demonstrates color addition. Primary colors of red, green, and blue are added to create secondary colors cyan, magenta, and yellow. The addition of red, green, and blue at equal intensities creates white. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Images, websites, and videos that are viewed on computers, TVs, and other electronic devices are defined with an additive color model. An LED display contains tiny pixels of red, green, and blue LEDs (an example of which is shown in Figure 27.2).

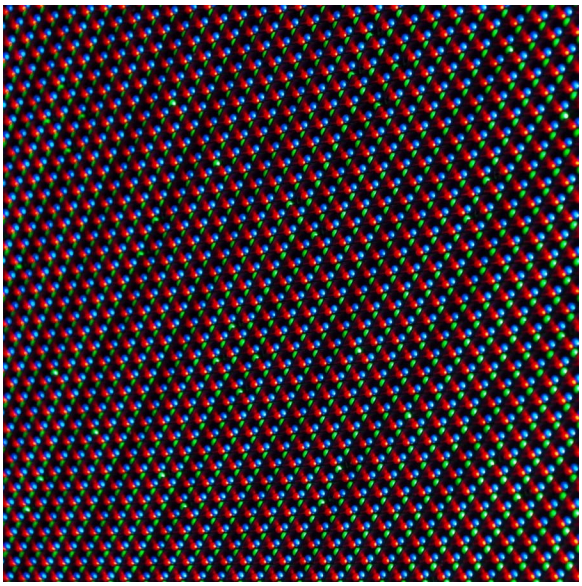


Figure 27.2 – A close-up image of an LED display shows separate red, green, and blue LEDs. MK38527 LED Display, by Martin Kraft, is licensed under CC BY-SA 3.0.

Many LCDs (liquid crystal displays) contain a white backlight that provides background illumination. This light is then sent through red, green, and blue filters to generate the colors you see on a television, smartphone or tablet screen, or computer monitor.

In the video below, Dr. Pasquale uses RGB light-emitting diodes to demonstrate color addition. The top dial changes the intensity of the red light; the center dial changes the intensity of the green light; and the bottom dial changes the intensity of the blue light. When Dr. Pasquale turns on the red and green LEDs, the result is yellow. Red and blue generates magenta light. Green and blue creates cyan light. Finally, Dr. Pasquale turns on all three colors and the result is white light.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When it comes to light and color addition, white, as mentioned, is generated when all primary colors are mixed at equal intensities. The color black is perceived when there is no light at all. In that respect, black can be considered the absence of colors.

Color subtraction

While color addition pertains to light transmission, color subtraction pertains to light reflection. Subtractive color models are used primarily in print. When you look at a color magazine, book, or poster, you are seeing the light waves that reflect off of the pages and into your eyes.

Color subtraction refers to the colors that are absorbed by pigments or other printed material. Cyan absorbs, or subtracts, the color red. Magenta absorbs the color green. Yellow absorbs the color blue. Secondary colors therefore subtract out primary colors. Through this color subtraction, other colors can be generated.

When cyan and yellow are mixed in a subtractive model, blue and red are subtracted and the result is green. When cyan and magenta are mixed, red and green are subtracted and the result is blue. When magenta and yellow are mixed, blue and green are subtracted and the result is red. When cyan, magenta, and yellow are mixed, the result is black. This is sometimes referred to as the CMY (cyan-magenta-yellow) model, and is shown in Figure 27.3.

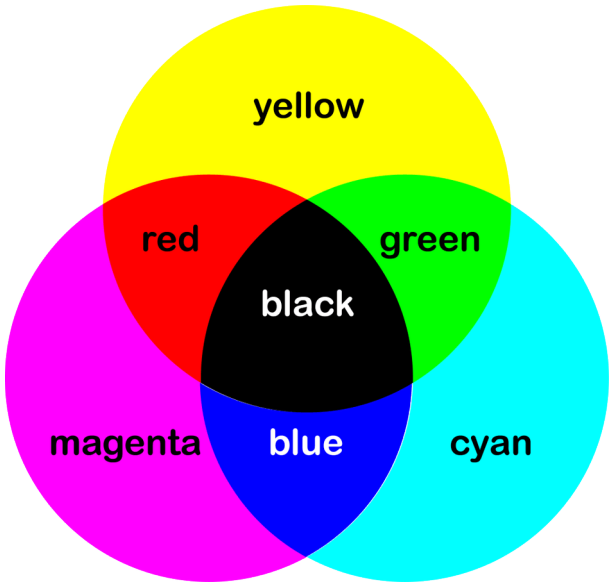


Figure 27.3 – The CMYK model demonstrates color subtraction. Primary colors of red, green, and blue are subtracted out of secondary colors cyan, magenta, and yellow. The subtraction of all colors creates black. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Color printers work by using separate cyan, magenta, and yellow inks. Usually, a separate black ink is used to generate black without using excessive amounts of cyan, magenta, and yellow ink. If you've ever heard or seen the term CMYK, that refers to subtractive colors used in printing: cyan, magenta, yellow, black.

While color subtraction is a model that explains color printing, it is not well-suited to explain color mixing of paints. The color of two mixed paints depends on the chemistry of the pigments themselves as well as the thickness of the paint, and requires very

complicated mathematics to approximate (source [PDF, 2 kB]).

Scattering

Selective transmission, reflection, and absorption explain the colors of most things we see on this planet. But some phenomena require the concept of scattering to explain. Scattering is a process in which the direction that light travels is changed without affecting the wavelength of the light.

For example: why is the sky blue on Earth? The Earth's atmosphere is composed of many particles of different gases, as we discussed earlier in this textbook. When light from the Sun travels through these particles, due to the size of the particles, short wavelength light is scattered more efficiently than longer wavelength light. Atmospheric scattering causes short wavelength light such as blue and purple to reach the surface of the planet. Our eyes are more sensitive to blue than purple, therefore we perceive the sky to be blue. The blue sky as seen from the Adirondack Mountains of New York State is shown in Figure 27.4.



Figure 27.4 – The sky appears blue due to scattering of light in the Earth's atmosphere. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Cloud particles, being larger than atmospheric particles, scatter longer wavelengths as efficiently as shorter wavelengths. This means that essentially all wavelengths of visible light are scattered equally well. This makes fair-weather clouds appear to be white, as seen in Figure 27.5.



Figure 27.5 – Clouds appear to be white due to their ability to scatter all wavelengths of visible light effectively. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

On overcast days, the sky appears gray, a muted shade of white. A gray overcast sky over Herrick Lake Forest Preserve is shown in Figure 27.6.



Figure 27.6 – An overcast sky appears gray, a muted shade of white. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

During sunrise and sunset, the Sun's shallow angle with respect to the horizon means that light from the Sun has to travel through more atmosphere than it does at other times of day when the Sun is high in the sky. When sunlight travels through so much atmosphere, so much blue light is scattered away that only the longer wavelengths of visible light are able to reach our eyes. The effect is a red sky. Sunset's striking colors, as seen from the College of DuPage parking lot, are shown in Figure 27.7.

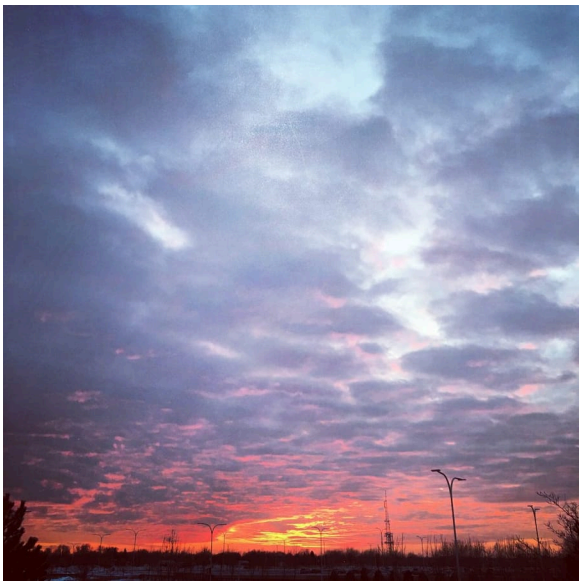


Figure 27.7 – At sunset, red and orange colors can be seen close to the horizon. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, a fish tank is filled with water. Dr. Fazzini shines a flashlight inside to demonstrate how light passing through the clear water looks from both the side and head-on. The clear water does not noticeably scatter any of the light as evidenced by the fact that the beam is not visible when viewed from the side. As a result, the light that passes through appears white. After filming that part of the video, milk was mixed with the water to act like the molecules in our atmosphere. Dr. Fazzini again shines the flashlight into the fish tank. The flashlight acts like the Sun, providing a white light source. When the flashlight is directed into the milk-water mixture, the visibility of the beam when viewed from the side demonstrates scattering of the light. The short wavelengths of light (notably blue) are scattered the most. When viewing the direct beam, the long wavelengths of light (orange and red) are what remain after the short wavelengths have scattered out. This is similar to what happens at sunrise and sunset.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

If the color of the sky is explained by our atmosphere, then what happens on celestial bodies with no atmosphere? On the Moon, for example, there is no atmosphere to scatter any wavelengths of light. The Sun appears as an orb in space, and the surrounding areas of space are black, as seen in Figure 27.8. Photos from Moon landings depict the darkness of space. Note that in Figure 27.8, the presence of glare is likely due to the astronaut's helmet and there additionally appears to be lens flare caused by the camera. None of those optical effects is due to atmospheric effects, as the Moon has no atmosphere!



Figure 27.8 – Because there is no atmosphere on the Moon, the sky appears black. The Sun appears as a white orb. Apollo 14 Lunar Module photographed against sun glare on lunar surface, by NASA Johnson Space Center, is licensed under Public Domain.

The atmospheric composition of Mars leads to a brownish pink sky color. Mars landers have taken color photos depicting this fascinating sky color. A color photograph of the Mars lander Ingenuity, surrounded by the Martian surface and sky, is shown in Figure 27.9.



Figure 27.9 – The Martian sky appears to be a brownish pink color due to the composition of the Martian atmosphere. Ingenuity at Airfield D, by NASA Jet Propulsion Laboratory, is licensed under Public Domain.

Further reading

- Color wheel – Learn about color theory and what colors look good together and why.
- Coblis Color Blindness Simulator – This simulator can show you how different images appear to individuals with different types of color vision deficiencies. (In particular, it is a helpful tool for ensuring that images and graphics are accessible.)
- Luftwerk Color Code – This art installation, which was exhibited at the College of DuPage's Cleve Carney Museum of Art in 2017, uses the effect of different color lighting on different color paints to achieve different artistic results. A video interview with the artists demonstrates many of these visual effects.

Practice questions

Conceptual comprehension

1. If you shine equal intensities of red light and green light onto the same spot, what color will you perceive? Why?
2. Using color addition, how would it be possible to generate colors such as orange or purple, which are not secondary colors of light?
3. Imagine you have a red filter and a blue filter. If you look through both filters simultaneously, what color will you see? How does this relate to the concept of color addition?
4. If you shine white light through a green filter, what color of light will appear to transmit through? Why? How would this change if the light is red, green, blue, or a mixture of those colors?
5. If you shine white light onto a surface covered with a cyan object, what color will the object appear, and why? How would this change if the light is red, green, blue, or a mixture of those colors?
6. Consider a scenario where you have a blue object under red light. What color will the object appear? Why?

Hands-on experiments

1. Use flashlights with colored filters or cellophane (red, green, blue). Shine each color onto a white surface and observe how they combine when overlapping. Then, shine each color onto different colored surfaces or objects and observe the results.
2. Place objects of different colors (such as toys, fabrics, or even colored candies) in direct sunlight. Observe how the colors appear. Then, bring the objects indoors and compare the colors that you see. Is indoor lighting the same or different from sunlight?

This page titled Chapter 27: Color is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 28: Reflection and refraction

Summary

Every time you look in a mirror, take a photo, use eyeglasses, or look through a window, you're experiencing reflection and refraction. Understanding how light travels enables us to grasp how the simple concepts of reflection and refraction affect our lives.

The principle of least time

Light travels from one point to another using the principle of least time. That is to say, light does not necessarily take the shortest distance between two points. Instead, it will travel whatever path takes the least amount of time. Sometimes, the path of least distance and the path of least time are the same. Other times, however, those two paths are not the same.

When light bounces off of a surface (such as a mirror), but otherwise travels through the same medium (such as air) throughout its journey, the path of least time is the same as the path of least distance. Because the speed of light through a single medium remains constant, the fastest path for light to travel will be the shortest path.

Consider light that starts at point A and bounces off of a mirror before traveling to point B (as shown in Figure 28.1). Any path that is not the shortest distance will take longer than the shortest path, due to the fact that time equals distance divided by velocity, and velocity remains constant through a single medium.

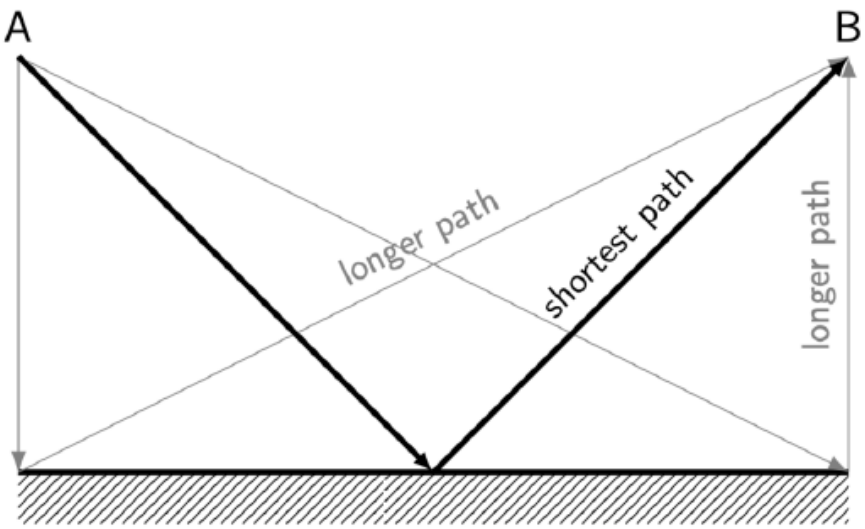


Figure 28.1 – When velocity remains constant, the path of least distance will also be equal to the path of least time. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

When light travels through different media, the path of least distance and the path of least time are no longer the same. We can use the lifeguard problem as an analogy to understand why this is so.

Consider a lifeguard on the beach. She sees a swimmer out in the water who needs help and may be drowning. The lifeguard needs to get to the swimmer as quickly as possible. She knows that she can run on the beach faster than she can swim through the water, so she wants to optimize her route to the swimmer to get to him as quickly as possible.

If the lifeguard takes a straight-line path to the swimmer, the path of least distance (shown in Figure 28.2), she will spend too

much time in the water. Because she's slower in the water than on the beach, this will not be the fastest route.

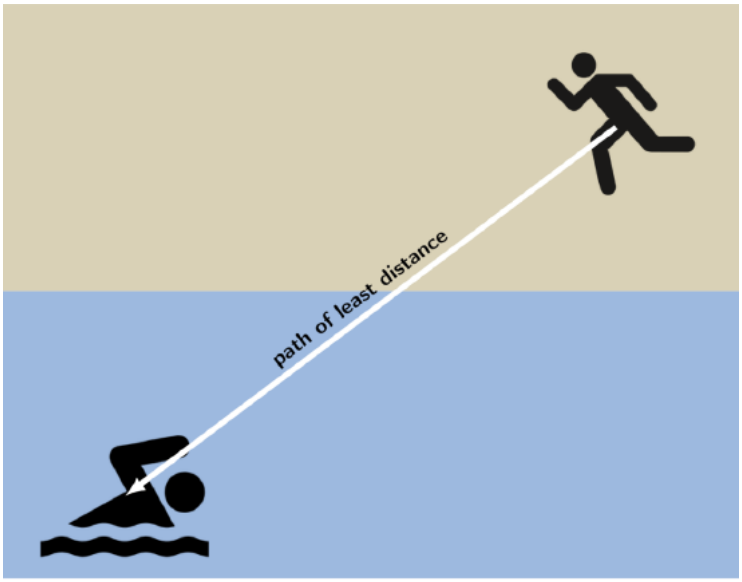


Figure 28.2 – The path of least distance spends too much time in the water where the lifeguard is slow. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Alternatively, the lifeguard could minimize the amount of distance she spends in the water (shown in Figure 28.3), but this would cause her to spend too much time on land, which is also not an optimal route.

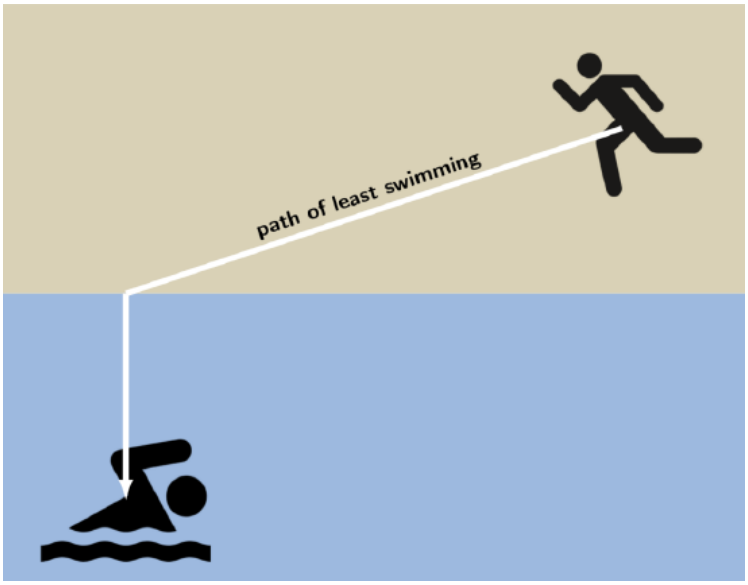


Figure 28.3 – The path of least swimming spends too much time on land; the total distance is too far and inefficient. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Depending on the lifeguard's exact speeds in the land and in the water, her path of least time will be an optimization of more time on the beach and less time in the water, without making the overall path longer than it needs to be. A possible path of least time for the lifeguard is shown in Figure 28.4.

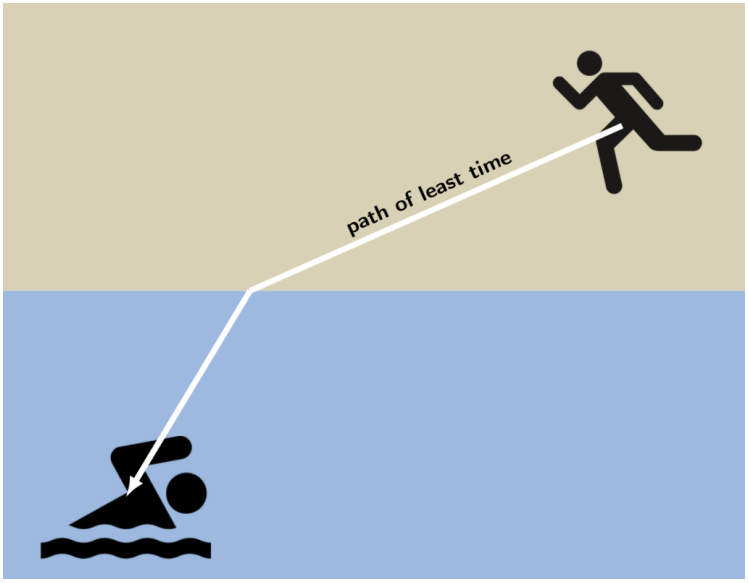


Figure 28.4 – The path of least time is the most efficient mix of spending more time on land and less time in water, without creating a path longer than it needs to be. The exact route would depend on the lifeguard's exact running and swimming speeds. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

When the speed of light changes throughout its journey from point A to point B, the light will bend in order to travel the path of least time.

Reflection

Reflection is what happens when waves bounce off of a surface. As we learned in the previous chapter of this textbook, selective reflection and absorption of different colors of light is what makes opaque objects look a certain color. When white light hits something blue, the blue light reflects (bounces) back to our eyes, and red and green light is absorbed by the object.

The law of reflection tells us how exactly each ray of light moves when it bounces off of a surface. Due to the principle of least time, and because the speed of light stays constant in a single medium, the path of least distance will be equal to the path of least time. (In this case, we are making the assumption that light will be traveling through a single homogeneous medium, such as air or water, as it travels to and bounces off of a surface.) This leads to the law of reflection: the angle of reflection is equal to the angle of incidence. (The incident light is the light that hits the mirror, and the reflected light is the light that bounces off of the mirror.) In equation form, the law of reflection can be stated as

$$\theta_r = \theta_i.$$

Angles are represented using the Greek letter theta (θ). In this equation, θ_r is the angle of the reflected light and θ_i is equal to the angle of the incident light.

When we talk about reflection and refraction, we measure angles with respect to the normal line. Wherever light is incident on a mirror, the normal line, at that point, is positioned at a right angle to the mirror's surface. On a planar (straight-line) mirror this is a line that is perpendicular to the mirror's surface everywhere on the mirror. On a curved mirror, the normal line will be perpendicular to the mirror's surface at the point where the light hits the mirror.

As stated above, when light hits a mirror at a spot, the angle of incidence is measured between the light ray and the normal line. The angle of reflection is equal to this angle of incidence. Therefore, if you know the angle of incidence, you can determine the angle of reflection by using the law of reflection: they are equal! In Figure 28.5, the mirror is planar and horizontal, and the normal line is depicted as a vertical black line at 0 degrees. The angle of incidence is measured between the normal line and the emitted light, which is 50 degrees. The angle of reflection is measured between the normal line and the reflected light, which is also 50 degrees.

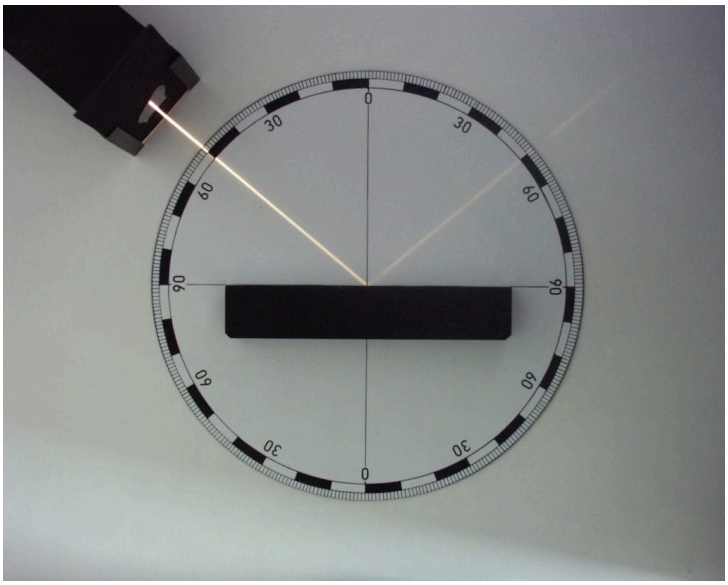


Figure 28.5 – The law of reflection states that the angle of incidence is equal to the angle of reflection. Both angles are measured with respect to the normal line. Light reflection, by Zátonyi Sándor, is licensed under CC BY-SA 3.0.

As previously noted, selective reflection is how different objects appear different colors. If reflection is occurring, why can't we see our reflection in all objects that allow light to bounce off of them? The answer is that, while light always obeys the law of reflection, sometimes surfaces are rough enough that the light will bounce back in scattered directions. This type of reflection is known as diffuse reflection, and is depicted in Figure 28.6. It should be stressed that each individual ray still obeys the law of reflection in which the angle of incidence equals the angle of reflection relative to the normal. It is just that for a rough surface the normal line varies from point to point on the surface. (In contrast, light that bounces off of a smooth, polished surface, such as a mirror, is known as specular reflection.)

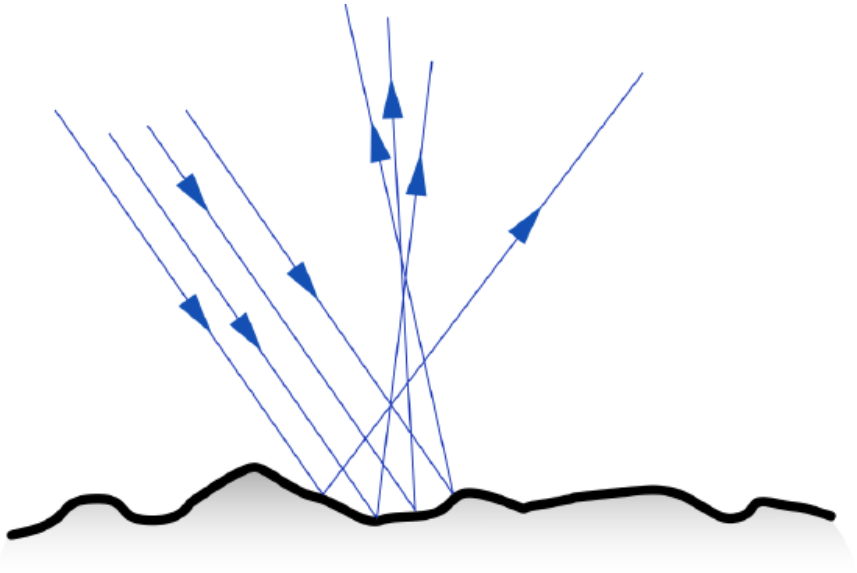


Figure 28.6 – Diffuse reflection follows the law of reflection, but scatters due to the roughness of the object's surface. Diffuse reflection, by MikeRun and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Mirrors

A mirror is a smooth reflective surface. Parallel rays of light that reflect off of a mirror can continue to travel parallel to each other, diverge (spread apart), or converge (come together) after reflection. These three scenarios are shown in Figure 28.7. Each ray of light reflects off of the mirror following the law of reflection. The converging or diverging property comes from the shape or curvature of the mirror.

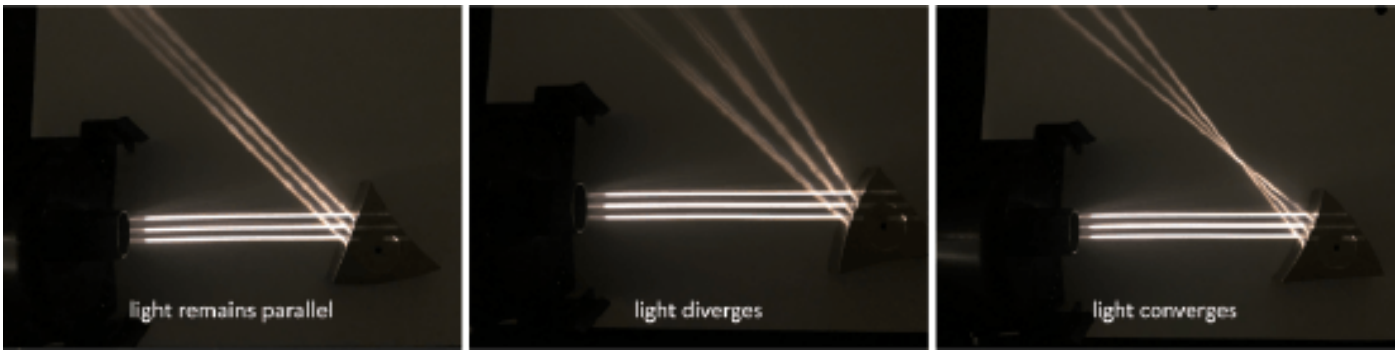


Figure 28.7 – Parallel light rays that are incident on a mirror can continue travelling parallel to each other after reflection (left), diverge (spread apart) after reflection (middle), or converge (come together) after reflection (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

When light reflects off of a mirror, the reflected light has properties that make it appear that those light rays were generated in a particular spot. For example, when we look in a bathroom mirror and see our reflection, that reflection occurs at a point where the reflected rays of light appear to have been generated. This reflection of ourselves, or of other objects, is known as an image.

An image is said to be virtual when the image is behind the mirror. A virtual image occurs when a mirror creates diverging rays of light upon reflection. Tracing these diverging rays backward (behind the mirror), the point where they come together is the point from which the reflected light appears to have been generated. The term “virtual” arises from the fact that no actual light from the object exists behind the mirror. A virtual image cannot be projected onto a screen.

An image is said to be real when the image is in front of the mirror. A real image occurs when a mirror creates converging rays of light upon reflection. The point at which the rays of light come together is the point from which the reflected light appears to have been generated. Since the reflected light actually passes through the point where the image is located, a real image can be projected onto a screen. Hence the term “real.”

Planar mirror

A planar mirror is a flat, smooth, reflective surface. You may use one in the mornings and evenings when you brush your teeth and hair. These types of mirrors are extremely common in our daily lives.

If parallel rays of light are incident on a planar mirror, they will continue to travel parallel to each other after reflecting off of the mirror. This is shown with three parallel rays of light in Figure 28.7 (left).

In a planar mirror, the image that appears is behind the mirror: a virtual image. If we trace the reflected rays of an object backward to where they appear to originate, they come together at points behind the mirror, creating a virtual image. Our eyes perceive the diverged light as coming from the virtual image. This virtual image appears upright and has the same size as the object. A ray diagram for a planar mirror is shown in Figure 28.8. The rays colored red in the diagram emanate from the tip of the object and the rays colored blue in the diagram emanate from the tail of the object. Based on their reflection off the mirror, the dashed lines point to where the image appears to have been generated.

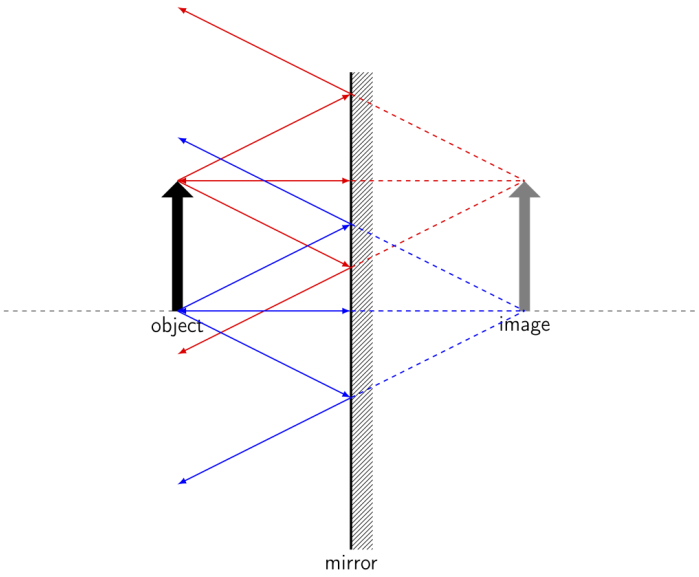


Figure 28.8 – A ray diagram of a planar mirror shows an image that is virtual, upright, and the same size as the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A virtual image that is upright and the same size as us corresponds to how we see our own reflections in a planar mirror. Dr. Pasquale took a selfie in a planar mirror; this photograph is shown in Figure 28.9. Note that her image is virtual (appears behind the mirror), upright, and is the same size that she is.



Figure 28.9 – A selfie of Dr. Pasquale taken in a planar mirror. Her image is virtual, upright, and the same size as she is. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Convex mirror

Mirrors can also be curved. A convex mirror is a mirror that's curved outward. If parallel rays of light are incident on a convex mirror, they will diverge after reflecting off of the mirror. This is shown with three parallel rays of light in Figure 28.7 (middle). A convex mirror has a focal point. If you were to draw lines tracing the diverged reflected light rays backward, the point where those rays come together is the focal point. Because a convex mirror causes light to diverge, the focal point is behind the mirror.

When we analyze light rays, we see that the object creates a virtual image behind the mirror that's upright and smaller than the object. A ray diagram for a convex mirror is shown in Figure 28.10.

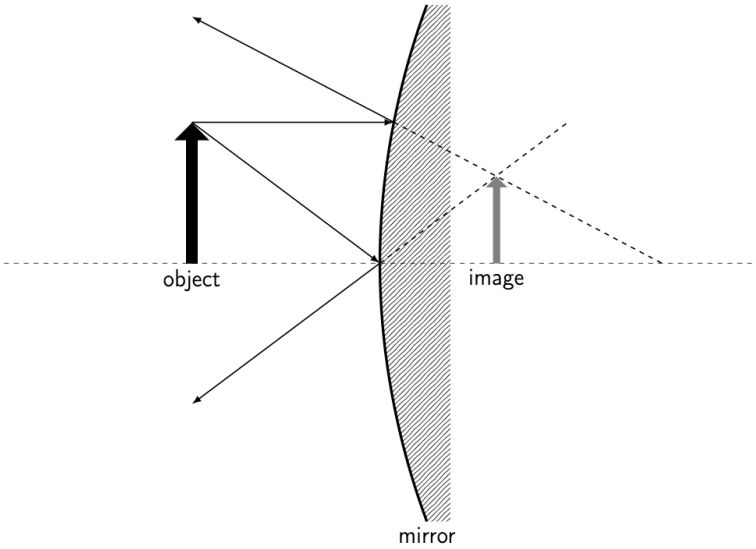


Figure 28.10 – A ray diagram of a convex mirror shows an image that is virtual, upright, and smaller than the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A virtual image that is upright and smaller than we are corresponds to how we see our own reflections in a convex mirror. Dr. Pasquale took a selfie in a convex mirror; this photograph is shown in Figure 28.11. Note that her image is virtual (appears behind the mirror), upright, and is smaller than she is.



Figure 28.11 – A selfie of Dr. Pasquale taken in a convex mirror. Her image is virtual, upright, and smaller than she is. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

There are many applications of convex mirrors in our lives. Perhaps you've seen them in a parking garage or other narrow space, to help you see what may be around a tight corner. A convex mirror seen in a parking garage in downtown Naperville is shown in Figure 28.12.



Figure 28.12 – A convex mirror seen in a parking garage in Downtown Naperville, Illinois. This mirror helps motorists see around tight spots to avoid collisions. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Convex mirrors are also used as the side mirrors in our cars. They show us a slightly smaller version of what's around us, allowing us to see an expanded view around the sides of our cars. This is also why there's usually a warning stating, "Objects are closer than they appear," as seen in Figure 28.13.



Figure 28.13 – The side mirror in a car is a convex mirror. A warning label typically states the "objects in the mirror are closer than they appear." Objects in mirror are closer than they appear, by Marcin Wichary, is licensed under CC BY 2.0.

The sculpture Cloud Gate (colloquially known as "The Bean") in Chicago, Illinois, is composed of convex mirrors on the outer surface. This is shown in Figure 28.14. Note that the images that appear in the sculpture are virtual, upright, and smaller than the objects that created them.



Figure 28.14 – Cloud Gate (also known as “The Bean”) in Chicago, Illinois, is composed of convex mirrors. Cloud Gate, Millennium Park, by Timothy Neesam, is licensed under CC BY-NC-ND 2.0.

Concave mirror

A concave mirror is a smooth reflective surface that’s curved inward. A concave mirror has a focal point, a place where all parallel rays will converge to a single point. Because a concave mirror causes light to converge after reflecting, the focal point is in front of the mirror. The convergence of three parallel beams of light on a concave mirror is shown in Figure 28.7 (right).

The image that’s created in a concave mirror can be virtual or real depending on where an object is positioned in relation to this focal point.

Consider an object that is between the focal point and the mirror. We can draw light rays and use the law of reflection to determine where those light rays will travel after bouncing off of the mirror. Our eyes perceive an image based on where these rays of light converge. In this case, they create a virtual image behind the mirror. The virtual image is upright and much larger than real life. A ray diagram corresponding to this scenario is shown in Figure 28.15.

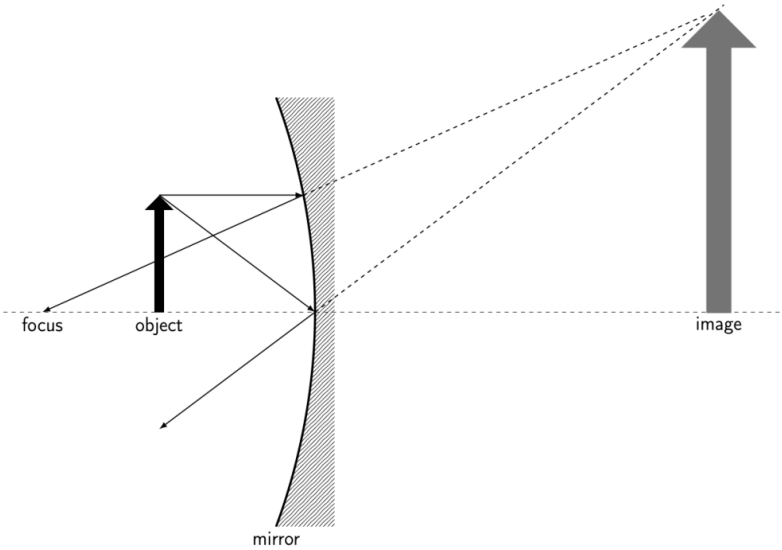


Figure 28.15 – A ray diagram of a concave mirror with the object between the focus and the mirror shows an image that is virtual, upright, and larger than the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Dr. Pasquale took a selfie while she was positioned between the focal point and the surface of a concave mirror. This photograph is shown in Figure 28.16. Notice that her image is virtual (appears behind the mirror), upright, and is larger than she is.



Figure 28.16 – A selfie of Dr. Pasquale taken in a concave mirror between the focal point and the mirror. Her image is virtual, upright, and larger than she is. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

You may have used a concave mirror in this fashion if you've used a makeup or shaving mirror. This greatly magnifies the appearance of your face and makes it easier to apply makeup or to shave.

As an object moves past the focal point, the image changes. Now, we can trace each ray of light and see that they converge on the same side of the mirror as the object, causing a real image. Again, the term real image simply means that the image is created on the same side as the object, due to light rays converging rather than diverging. The real image generated in this case is upside down. The image may be larger than, smaller than, or the same size as the object, depending on how far away from the focal point the object is positioned. A ray diagram of a concave mirror with the object past the focal point is shown in Figure 28.17. When the object is close to the focus (Figure 28.17, left) the image is larger than the object. As the object moves farther away from the focus (Figure 28.17, right) the image becomes smaller than the object.

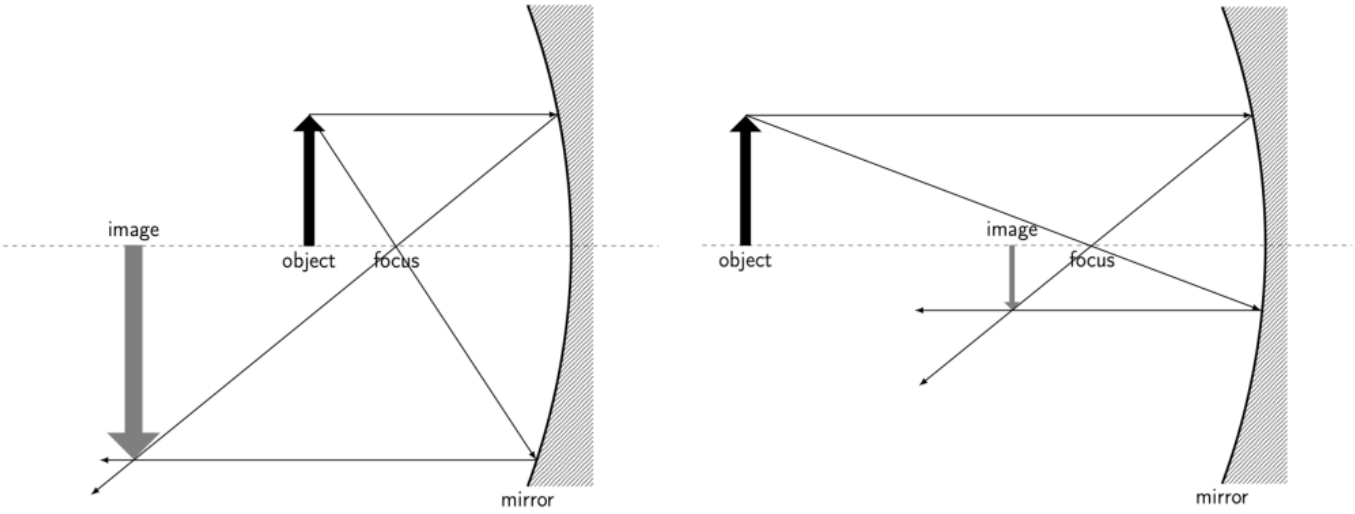


Figure 28.17 – A ray diagram of a concave mirror with the object beyond the focal point of the mirror. When the object is near the focus, the image is real, inverted, and larger than the object (left). When the object is farther away from the focus, the image is real, inverted, and smaller than the object (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Dr. Pasquale took a selfie while she was positioned past the focal point of a concave mirror. This photograph is shown in Figure

28.18. Notice that her image is real (appears in front of the mirror), upside down, and is larger than she is.



Figure 28.18 – A selfie of Dr. Pasquale taken in a concave mirror beyond the focal point. Her image is real, inverted, and larger than she is. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

It's possible you've seen this effect if you've looked in a funhouse mirror. And if you step far enough away from a makeup or shaving mirror, you'll notice that you will pass the focal point and go from being upright and magnified to upside down and magnified. If you continue walking backward, your image will go from being magnified to the same size, to smaller than you are.

Because a concave mirror generates a real image by focusing light in front of the mirror, that image can be projected onto a screen or seen by others as an optical illusion. For example, in Figure 28.19 (left), a light bulb appears to be screwed into the socket placed on top of a cardboard box. However, when viewed from the side, as in Figure 28.19 (right), the light bulb (object) is actually upside down below the cardboard box. A concave mirror creates the inverted image of the light bulb making it appear right side up on top of the box.



Figure 28.19 – A light bulb appears to be screwed into a socket connected to a cardboard box (left). When viewed from the side, the actual light bulb (object) is actually upside down and connected under the box. A concave mirror creates a real image of the light bulb seen in this optical illusion. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Planar, concave, and convex mirrors are used in reflecting telescopes. A telescope is used to collect light from far away celestial objects and focus it to a point, which is accomplished by using curved mirrors. Typically, a concave mirror is used to collect the

incoming light to a point. Either film, a digital recording sensor, or an eyepiece can be placed at that focus to record or view the celestial object. Among many other telescopes that use mirrors is the Hubble Space Telescope.

One-way mirrors

A typical household mirror is simply a very smooth reflective metal (possibly silver and/or aluminum) that's been highly polished and covered with glass. If you touch the mirror, you'll notice a gap between your finger and the reflection. (Dr. Pasquale has demonstrated this in Figure 28.20.) This gap exists because of the presence of the glass protecting the polished metal surface.

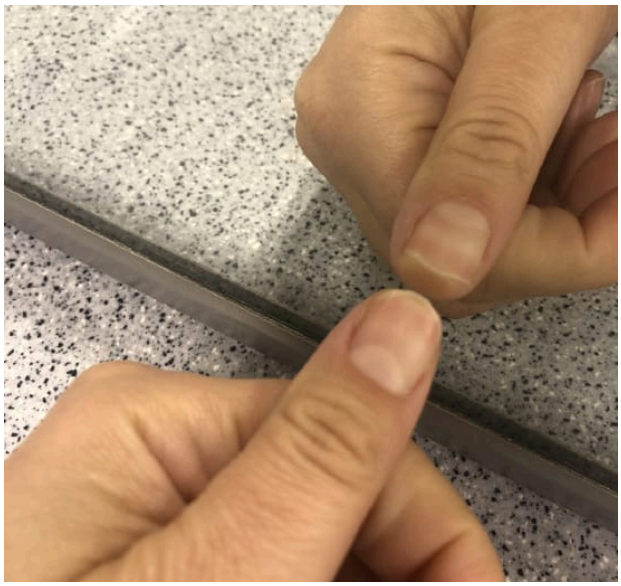


Figure 28.20 – A typical household mirror has a highly polished reflective surface covered by glass. Because of the glass, there will be a gap between an object and the image. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Perhaps you've heard of a one-way mirror. A one-way mirror is simply a piece of glass that separates a dark area and a light area. These surfaces appear to look like a mirror if you're on the light side of the glass, but they look like a window if you're on the dark side of the glass. In fact, windows in buildings and homes also appear this way at night when it's dark outside but interior lighting is used.

A one-way mirror isn't really much of a mirror at all. On the light side, we can easily see our reflection because the light from the reflection outweighs the small amount of light passing through from the dark side of the glass (Figure 28.21, left). On the dark side, we simply see the light that transmits through from the bright side (Figure 28.21, right). There is too much transmitted light to enable somebody on the dark side of the glass to see their reflection. Instead, they see whatever is on the other side of the glass.

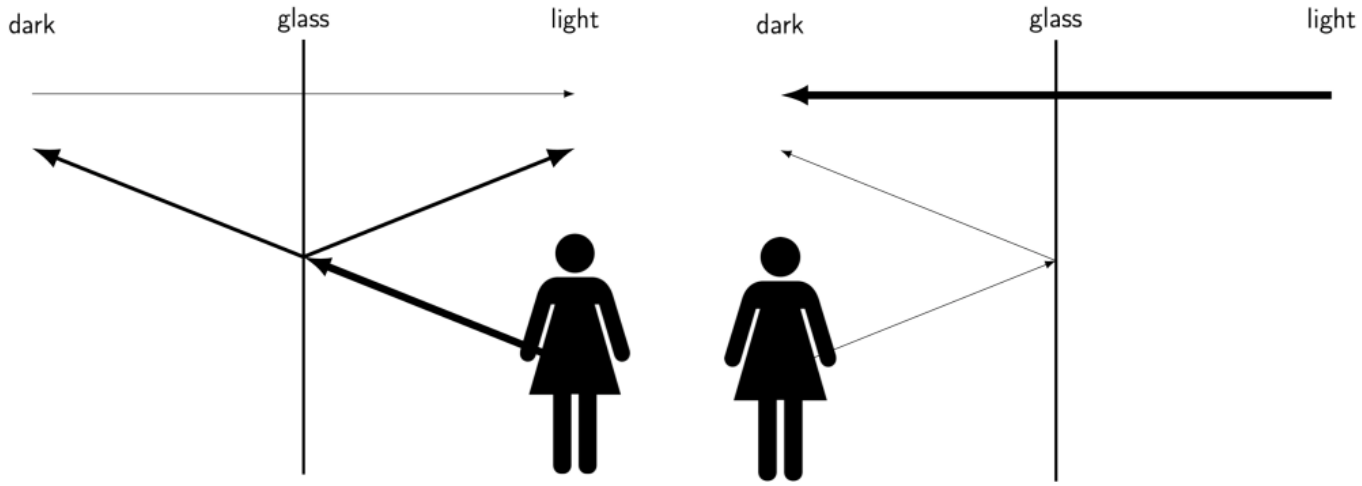


Figure 28.21 – A person standing on the bright side of a window will see their reflection, as the reflected light is much stronger than any light coming through the window from the dark side (left). A person standing on the dark side of a window will only see light coming through from the bright side, as the reflected light is much weaker than the light coming through from the bright side (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

You can identify a one-way mirror by touching the glass. Because there is no metal reflecting surface, you'll notice there is no gap between your finger and its reflection. This is demonstrated by Dr. Pasquale in Figure 28.22. (Note that this cannot distinguish a one-way mirror from a "normal" mirror that does not have a glass coating over the polished reflective surface.)



Figure 28.22 – A one-way mirror (in this case, a window separating a dark area from a light area) creates a reflection on the surface of the glass. In this case, there is no gap between an object and its image. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Refraction

Refraction is a bending of light that occurs when light travels through different media. Because of the principle of least time, light will bend when it travels from one transparent medium into another. For example, light will bend when it travels from air to a medium such as water or plastic where the speed of light is slower than it is in free space.

The extent to which the speed of light slows down is known as the index of refraction. (Different media can be characterized by their own index of refraction.) Index of refraction has a symbol of the lowercase letter n, and is equal to the ratio of the speed of light in free space and the speed of light in that medium. In equation form,

$$n = \frac{c}{v},$$

where c is the speed of light in free space (3×10^8 m/s) and v is the speed of light in that particular medium.

The index of refraction of a medium is an indication of how much light will bend when it enters that medium from free space. Something with a smaller index of refraction will bend light much less than something with a higher index of refraction. For context, water has an index of refraction of 1.3. Diamond has a relatively high index of refraction of 2.4.

To quantify the bending of light, we look at the normal line at the point where light hits the interface of the two media. (Remember: the normal line is drawn at a 90 degree angle to the medium at that point.) Light enters at an angle of incidence, which is measured with respect to the normal line. This is shown in Figure 28.23, which depicts light traveling from point P to point O in a medium with index of refraction n_1 and speed of v_1. The angle of incidence is depicted as θ_1, measured with respect to the normal line. The light bends at the interface and travels from point O to point Q in a new medium with index of refraction n_2 and speed of v_2. The angle of refraction is depicted as θ_2, measured with respect to the normal line.

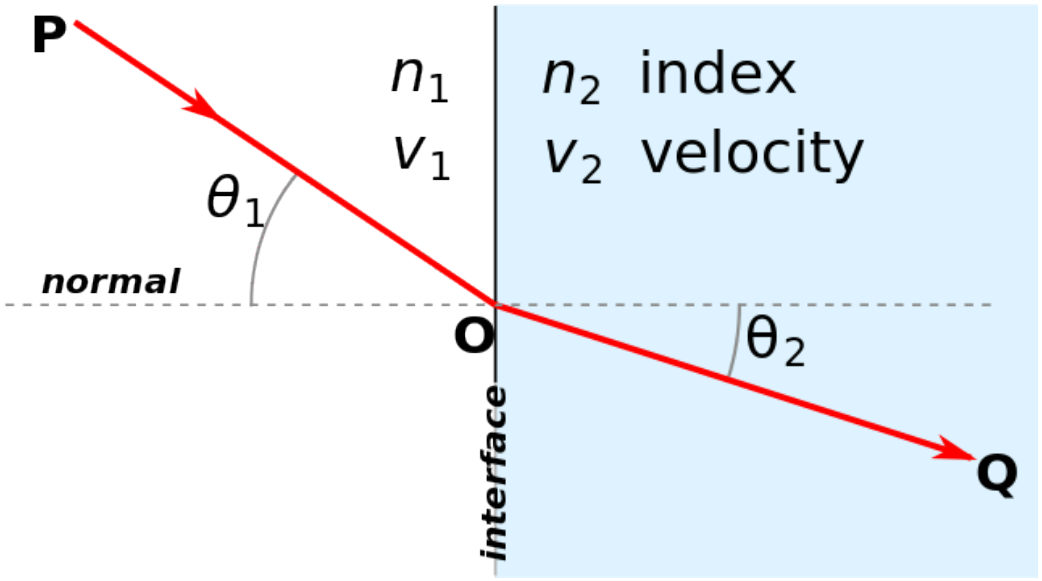


Figure 28.23 – Light traveling through two media will bend at the interface. The angle of incidence and angle of refraction are both measured with respect to the normal line. Snells law, by Cristan, is licensed under Public Domain.

When light travels from a lower to higher index of refraction (for example, from air to glass), the angle of refraction will be smaller than the angle of incidence. When light travels from a higher to a lower index of refraction (for example, from water to air), the angle of refraction will be larger than the angle of incidence.

The exact amount of bending can be explained by Snell's law of refraction. Note that Snell's law of refraction contains trigonometry, which you are not expected to be familiar with in this textbook. Therefore, we will only consider Snell's law qualitatively. In equation form, Snell's law of refraction states that

$$n_i \sin\theta_i = n_r \sin\theta_r,$$

where n_i is the index of refraction of the first medium, θ_i is the angle of incidence, n_r is the index of refraction of the second medium, and θ_r is the angle of refraction.

Because the angle of incidence is measured with respect to the normal line, it can only range between 0 degrees and 90 degrees. This means that as the value of the angle of incidence increases, the sine of the angle of incidence increases as well. Because the product of index of refraction and the sine of the angle is constant on both sides of the equation, if one variable is increased on one side of the equation, the other variable on that side of the equation must decrease to compensate. In other words: as the index of refraction of the second medium increases, the angle of refraction must decrease to keep the product constant (or vice versa).

This bending of light explains some things you may have seen in your everyday life. For example, when a straw is placed in a glass of water, the refraction of light in the water causes the straw to look broken. This is shown with a pen in Figure 28.24. Note that

refraction occurs twice: when light travels from air to the glass, and when it travels from the glass to the water.



Figure 28.24 – A red pen appears to be broken; this is simply an optical illusion caused by the refraction of light between air and the glass, and between the glass and the water. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Refraction is also used to help our eyesight (among other things). This will be explained in more detail later in this chapter.

Dispersion

Although Snell's law of refraction seems to indicate that the amount of bending that occurs in a material is independent of wavelength, in fact, different wavelengths of light will bend by differing amounts through different media. The index of refraction of a material is not just a single number, but is a function that varies based on the light wavelength. This fact can be used to split light into its constituent colors. This property is known as dispersion.

A prism splits white light into a rainbow due to dispersion (recall from the previous chapter of this textbook that white light is a mixture of all colors of light). Each color of light bends different amounts. Red bends the least through glass, and violet the most. This difference in bending is what causes a rainbow to form. A photograph of a prism splitting white light into a rainbow is shown in Figure 28.25.

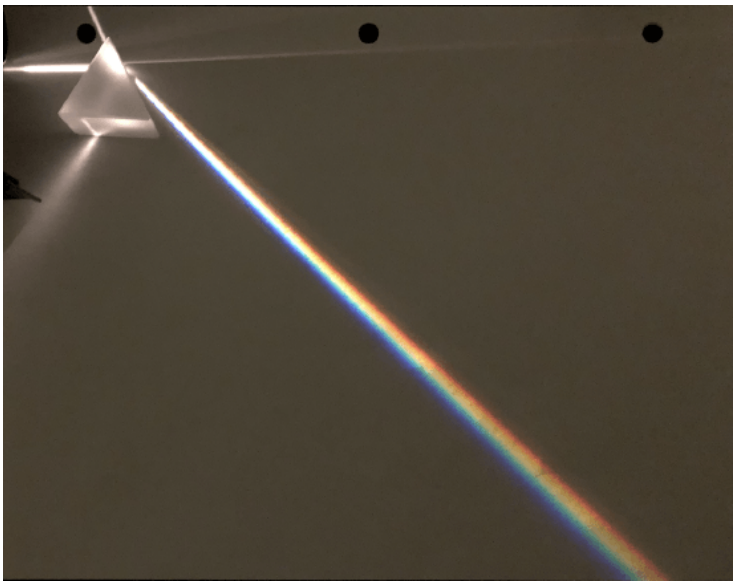


Figure 28.25 – A prism will split white light into a rainbow of colors. This is due to the property of dispersion. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Dispersion also occurs in raindrops and is the physical cause of rainbows we see in the sky. Rainbows form from a combination of reflection, refraction, and dispersion. In order to see a rainbow in the sky, the Sun needs to be behind you and atmospheric water droplets need to be in front of you. Some of the white light from the Sun refracts into the water droplet and disperses into the familiar spectrum of colors within the droplet. These rays strike the back side of the droplet, where some of the light reflects back into the droplet and continues to disperse. Some of the dispersed light that returns to the front side of the droplet refracts back out toward the observer. Since the index of refraction of water is greater for violet than for red light, the violet emerges on top and the red light emerges at the bottom. In that case, why do the rainbows we see have the red on top of the bow and violet on the bottom? The reverse ordering of the colors arises from the light returning from different droplets.

If red light from droplets reaches your eyes, then the violet light passes over your head. If violet light is reaching your eyes, then the red light is below your head. The rainbow you see is due to dispersed light returning from different droplets. Your eyes intercept the red light from droplets higher in the sky and the violet light from droplets lower in the sky as depicted in Figure 28.26.

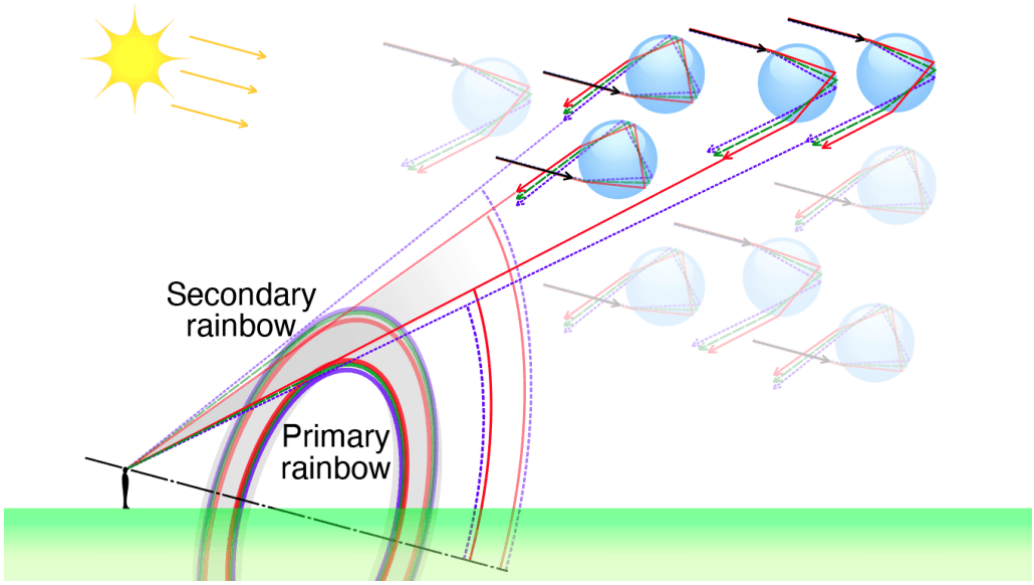


Figure 28.26 – A primary rainbow is generated by the refraction of sunlight through spherical water droplets. A secondary rainbow can be generated when light reflects twice before refracting. Rainbow principle, by Cmglee and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Note that you can often see a secondary rainbow above the primary rainbow. The secondary rainbow comes about from a double reflection within the droplet. Some of the white light from the Sun that enters near the bottom of a droplet can undergo a reflection twice within the droplet and emerge near the top. Note that the ordering of the emerging color spectrum is reversed with red emerging on top and violet emerging on the bottom. This results in the reversed ordering of the spectrum when the double-reflected light from different droplets reaches your eyes. This phenomenon, known as a double rainbow, is shown in Figure 28.27.



Figure 28.27 – A photograph of a double rainbow. Note the primary rainbow has colors of ROYGBIV from top to bottom, and the secondary rainbow has these colors in reverse order. Rainbow over the Lipno reservoir, by Alexis Dworsky, is licensed under CC BY 2.0.

Total internal reflection

According to Snell's law, light that travels from an object with a high index of refraction to a medium with a lower index of refraction will bend the refracted light outward. (The angle of refraction will be larger than the angle of incidence.) At a certain point, known as the critical angle, light that is incident on an interface between a high index of refraction medium and a low index of refraction medium will bend at a 90 degree angle, exactly along the interface. Light that is incident at angles greater than the critical angle will not refract at all, but will instead completely reflect off of the interface. This phenomenon is known as total internal reflection.

Total internal reflection is demonstrated by Dr. Pasquale in the video below. Light enters into a piece of plastic from a light source. At the interface between the plastic and the air, some of the light is bent outward into the air (refracted) and some of the light is reflected off of the interface. As Dr. Pasquale rotates the setup to increase the angle of incidence, eventually, the refracted light bends at a 90 degree angle, and then vanishes. This is the point of total internal reflection.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Total internal reflection is used in fiber optics. Light rays are focused into a long fiber of glass with a laser, and total internal reflection causes those light rays to remain trapped inside of the fiber, allowing them to travel long distances as they carry telecommunications signals from one place to another.

An analogous process is shown in Figure 28.28, where red (left), green (middle), and violet (right) laser pointers were shone into

a piece of curved plastic. You can see that when the plastic curves, because it has a higher index of refraction than air, the light from the laser pointers remains inside the plastic due to total internal reflection.

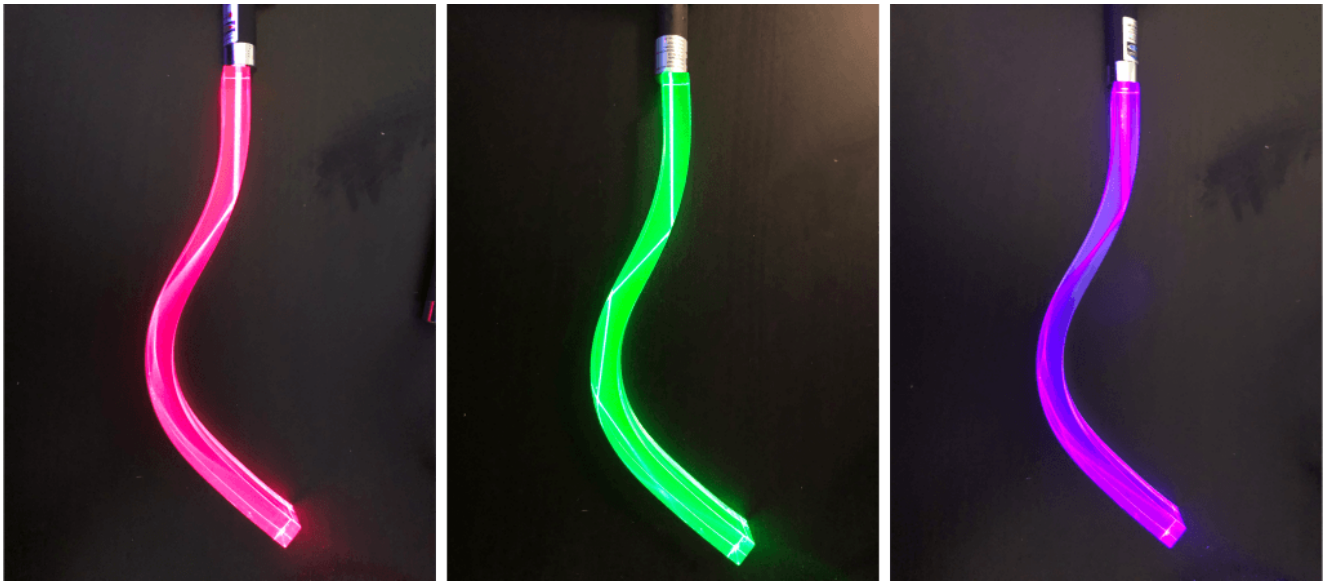


Figure 28.28 – Total internal reflection is demonstrated by shining a red (left), green (middle), and blue (right) laser pointer into a curved piece of plastic. This is similar to how fiber optic cables send light over long distances. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Diamonds are cut with angles such that much of the light that travels into the diamond is trapped inside with total internal reflection, causing them to sparkle and shine.

If you've ever been under water and looked upward, total internal reflection is the reason why the surface of water from underneath tends to look shiny, somewhat like a mirror. A photograph of this "mirror-like" property of water's surface is shown in Figure 28.29. (Note that the fish at the top of the image is a reflection of the fish at the bottom of the image.)



Figure 28.29 – When viewed from underneath, water's surface appears mirror-like due to total internal reflection. Black triggerfish, by Jan Derk, is licensed under Public Domain.

Lenses

A lens is a piece of transparent material that causes light to bend and either converge or diverge after traveling through it. Parallel

rays of light that refract through a lens can diverge or converge after refraction. These two scenarios are shown in Figure 28.30. Each ray of light follows Snell's law at both interfaces between the air (or other medium) and the lens. The converging or diverging property comes from the shape or curvature of the lens.

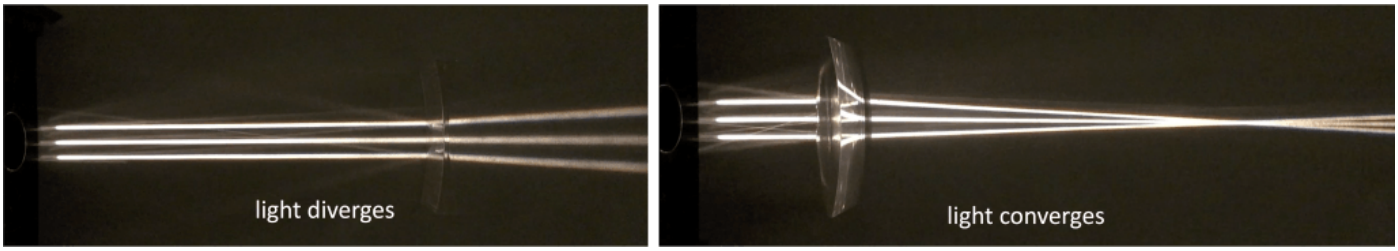


Figure 28.30 – Parallel light rays that are incident on a lens can diverge (spread apart) after refraction (left) or converge (come together) after refraction (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Just as mirrors cause reflected light to appear to have been generated from a particular spot, lenses cause refracted light to appear to have been generated from a particular spot. Our eyes perceive the bent light as having come from that spot, known as an image. An image is said to be virtual when the image is on the same side of the lens as the object, and the image is said to be real when the image is on the opposite side of the lens as the object. (This is the opposite of the way we use the terms "virtual" and "real" when discussing mirrors.) This again has to do with whether the lens causes the light rays to diverge or converge beyond the lens. Light rays that converge beyond the lens form a real image that can be projected on a screen. The diverging rays beyond the lens shown in Figure 28.30 (left) do not actually come from the focal point on the left side of the lens. Hence, the image cannot be projected onto a screen and is therefore virtual as described below.

Biconcave lens

A biconcave lens causes parallel light waves to diverge. Biconcave means that both sides of the lens are curved inward. We can determine that divergence occurs by applying Snell's law of refraction at each interface between air and the lens. If we trace the diverged light rays backward, we can find a focal point. This is known as a virtual focal point because it's in front of the lens. The virtual focal point for the lens depicted above in Figure 28.30 (left) is shown below in Figure 28.31.

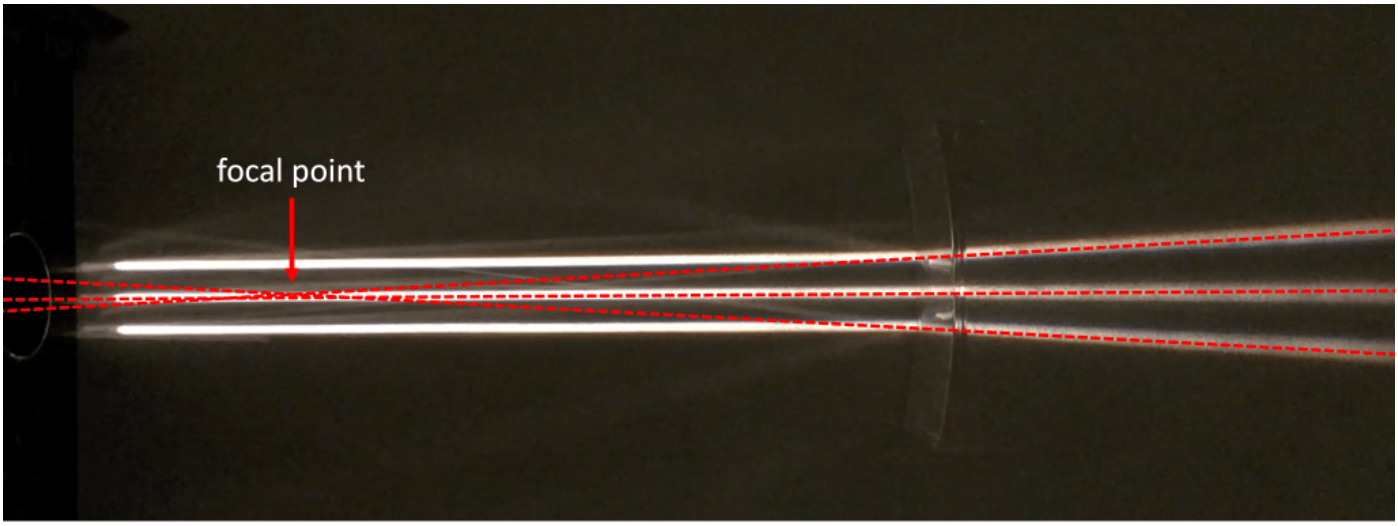


Figure 28.31 – The virtual focal point of a diverging lens can be found by tracing the diverging rays of light backward to where each ray intersects. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

An image viewed through a concave lens will be upright and smaller than the object that creates the image. It is known as a virtual image as it appears on the same side of the lens as the object. A ray diagram for a biconcave lens is shown in Figure 28.32.

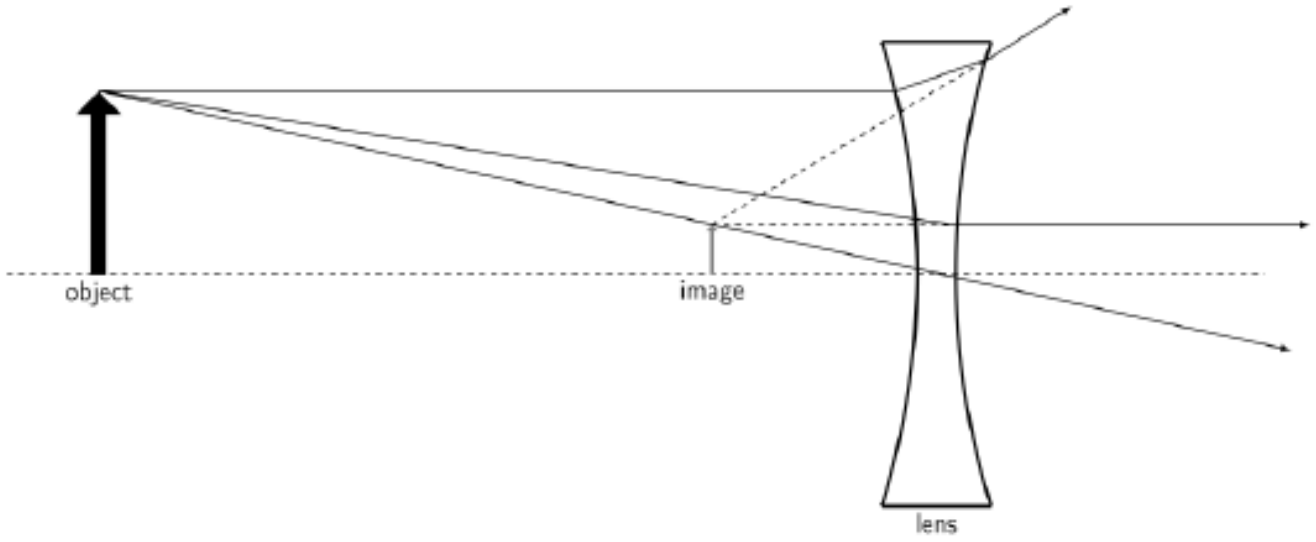


Figure 28.32 – A ray diagram of a biconcave lens shows an image that is virtual, upright, and smaller than the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A photograph of the image of the word “PHYSICS!” is seen through a biconcave lens and shown in Figure 28.33. Note that the image is upright and smaller than the object.

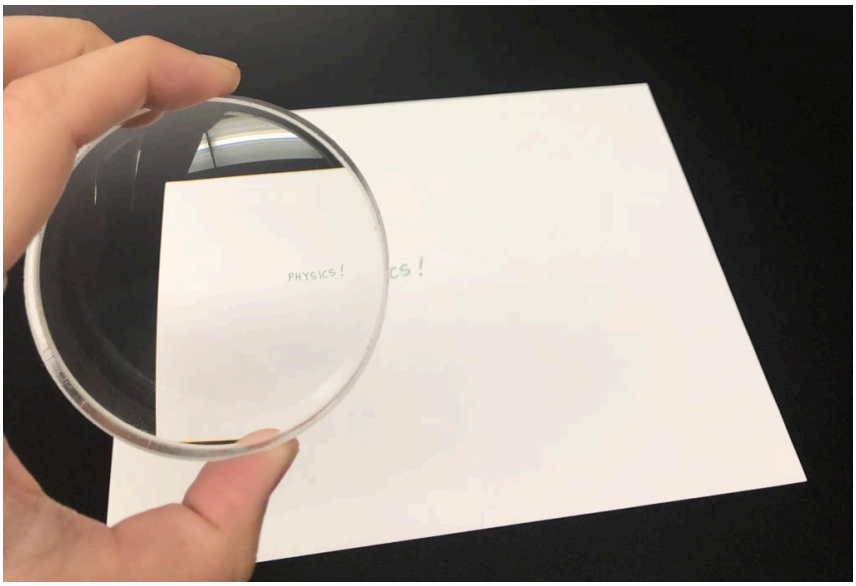


Figure 28.33 – An image of the word PHYSICS! viewed through a biconcave lens. Note that the image is virtual, upright, and smaller than the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Eyeglasses use refraction of light through glass to focus light on the retina. People who are nearsighted (such as Dr. Pasquale) have eyes that focus light in front of their retina. Eyeglasses are designed to bend the light and cause it to focus on the retina instead. The process to correct nearsightedness uses diverging lenses such as biconcave lenses. The diagram of this process is shown in Figure 28.34.

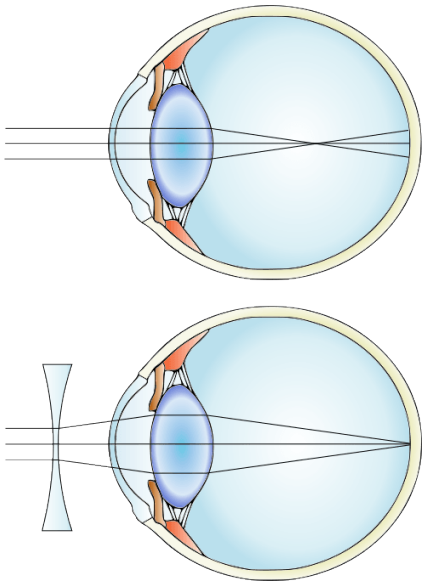


Figure 28.34 – A myopic eye will focus light rays in front of the retina (top). A corrective lens is used to bend the light so that it focuses on the retina (bottom). Myopia and lens correction, by Gumenyuk I.S., is licensed under CC BY-SA 4.0.

Biconvex lens

A biconvex lens causes parallel light waves to bend inward, or converge (as shown above in Figure 28.30, right). Biconvex means that both sides of the lens are curved outward. Because light that travels through a biconvex lens converges, the focal point is said to be real. The focal length is the distance between the lens and the focal point. The image that's created in a biconvex lens has to do with where an object is positioned in relation to this focal point.

When an object is viewed through a biconvex lens between the focal point and the lens, the image will be virtual, upright, and larger than the object. A ray diagram for a biconvex lens in this scenario is shown in Figure 28.35.

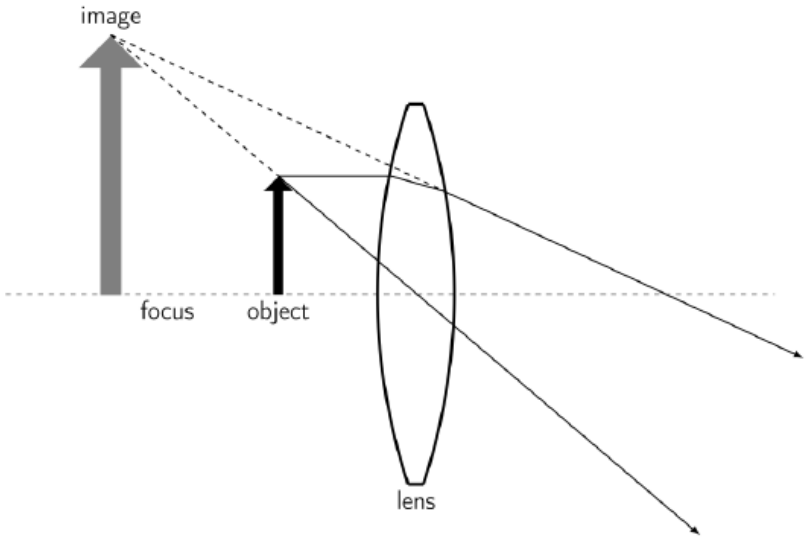


Figure 28.35 – A ray diagram of a biconvex lens with the object between the focal point and the lens shows an image that is virtual, upright, and larger than the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A photograph of the image of the word "PHYSICS!" is seen through a biconvex lens between the focal point and the lens, and shown in Figure 28.36. Note that the image is upright.

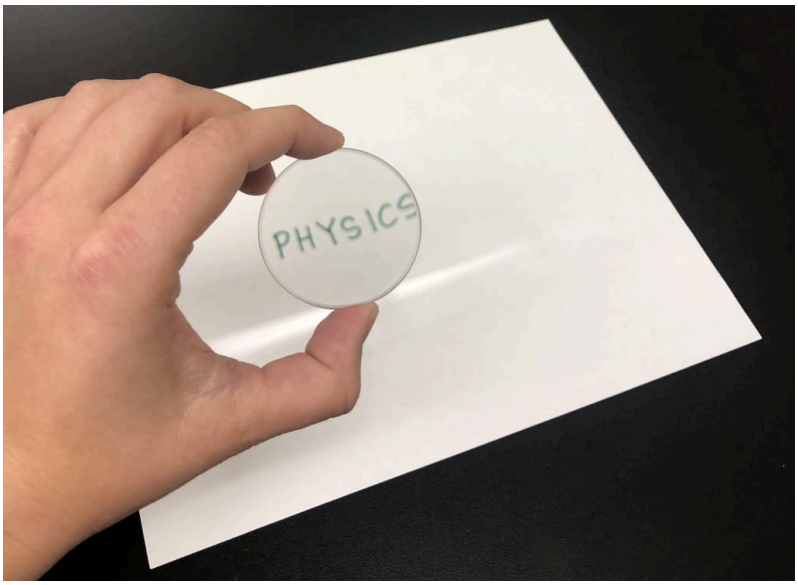


Figure 28.36 – An image of the word PHYSICS! viewed through a biconvex lens between the focal point and the lens. Note that the image is virtual, upright, and larger than the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

As the object moves away from the focal point, the image changes. When the object moves beyond the focal point, the image becomes real, inverted, and magnified. As the object continues to move farther away, it will eventually become reduced instead of magnified. A ray diagram of a biconvex lens with the object past the focal point is shown in Figure 28.37. When the object is close to the focus (Figure 28.37, left) the image is larger than the object. As the object moves farther away from the focus (Figure 28.37, right) the image becomes smaller than the object.

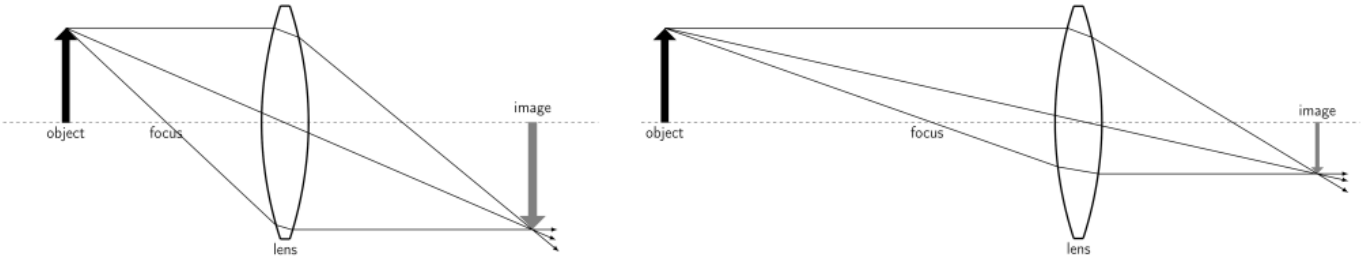


Figure 28.37 – A ray diagram of a biconvex lens with the object beyond the focal point. When the object is near the focus, the image is real, inverted, and larger than the object (left). When the object is farther away from the focus, the image is real, inverted, and smaller than the object (right). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A photograph of the image of the word “PHYSICS!” is seen through a biconvex lens beyond the focal point and shown in Figure 28.38. Note that the image is inverted and larger than the object.

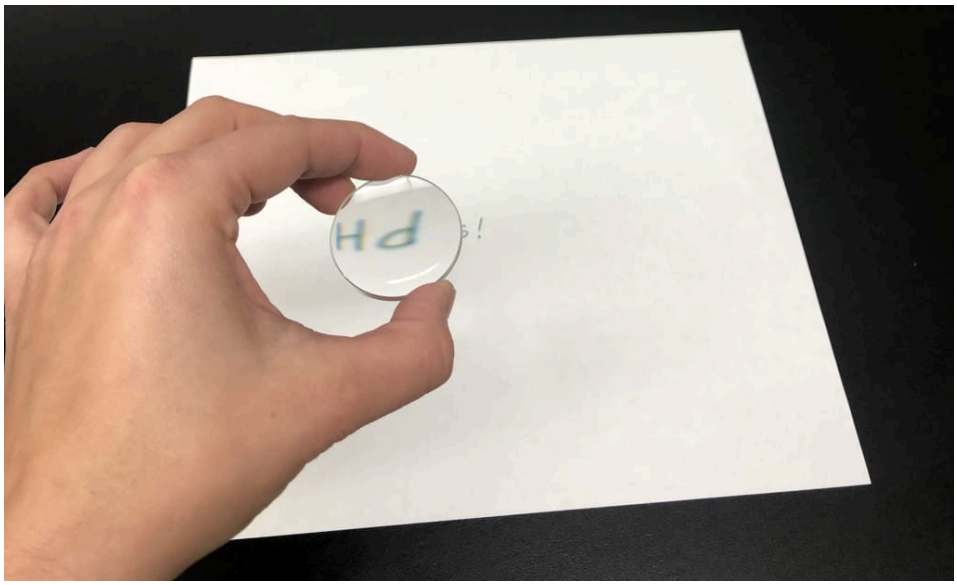


Figure 28.38 – An image of the word PHYSICS! viewed through a biconvex lens past the focal point. Note that the image is real, inverted, and larger than the object. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

An application of biconvex lenses is in correcting hyperopia (farsightedness). People who are farsighted have eyes that focus light behind their retina. For example, Dr. Fazzini needs to wear reading glasses or "readers" that use a converging lens (such as a biconvex lens) to correct hyperopia by focusing light onto the retina. The diagram of this process is shown in Figure 28.39.

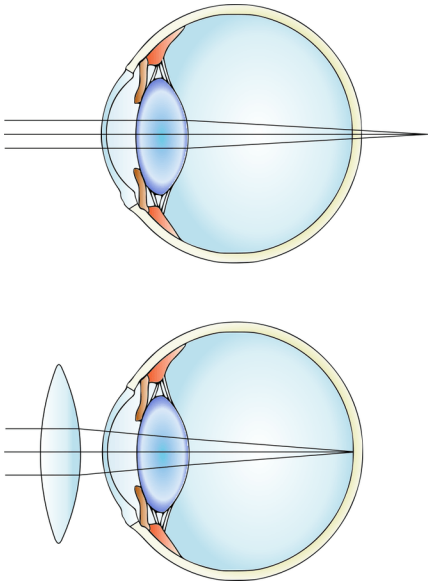


Figure 28.39 – A hyperopic eye will focus light rays behind of the retina (top). A corrective lens is used to bend the light so that it focuses on the retina (bottom). Hypermetropia color, by Гуменюк И.С., is licensed under CC BY-SA 4.0.

Other lens applications

Because lenses have the ability to reduce and enlarge the size of objects when creating images, they are widely used in many optical applications where small or distant objects need to be enlarged.

Optical microscopes use visible light to enlarge an object so it can be seen with more detail. A compound microscope uses two lenses (an eyepiece lens and an objective lens) to create an enlarged, upright image, as shown in Figure 28.40. Some compound microscopes can project the image onto a sensor for recording as a digital image.

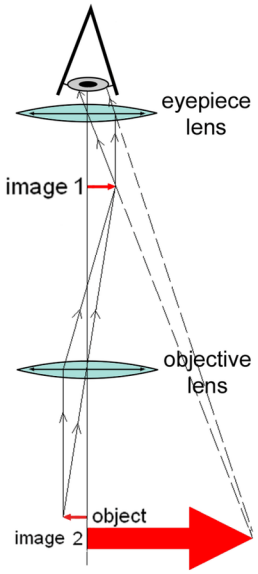


Figure 28.40 – A compound microscope creates an enlarged image by using two lenses. Microscope compound diagram, by Fountains of Bryn Mawr, is licensed under CC BY-SA 3.0.

Refracting telescopes make use of lenses to make images of distant objects in the night sky. A ray diagram for a simple refracting telescope is shown in Figure 28.41. Parallel rays of light from a distant celestial object are focused to a point by an objective lens, creating a real image in between the objective and eyepiece lenses. This image becomes the object for the eyepiece lens, which creates the magnified virtual image viewed by the user of the telescope.

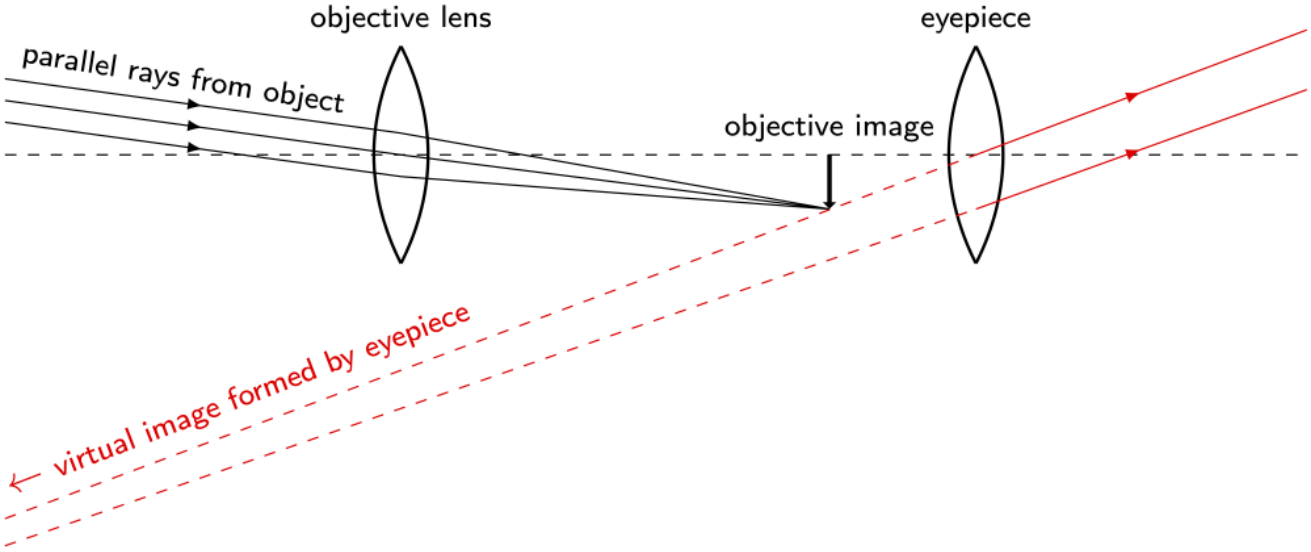


Figure 28.41 – A refracting telescope generates a virtual image of a distant celestial object after light passes through objective and eyepiece lenses. This figure, created by Alyssa J. Pasquale, Ph.D. and David R. Fazzini, Ph.D., is licensed under CC BY-NC-SA 4.0.

Refracting telescopes are ultimately limited in size by the ability to manufacture large lenses. (In addition, the need to mount lenses on their edges so as not to obstruct the field of view can cause gravitational issues with large lenses.) They also suffer from chromatic aberration, which is a type of lens defect described below. Refracting telescopes have been mostly replaced by reflecting telescopes, which use a collection of mirrors instead of lenses to create an image of distant objects.

Cameras use lenses and mirrors to focus light onto film (or other light-sensitive material) or a digital sensor to create an image. There are many types of camera lenses, each of which are typically composed of multiple lenses. The camera lens shown in Figure 28.42 has been cut in half to demonstrate this property. Whether a camera lens acts as a telephoto (zoom) lens or a wide-angle lens (or something else) has to do with the lenses that are used.

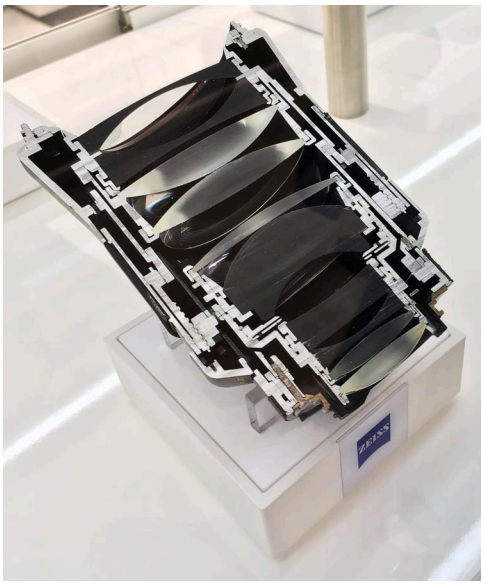


Figure 28.42 – A camera lens is itself composed of multiple lenses to generate an image. Zeiss Otus 55mm f1.4 cut, by Laserlicht, is licensed under CC BY-SA 4.0.

Optical aberrations

While lenses and mirrors are built to create images, there can be issues with lenses and mirrors due to their composition and construction. A distortion or aberration occurs when light is not focused at a single point. Instead, light rays are spread out around the focal point.

Spherical aberration

Both lenses and mirrors can both suffer a type of distortion known as spherical aberration. To obtain a single focal point, a parabolic mirror or lens shape is required. However, a spherical shape is much easier to manufacture. This aberration is not prevalent for light rays that are incident close to the center of the lens (or mirror). However, when light is incident far from the center of the lens or mirror, the difference in focal point will be quite pronounced.

A parabolic lens without spherical aberration is shown in Figure 28.43 (top), while a spherical lens that does suffer from spherical aberration is shown in Figure 28.43 (bottom).

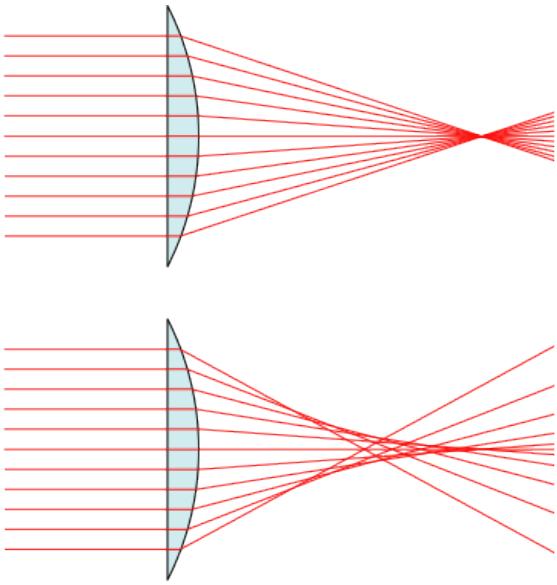


Figure 28.43 – In a parabolic lens, parallel light rays will focus to a single focal point (top). In a spherical lens, parallel light rays will instead spread out around the focal point (bottom), which is known as spherical aberration. Spherical aberration 2, by Mglg, is licensed under Public Domain.

A mirror with spherical aberration is shown in Figure 28.44.

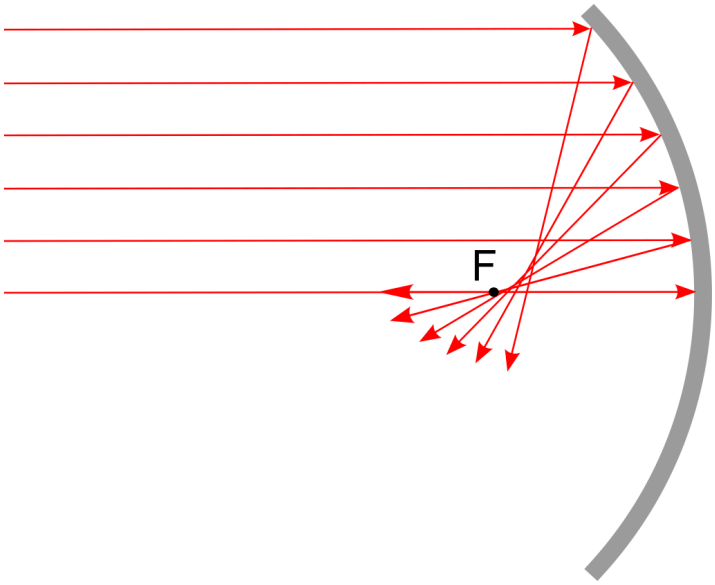


Figure 28.44 – In a spherical mirror, parallel light rays will spread out around the focal point instead of focusing to a single point. Spherical aberration of a concave spherical mirror, by Jean-Jacques MILAN and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Chromatic aberration

Chromatic aberration occurs when different colors of light focus at different focal points. This is due to dispersion. Through glass or plastic lenses, red light will bend less than blue light, all other things being equal. An exaggerated depiction of chromatic aberration is shown in Figure 28.45. Lenses suffer from chromatic aberration regardless of their shape. However, mirrors do not because the law of reflection holds regardless of the wavelength of the light.

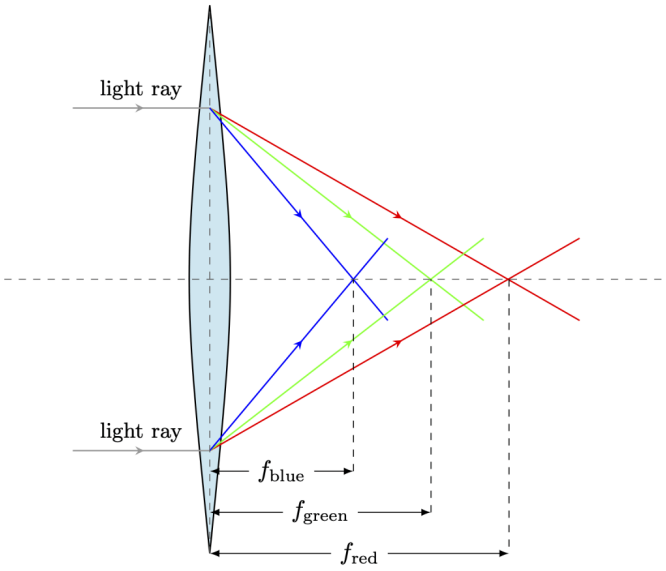


Figure 28.45 – Chromatic aberration occurs due to dispersion. Red light will bend less than blue light through a glass or plastic lens. Lens Color Error, by Alexandros Tsagkaropoulos, is licensed under CC BY-SA 4.0.

Further reading

- Yerkes Observatory – The Yerkes Observatory in Lake Geneva, Wisconsin, contains two telescopes: one refracting and one reflecting telescope. The refracting telescope at Yerkes was the largest in the world as of the time of its dedication in 1897. Since then, many notable scientists have made use of the telescope to conduct scientific research.

Practice questions

Hands-on experiments

1. Shine a flashlight onto a mirror at different angles. Observe the angle of incidence and the angle of reflection (measured with respect to the normal line). Use a ruler to measure and compare these angles.
2. Fill a clear glass with water. Place a pencil in the glass and observe how it appears bent or displaced. Does the bent appearance change based on the angle from which you observe the glass? (Look from above, below, and from the sides of the glass.)
3. Place a coin at the bottom of a clear glass. Observe the coin from the side of the glass and then from above the water level. How does refraction change where the coin appears to be located?

This page titled Chapter 28: Reflection and refraction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 29: Light waves

Summary

Light has many fascinating properties and interacts in many meaningful and useful ways. By treating light as a wave and understanding its wave properties, we can learn about even more properties and applications of light in our lives.

Light models

So far, we have discussed how light travels by treating it like a ray. A beam of light travels from one place to another, and along the way it might bounce (reflect) or bend (refract). Possibly it might transmit or be absorbed. This simple model of light explains a lot: mirrors, lenses, rainbows, and other phenomena we've discussed in the past few chapters of this textbook. This model of light is known as ray optics.

Ray optics is great in its simplicity, but there are aspects of light that cannot be explained by this model. We can expand our model to include the wave nature of light. This new model is known as wave optics. Treating light as a wave allows us to understand and quantify phenomena such as interference and diffraction.

If we wish to expand our view of light even more, we can use the model of electromagnetic optics. This takes the wave nature of light into account, but also includes the fact that light is composed of electric-field and magnetic-field vectors oscillating in space. An electromagnetic approach to optics explains and quantifies phenomena such as light polarization.

Finally, a fully quantum model of light, quantum optics, can be used to explain all of the above phenomena, plus phenomena such as the photoelectric effect, lasers, semiconductors, and more. This is the most complicated and mathematically rigorous model of light, used to explain complex phenomena when the other models fall short. We will discuss some of these fascinating topics in a later chapter of this textbook, but will not go into a full mathematical treatment of quantum optics.

light model	**phenomena**
ray optics	reflection, refraction, dispersion
wave optics	interference, diffraction
electromagnetic optics	polarization
quantum optics	photoelectric effect, lasers, semiconductors

Huygens' principle

In 1690, a mathematical theory of light by Christiaan Huygens was published. Huygens' proposal of how light propagates is known as Huygens' principle. Huygens' principle states that every point on a wave can be treated as a source of a spherical wave that

itself propagates through time and space. Each of these individual spherical waves interacts with the others through interference. Each individual spherical wave amplitude is added to the amplitude of every other individual spherical wave, and that sum total represents the total wave. Note that Huygens' principle is a simplified model that we can use to understand some concepts, but it is not a complete theory of optics or electromagnetism.

For example, a plane wave is a light wave that travels through space with all of the wave crests and troughs aligned in straight lines. If we want to draw a plane wave, we can draw a line where the wave crests are, and note that they form parallel lines in space. A plane wave is a very simple type of wave to analyze mathematically, and is a nice approximation that physicists and engineers can use in optics calculations. Huygens' principle says that we can represent each point on each wave crest in the plane wave as a spherical wave. The interference of those spherical waves causes what we see as a plane wave. This is shown in Figure 29.1. Note that the figure shows only five representative points. However, Huygens' principle states that every point (of which there are an infinite number) acts as a source of a spherical wave whose constructive interference effect gives rise to the next wavefront.

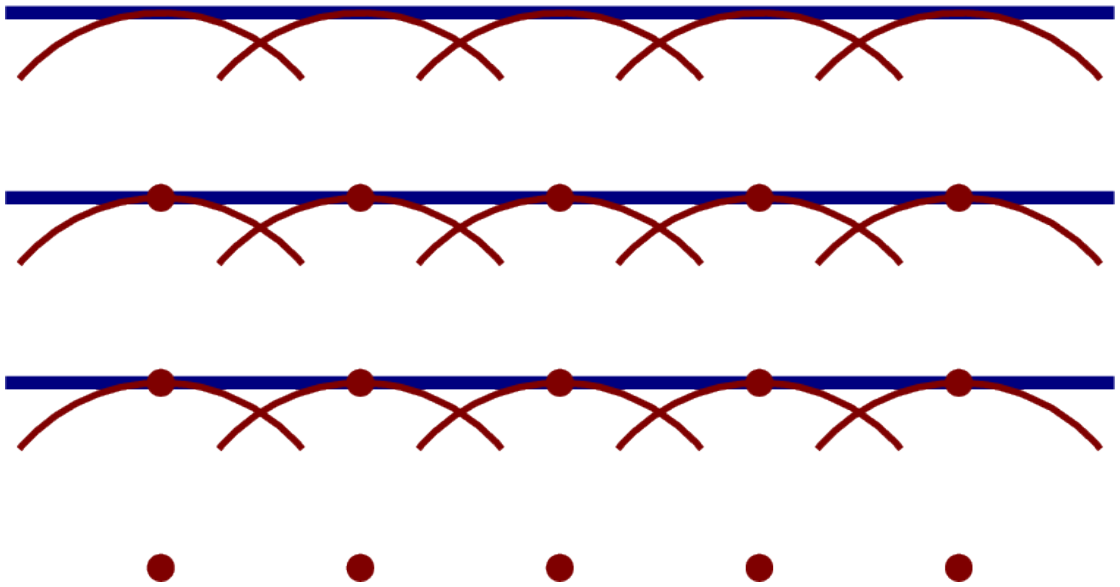


Figure 29.1 – Huygens' principle can be used to explain the shape of a plane wave. Every wave front is generated by constructive interference of spherical waves at every other wave front. Huygens–Fresnel principle, by Izaak Neutelings, is licensed under CC BY-SA 4.0.

Huygens' principle is particularly useful in describing what happens when light passes through a narrow opening. A plane wave traveling through a narrow opening will curve at the edges when it passes through. Because Huygens' principle describes light as a collection of spherical waves, it provides a description of that spherical shape at the edges. This is shown in Figure 29.2.

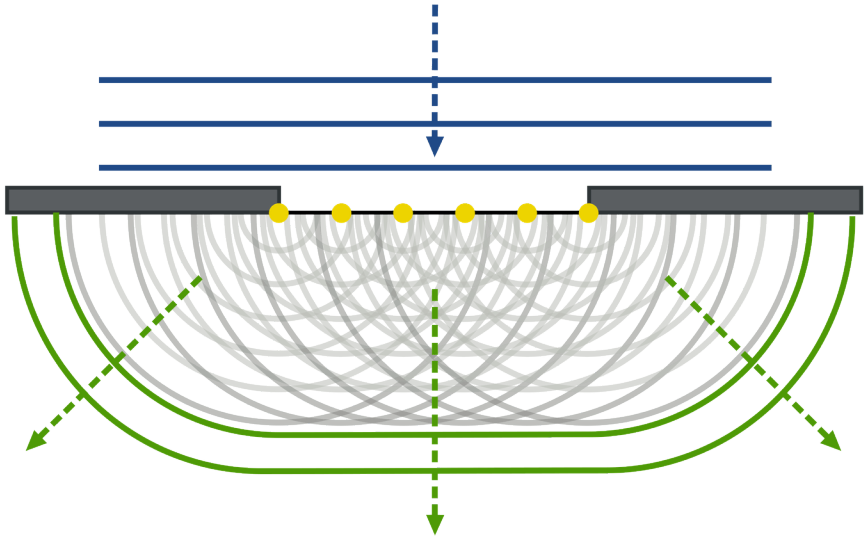


Figure 29.2 – Huygens' principle can be used to explain the bending of light as a plane wave travels through a narrow opening. Wave diffraction in the manner of Huygens and Fresnel, by Arne Nordmann, is licensed under CC BY-SA 3.0.

The amount of curving relates to the wavelength of light traveling through the opening as well as the width of the opening. The larger the width of the opening as compared to the wavelength of the wave, the less pronounced the curving of light will be. The smaller the width of the opening as compared to the wavelength of the wave, the more pronounced the curving of light will be.

Diffraction

Diffraction is a bending of light (or, in general, any wave) as it travels around a corner or through a narrow opening, as previously discussed using Huygens' principle as a model. Diffraction can occur with all waves, but only creates a noticeable effect when the width of the opening is on the same size scale as the wavelength of the wave.

Diffraction occurs not only with visible light, but all waves, including water waves. In addition, diffraction occurs with all types of electromagnetic waves, not just visible light. Radio waves, for example, will diffract around buildings, mountains, and other terrain.

The amount of bending that occurs in diffraction has to do with the wavelength of the light. The longer the wavelength, the more bending will occur, with all other things being equal. In this manner, diffraction can be used as a technique to split polychromatic light (light composed of multiple wavelengths of light) into its constituent components.

When light travels through a narrow opening and gets diffracted, that light can then be collected and measured. The intensity of this light creates a pattern, known as a diffraction pattern. When light diffracts through a single slit, the diffraction pattern appears as shown in Figure 29.3. There is a broad peak in intensity opposite the opening, with some smaller peaks appearing at regular intervals, eventually reducing in intensity to zero. The mathematics behind this pattern is complicated, but in general terms, the broadness of the peak and distance between peaks have to do with the width of the slit, the wavelength of the light, and the distance away from the opening where the light intensity is measured.

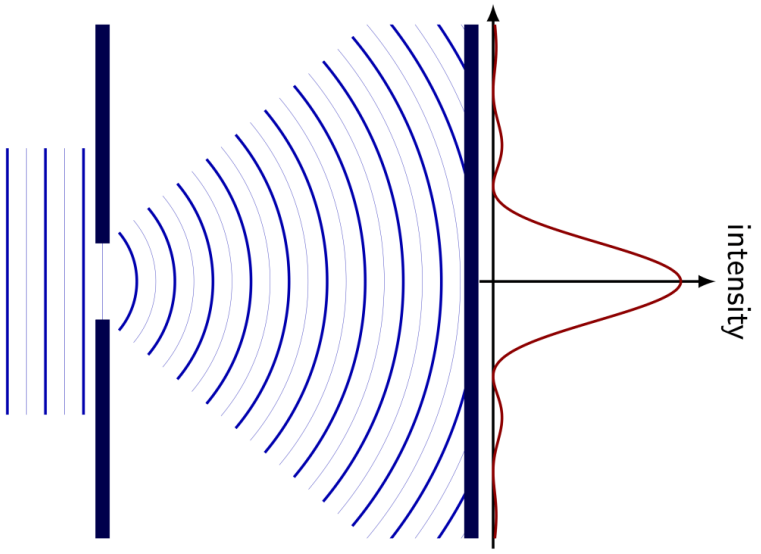


Figure 29.3 – Diffraction through a single slit creates a diffraction pattern that arises due to the curving of the wave as it travels through the opening. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

When a second slit is introduced close to the first, it is known as a double slit. The diffraction pattern now shows a more pronounced periodic spacing between areas of constructive interference when measured. This particular pattern arises due to a combination of diffraction from the slits and the interference between the two waves. This is shown in Figure 29.4.

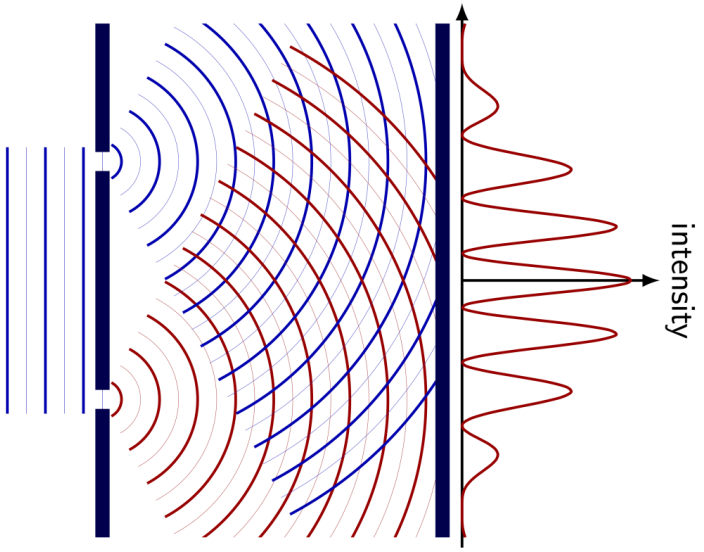


Figure 29.4 – Diffraction through a double slit creates a more pronounced periodic spacing between peaks in the diffraction pattern. This is due to a combination of interference between the two spherical waves and the diffraction from the slits. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

A diffraction grating is a component that has regularly spaced narrow openings for light to pass through. By spacing these openings periodically, light at different wavelengths will get diffracted through at different angles. This technique is particularly good at splitting polychromatic light into its constituent components because the periodic spacing between the openings will create regions of constructive and destructive interference for each wavelength.

The exact locations where constructive and destructive interference occurs depends on the wavelength of the light. It turns out that the longer the wavelength, the farther apart the regions of constructive and destructive interference. The graph in Figure 29.5 (Download this data [XLSX, 25 kB]) shows the intensity of a double-slit diffraction pattern for red, green, and blue light as it would appear on a screen a certain distance away from the openings. Red light (which has the longer wavelength) has constructive interference peaks spaced farther apart than green light (which has a shorter wavelength than red). Blue light (which

has a wavelength shorter than the two other colors) has the smallest spacing between the interference peaks.

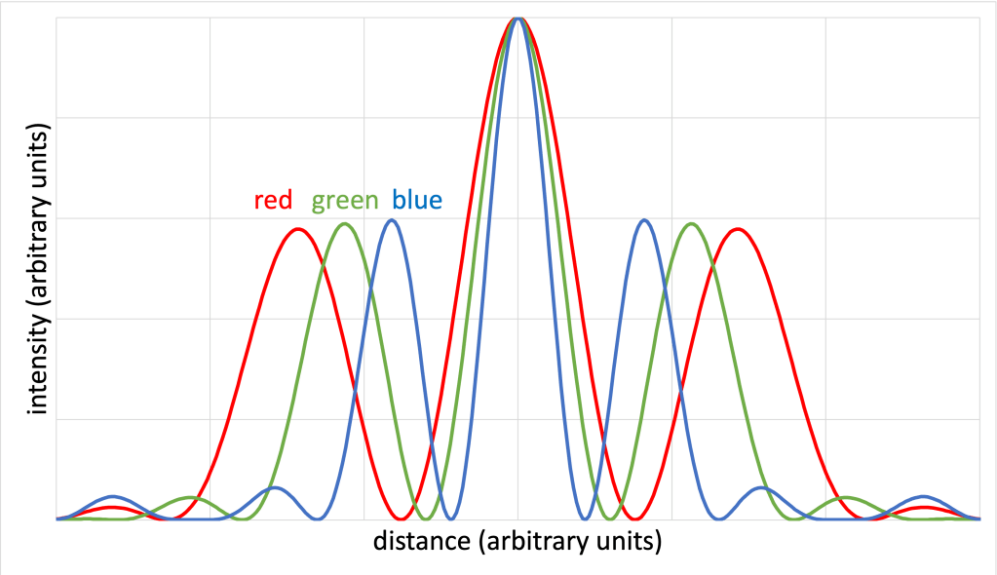


Figure 29.5 – The intensity of a diffraction pattern for light passing through a double slit changes based on wavelength. Red light has the largest spacing between constructive interference peaks, then green, and blue has the smallest spacing between constructive interference peaks. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

While both diffraction gratings and prisms are capable of splitting polychromatic light into its individual components, a diffraction grating typically spreads the components of light into a greater range of angles than a prism. This allows us to more accurately determine the intensity of each wavelength of light than is possible when using a prism. (In other words, because the light is spread over a greater number of angles, it is possible to get more detail in between each particular angle.)

In the video below, Dr. Pasquale demonstrates the effect of different color (wavelength) light as it is sent through a diffraction grating that has 1000 openings per millimeter. First, she shines a blue laser pointer through the diffraction grating. On a screen 1 meter away, we can observe each of the bright spots indicating points of constructive interference. Dr. Pasquale used a ruler to measure the distance between each spot, which is 46 cm.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

When she then shines a green laser pointer through the diffraction grating (keeping all other variables the same), the spots are now spread farther apart: 65 cm. This is because the larger wavelength of green light as compared to blue light, given the same spacing on the diffraction grating, will cause the light to spread more.

Repeating the experiment with a red pointer gives the largest distance between spots: 90 cm. Red has the longest wavelength of visible light, and will therefore spread the most.

In the video below, Dr. Pasquale repeats this demonstration using a different diffraction grating. This one contains 500 openings per millimeter, which means that each opening is wider than in the previous demonstration. This time, the blue laser pointer spreads apart 20 cm, the green laser pointer spreads apart 27 cm, and the red laser pointer spreads apart 36 cm. The same relationship between wavelength and diffraction angle is demonstrated. The larger the wavelength, the larger the diffraction angle (assuming the size of the diffraction grating is kept constant).

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

It's also possible to shine a white light source on a diffraction grating and see what happens. Because white light is composed of all other colors of light, we can observe each color getting split apart due to the different amounts of bending based on wavelength. When shining white light through a diffraction grating, the longer wavelengths bend the most, and the shorter wavelengths bend the least. This results in a rainbow pattern. A photograph of an incandescent light bulb viewed through a diffraction grating is shown in Figure 29.6. Note that the constructive interference occurs to the right of the bulb in the photograph, showing the composition of the white light. (In particular, an incandescent light bulb contains a continuous spectrum of colors from red to violet, which will be described in the next chapter of this textbook.)



Figure 29.6 – An incandescent light bulb seen through a diffraction grating shows the spreading out of all of the constituent colors that compose the white light. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Diffraction gratings have lots of uses in physics and engineering. For example, when Dr. Pasquale was a graduate student, she would use a diffraction grating to determine how well different devices she designed in a clean room would emit different wavelengths of light. By focusing white light on the device, and taking the emitted light and sending it through a diffraction grating, each wavelength of light would be split apart. A recording device known as a CCD was used to measure the intensity of each wavelength. She could then use that recorded data to see what wavelengths of light her device emitted.

Scientists also use diffraction gratings to determine the composition of stars, including our Sun. By looking at the different wavelengths of light that stars emit, and understanding the concept of light emission that we'll discuss later in this textbook, scientists can determine the elemental composition of stars.

One last thing about diffraction. Because light bends when it interacts with an opening, diffraction limits our ability to see really small things. This means that most imaging devices, such as cameras and microscopes, can only focus on things that are larger than about one half of the wavelength of light that's used. If the smallest wavelengths of visible light are 400 nm, that means the smallest things that can be seen using a light microscope or camera must necessarily be larger than 200 nm. We cannot use visible light waves to see things smaller than that!

Thin-film interference

When light waves reflect at the top and bottom interfaces of a thin layer of material, an interference effect will occur. This phenomenon is known as thin-film interference. Light will reflect off of both the top and bottom surfaces of the thin film. Under certain conditions of wavelength and thickness, it is possible for these two reflections of light to constructively interfere. This is shown qualitatively in Figure 29.7.

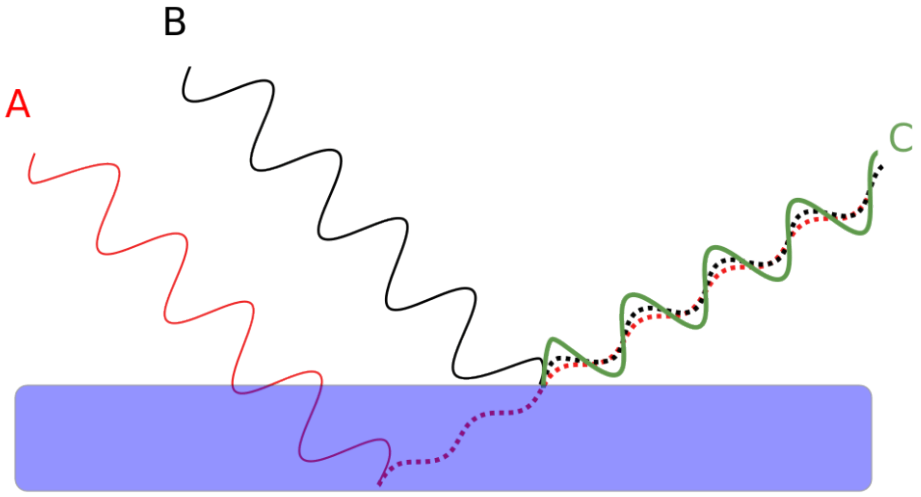


Figure 29.7 – Constructive interference occurs in a thin film when the light waves that reflect off of the top and bottom surfaces are in phase with each other after reflection. Constructive phase interaction, by Jhbdel, is licensed under CC BY-SA 3.0.

It is also possible for destructive interference to occur in thin films. This occurs when the top and bottom reflections of light are out of phase with each other. This is shown qualitatively in Figure 29.8.

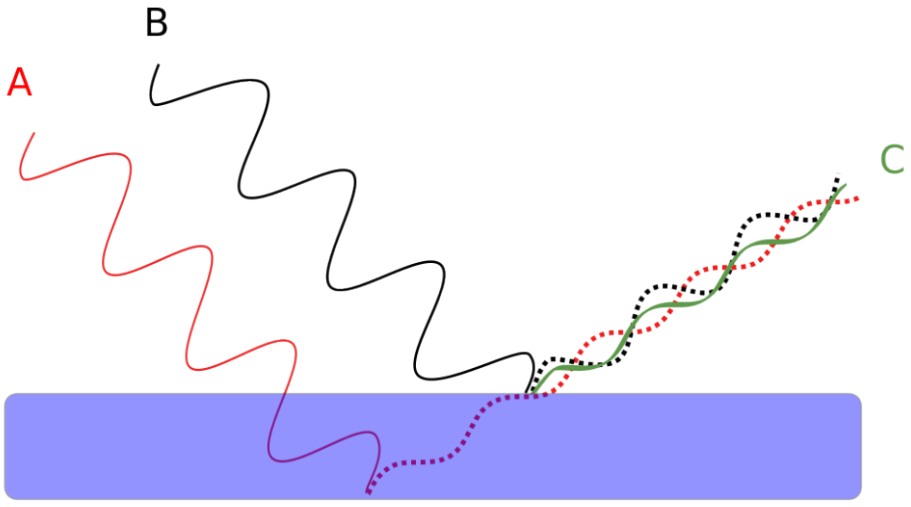


Figure 29.8 – Destructive interference occurs in a thin film when the light waves that reflect off of the top and bottom surfaces are out of phase with each other after reflection. Destructive phase interaction, by Jhbdel, is licensed under CC BY-SA 3.0.

The relationship between the thickness of the film and wavelength to create conditions of constructive and destructive interference is complicated. However, generally speaking, the wavelength of light that will be affected has to do with the thickness of the film, the index of refraction of the film, and the angle that the light hits the film.

This is demonstrated in Figure 29.9. Dr. Pasquale deposited glass with different thicknesses onto silicon, which is highly reflective to visible light. In both photographs, each sample appears to be a different color based on the thickness of the glass. The colors of the samples in the left and right images are slightly different due to the difference in angle that the light hits the samples (the only difference between the two images is the viewing angle). Both photographs contain a sample that is only silicon with no glass on it, in the upper left, for reference. In addition, a sample that is only glass is shown in the bottom right, also for reference. (As these two reference samples do not contain any films on them, they do not look different at different viewing angles.)

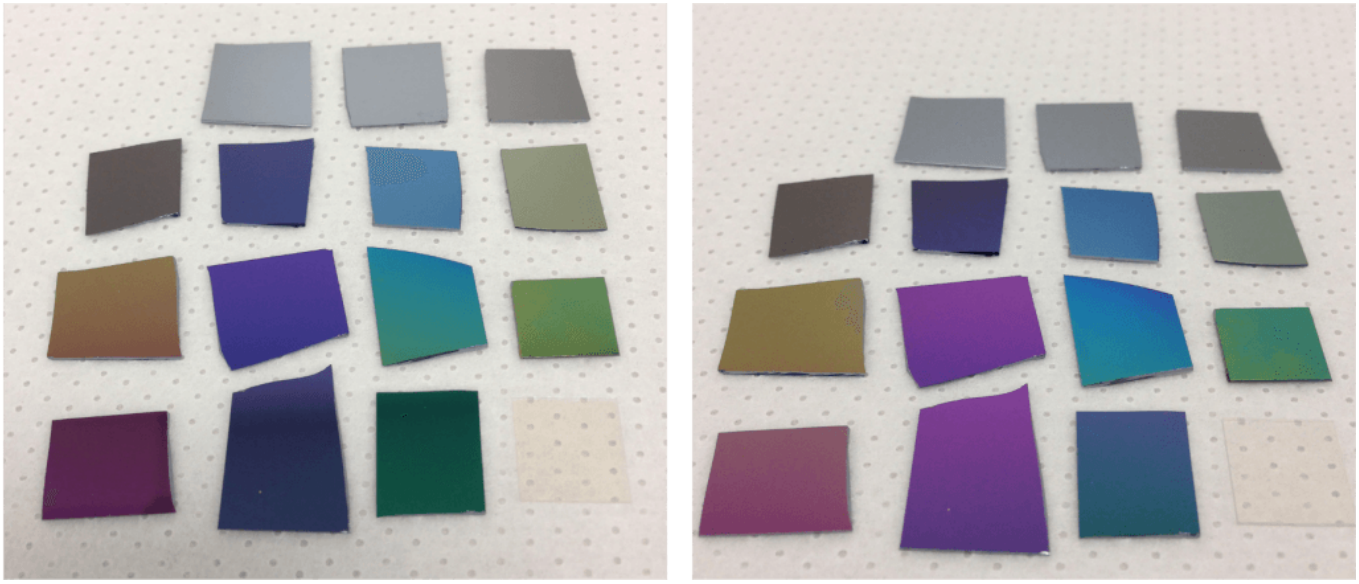


Figure 29.9 – Constructive interference based on the thickness of a thin film will cause films to appear to be a single color. These two photographs show the same samples seen at different viewing angles, showing how the angle affects the color that is generated in this interference process. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This thin-film interference effect is very useful in optical applications. By layering multiple thin films, it is possible to create an anti-reflection coating. These are used frequently on eyeglasses, cameras, microscopes, and other optical devices.

Single thin films can be used to determine optical flatness. Using a light source that emits a single wavelength of light (known as monochromatic, as it only has one wavelength of light present), the presence of interference fringes (patterns of dark and light spots) can be used to determine the flatness of a gap between two surfaces.

If the gap between two planar surfaces is completely flat, the interference fringes will appear as periodic parallel lines. This is demonstrated in the video below where Dr. Pasquale uses a monochromatic light source shining on two stacked pieces of glass. When she pushes on the upper piece of glass, she changes the thickness of air between the two pieces of glass, and the interference fringes change their pattern.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

If instead the gap between a flat object and a spherical surface is illuminated with monochromatic light, the effect is known as "Newton's rings."

If instead of monochromatic light, polychromatic or white light is used to illuminate a thin film, the effect can be very colorful. This occurs, for example, on oily water puddles (as shown in Figure 29.10). In a thin puddle of oily water, a thin layer of oil will sit on top of the water. As the thickness of the oil changes at the edges of the puddle, the changing thickness and light interference

result in a rainbow. This is also seen in soap bubbles and other similar phenomena.



Figure 29.10 – Thin-film interference causes a rainbow pattern to appear in a thin layer of oil on a puddle of water. Dieselrainbow, by John, is licensed under CC BY-SA 2.5.

Polarization

As mentioned, modeling light as a wave allows us to understand properties of light such as interference and diffraction. Expanding this model to encompass electromagnetism allows us to understand the property of polarization. Recall that light is composed of electric fields and magnetic fields oscillating at right angles to each other. (Note that the field components do not actually move up and down; instead, their intensities vary along dimensions that are at right angles to each other.) Light is a transverse wave, which means the direction of motion of the light wave is at right angles to these fluctuations in the intensity of the electric and magnetic fields.

Polarization is the property of transverse waves which allows them to have their electric field components oscillating in a single direction. Essentially, the electric fields that oscillate in a perpendicular direction to a device known as a polarizing filter can be blocked, allowing only a single orientation of waves to travel through space.

Note that polarization is not possible in longitudinal waves. To block the changes in intensity in a longitudinal wave, the entire wave motion would need to be blocked. The fact that light can be polarized means that electromagnetic waves are transverse.

An analogy that can be used to conceptualize this is shaking a long rope up and down and seeing if it passes through a picket fence. When the orientation of the rope is parallel to the picket fence, it passes through (Figure 28.11, top). However, when the orientation of the rope is perpendicular to the picket fence, it is completely blocked (Figure 28.11, bottom).

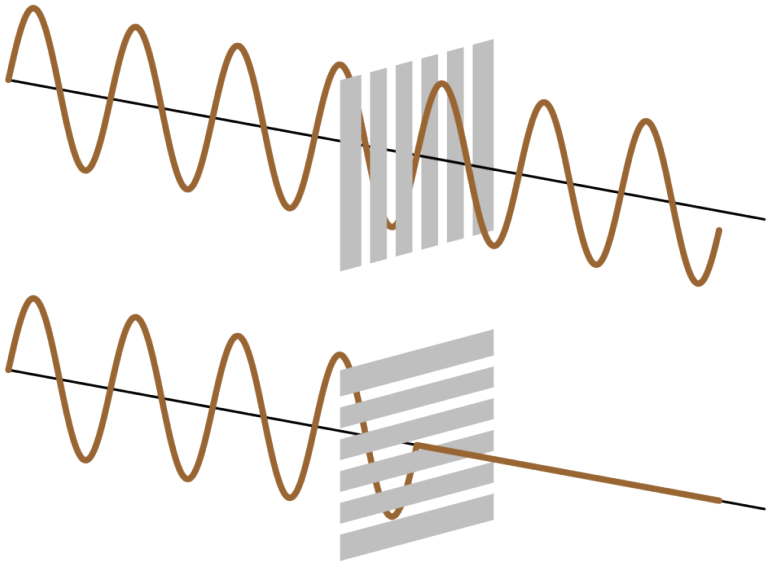


Figure 29.11 – An analogy for light polarization is considering how an oscillating rope would pass through a picket fence. When the slats in the fence are parallel to the oscillation direction, the rope will pass through (top). When the slats in the fence are perpendicular to the oscillation direction, the rope oscillation will be blocked. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Many light sources generate light waves with all orientations of electric field oscillations, known as unpolarized light. This light can be passed through a linear polarizing filter, which only allows light of a certain orientation to pass through (all perpendicular orientations are blocked). This polarized light is known as linearly polarized, as the electric field of the light oscillates in a linear direction. Figure 29.12 shows unpolarized light passing through a polarizer and becoming linearly polarized. A second polarizer takes that linearly polarized light and once again changes its polarization to a different linear orientation.

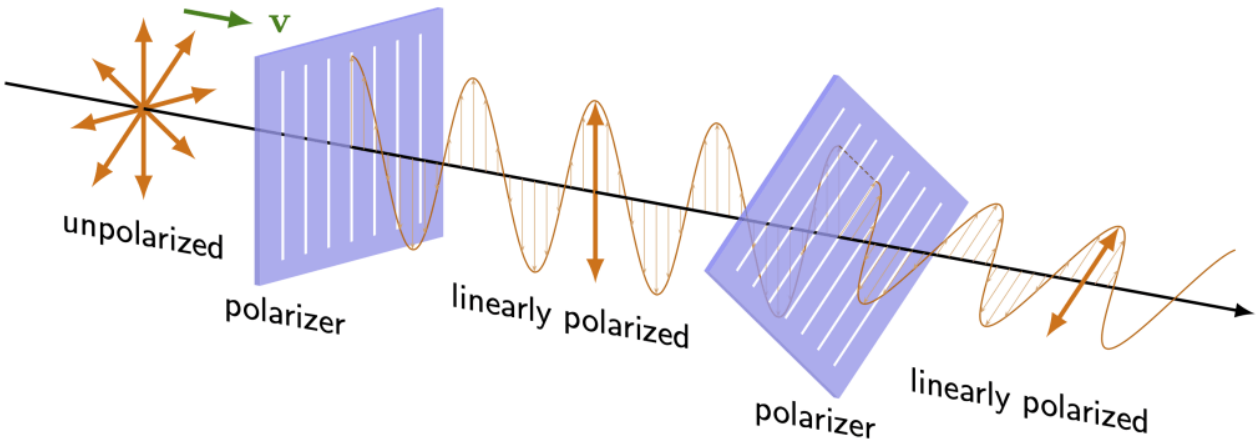


Figure 29.12 – Unpolarized light can become polarized. That light can then be sent through a second polarizing filter and become polarized in a different orientation. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

In general, the component of light that is parallel to a polarizer's orientation will be able to pass through. Qualitatively, this means that, for two polarizers at different angles from each other, intensity will be reduced every time light passes through a new polarizer. (This textbook will not consider this value quantitatively, as it requires trigonometry to do so.) Figure 29.13 demonstrates how light exhibiting a single orientation of polarization can be broken down into components that are parallel and perpendicular to a polarizer. The polarizer blocks the perpendicular component. (Earlier in this textbook when discussing work, we had to calculate the parallel component of a force. This is somewhat similar in that we are calculating the parallel component of a particular orientation of light.)

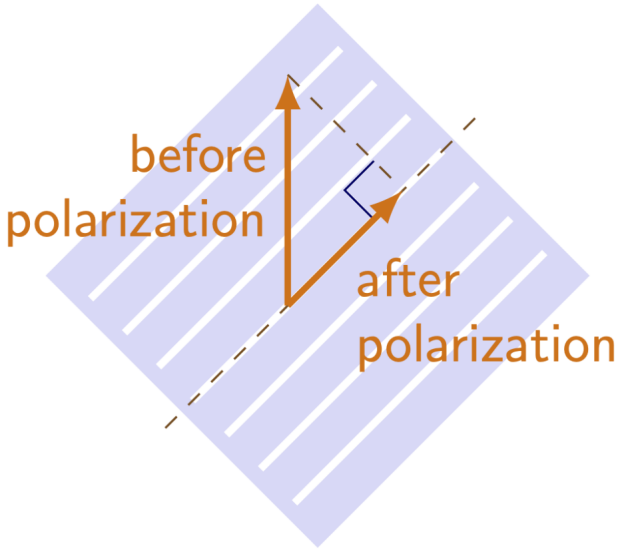


Figure 29.13 – The component of light that is oriented parallel to a polarizer will pass through the polarizer. This means that intensity will decrease every time light passes through a polarizer that's at a different angle. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

If a polarizer is completely perpendicular to the orientation of light, then there is no component that can pass through (as there will be no parallel component), and all light will be blocked. Any time two subsequent polarizers are placed at 90 degree angles to each other, light will be blocked. This is depicted in Figure 29.14.

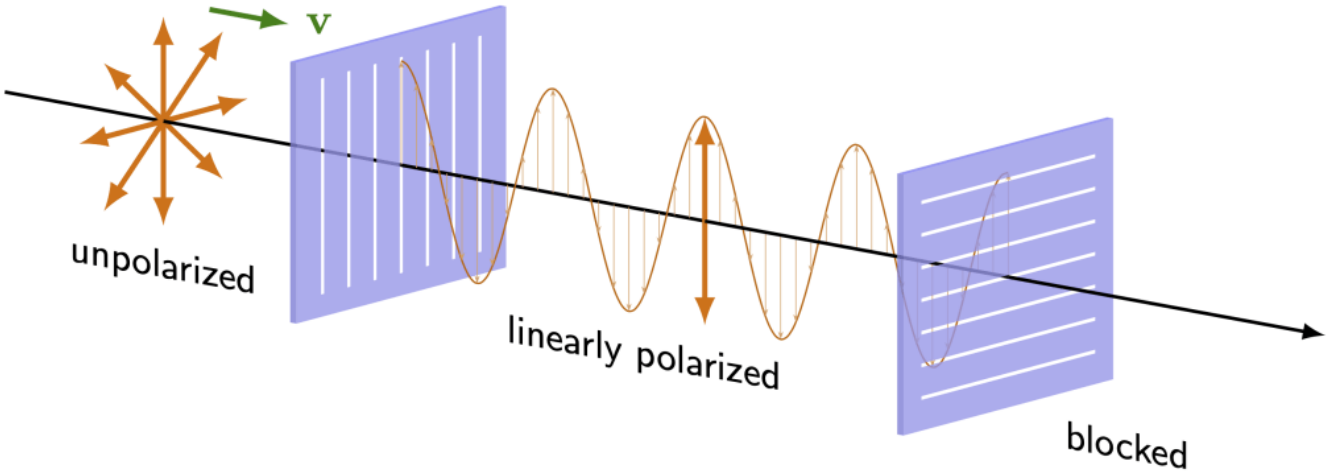


Figure 29.14 – Two polarizers at perpendicular orientations will block all light from passing through. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

However, if a third polarizer is placed between the two perpendicular polarizers, it is once again possible to have light pass through to the output. This is depicted in Figure 29.15. This is because there is always a parallel component to pass through each of the linear polarizers. The only way to completely block light with a polarizer is to have two subsequent polarizers at right angles to each other.

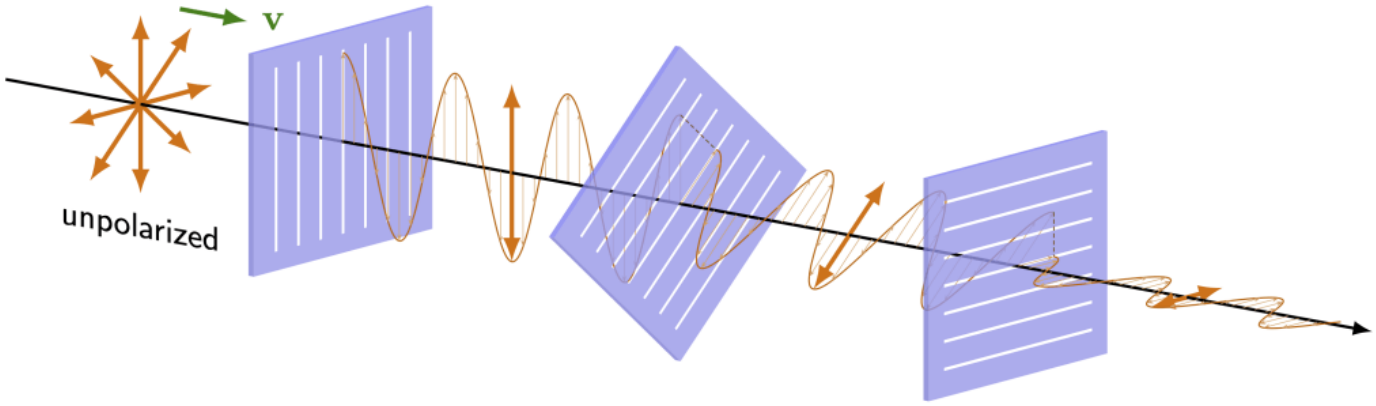


Figure 29.15 – A third polarizer placed at a non-perpendicular angle between two crossed polarizers will allow light to transmit through to the output. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

Dr. Pasquale demonstrates this effect in the video below. The demo starts with a single linear polarizing filter (filter A) between a light bulb and the video camera. A second filter (filter B) at a perpendicular orientation to filter A is then placed into the path, blocking the light from the light bulb. Dr. Pasquale rotates filter B through 180 degrees of rotation, demonstrating that the light will become brighter until filters A and B are parallel, and then become blocked again when filters A and B are perpendicular.

Then, Dr. Pasquale places a third polarizing filter (filter C) in between filters A and B. This filter is placed at a random angle (not perpendicular to either A or B) and light from the light bulb is no longer blocked. Dr. Pasquale rotates filter C through a broad range of angles, showing that it will block light when perpendicular to either filter A or filter B. When Dr. Pasquale removes filter C, the light is once again blocked.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Brewster's angle

When unpolarized light reflects off of a surface at a particular angle, known as Brewster's angle, the reflected light will be polarized 100% in a direction parallel to the reflecting surface. The light polarized perpendicularly does not reflect. At angles greater than the Brewster angle, the reflected light will still have both components, but the perpendicular component of polarization will always be less than the parallel component. This is shown in Figure 29.16. (In the figure, the ray of light containing dots and arrows depicts unpolarized light that contains some polarizations parallel to the screen/page, and some that are perpendicular to the screen/page. The reflected light, containing only dots, is polarized perpendicular to the screen/page. The refracted light contains a mix of both polarizations, but favors light parallel to the screen/page.)

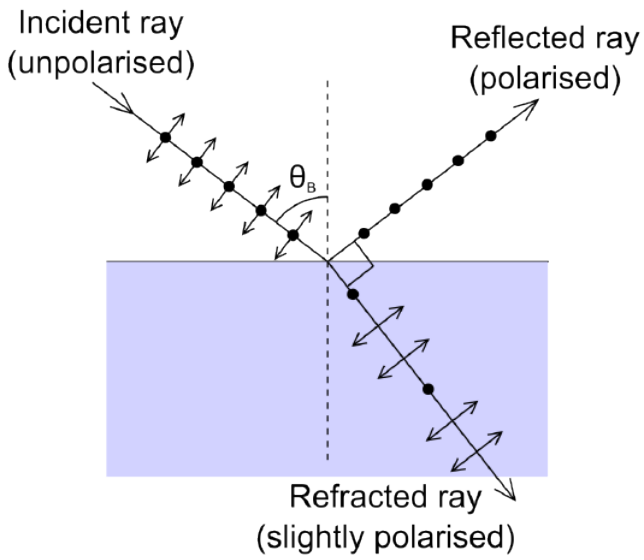


Figure 29.16 – Light reflecting at the Brewster angle will contain a single polarization. Brewsters-angle, by Pajs, is licensed under Public Domain.

This means that any polarizing filter that blocks that reflected light will partially eliminate glare from reflections that occur at angles greater than the Brewster angle. This is what polarized sunglasses are designed to do. Polarized sunglasses block the polarization of light that is not reduced by this effect, which can cut down on glare.

When wearing polarized sunglasses, it can be very obvious when you look at something that generates polarized light, such as LCD screens, as they will appear not to emit any light. By tilting your head 90 degrees while wearing polarized sunglasses, you can see polarization effects. Not only do LCD screens and monitors have polarization effects, but so do heat-treated windshields in cars.

Birefringence

Some materials have an index of refraction that changes based on the polarization of light. This means that light in one polarization will bend more than light at the cross polarization. These materials are known as birefringent.

A birefringent crystal placed over the word “Birefringence!” shows this property in Figure 28.17 by demonstrating the double image created by the crystal. One of the images is polarized in one orientation, and the other image is polarized in the opposite orientation.



Figure 29.17 – The word “birefringence” appears twice when viewed under a birefringent crystal. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Pasquale demonstrates the polarization of each component of the double image by rotating a polarizing filter over the piece of calcite. Depending on the orientation of the polarizing filter, one or the other of the images can be blocked.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Practice questions

Hands-on experiments

1. If you own polarized sunglasses, use them to find examples of polarized light. Either rotate your head through 90 degrees while wearing them, or rotate the sunglasses through 90 degrees in front of your eyes. Notice if anything changes.
2. Use a flashlight to bend light around a corner using diffraction. The instructions at this link will explain how.

This page titled Chapter 29: Light waves is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 30: Light emission

Summary

From the lights we use in our homes, to the lights illuminating our smartphone screens, to the Sun in the sky, light emission is a very big part of our lives. This chapter will explore different types of light-emission processes from incandescence to lasers.

Light emission

Recall from a previous chapter that electrons move about in shells around the nucleus of an atom. These shells, also called orbitals, are characterized by the amount of energy of each electron. Each atom has a unique set of energy levels that each electron can occupy. The lowest energy level is known as the ground state. An electron can be bumped into higher energy levels, in which case the electron is said to be in an excited state.

We can model these different energy levels using an energy-level diagram, similar to that shown in Figure 30.1. An electron can only reside in one of the energy levels of an atom at any given time, and cannot have an energy that is in between levels. The energy levels are therefore discrete: they only have certain allowed values for their energies.

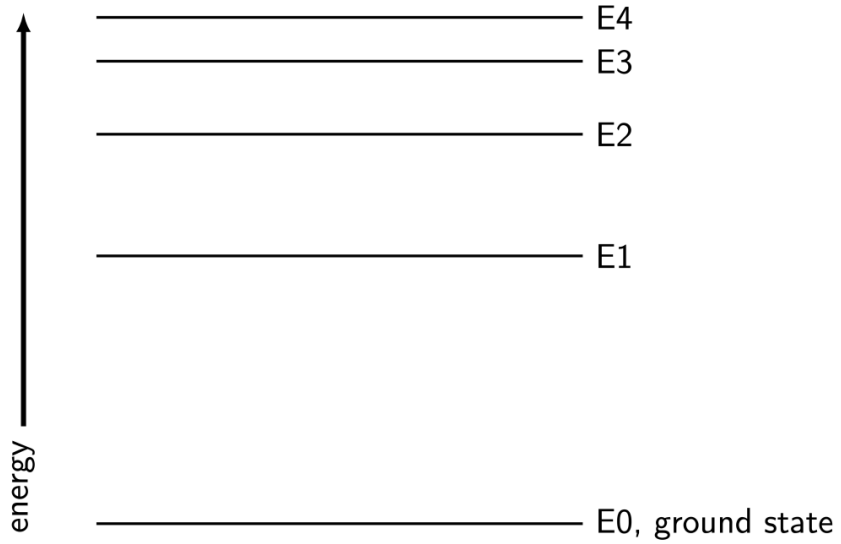


Figure 30.1 – An energy-level diagram depicts all of the different energy levels in which electrons can reside. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The lowest energy that an electron can occupy is called the ground state, often designated by E_0. The higher available energy levels are called excited states. For example, the energy level diagram in Figure 30.1 shows the first four excited states above the ground state.

Electrons can jump from one energy level to another. When an electron gains energy equal to the exact difference in energy between the (lower) starting level and the (higher) ending level, it can jump up one or more levels. Electrons can gain energy from light waves or from collisions with other electrons or subatomic particles. When electrons gain energy and move up one or more energy levels, this process is known as electron excitation.

Eventually, electrons that are in an excited state will relax back to the ground state. This process is called deexcitation. The deexcitation process can happen either as a result of a collision with another atom or with the emission of light. Either way, energy

needs to be conserved. In this chapter, the light emission process is considered. The energy of that light will be equal to the exact difference in energy between the (higher) starting level and the (lower) ending level. The frequency of the emitted light is proportional to the energy of the emitted light. The equation we can use to determine the energy difference between such levels is

$$E = hf,$$

where E is the energy of the light, f is the frequency, and h is the constant of proportionality known as Planck's constant. Planck's constant is equal to

$$h = 6.626 \times 10^{-34}\ \text{Js}.$$

Each element has a unique configuration of energy levels. This means that each element has a unique spectrum of light that can be emitted during the deexcitation process. It's possible for electrons to drop between any two levels, meaning that many different frequencies of light can be generated by this process. An atom with four energy levels (one ground state and three excited states) can represent up to six possible transitions, and six different frequencies of emitted light, as shown in Figure 30.2.

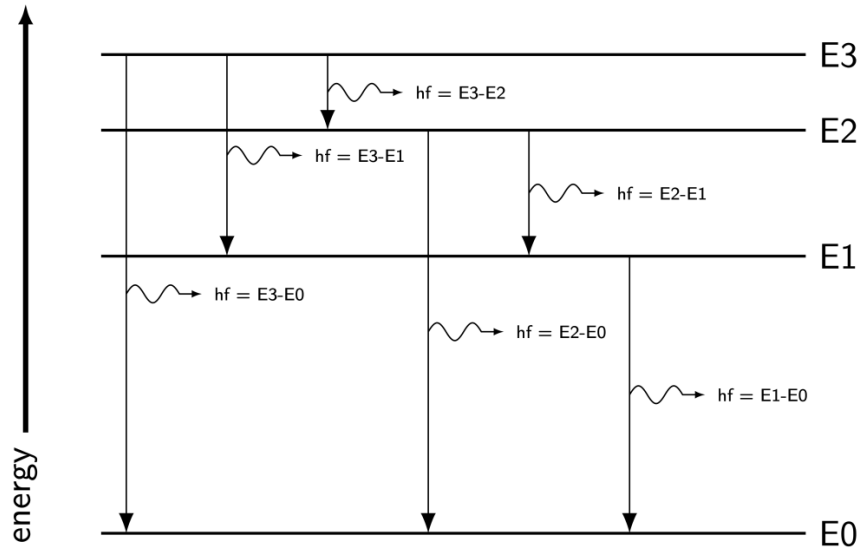


Figure 30.2 – A substance with four energy levels can support up to six different electron transitions, each with its own wavelength of light that will be emitted during that deexcitation process. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The collection of frequencies of light emitted by a substance via this deexcitation process is known as its emission spectrum. The measurement and analysis of these spectra is known as spectroscopy. Using spectroscopy, it's possible to determine the composition of gases (for example, stars, among other things) by examining the frequencies of light they emit. These frequencies can be compared to known emission spectra of different elements, such as hydrogen and helium, to determine the exact elements that are present in stars and other celestial bodies. This is particularly useful as it is not feasible to obtain samples from stars to bring back to Earth to measure using other methods!

A gas discharge lamp is a type of light source that generates light using electron excitation and deexcitation. A gas discharge lamp contains a tube filled with a gas. When a large voltage is applied to that gas, it is given enough energy to ionize the gas. Ionized gas contains many mobile electrons, which can create current. This current causes collisions between electrons and other ions of the gas, which excites electrons. When the electrons deexcite, light is emitted.

The spectra of gas discharge lamps can be measured with a diffraction grating. The spectra for a few different gas discharge lamps, and photographs of the light they produce, are shown in Figure 30.3. The distinct color of each lamp comes from the combination of all of the wavelengths of visible light that they emit.

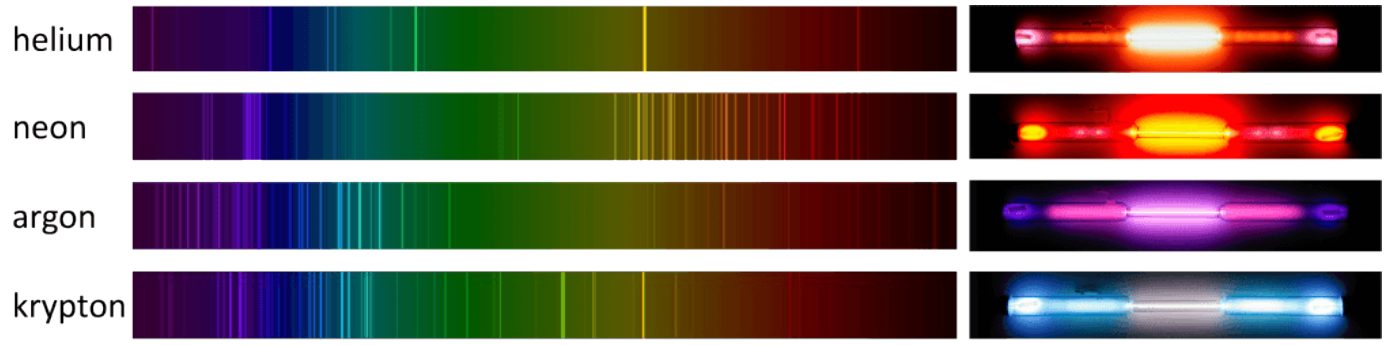


Figure 30.3 – Helium, neon, argon, and krypton (among other gases) can be used in gas discharge lamps. Each has its own unique spectrum of light emission, which creates light in its own unique color. Helium spectrum and neon spectrum, created by Teravolt, are licensed under Public Domain. Argon spectrum, created by Abilanin, is licensed under CC BY-SA 3.0. Krypton spectrum, created by Mrgoogfan, is licensed under Public Domain. Helium discharge, neon discharge, argone discharge, and krypton discharge, created by Alchemist-hp, are licensed under GFDL 1.2.

Before LED lighting was ubiquitous, many stores would use “neon lights” for their signs. These neon lights could contain neon, a noble gas that emits an reddish orange color, or other gases that emit different colors. “Neon” lights are shown in Figure 30.4. Note that each color is generated with a different gas.



Figure 30.4 – “Neon” lighting is not just composed of neon gas discharge. Each color “neon” light comes from a different element. Coloful light, by Violetbonmua, is licensed under CC BY-SA 3.0.

Sodium vapor lights were commonly used on highways, as they are a relatively efficient way to generate high intensity lighting, perfect for driving at night. Sodium vapor lamps can be identified by their unique yellow hue, as shown in Figure 30.5. Many of these lamps are being replaced by LEDs, which are capable of generating a more pleasant white color.



Figure 30.5 – Sodium vapor lamps, which have a distinct yellow color, were very commonly used on roadsides and highways. Sodium-vapor street light, by Famartin, is licensed under CC BY-SA 4.0.

X-rays

In 1895, a new type of electromagnetic wave was discovered when a beam of electrons emitted from a cathode-ray tube caused a nearby piece of paper to fluoresce (a process that will be described later in this chapter). These electromagnetic waves were called “x-rays” due to the fact that their characteristics were unknown at the time. That name has persisted, and we continue to call this type of electromagnetic radiation x-rays even today. While not much was known about the properties of x-rays at the time of their initial discovery, it was known that the rays did not contain any electric charge, due to the fact that they did not deflect in the presence of electric or magnetic fields.

It is now known that x-rays are a type of high-energy electromagnetic radiation produced when highly energetic electrons are quickly slowed when they strike the positive electrode (called the anode) in a cathode-ray tube. The slowing of the electrons produces x-rays consisting of a continuum of energies over the frequency range of x-rays. Maxwell’s equations (the equations used to determine the properties of electromagnetic waves) state that any time a charged particle accelerates (speeds up, slows down, or changes direction), light will be emitted. A broad range of energies is generated due to the collisions between the electrons and the anode. Some electrons may slow down with a single collision (in which a lot of energy is released at once), and others may slow down with multiple collisions.

In addition to the continuous spectrum generated by collisions with the anode, these collisions can knock the electrons in the anode’s atoms into excited states. As those electrons deexcite, they produce x-rays as well. Because the deexcitation process occurs at only specific wavelengths for different materials, these show up on an emission spectrum as very discrete frequencies or energies. A typical x-ray spectrum is shown in Figure. 30.6. The broad background represents the braking radiation (x-rays generated by collisions with the anode), and the discrete spikes represent the emission lines from electron deexcitation (x-rays generated by the deexcitation of excited anode electrons).

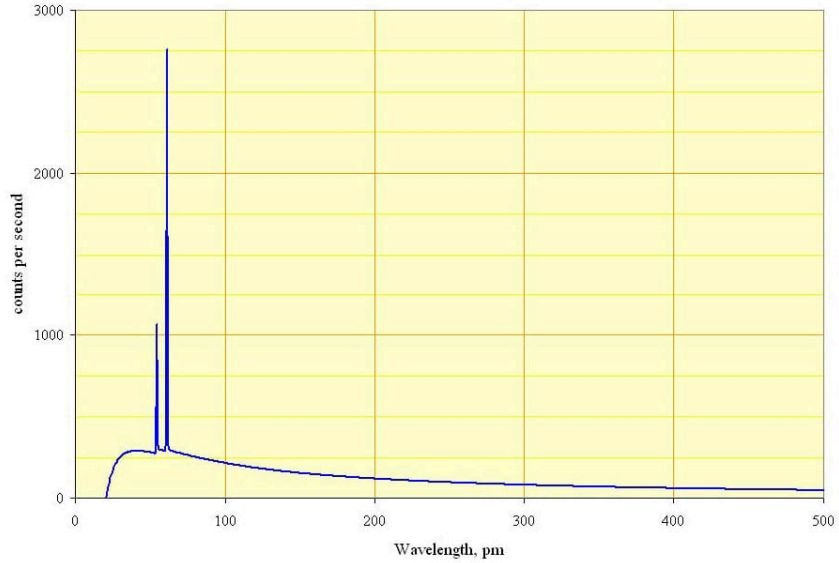


Figure 30.6 – The wavelengths of light generated in a cathode-ray tube in the x-ray region of the electromagnetic spectrum consists of braking radiation (continuous background) as well as emission from electron deexcitation (discrete spikes). Tube spectrum, by LinguisticDemographer, is licensed under Public Domain.

Light absorption

While an emission spectrum showcases the light emitted by a material, an absorption spectrum shows the wavelengths of light that are absorbed by a material. If white light is shined upon a sample, the sample can absorb particular frequencies (given by its energy-level diagram) causing excitations in the material. Those frequencies will be subtracted out from the white light, and appear as black lines in the absorption spectrum, as shown in Figure 30.7.



Figure 30.7 – An absorption spectrum appears as black lines at frequencies where light causes excitation in electrons. Spectral lines absorption, by Stkl, is licensed under Public Domain.

Incandescence

Incandescence is a form of light emission previously discussed in this textbook in the context of heat transfer. All objects emit electromagnetic radiation with a frequency that corresponds to the temperature of that object. Relatively cool objects such as humans, plants, and animals emit light in the infrared portion of the electromagnetic spectrum. Hotter objects will emit light with higher frequencies. At a high enough temperature, objects will emit visible light.

Unlike an emission spectrum, the incandescence spectra of objects is a continuous spread of frequencies (and therefore energies) as opposed to discrete frequencies as seen in the deexcitation of gas atoms. Furthermore, the incandescent spectrum is the same at a given temperature regardless of the material. The incandescent spectrum for different temperatures is shown in Figure 30.8. Notice that cooler objects emit light at wavelengths longer than visible. Hotter objects overlap with the visible portion of the electromagnetic spectrum, which will appear white if enough colors are emitted by the object. The discovery of the correct mathematical expression for the intensity of the incandescent spectrum as a function of frequency was made by Max Planck in 1900.

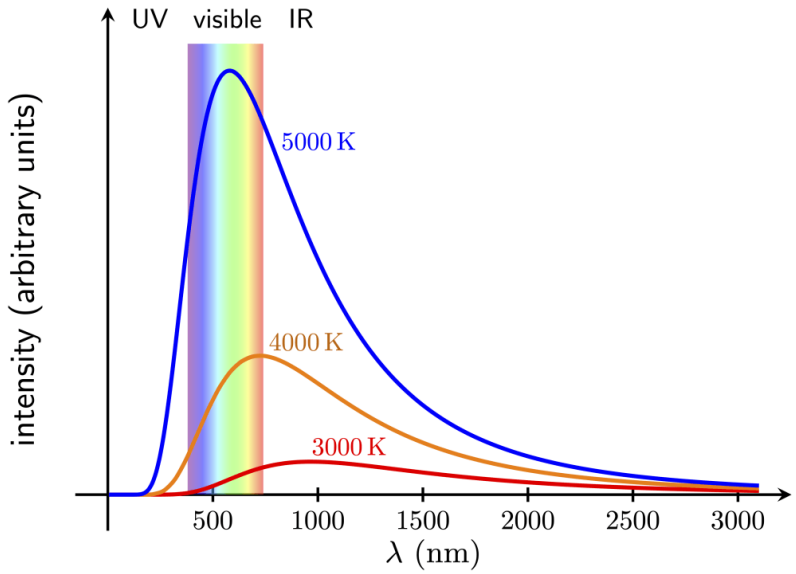


Figure 30.8 – Cool objects will emit light in the infrared or red region of the electromagnetic spectrum. As objects become hotter, their incandescence spectra will start to overlap more with visible light, eventually appearing white or blue. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

In this manner, we can use the color of an incandescent object to determine its temperature. An incandescent object that appears red or orange is cooler than an incandescent object that appears white or blue.

Incandescent light bulbs use the property of incandescence to generate white light. A very thin filament is heated by passing electric current through it. This heat causes the filament to glow with multiple overlapping wavelengths of visible light (shown in Figure 30.9). The mixture of all frequencies of visible light creates white light.



Figure 30.9 – An incandescent bulb emits light through heating a filament, which glows white when it is hot enough. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Using a diffraction grating, we can see a continuous rainbow spectrum emanating from an incandescent light bulb, shown in Figure 30.10.

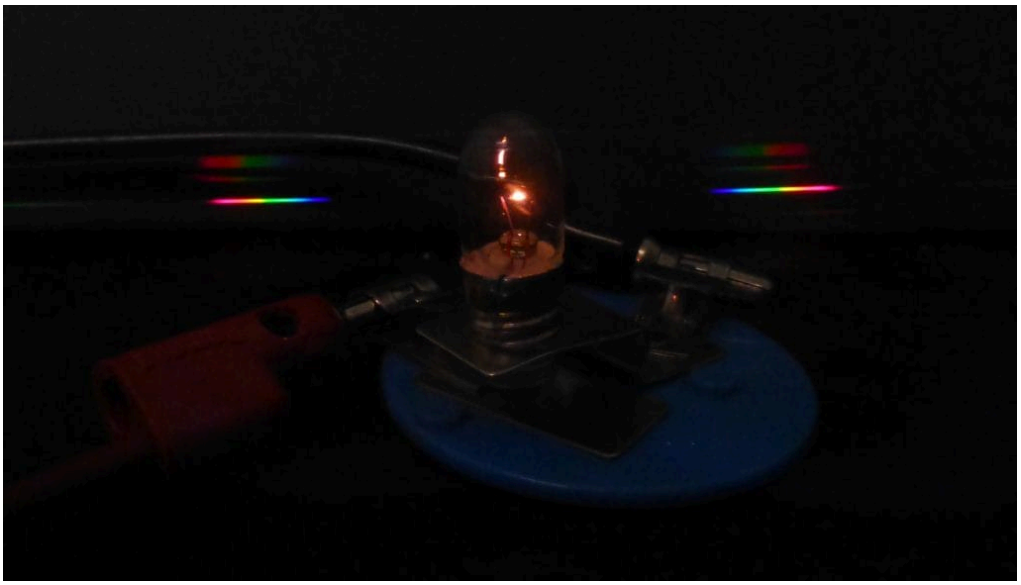


Figure 30.10 – When viewed through a diffraction grating, an incandescent bulb shows a continuous spectrum (rainbow) of light from red to blue. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Today, we use incandescent bulbs less and less in our homes and work environments, as the process of incandescence to generate light is very inefficient. (In fact, incandescent light bulbs have been or are being phased out in many countries around the world.) The power consumed by incandescent light bulbs is quite high compared to LED and fluorescent lights. Most of the electrical energy used to light an incandescent bulb is wasted as heat.

In the video below, Dr. Pasquale uses a hand-crank generator to light up different bulbs. First, she cranks the generator to light up an incandescent bulb. This takes rather a lot of energy to do, and the bulb lights dimly. Then, she turns the crank to light the LED bulb. This bulb lights right away, as it is very easy to illuminate. An LED bulb requires much less energy to function. However, note that the LED only illuminates every half cycle when powered by an AC generator. This problem can be fixed in home lighting by converting AC to direct current (DC), as previously discussed in this textbook.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Other examples of incandescence in our daily lives are seen in electric stoves and toasters. Both of these objects contain resistive heating elements. Electric current is sent through the heating elements, which then warm up due to their resistance. At cool temperatures, they will glow red, as seen in Figure 30.11 for a hot toaster. As they warm up, they will glow orange, or possibly even yellow if very hot. Note that when the electric stove or toaster is turned off, both still emit light via incandescence! However, we can no longer see it. As the objects are likely at room temperature, we can expect that light to be in the infrared region of the electromagnetic spectrum, and therefore not visible.



Figure 30.11 – The resistive heating elements in an electric toaster incandesce and emit visible light when hot. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Fluorescence

Fluorescence is a process of light emission where an object absorbs light at a high frequency, causing electron excitation. During the subsequent deexcitation process, light is then emitted at a lower frequency. During this process, a minor amount of heat is generated, which is what causes that change in frequency between the absorbed and emitted light. Energy is conserved. An energy-level diagram depicting this process is shown in Figure 30.12.

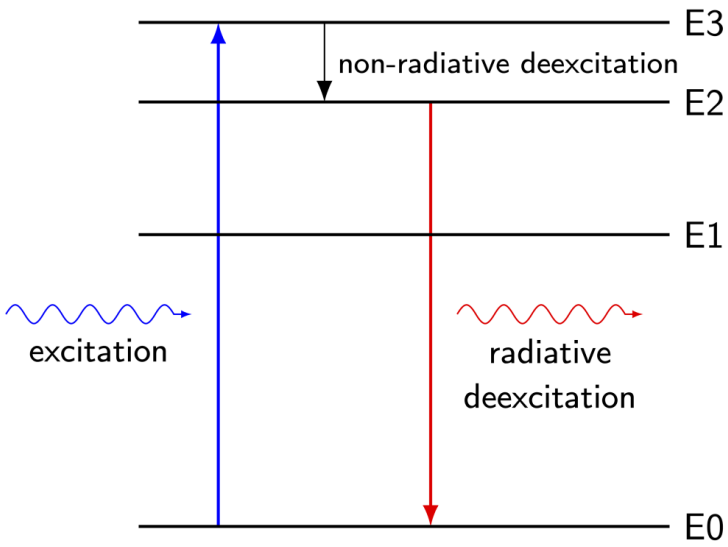


Figure 30.12 – An energy-level diagram depicting the fluorescence process shows the absorption of high-energy light, followed by a non-radiative deexcitation process and the emission of low-energy light. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Fluorescence has many applications in the medical field. A fluorescent dye can be attached to certain types of cells or other biological materials. When excited with the high frequency light, the fluorescent dye will emit light at a different frequency. This is particularly useful because the high frequency light is usually bright enough to overwhelm any other light. By filtering that high frequency out, only the fluorescent light can be collected, allowing scientists and doctors to see only the parts of samples that are attached to the fluorescent dye. They emit at a different frequency and would not be blocked by that filtering process. In this manner, particular types of cells can be imaged (endothelial cells imaged by fluorescence are shown in Figure 30.13), biological

processes can be viewed, and the presence or volume of cancer or bacteria can be measured.

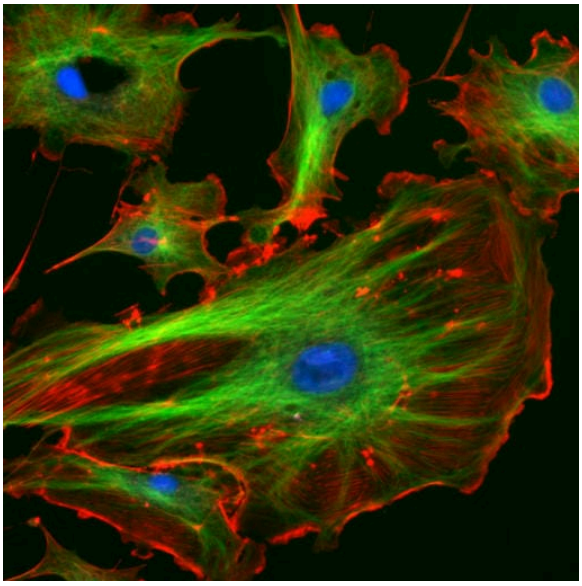


Figure 30.13 – A fluorescence micrograph of endothelial cells stained with three different fluorescent agents as viewed under UV light. FluorescentCells, by National Institutes of Health, is licensed under Public Domain.

Fluorescence is also present in nature. Some types of animals and fish exhibit fluorescence. Even plants contain a fluorescent molecule: chlorophyll. Some minerals in rocks are fluorescent as well. We can see these colorful fluorescent emissions when shining a UV light on the minerals, as demonstrated in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Fluorescent components are also used in everyday items such as highlighter pens (to create the bright “neon” colors) and laundry detergents (to act as an optical brightener).

Phosphorescence

Phosphorescence is a light emission phenomenon akin to fluorescence. Similarly to fluorescence, a phosphorescent material will become excited by high frequency light and re-emit lower frequency light. However, in phosphorescence, this process of re-emission is extended for a time period even after the exciting source is removed. Sometimes it is a few seconds, sometimes it can be several minutes or hours. In other words, phosphorescence is essentially the same thing as fluorescence, except that the electrons in the excited state remain there for an extended period of time rather than dropping back down immediately after the source of the excitation is removed.

Many “glow-in-the-dark” paints are phosphorescent. When illuminated during the day with sunlight or room lighting, the electrons in the material gain energy. At night, after the illumination is removed, the glow-in-the-dark paint will slowly re-emit the lower frequency light over the span of several minutes or even hours. This is demonstrated in the video below. Dr. Pasquale has a light shining on a glow-in-the-dark star. After turning off the light, the star continues to glow.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Chemicals known as phosphors may emit light using the process of fluorescence, and some may emit light using the process of phosphorescence. Phosphor coatings are very common when high energy light is used to generate white light. By mixing multiple phosphorescent or fluorescent chemicals together, a mixture of red, green, and blue light can be generated to create white light for use in home lighting.

Fluorescent lights

Fluorescent lights use the properties of light emission and phosphorescence to generate light. A mercury vapor inside of a glass tube is excited into a plasma by the application of high voltage. Mercury plasma emits light in the UV region of the electromagnetic spectrum. On its own, this UV light would be useful for sterilization or other medical uses, but would not be a good light source for our homes due to the presence of ionizing radiation.

The surface of a fluorescent light tube is coated with a phosphor. This phosphor absorbs the high-energy UV light and generates other colors of light. The exact chemicals used in the phosphor coating depend on the color temperature of the light. In what may be considered confusing language from a physics perspective, blue colors are known as “cool” and yellow colors are known as “warm” in color theory. Often, this nomenclature will be used on boxes of indoor lighting. However, “cool” lights from the blue part of the visible spectrum are actually at a higher temperature (typically 5000 K) than “warm” lights from the yellow or red part of the visible spectrum (typically 3000 K). When you go to the hardware store to purchase fluorescent bulbs, you will see different color temperatures available, depending on the lighting effect that you prefer. An image of two light fixtures, one fitted with “warm” (3000 K) and one fitted with “cool” (5000 K) bulbs, is shown in Figure 30.14.



Figure 30.14 – A “cool” bulb appears bluish-white, whereas a “warm” bulb appears yellowish-white. The color temperature comes from the phosphor used in the bulb coating. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Because the mercury vapor emits light at discrete frequencies, and the phosphor coating also emits light at certain frequencies (although the phosphor emission has a broader spectrum), when using a diffraction grating to look at the spectrum of a fluorescent bulb, the combination of both of these effects can be seen. Each of the lines of the mercury vapor emission spectra can be seen, with some blurring between those lines due to the broad phosphor emission spectrum. This is shown in Figure 30.15.

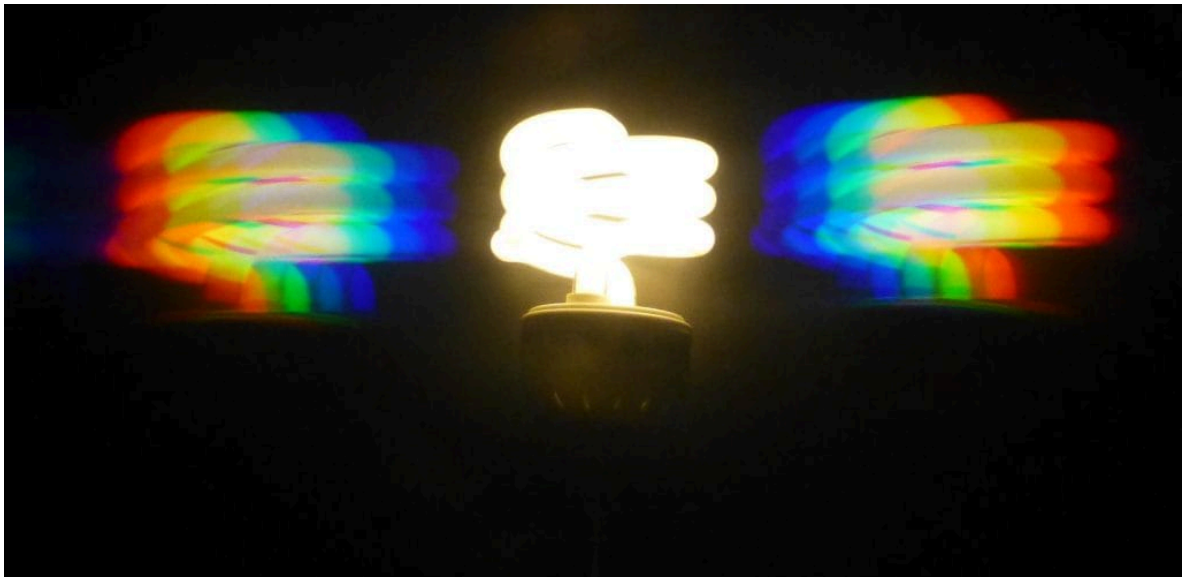


Figure 30.15 – A diffraction grating shows the composition of a fluuorescent lamp. There are discrete colors blurred together by the phosphor coating. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Light-emitting diodes

Light-emitting diodes (LEDs) use two layers of semiconducting material stacked together to create current when a voltage is applied in a certain direction. A diode, discussed previously in this textbook, acts like a one-way valve that only allows current to flow in one direction, if voltage is applied in the correct orientation. An LED is a special type of diode that emits light when current flows.

Every light-emitting diode will generate light at a particular frequency, given by the exact semiconductor that's used to create it. Because of this, light-emitting diodes are monochromatic. However, engineers and scientists have created LEDs that can generate frequencies from UV to infrared. A selection of several different colored LEDs are shown in Figure 30.16. These are small LEDs used for electronics and hobbyist projects. (Perhaps you have seen LEDs similar to these that are used as indicator lights in appliances or other electronics.)

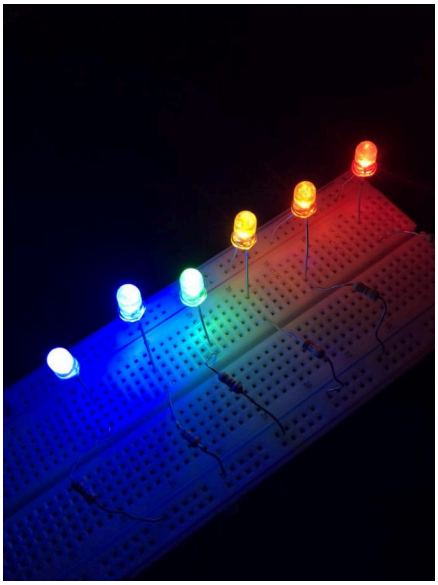


Figure 30.16 – Small LEDs used for electronics projects come in many different colors. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Monochromatic light may be appropriate for indicator lights or for use in specific electronic projects, but how can LEDs be used

to generate white light, which is required for home lighting? In fact, LEDs can be used to generate white light using two different methods.

The first method for generating white light from an LED is to package three separate LEDs together: one red, one green, and one blue. This is known as an RGB LED. A close-up photograph of an RGB LED, showing the separate red, green, and blue diodes, is shown in Figure 30.17.

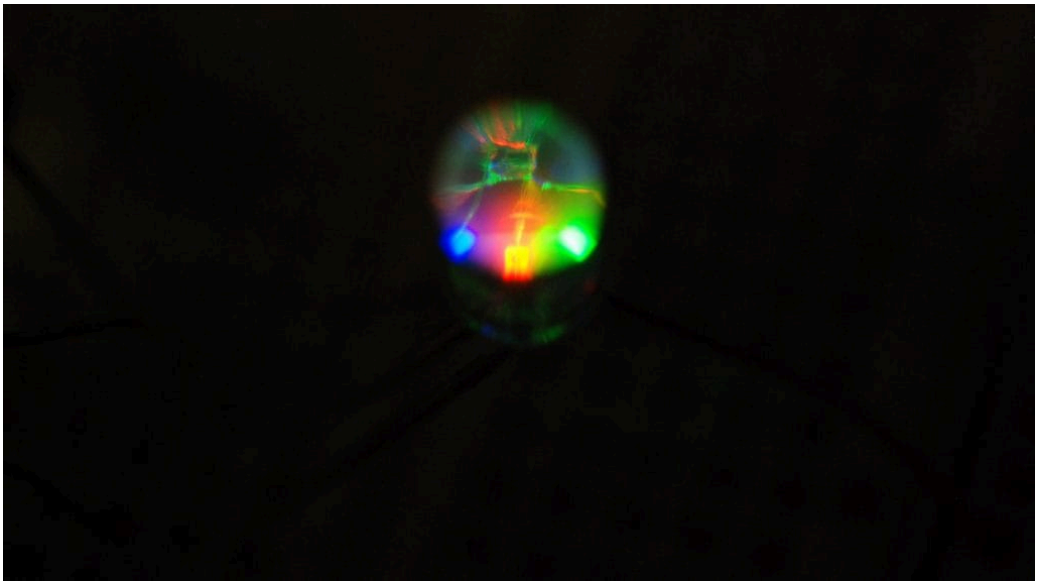


Figure 30.17 – A close-up photograph of an RGB LED shows separate red, green, and blue light-emitting diodes inside the package. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

By combining red, green, and blue light together using the color addition process described earlier in this textbook, white light can be generated. RGB LEDs are also used in tunable light bulbs (whose color can be changed by the user). The addition of red, green, and blue LEDs to generate white light is demonstrated in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The second method for generating white light is similar to that used in fluorescent bulbs. An ultraviolet LED is packaged in a piece of epoxy that is coated with a phosphor. Depending on the exact phosphorescent chemicals used, different color temperatures of white LED bulbs are achieved.

A household light fixture containing integrated LED lights is shown in Figure 30.18. The light fixture contains electronics in the center to convert the high voltage AC signal to a lower value DC voltage appropriate for lighting the LEDs (which is also adjustable for use with a household dimmer switch). The LEDs themselves are arrayed in a circular pattern around the edge of the fixture. Each individual LED is rather small, so many of them are used to generate bright light for our homes.



Figure 30.18 – A household light fixture containing integrated LEDs has electronics for converting AC to DC as well as multiple individual light-emitting diodes. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

An LED light bulb contains similar circuitry inside the bulb casing. This is why LED light bulbs tend to be heavier than similarly sized incandescent light bulbs. An LED light bulb with the casing broken to display the interior components is shown in Figure 30.19.



Figure 30.19 – An LED light bulb contains electronics inside the bulb casing. LED-E27-Light-Bulb-1134, by David R. Tribble, is licensed under CC BY-SA 4.0 .

Photovoltaic panels, or solar panels, function by the reverse process. Light that's absorbed by a semiconducting material will generate electric current. This is an excellent method of creating energy that does not require fossil fuels, and uses a process totally different from the generators discussed in a previous chapter.

Lasers

The term laser is an acronym for light amplification by stimulated emission of radiation. A laser contains a light source to start this light amplification process. (A laser pointer will likely contain a light-emitting diode as a light source.) This light source shines into an amplifying medium. This can be a gas or even a solid material that amplifies light by means of stimulated emission. Stimulated emission occurs when an electric field is used to provoke deexcitation of many electrons, all of which emit photons that are

identical to each other. While this is a very complicated process, in general it means that the material is capable of generating more light from the initial light source. (There is an energy source used; a laser does not violate conservation of energy!)

In addition to the amplifying material, a laser contains two mirrors on each end to cause the light waves to travel back and forth throughout the material. This does two things: first, it causes the light to spend more time in the amplifying material, creating more light by stimulated emission. Second, it causes the light waves to interfere, creating light waves that are all in the same phase and have the same wavelength using the process of constructive interference. Because the output light all has the same wavelength, a laser is a monochromatic light source. Different types of lasers will emit at different frequencies. The output from six different lasers is shown in Figure 30.20.

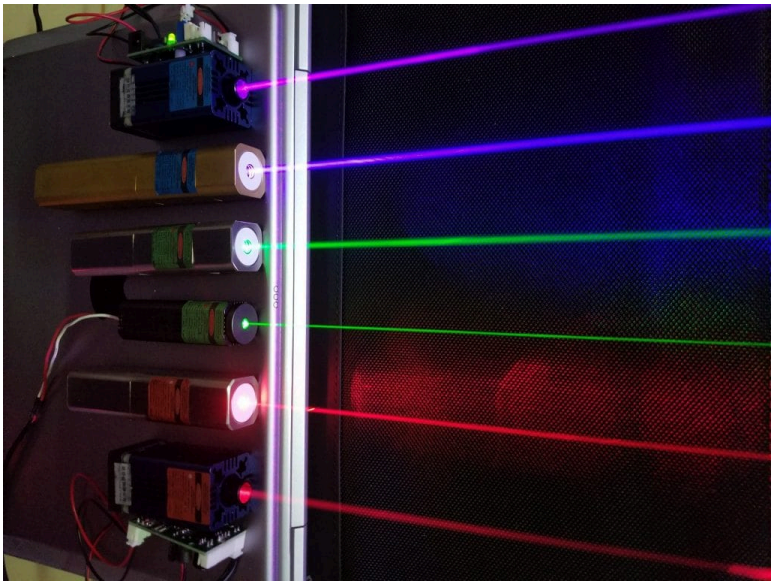


Figure 30.20 – A laser is capable of generating very intense monochromatic light. Each type of laser can produce light at different frequencies. Lasers, by 彭嘉傑, is licensed under CC BY 2.5.

The output of a laser contains light waves that are very high-intensity and also completely identical. (Because of the high intensity output, it is extremely important to exercise caution when using lasers, as they can cause vision loss or even blindness if used incorrectly.) This type of light is known as coherent light. Coherent light means that all of the waves have the same frequency and phase.

The properties of laser light make them extremely useful in many scientific, medical, and industrial applications. In communications, lasers are used to send optical data along fiber optic cables (using the property of total internal reflection to remain inside the cable). They are also used to read or write from optical discs such as CDs and DVDs. Industrial uses of the high intensity output generated from lasers include laser welding and cutting. The straight lines projected by lasers have use in construction and surveying. Medical uses of lasers include cancer treatment and laser eye surgery. In addition, scientists and engineers frequently use lasers in scientific and medical research.

Further reading

- December 1958: Invention of the Laser – This article from the APSNews “This Month in Physics History” series covers the invention of the laser.

Practice questions

Conceptual comprehension

1. If you are able, go outside on a clear day and look at the sky (preferably with a telescope). If you do not have access to an unobstructed sky or telescope, take a look at the January 16, 2024 NASA Astronomy Picture of the Day. What does it mean for some of the stars to appear red and for some of the stars to appear blue?

Numerical analysis

1. A hydrogen atom requires the absorption of a light wave containing 2.18×10^{-18} J of energy to ionize its electron. Calculate the wavelength and frequency of a light wave with that value of energy.
2. Consider the energy-level diagram shown in Figure 30.21. Calculate the...
    1. ...total number of electron transitions.
    2. ...amount of energy required to jump from the ground state (E0) to E2.
    3. ...wavelength of light corresponding to the energy difference between the ground state (E0) and E2.
    4. ...largest energy transition.
    5. ...wavelength of light corresponding to the largest energy transition.
    6. ...smallest energy transition.
    7. ...wavelength of light corresponding to the smallest energy transition.



Figure 30.21 – A simplified energy-level diagram corresponding to an atom of helium. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

This page titled Chapter 30: Light emission is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

VII: QUANTUM AND NUCLEAR PHYSICS

Chapter 31: Light Quanta
Chapter 32: Quantum physics
Chapter 33: Nuclear physics and radioactivity
Chapter 34: Fission and fusion

This page titled VII: Quantum and nuclear physics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 31: Light Quanta

Summary

At this point, this textbook has provided some convincing evidence that light exhibits many wave-like properties. It can interfere, it can diffract, and it can be polarized.

This chapter will take that understanding of light and turn it on its head. What if light isn't only a wave, but also a particle? Hopefully this chapter will show you how light can be treated as both a particle and a wave in different situations. In fact, both the wave and particle aspects of light are necessary in order to formulate a complete understanding of light. The quantum nature of light is truly fascinating!

Quantum physics

By the end of the 19th century, humankind had developed what is known as classical physics. Classical physics describes the natural phenomena we've discussed so far in this textbook: physics that uses Newton's laws of motion, Maxwell's equations, and the laws of thermodynamics. These laws seemed to form a complete understanding of how the universe worked in terms of particles and waves.

While some mysteries (such as the spectral distribution of incandescence and the observation of ultraviolet light ejecting electrons from metals) remained, these were expected to be soon resolved with a proper application of the classical laws.

At this same time, scientists began to make other startling discoveries. These included the discovery of x-rays, radioactivity, and the electron. Attempts to explain these new observations with laws of classical physics failed. Classical physics led to very deep contradictions between what the laws predicted and what was experimentally observed.

It should be stated at the onset that the classical laws of physics work very well for objects that are larger than atoms and molecules. We have seen a lot of convincing evidence for this so far in this textbook! The strange new effects that were observed occurred when doing science on a very small scale: light interacting with small particles such as atoms, protons, or electrons.

What was needed was a new set of physical laws to explain the observed experimental results that occur on a very small scale. This led to the quantum revolution. We call these new laws quantum physics.

Of course, any new scientific theory has to also account for the correct predictions of the old theory. That is, in addition to explaining new experimental observations, the new theory has to also explain observations for all the situations in which the old theory was valid. This notion was put forth by Niels Bohr in the early 1920s and is called the correspondence principle. The correspondence principle requires that quantum physics and classical physics produce the same result when quantum physics is applied to large-scale objects.

At the heart of quantum physics is the concept that properties that we consider to be continuous on the macro (large) scale become quantized when things become extremely small. A quantity that is quantized is something that only takes on certain well-defined values. For example, the number of people in a room is a quantized property. There can be a room filled with 40 people, or 28, but not four and a half people. The number of pennies in a jar is quantized, as is the number of atoms in a person, the charge of an atom, or the energy of electrons in an atom (as discussed in the previous chapter of this textbook).

The idea is that quantum physics can be used to explain all phenomena: those happening on the very small scale as well as those happening on a very large scale. However, because quantum physics is so mathematically complicated, and because classical physics works so well on the macro scale, there is no need to use quantum physics when calculating quantities on the macro scale. Because quantum equations will produce the same results as Newton's laws (for example), using Newton's laws is perfectly sufficient when calculating velocity, acceleration, momentum, and other kinematic quantities for any interaction that is large enough for us to see with our eyes.

It's also important to point out that the theoretical models we currently use to understand the universe, whether it be quantum physics, Newtonian physics, or general relativity, are excellent models that have undergone rigorous scientific inquiry and scrutiny. But that doesn't mean that these models are not subject to adjustment and change as we learn more about the universe. Either way, the correspondence principle dictates that any modifications must still yield the same results for situations where the previous theory worked.

This chapter focuses on quantum physics as it pertains to light in particular. The next chapter of this textbook will expand on this for a more general discussion on quantum phenomena.

Wave-particle duality

At this point, we have seen some convincing evidence that light is a wave. Light can undergo diffraction, interference, and polarization. These are all phenomena that would not happen if light was not a wave. Furthermore, the fact that light can be polarized indicates that light is a transverse wave.

The photoelectric effect

In 1905, Albert Einstein published a paper describing a phenomenon that had an astounding implication: that light acts not like a wave, but like a particle! This paper was on the photoelectric effect, and his explanation of this effect was the reason that Einstein received a Nobel prize in 1921.

The photoelectric effect is the emission of electrons when light of a particular frequency (especially UV light) strikes a metal target. By itself, this observation was not groundbreaking. It is the details of how the photoelectric effect works that upended our understanding of classical physics and the treatment of light as a wave.

In this video below, a piece of zinc has been loaded up with electrons by rubbing it with a charged piece of PVC. Placing the zinc on the electroscope visually demonstrates the presence of these electrons. When Dr. Pasquale shines a UV light source on the zinc, electrons are ejected from the metal. We can tell that this occurs because the electroscope rapidly goes back to its neutral state.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Classical physics predicts that energy is transferred from the light wave to the electron. It predicts that when a sufficient amount of energy is built up, an electron will eventually be emitted from the metal. This predicts that the intensity of light would be the indication of electron ejection. Strong light would eject electrons quickly, and dim light would eject electrons with a time delay as the electrons gradually build up more energy.

We can test this classical prediction by taking that same piece of negatively charged zinc and illuminating it with a bright white light source. If the classical model were correct, this would also cause the zinc to quickly eject the electrons and cause the electroscope to relax back to its neutral state. This is demonstrated in the video below. When Dr. Pasquale turns on a bright light source, which is much brighter than the relatively dim UV source used earlier, there is no change in the presence of electrons. No matter how long she waits, the electroscope remains negatively charged. (At the end of the video, Dr. Pasquale removes the excess negative charges by grounding the electroscope.)

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

These results must have another explanation. The classical model does not hold up here.

A summary of the classical predictions and the observed experimental results for the photoelectric effect (the ejection of an electron after monochromatic light is absorbed by a metal) is provided in the table below.

The kinetic energy of the ejected electrons should increase when light of greater intensity is used.	The kinetic energy of the ejected electrons is the same regardless of the light intensity.
The kinetic energy of the ejected electrons should not be affected by the frequency of the light. Only the intensity of the light should affect the kinetic energy.	The kinetic energy of the ejected electrons decreases to zero at a threshold frequency, below which no electrons are ejected, regardless of the light intensity.
Dim light should require a longer time for electrons to be emitted (because more time is needed for the electrons to acquire the needed energy to escape).	Although fewer electrons are ejected, there is no measurable time delay when using dim light compared to bright light.

Einstein proposed that the results of the photoelectric effect could be easily explained by treating light as though it interacted with the metal in discrete bundles or quanta of energy. These packets of electromagnetic energy later became known as photons. (It was not until 1926 that the term photon was coined.) In short, a photon is a particle of light.

Quantum theory predicts that the intensity of the light does not matter: the frequency of the light does. Each individual photon that hits the zinc has energy given by Planck's constant times the frequency (recall $E = hf$ from the previous chapter). In this manner, bright light at frequencies that are lower than a threshold energy will never cause electron ejection, which is what we observed when shining bright white light on the zinc. Each photon of white light was incapable of ejecting an electron, as white light is composed of frequencies of light that are lower than that of UV light.

Similar to the photoelectric effect is the photovoltaic effect. The photovoltaic effect occurs when a photon hits a material and causes an electron to become excited such that it can now give rise to electric current. The photovoltaic effect is used to generate electric power in solar panels.

Both the photoelectric and photovoltaic effects indicate that light is a particle. But effects such as interference and polarization indicate that light is a wave. This is known as wave-particle duality. Wave-particle duality is the concept that something like light is both a wave and a particle. Both wave and particle descriptions are required to accurately depict quantum phenomena such as light. Generally speaking, when light travels through space, it acts like a wave. When light interacts with matter, it acts like a particle.

Asking if light is a particle or a wave is like asking if a spork (Figure 31.1) is a fork or a spoon: it's both! So is light: both a wave and a particle.



Figure 31.1 – Is a spork a fork or a spoon? It's both! Titanium spork, by Wtshymanski, is licensed under CC BY-SA 3.0.

As Einstein stated: "It seems as though we must use sometimes the one theory and sometimes the other, while at times we may use either. We are faced with a new kind of difficulty. We have two contradictory pictures of reality; separately neither of them fully explains the phenomena of light, but together they do."

Matter waves

In 1924, Louis de Broglie, a French graduate student, pondered that if entities that we typically consider to be waves can act like particles, perhaps it is possible that entities that we typically consider to be particles can act like waves. In 1929, de Broglie won the Nobel prize for his hypothesis after the first experimental evidence for the wave nature of electrons was obtained by Clinton Davisson and Lester Germer in 1927.

As with all quantum properties, this effect (particles acting like waves) is only noticeable on a very small scale. Macroscopically, the wave-like nature of particles is a negligible effect that we can ignore without affecting the outcome of our experiments or calculations. On the atomic scale, the wave-like nature of particles cannot be ignored.

We use the de Broglie equation to calculate the wavelength of matter waves. The de Broglie equation states that

$$\lambda = \frac{h}{p}.$$

The wavelength of the matter wave is equal to Planck's constant divided by the momentum of the object.

Consider a macroscopic object: a 150 gram baseball, thrown at a fast pitch of 45 meters per second. First we can calculate the momentum of the baseball. Remember that momentum equals mass times velocity, and we need to be careful that our mass is expressed in units of kilograms. The momentum of the baseball is

$$\begin{aligned} p &= (0.15\ \text{kg})(45\ \text{m/s}) \\ &= 6.75\text{kgm/s}. \end{aligned}$$

The wavelength of the baseball is Planck's constant divided by this momentum. It can be calculated as

$$\begin{aligned} \lambda &= \frac{6.6 \times 10^{-34}\text{Js}}{6.75\text{kgm/s}} \\ &= 9.8 \times 10^{-35}\ \text{m}. \end{aligned}$$

Considering that a baseball has a diameter of 7.25 cm (0.07 m), the size of the wavelength of the baseball is essentially zero compared to the size of the baseball itself. This wavelength compared to the size of the baseball is roughly the same as comparing the size of an atomic nucleus to that of the Milky Way galaxy! In short, the wavelength of the baseball is so small that we will never notice its wave-like effects such as diffraction or interference.

Now let's consider an electron, a decidedly subatomic object. An electron has a mass of 9.1×10^{-31} kg, and for this example we can assume that it travels at 1% of the speed of light. The momentum of the electron will therefore be

$$\begin{aligned} p &= \left(9.1 \times 10^{-31} \text{ kg}\right)\left(3 \times 10^{6} \text{ m/s}\right) \\ &= 2.7 \times 10^{-24} \text{kgm/s}. \end{aligned}$$

The wavelength of the electron is Planck's constant divided by the momentum, which is 2.4×10^{-10} m, or 0.24 nm. While this is definitely a small number, it is equivalent in length scale to the wavelength of x-rays. And considering that the diameter of an entire atom is about the same size, the wavelength of an electron is definitely noticeable on this length scale. In fact, it is the wave behavior of the electrons that give atoms their size! (An atom does not have a rigid boundary or structure to it; the size of an atom is therefore defined by the distance between the nucleus and the extent of the wave motion of the electrons, a concept known as the atomic radius.)

All of this discussion leads to further questions. When we speak of matter waves, what exactly is oscillating? How do these waves propagate? Even de Broglie himself did not know. This will be addressed in the next chapter of this textbook.

The double-slit experiment

What experimental evidence might convince you that subatomic particles such as electrons act like waves? The Davisson–Germer experiment confirmed the wave-like nature of electrons. This experiment measured a diffraction pattern generated by electrons as they were scattered by a periodic crystal structure (which acts in a similar manner to a diffraction grating).

The data measured in this experiment is shown in Figure 31.2, which plots the current (which is proportional to the number of electrons hitting a screen some distance away from the periodic crystal) on the y-axis and the scattering angle on the x-axis. The data demonstrates a diffraction pattern, which would not occur if the electrons acted as particles instead of waves.

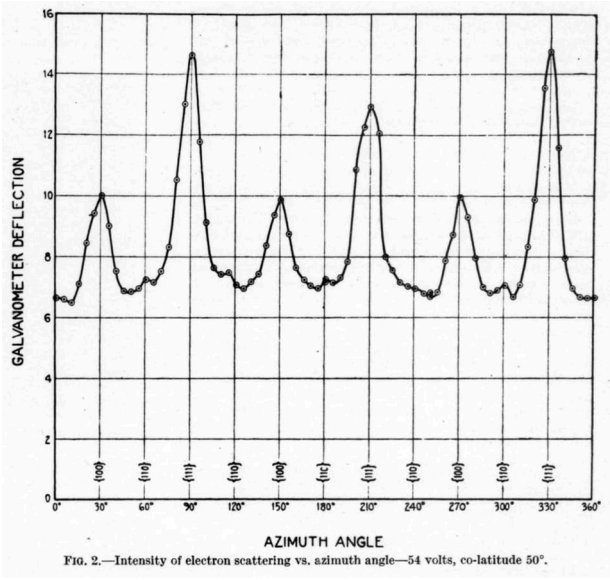


Figure 31.2 – The experimental data from the Davisson-Germer experiment shows that electrons act as a wave. They create a diffraction pattern on a detector. The Scattering of Electrons by a Single Crystal of Nickel, by C. Davisson, L. H. Germer, is licensed under Public Domain.

Recall that diffraction of waves occurs any time those waves travel through a narrow opening. If there are two small openings, this will lead to an interference pattern that arises due to a mixture of diffraction from each slit and interference, leading to bright and dark spots on a screen some distance away.

Diffraction experiments have shown that even when a light source is so low that single photons are sent through a double slit one at a time, the interference pattern continues to be observed on a screen some distance away. It should be pointed out that it is not that the entire interference pattern flashes with each photon. Rather, the individual photons hit the screen as particles and make small dots. It is only after recording the individual dots of each photon that we see the development of the interference pattern. This is an effect that cannot be explained using classical models of physics.

If objects that are completely particle-like are sent through one or two slits, we would likely expect that they would create an image that resembles the size and shape of the opening, somewhat like spraying paint through a stencil.

It turns out that when electrons are sent through a double slit, they also generate a wave-like interference pattern when measured some distance past the double slit. This is not what we would expect if electrons only acted like particles. And just as with single-photon experiments, when electrons are shot through the double slit one at a time, they will, over time, create an interference pattern. Not only do electrons act like waves, but they do so in a way that cannot be explained by classical physics. This effect is demonstrated in the video below, which shows the accumulation of multiple electrons on a sensor. Over time, the interference pattern created by the electrons' wave-like nature becomes clear.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

The uncertainty principle

Hopefully by now you're convinced that all objects are wave-like, and that this wave-like property is especially pronounced on the scale of the subatomic. Inherent in all wave-like systems is the uncertainty principle, which theorizes the extent to how well we can know information about related properties. This is not a limitation of current technology, but a fundamental property of wave-like systems.

For example, position and momentum are linked in uncertainty. The uncertainty principle between these two quantities can be expressed as an equation:

$$\Delta x \Delta p \geq \frac{\hbar}{2},$$

where uncertainty is represented by the capital Greek letter delta, followed by the symbol used for the variable that there is uncertainty in. The quantity

$$\hbar = \frac{h}{2\pi},$$

where h is Planck's constant.

To illustrate this, consider a long train of waves such as shown in Figure 31.3 (top). This wave appears to have a very well-defined wavelength. If the wave represented a matter wave of some particle, we could use the de Broglie equation to find that this particle has a very well-defined momentum. However, the wave is spread over a very wide region of space. In other words, the position of the particle represented by this wave can be anywhere since the wave is so spread out.

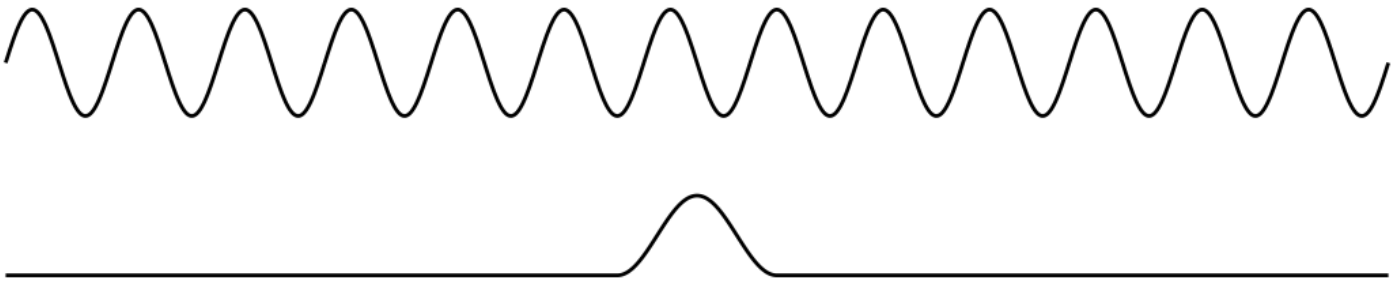


Figure 31.3 – A long train of waves exhibits a very clear wavelength (which can be used to calculate momentum), but has no obvious position (top). A single pulse of a wave contains a very clear position, but no clear wavelength or momentum (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

On the other hand, consider the wave pulse shown in Figure 31.3 (bottom). In this case, the location of the wave appears to be localized in a small region of space. For a matter wave of this shape, the position of the particle represented by this wave is well-defined. However, because there is not a well-defined wavelength, the momentum cannot be measured precisely.

Suppose you wanted to track the trajectory of a subatomic particle such as an electron. In order to "see" the electron, you need to shine light on it. However, the individual photons striking the electron can and do alter the original behavior of the electron. You could use low-energy photons to do this so that you would not alter the momentum of the electron very much. However, low-energy photons have a long wavelength. This means that there will be a greater spread in the uncertainty of the position of the electron. If you want to be more precise in determining the position of the electron, you will need to use light that has a shorter wavelength. However, photons that have a short wavelength have a higher frequency and therefore carry more energy. These photons will give the electron a bigger kick and therefore create greater uncertainty in the momentum of the electron.

This exchange is unavoidable. It is a fundamental property of wave-like systems that there is a lower limit with which the precision that complementary variables (such as position and momentum) can be measured. Because the product of both uncertainties has a lower limit, there is a limit to the precision of any measurements we can make on these two properties. If you know the position of something with a high level of accuracy, then the accuracy with which you can measure momentum will be limited. To the extreme: if you know with perfect certainty the exact location of a wave, then it is impossible to know the momentum of that wave, and vice versa. The uncertainty can never be zero for both quantities simultaneously.

Two other complementary properties that are linked in uncertainty are energy and the time at which the object had that energy. We can state that uncertainty in these two variables as

$$\Delta E \Delta t \geq \frac{\hbar}{2}.$$

In order to measure the energy of a particle, we need to track it. To do this, we could measure the time it takes to travel a certain distance. Since the particle is spread out in space due to its wave-like properties, tracking it for a short time does not give a very precise measure of that distance. Therefore, the energy cannot be determined precisely. To get a more precise measure of the energy, we have to allow the particle to move through space for a longer time. Again, this exchange cannot be avoided. The more precisely we measure one variable, the more imprecisely the complementary variable can be determined.

It's important to point out that uncertainty doesn't just exist on the quantum scale, but also on the macro scale. However, because $\hbar$ is so small, this effect is trivial for most objects we can see with our eyes. If you wanted to track the trajectory of a ball thrown through the air, you would still need to observe the light reflecting off the ball into your visual receptors. In this case, the interaction of the individual photons striking the ball will not alter the trajectory of the ball in any significant way. That is, the light does not affect macro-scale experiments in any noticeable way.

Further reading

- Bach, R., Pope, D., Liou, S.H., & Batelaan, H. (2013). Controlled double-slit electron diffraction. *New Journal of Physics*, 15(3), 033018. doi.org/10.1088/1367-2630/15/3/033018
- The double-slit experiment – This article from Physics World magazine explains a double-slit experiment conducted with electrons.

Practice questions

Numerical analysis

1. Calculate the wavelength of an Olympic sprinter having a mass of 94 kg who completes a 100 m run in 9.58 seconds.
2. Calculate the wavelength of a cyclist who, along with her bicycle, has a mass of 100 kg and rides at a speed of 11.2 m/s.
3. Calculate the wavelength of an ice hockey puck with a mass of 160 g moving at a speed of 51 m/s.
4. Calculate the wavelength of an electron having a mass of 9.1×10^{-31} kg and a speed of 1.45×10^{8} m/s.
5. Calculate the wavelength of a neutron having a mass of 1.67×10^{-27} kg and a speed of 2290 m/s.
6. Calculate the wavelength of a helium atom having a mass of 6.64×10^{-27} kg and a speed of 1780 m/s.
7. Calculate the wavelength of a sodium atom having a mass of 3.82×10^{-26} kg and a speed of 870 m/s.

This page titled Chapter 31: Light Quanta is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 32: Quantum physics

Summary

The wave-like nature of matter brings about some bizarre phenomena that we don't experience on the macro scale in classical physics. This chapter will introduce you to some of the major topics in quantum physics.

The Bohr model of the atom

Earlier in this textbook, we discussed atomic models. From Democritus to Thomson to Rutherford, we learned how each of the atomic models came to be, and what experimental evidence led to changes in how we understand atoms over the past several centuries.

As discussed earlier in this textbook, the emission spectrum of an element is the collection of frequencies of light emitted via deexcitation of excited electrons in the gaseous form of that element. If instead of measuring the light emitted from the gas, we shine a continuous spectrum of light onto that gas, those same frequencies can cause excitations that remove light and produce dark lines in the continuous spectrum. This is called an absorption spectrum.

While these spectra have been observed since the early 1800s, it was not until 1885 that Johann Balmer discovered a simple empirical formula that correctly calculated the visible wavelengths from the frequencies observed in the spectrum of hydrogen. (Hydrogen, the simplest atom, has an emission spectrum with fewer lines and more order than any other elemental atom.) The series of visible lines later became known as the Balmer series in his honor. Although Balmer did not know why his formula worked, his formula was able to correctly predict the existence of other spectral lines (in the non-visible region of the electromagnetic spectrum) that were yet to be observed.

It was generally believed that atomic spectra were clues to the structure of atoms in much the same way that particular resonant frequencies are related to a musical instrument.

In 1913, Danish physicist Niels Bohr made use of Ernest Rutherford's nuclear model of the atom and put forth a new model of the atom and found that he could derive Balmer's formula using a blend of classical physics and quantum assumptions that Bohr made. (These assumptions turned out to be correct, but at the time, Bohr had no basis for making them.) This model is known as the Bohr model of the atom.

In his model of the hydrogen atom, Bohr proposed that the electron was held in a circular orbit around the nucleus (a single proton) by the electric force, analogously to a planet held in orbit around its sun by gravity. However, unlike the planet, which can orbit at any arbitrary distance, the electron could only orbit at certain discrete distances with only certain allowed energies. This is where the quantization concept enters the picture. Bohr's model of electrons orbiting atoms in discrete orbits given by energy levels was prompted by the measurements of atomic emission spectra that exhibit discrete frequencies.

The classical model of an electron orbiting a nucleus would indicate that the electron is accelerating. (As discussed, an object that moves in a curved path necessarily accelerates.) Maxwell's theory of electromagnetism states that accelerating electrons emit light (which was discussed in chapter 30 in this textbook in the context of the generation of x-rays). Because Bohr's model predicts that electrons travel in a curved path, it would be reasonable to expect that those electrons would emit light. However, it is known that electrons traveling in orbits do not emit light. Bohr simply asserted that they do not radiate while traveling in these atomic orbits, without giving an explanation for this discrepancy between his model and observed reality. However, Bohr's model did account for electrons making a "quantum leap" from one allowed orbit to another, by either absorbing or emitting a photon that has energy equal to the energy difference between the two orbits.

Electron waves

The shortcomings of a purely "planetary" and particle-like model of electrons can be explained by the wave nature of electrons. Recall that the double-slit experiment experimentally demonstrated that electrons exhibit wave-like behavior. That's true when

electrons are orbiting an atom as well as when electrons travel freely through space.

Each electron will form a standing wave as it moves about the nucleus. Only waves that have a particular radius, given by the constructive interference of waves, can be supported in shells around the atom. This video below shows an analogy of how this works. If each shell is constituted by an integer value of the electron's wavelength, which would support constructive interference, then those shells have very well-defined sizes. This is analogous to the standing waves we discussed previously in this textbook.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In fact, electron waves are more complicated than this, but this conceptual understanding helps to describe the quantum picture of electrons, and addresses the drawbacks of the classical model of electron orbits. The actual electron orbitals take on very bizarre shapes that include lobes and donuts in highly directional orientations based upon the solutions to an equation that will be discussed later in this chapter. Figure 32.1 illustrates some of these orbital shapes.

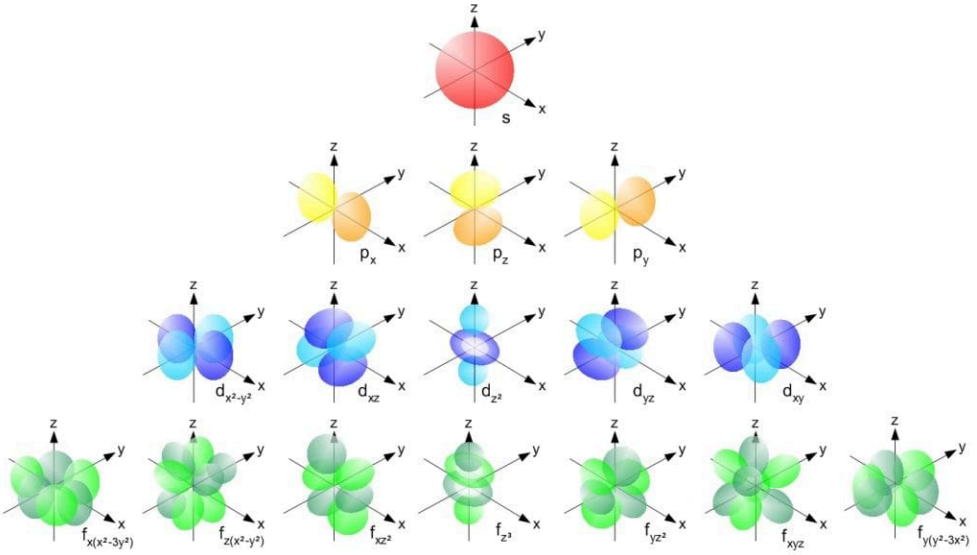


Figure 32.1 – A representation of several of the atomic orbitals of a single electron in a single hydrogen atom. These are defined by solving a quantum-mechanical equation known as the Schrödinger equation. Single electron orbitals, by haade, is licensed under CC BY-SA 3.0.

Quantum mechanics

Quantum mechanics is the study of mechanics (the description of the motion of physical objects of matter) when things become so small that classical physics breaks down and stops working adequately. Quantum objects have three major differences from classical objects.

1. The properties of quantum objects (such as energy and momentum) appear in discrete, quantized values.
2. Quantum objects exhibit wave-particle duality.
3. Quantum objects have fundamental limits to the degree of precision that can be obtained when measuring certain quantities, given by the uncertainty principle.

How do we measure quantities using quantum mechanics? When considering a particle in a classical situation, the future behavior of the particle can be obtained with absolute certainty using Newton's laws (particularly $\vec{F} = m\vec{a}$). When considering

any distribution of charges or currents, the electric and magnetic fields can be found using Maxwell's equations. At the heart of quantum mechanics is an equation proposed by Austrian physicist Edwin Schrödinger in 1926. This equation (known as the Schrödinger equation) is the means by which the future behavior of quantum particles is predicted. It should be mentioned that Newton's laws, Maxwell's equations, and the Schrödinger equation cannot be derived from basic principles because they are basic principles. The validity of Newton's laws, Maxwell's equations, and the Schrödinger equation comes from their ability to make predictions consistent with experimental observations.

In the search for the Schrödinger equation, there are certain properties the equation is expected to have.

1. It needs to be consistent with conservation of energy: $ME = KE + PE$.
2. It has to be consistent with de Broglie's hypothesis of the wave nature of matter $\lambda = \frac{h}{p}$.
3. It needs to give results that are consistent with the observed experimental results.

While it is possible to generate equations that are consistent with the first two points (obeying conservation of energy, and being consistent with de Broglie's matter wave hypothesis), only the Schrödinger equation (presented below) satisfies all three of these points as it also gives results that are consistent with known experimental data. (Source: Krane, Kenneth S., *Modern Physics*. John Wiley, 1983.)

We now return to the question of what is oscillating when it comes to de Broglie's matter waves. What we find in quantum mechanics is that objects can be described with a mathematical quantity known as the wave function, denoted by the Greek letter psi (ψ). It is interesting to note that de Broglie and even Schrödinger did not know how to interpret the wave function. It was Max Born, a German physicist, who in 1926 found that the wave function was related to the probability of finding the particle in a particular region of space in a particular time frame. (More accurately, Born found that it was the magnitude of the square of the wave function that provided these probabilities.) Furthermore, the wave function could be used to calculate the probability of the outcome of different measurements that can be made on a quantum object. For example, the wave function can be used to determine the probability that a quantum particle (such as an electron) will have a particular energy or momentum within a certain range. As bizarre as it may sound, it is waves of probability that are actually oscillating!

It is important to point out that while the Schrödinger equation will be shown below for demonstration purposes, it is not expected that you will use this equation at all, as it is extremely complicated! The mathematics in this equation goes way beyond the math that you are expected to have learned at this stage. In a way, that's the point: quantum mechanics is a very difficult model of physics that we don't use on the macro scale because it is not necessary! Why use this complicated equation when we can use simple equations like $\vec{F} = m\vec{a}$ and get the same results?

Schrödinger's equation is

$$\left[\frac{-\hbar^2}{2m}\nabla^2 + U\right]\psi = i\hbar\frac{\partial\psi}{\partial t}.$$

The left-hand side of the equation contains terms that have to do with how the wave function changes in space. The first term essentially deals with the kinetic energy and the second term deals with the potential energy. The right-hand side of the equation contains terms that have to do with how the wave function changes in time and relates to the total energy of the object. We see Planck's constant divided by two pi ($\hbar$) twice. m is the mass of the quantum object. The upside-down triangle (∇) and squiggly d (∂) are calculus operators. i is the square root of negative one, and leads to wave properties when doing calculus.

To reiterate, instead of obtaining exact values of energy, position, and momentum, like we can calculate in classical physics, the wave function provides us with probabilities. A wave function for an electron orbiting a nucleus will give us the probability of finding an electron existing at any one orbital state around the nucleus of an atom.

To illustrate this idea, Figure 32.2 shows a contrived wavefunction ψ (Figure 32.2, top) and the square of the wavefunction ψ^2 (Figure 32.2, bottom). The tall regions near the center of the ψ^2 graph indicate a high probability of finding the particle slightly off center of the region. The smaller peaks indicate that the particle has a lower probability of being found near the outer edges of the region. The five locations where ψ^2 is zero indicate the particle will never be found at those locations.

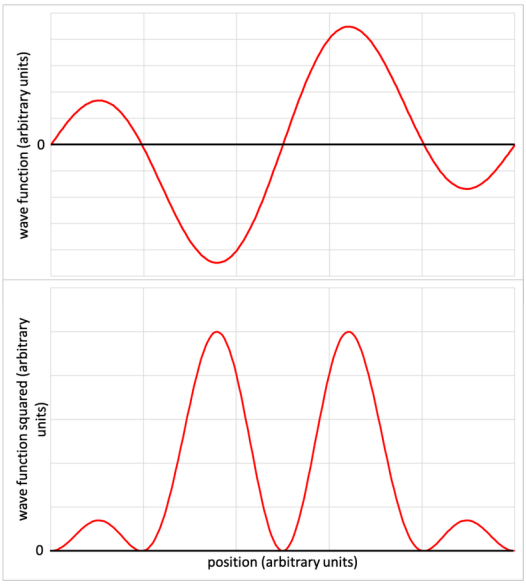


Figure 32.2 – An example of a wave function, which can be calculated with Schrödinger's equation (top). The wave function squared represents the probability of finding a quantum particle at a particular position (bottom). This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

As far as atoms go, Schrödinger's equation has only been solved exactly for the hydrogen atom, which consists of one proton and one electron. It can also be solved exactly for individual particles such as a single proton, or electron, in a few simple environments. When using Schrödinger's equation to model complicated particles or atoms that have more subatomic particles than hydrogen, computers need to be used to numerically calculate wave functions.

The major limitation of Schrödinger's equation is that it is very complicated to use, to the point where solving it analytically (in essence, solving an equation "by hand" by using mathematical operations) quickly becomes impossible. Also, the wave function only tells us probabilistic information about quantum particles. The quantum universe is indeed much different from the classical universe!

The observer effect

The observer effect occurs when measuring a quantum process. When making an observation, we influence that process in some way and therefore affect the outcome. This doesn't only happen in quantum systems, it also happens in classical systems. To measure the tire pressure in a car or bicycle tire, some of the air in the tire is let out when connecting the tire pressure gauge. This changes the measurement. Whatever pressure is measured is going to be slightly less than the pressure that existed in the tire before it was measured. Another example is measuring the temperature of a substance with a thermometer. As the substance and thermometer reach equilibrium, some energy must transfer between the two. This changes what the temperature of the substance was (even if only slightly) before the thermometer was introduced.

Most of the time, our influence on the outcome of a classical or macro-scale measurement is negligible. When using an ultrasonic sensor to measure the speed of a cart in a physics experiment, a sound wave is bounced off of the cart. There is going to be energy and momentum transfer between the sound wave and the cart, but the speed and momentum of the cart will not change to any noticeable degree as a result of using sound waves to measure them. We can essentially ignore the observer effect in cases such as this.

In particle physics, measuring the speed of an electron requires hitting it with a photon. Unlike the sound wave and the cart, an electron is a subatomic particle that will experience non-negligible changes when it interacts with a photon. When measuring the electron by hitting it with a photon, the electron can be affected to a significant degree.

There is a quote in the book Perpetual Motion by Alec Thompson Stewart that does a great job summing this up:

> "Perhaps the most gentle method of searching for an electron in an atom in order to measure its speed is to shoot another electron through the atom. If the exploring electron hits an electron in the atom, it is deflected and from the angle it may be possible to obtain the velocity of the electron it hit. From the atom's point of view, this technique is

not exactly subtle. It would be like closing one's eyes and driving full speed through a street intersection to find out whether traffic was moving on the cross street. If a collision occurred, one could assume that there was a car on the cross street."

When we measure a quantum object, we affect its outcome and cause the wave function to take on one particular value. This is known as collapsing the wave function. Whereas prior to our measurement an electron could have had one of many different values for momentum (for example), after measurement, we know exactly what that value is. The wave function has collapsed.

Relevant to this topic is a thought experiment known as Schrödinger's cat. First, it is important to point out that this is a thought experiment, and was not actually performed, nor were any cats harmed. It only serves as a metaphor to illustrate some of the bizarre interpretations of quantum mechanics. The thought experiment takes a cat, a flask of poison, a Geiger counter, and a radioactive source in a box. The radioactive source has a 50% chance of decaying and emitting a particle within a certain time frame that can be measured by a Geiger counter. If the decay particle is measured by the Geiger counter within the time frame, the flask of poison is shattered and the poison is released, killing the cat. If the detector does not record the particle within the time frame, the mechanism is prevented from shattering the flask. This is depicted in Figure 32.3.

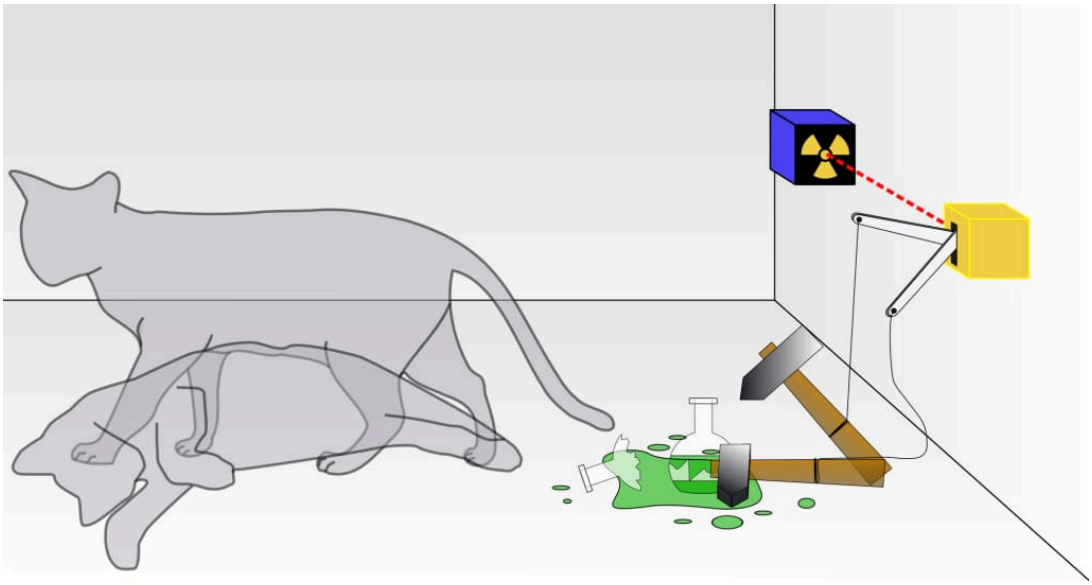


Figure 32.3 – A depiction of the Schrödinger's cat thought experiment, containing a flask of poision, a radioactive source, a Geiger counter, and a possibly unlucky cat. Schrodingers cat, by Dhatfield, is licensed under CC BY-SA 3.0.

The question is: after the allowed time interval, is the cat alive or is the cat dead? This determination cannot be made without opening the box and taking a look. One interpretation of quantum physics states that, after the time frame, the cat is both alive and dead. It exists in a so-called superposition of both states. It is not until the box is opened that we observe the cat to be in one or the other state, collapsing the wave function of the cat into either the "alive" state or the "dead" state with 50% probability of either state being the outcome.

There are a lot of arguments and interpretations of this thought experiment, and it has led to rigorous debate between physicists.

Quantum tunneling

Consider throwing a bouncy ball at a wall. Assuming you don't throw the ball so fast that it smashes through the wall, it will predictably bounce back to you. This is a classical particle: a large object that obeys the laws of classical physics. Its wave nature is so trivial that it can be completely ignored.

If instead you throw an electron at a barrier, it is possible that the electron will tunnel through. During that process, the barrier won't be damaged or changed, but the electron will simply pass through it. This is due to the wave-like nature of quantum objects. This phenomenon is known as quantum tunneling.

The simulation in the video below shows the wave function of a quantum object as it travels toward a barrier. While most of the object reflects off of the barrier, a small amount tunnels through.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

If you think this sounds bizarre, consider that the solid-state memory used in your USB drives, smartphone, and possibly computer or laptop, uses quantum tunneling to store data in its flash memory. Electrons tunnel through an insulating barrier and get stored to save your files, text messages, and photos as binary data. This is a real phenomenon that only occurs with regularity on the quantum level.

Quantum entanglement

Quantum entanglement describes the process during which two or more quantum objects interact in such a way that their quantum properties become interrelated. For example, two photons can be generated such that one has a vertical polarization and the other has a horizontal polarization. The two photons have mutually perpendicular polarizations. This means that if you were to measure the polarization of one of the photons, you would instantly know the polarization of the other, as the two are perpendicular to each other.

The aspect of quantum entanglement that is particularly fascinating has to do with the implications of having simultaneous knowledge of both particles. Imagine that a physicist named Alice entangles two photons in polarization. Alice keeps one photon traveling in an optical setup nearby, and sends the other one to a friend (Bob) on another planet, or even on the other side of the Earth. If Alice measures the polarization of her photon, she instantly knows the polarization of the other photon. Scientists were initially concerned that this seems to imply that knowledge moves faster than the speed of light. If the second photon is on Jupiter, Alice knows its polarization without having to travel all the way there to measure it!

The fact is, however, that by the time Alice calls Bob to tell them the results of the polarization measurement, the transmission of that information to Bob will obey the universal speed limit. Bob will not know the outcome faster than light speed, as Alice's message must travel at or below the speed of light to reach them. While Alice may know information about both particles, that information cannot be transmitted faster than the speed of light.

This is an active area of research for use in quantum computing. Due to the amount of information that can be encoded into entangled quantum particles, the hope is that quantum computing can solve problems that would take classical computers an enormous amount of time to solve.

Further reading

- The Schrödinger equation explained – This short video describes the Schrödinger equation.
- Rigden, J. S. (2002). *Hydrogen: The Essential Element*. Harvard University Press
- What is quantum entanglement? – This article from *Astronomy* magazine explains the concept of entanglement.
- Einstein's 'spooky action at a distance' spotted in objects almost big enough to see – This article from *Science* magazine discusses the entanglement of particles that are almost on the macro scale.

Practice questions

Numerical analysis

1. Consider the contrived wave function for a hypothetical particle shown in Figure 32.4. (Download this data [XLSX, 9 kB]) Determine the probability of finding the particle in the following ranges:
    1. 0 < x < 0.5 nm
    2. 0.5 nm < x < 2.5 nm

3. 1 nm < x < 2 nm
4. x < 0

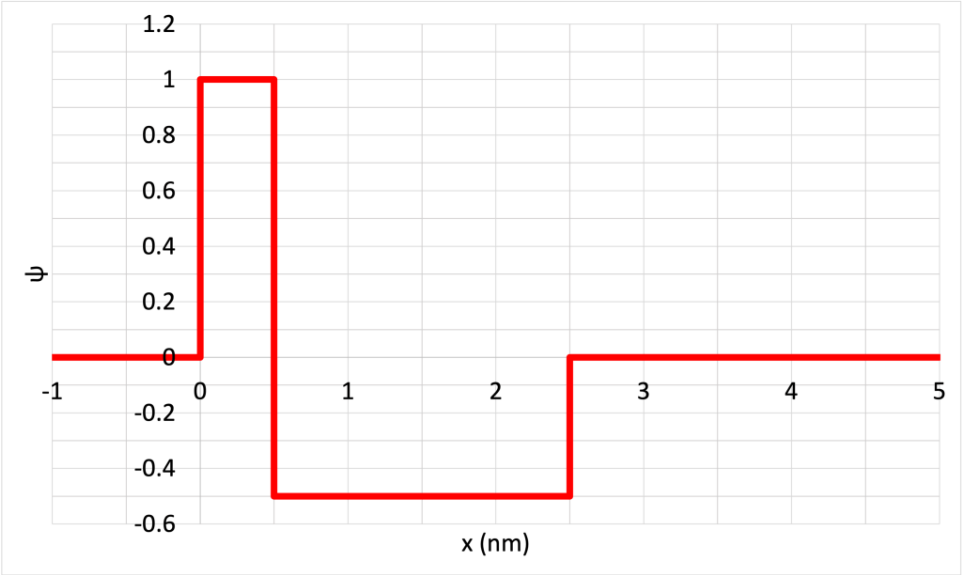


Figure 32.4 – A graph showing the wave function of a hypothetical quantum particle. This figure, created by Alyssa J. Pasquale, Ph.D. and David R. Fazzini, Ph.D., is licensed under CC BY-NC-SA 4.0.

2. Consider the contrived wave function for a hypothetical particle shown in Figure 32.5. (Download this data [XLSX, 9 kB]) Determine the probability of finding the particle in the following ranges:
   1. 0 < x < 1 nm
   2. 1 nm < x < 2 nm
   3. 3 nm < x < 4 nm
   4. 1.5 nm < x < 3.5 nm
   5. 0 < x < 5 nm

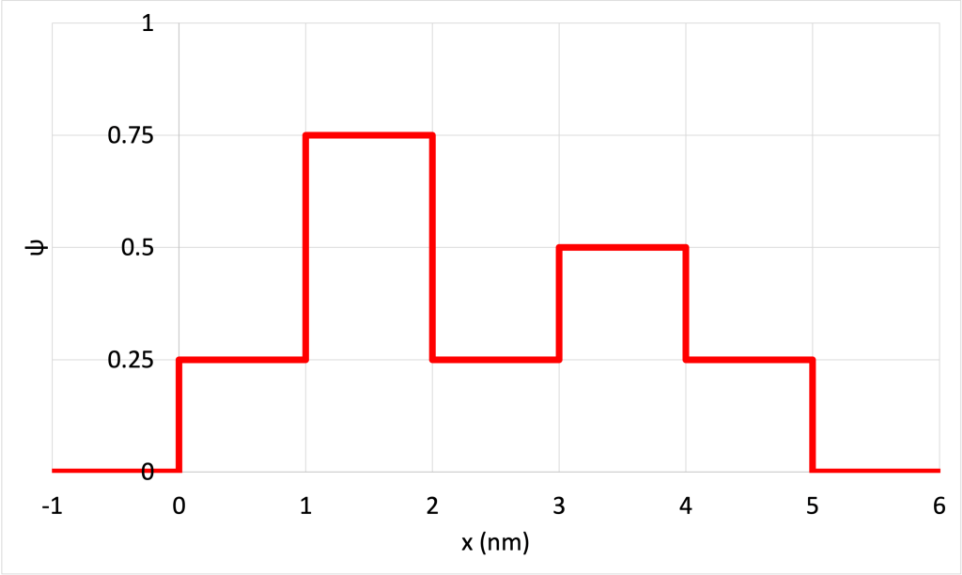


Figure 32.5 – A graph showing the wave function of a hypothetical quantum particle. This figure, created by Alyssa J. Pasquale, Ph.D. and David R. Fazzini, Ph.D., is licensed under CC BY-NC-SA 4.0.

3. Figure 32.6 shows a contrived plot of ψ^2 for a hypothetical particle. (Download this data [XLSX, 9 kB]) Determine the probability of finding the particle in the following ranges:
   1. 0 < x < 1 nm
   2. 1 nm < x < 2 nm

3. 0.5 nm < x < 2 nm
4. 1.5 nm < x < 3 nm
5. x > 3 nm

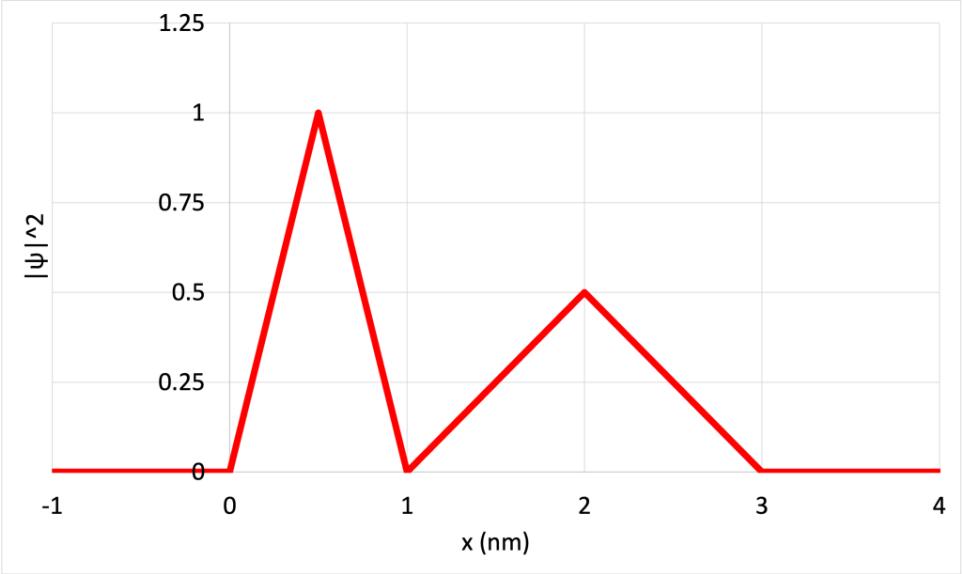


Figure 32.6 – A graph showing the wave function of a hypothetical quantum particle. This figure, created by Alyssa J. Pasquale, Ph.D. and David R. Fazzini, Ph.D., is licensed under CC BY-NC-SA 4.0.

This page titled Chapter 32: Quantum physics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 33: Nuclear physics and radioactivity

Summary

While we have spent a lot of time discussing the importance of the mobile part of an atom, the electrons, we haven't really delved into the nucleus yet. This chapter will discuss the inner workings of the nucleus and how it gives rise to radioactivity, with both positive and negative consequences.

The atomic nucleus

Recall that the results of Ernest Rutherford's gold foil experiment in 1911 led to the discovery of the atomic nucleus. The nucleus was found to have a positive charge, contain almost all the mass of the atom, and have a diameter that was about 10^{-5} the diameter of the atom itself (for very light atoms) and about 10^{-4} (for very heavy atoms). With this discovery, the atom came to be viewed as a very tiny positively charged nucleus with electrons orbiting around that nucleus in a planetary fashion (in that electrons move around the nucleus, similar to how planets move around a star) as it is most commonly depicted in textbooks. This planetary model is depicted in Figure 33.1. However, such diagrams, such as the one shown in Figure 33.1, are not to scale. To put the scale of an atom in perspective, if the atomic nucleus was the size of a baseball, then the atom itself would be about 8 km (or about 5 miles) in diameter. In other words, atoms are mostly empty space!

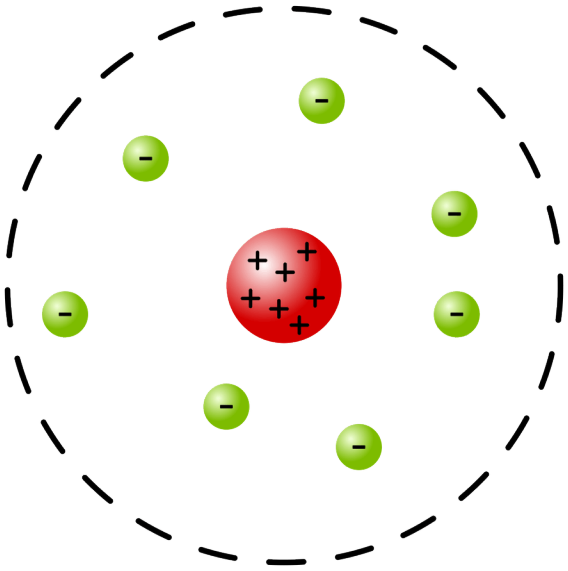


Figure 33.1 – The Rutherford model predicted a positively charged nucleus with electrons traveling around it in a planetary fashion. This depiction of the atom is not to scale! Rutherford atomic planetary model, by Bensteele1995, is licensed under CC BY-SA 3.0.

Subsequent experimentation by Rutherford later showed that hydrogen nuclei were present in all atomic nuclei. As a result, Rutherford is also credited with the discovery of the (positively charged) proton in 1917. Although we now know that the atomic nucleus is composed of protons and neutrons, it was not until 1932 that the (electrically neutral) neutron was discovered. So the planetary model predicted by Rutherford was modified again to include an atomic nucleus, composed of protons and neutrons, which was surrounded by electrons. Collectively, protons and neutrons are referred to as nucleons.

If there are positively charged protons in the nucleus, the question arises, what keeps the protons clumped together? Recall that like charges repel. Positive charges repel other positive charges. Why don't nuclei repel themselves apart due to the electrostatic repulsion among the protons? The answer requires that there be another kind of interaction whose attractive strength overpowers the electrostatic repulsion of the protons. This new force is called the strong nuclear force. The nuclear strong force is the strongest (hence its name) of the four known fundamental forces of nature. (The other three are electromagnetism, the weak

nuclear force, and gravity.) The table below shows the basic properties of each of the four fundamental forces.

strong nuclear	binds quarks together and binds nucleons together (attractive)**	1	10^{-15}
electromagnetism	holds atoms together and binds atoms into molecules (attractive and repulsive)	~0.01	infinite
weak nuclear	responsible for particle decay and radioactivity (neither attractive nor repulsive)	~10^{-7}	10^{-18}
gravity	holds large-scale structures (solar system, galaxies, etc.) together (attractive)	~10^{-40}	infinite

* Compared to the strong nuclear force.

** There are instances where the strong force can be repulsive, but the discussion of the details is beyond the scope of this textbook.

Other than the fact that the strong force is literally holding all of our constituent atomic nuclei together, we do not directly experience the strong force in our daily lives. It is not something that we can feel like gravity, or experience like electromagnetism. Particles that interact through either electromagnetism or gravity can be any distance apart. On the other hand, the strong nuclear force is extremely short-ranged and only attracts the nearest neighbor nucleons.

In small nuclei, the strong nuclear force will easily keep the protons together despite their electrical repulsion. The presence of the electrically neutral neutrons in atoms also helps with atomic stability by creating a buffer between the repeling protons. If an atom has lots of protons, the size of the atomic nucleus becomes quite large. At a certain point, the size scale of an atomic nucleus may be larger than the length scale that the strong force can operate. These atomic nuclei are unstable, and undergo radioactive decay. Elements with more protons than lead (elements with an atomic number greater than 82) all undergo spontaneous decay.

Consider a crude analogy: imagine stacking fluffy pillows one on top of the other into a tower. Each added pillow increases the potential energy of the system. It is easy to stack the first few pillows. However, as the stack gets taller, it becomes harder to stack the next pillow due to the structural instability of the tower. The support forces eventually give way to gravity: the tower tumbles and releases the potential energy as the pillows fall.

As discussed previously in this textbook, isotopes are elements that have the same number of protons but different numbers of neutrons. Many elements have multiple isotopes. Some of those isotopes may be stable, and some may be unstable. Interestingly, neutrons themselves are intrinsically unstable and decay spontaneously. It is only in the environment of the nucleus that the strong force of neighboring nucleons keeps the neutrons stable. However, a nucleus with too many neutrons will be unstable and undergo decay. When the strong force is no longer able to hold the mutually repulsive protons and the neutrons together, the atomic nucleus decays with the release of energy. This nuclear decay is referred to as radioactivity.

Radioactivity

Just a few months after the discovery of x-rays, Henri Becquerel serendipitously discovered a new form of radiation when uranium salts left in a drawer caused a nearby photographic plate to become exposed (this photographic plate is shown in Figure 33.2).

Unlike other high-energy particles such as x-rays (which are also capable of exposing a photographic plate in this manner, and which are purposefully generated using a cathode-ray tube), this new radiation occurred spontaneously. Furthermore, scientists performed experiments and learned that whatever was emitted by the uranium could not be mere x-rays, because the particles emitted by the uranium could be deflected by electric and magnetic fields. This means whatever the uranium was spitting out contained an electric charge. This was the discovery of spontaneous radioactivity. It was subsequently found that there are three primary types of radiation. These three types were named after the first three letters of the Greek alphabet: alpha (α), beta (β), and gamma (γ). These will be discussed in turn starting in the next section.

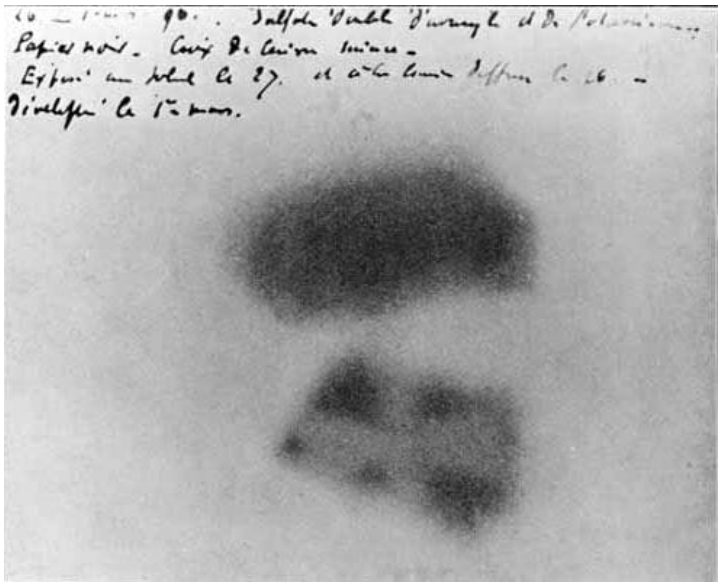


Figure 33.2 – A photographic plate kept in a drawer with a sample containing uranium on top was exposed and created an image due to the high-energy particles emitted by the uranium. Becquerel plate, by Henri Becquerel, is licensed under Public Domain.

It should be pointed out that x-rays are an atomic phenomenon resulting from electron deexcitations from higher shell electrons dropping into lower shells of the electron cloud surrounding the nucleus. However, radioactivity is a nuclear phenomenon resulting from drastic changes taking place within the nucleus of an atom. It is also interesting to note that these nuclear transitions were discovered more than a decade before the nucleus itself was discovered!

The discovery of radioactivity initially unleashed a lot of new knowledge about atoms and isotopes. In particular, the study of radioactivity and radioactive elements was expanded by the work of Marie Curie, who received two Nobel Prizes for her work in the discovery of radioactivity and isolation of the elements radium and polonium.

Before we learned about the negative effects of radiation exposure, the discovery of radioactivity also led to an enthusiasm for radioactive products (many of which were pseudoscientific), such as radium-infused water (shown in Figure 33.3), that people would consume, not realizing that radioactivity is dangerous and can cause cancer and other health hazards in high dosages.



Figure 33.3 – Radithor was a pseudoscientific “health” product containing radium-infused water. Radithor bottle, by Sam LaRussa, is licensed under CC BY-SA 2.0.

Not just pseudoscientific products contained radioactivity. Some commercial products were designed with transuranic elements, even before the fission process was fully understood. Uranium glass is a type of glass that uses uranium to achieve a brilliant green coloration. Some red/orange glazes used in ceramics and tiles used uranium to obtain the coloring. Fiesta brand dinnerware manufactured in the middle of the 20th century contained uranium, and is known to be radioactive. A radioactive Fiesta brand dinner plate is shown in Figure 33.4.



Figure 33.4 – The ceramic glaze used in mid-century Fiesta brand dinner plates contains uranium. The naturally unstable uranium nuclei emit radioactive particles. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In the video below, Dr. Fazzini and Dr. Pasquale use a Geiger counter to measure the decay particles emitted by the uranium in the ceramic glaze of a Fiesta brand plate. When a lead brick is placed between the plate and the Geiger counter, the reading on the Geiger counter goes back to regular background levels of radiation.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Fortunately, our knowledge of radioactivity and radiation is much more thorough now that we've had more than 100 years to learn more about it.

As we'll discuss to a small degree in this chapter, there are many beneficial uses of radioactivity in our lives. Medical sources of radiation can be used to diagnose and treat otherwise lethal conditions such as cancer. Carbon dating is used to determine the age of organic artifacts back to about 50,000 years. Uranium dating can be used to determine the age of ancient rocks that are billions of years old. Scientists use radioactivity to create new atoms and isotopes. Smoke detectors use a radioactive source to detect the presence of smoke. The interior of the Earth remains hot due to the energy released in the radioactive decay of unstable nuclei deep in the Earth (this is the source of geothermal energy).

Alpha decay

Alpha decay is a radioactive decay process wherein a nucleus emits a particle known as an alpha particle. This particle was found to have a positive electric charge and could be blocked using just a piece of paper. It was not until 1909 that Rutherford discovered that an alpha particle was a helium nucleus. It is now known that a helium nucleus consists of two protons and two neutrons. The strong force happens to hold these four nucleons into a very tightly bound entity. When alpha decay occurs, it is generally easier for the unstable nucleus to eject an entire helium nucleus rather than a single proton. (Although proton emission can occur, it is extremely rare.) The process of alpha decay is depicted in Figure 33.5.

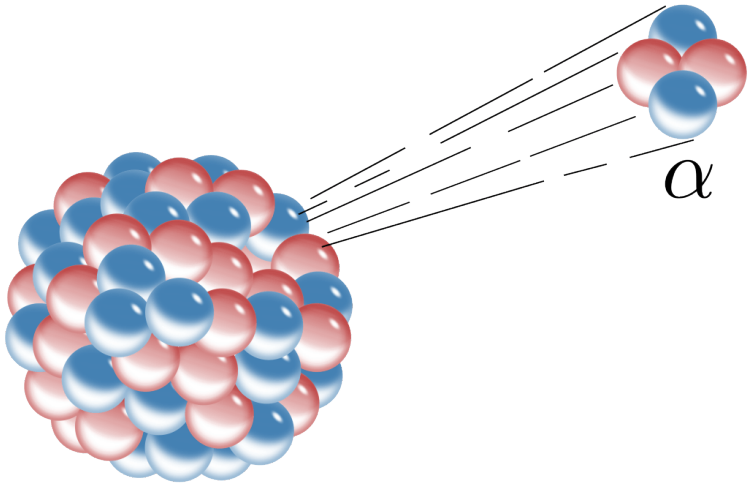


Figure 33.5 – Alpha decay occurs when an unstable nucleus emits an alpha particle (composed of two protons and two neutrons). Alpha decay, by Inductiveload, is licensed under Public Domain.

During the alpha decay process (and, as will be discussed, all of the radioactive decay processes), the total charge and total energy of all particles is conserved, as is the total number of nucleons in all of the particles.

The composition of alpha particles was determined using the magnetic deflecting force described earlier in this textbook. This deflection is shown in Figure 33.6.

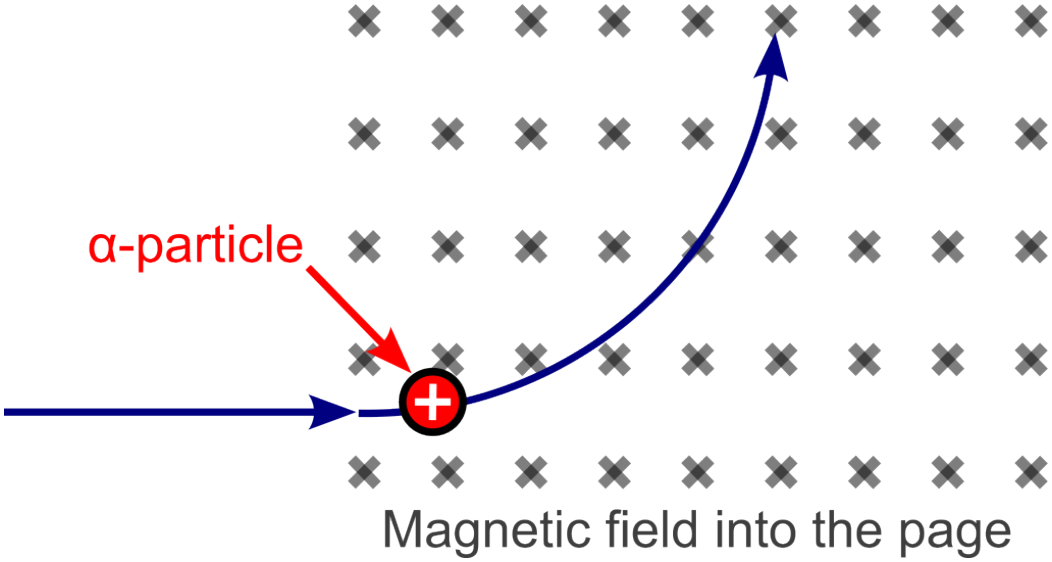


Figure 33.6 – The magnetic deflection force can be used to explain the movement of an alpha particle through a magnetic field. Alphaparticlemagnetic, by Stannered, is licensed under CC BY-SA 3.0.

The amount of deflection that a charged particle undergoes gives information about its charge and mass. Because of inertia, a heavy particle will deflect less than a light particle. The direction of deflection has to do with the sign of its electric charge. In 1909, American physicist Robert Millikan measured the elementary charge of the electron. Once the elementary charge was known, the masses of the deflected particles could be determined. (This is because the particle, traveling in a circular path, requires a centripetal force, contributed by the magnetic deflection force. Knowledge of the elementary charge is necessary to calculate the mass of the charged particle once the radius of the circular path is known.) The deflection direction of alpha particles (which can be determined by the right-hand rule) is consistent with that of a positive charge.

Because alpha particles are so massive, they are very likely to lose energy when coming into contact with any other atoms. So while they are dangerous, they can be blocked by anything more massive than, say, a sheet of paper. It's therefore relatively easy to shield against alpha particle exposure.

An example of alpha decay occurs when uranium-238 emits an alpha particle and becomes thorium-234:

$$^{238}_{92}\mathrm{U} \rightarrow ^{234}_{90}\mathrm{Th} + ^{4}_{2}\mathrm{He}.$$

The atomic mass number is conserved. There are 238 nucleons in the original atom of uranium. The final atom of thorium has 234 nucleons, and the other 4 are present in the alpha particle. Atomic number (charge) is also conserved. The original uranium atom has 92 protons. The final thorium atom has 90 protons, and the alpha particle has the remaining 2.

In general, the alpha decay of any isotope is depicted by the equation:

$$^{A}_{Z}\mathrm{X} \rightarrow ^{A-4}_{Z-2}\mathrm{X'} + ^{4}_{2}\mathrm{He},$$

where A and Z are the atomic mass number and atomic number of the unstable nucleus, respectively. The letter X represents the nuclides (nuclei of a particular isotope) before the decay, while $\mathrm{X'}$ represents the nuclides after the decay. The nuclide on the left side is called the parent nucleus while the one on the right side is called the daughter nucleus.

A cloud chamber is a device that can be used to track the byproducts of radioactive decay. In the video below, lead-210 that is placed in the center of the chamber undergoes alpha decay, which generates condensation in an alcohol vapor inside the chamber and can be visualized as trails of clouds emanating from the source.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Beta decay

Beta decay is a radioactive decay process wherein a nucleus emits an electron. Before the identity of this particle was known, experiments with magnetic fields showed that this byproduct of beta decay contained a negative charge, had the same mass as an electron, and was more penetrating than alpha particles. This may seem peculiar since we do not find electrons in the nucleus. What is going on here?

Beta decay is a process whereby a neutron in a parent nucleus converts to a proton, and while doing so emits an electron and another subatomic particle now known as an anti-neutrino. This decay is depicted in Figure 33.7. During this process, the total charge and energy of all particles is conserved, as is the total number of nucleons in all of the particles. The proton remains in the daughter nucleus after the electron is shot out. This particular process is referred to as β^- decay due to the emission of the negatively charged electron from the nucleus.

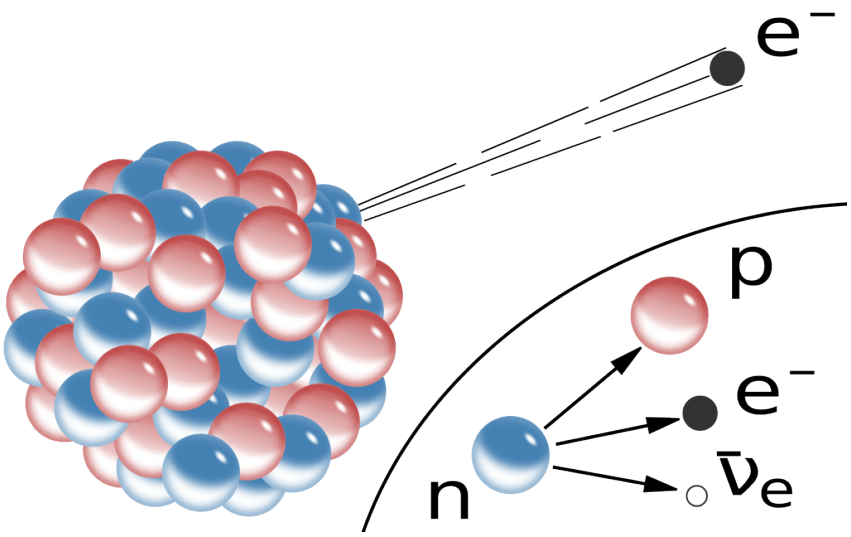


Figure 33.7 – Beta decay occurs when an unstable nucleus emits an electron and an anti-neutrino. Beta decay, by Inductiveload, is licensed under Public Domain.

Because they are smaller than alpha particles, beta particles are more penetrating than alpha radiation. Therefore, they require thicker shielding to prevent exposure. However, it is possible to protect against beta particles by using a few millimeters of aluminum, for example.

An element that undergoes β^- decay is thorium-234. An atom of thorium-234 will emit an electron and become protactinium-234:

$$^{234}_{90}\mathrm{Th} \rightarrow ^{234}_{91}\mathrm{Pa} + ^{0}_{1}\mathrm{e}.$$

Note that the atomic mass number is conserved. There are 234 nucleons in the original atom of thorium, and 234 nucleons in the final atom of protactinium. The atomic number (charge) is also conserved. There are 90 protons in thorium. There are 91 protons in protactinium, balanced by one electron.

In general, β^- decay can be summarized by the equation:

$$ {}^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z+1}\mathrm{X'} + {}^{0}_{-1}\mathrm{e}. $$

(The notation associated with the electron is to remind you that the electron consists of zero nucleons and has an electric charge of -1. Also, the anti-neutrino is omitted from the equation for simplicity.) After this process, the number of protons has increased by one at the expense of the neutron that decayed, but the total number of nucleons is unchanged.

In 1934, an “opposite” type of beta decay was observed by Irene and Frederic Joliot-Curie (a husband and wife who were the daughter and son-in-law of Marie and Pierre Curie). In this decay, a proton in a parent nucleus converts to a neutron that remains in the daughter nucleus, while a positron and a neutrino are ejected from the daughter nucleus. As discussed earlier in this textbook, a positron has the same mass as an electron but has a positive electric charge. Consequently, this decay process is referred to as β^+ decay or positron emission.

Gamma decay

An unstable nucleus that has just undergone alpha or beta decay may be left in an excited state. Just as there are electron energy levels in which electrons can become excited and decay to a lower energy level, protons and neutrons within the nucleus are also arranged in energy levels. A nucleus in an excited state can decay to a lower energy level with the emission of a photon. This nuclear decay process is known as gamma decay. And just as an atom emits a photon when an electron drops to a lower orbital energy level, a nucleus can deexcite and emit gamma radiation. As the energy levels in the nucleus are thousands of times more energetic than the x-rays emitted from atomic electron transitions, gamma radiation is electromagnetic radiation that is thousands of times more energetic than x-rays. In short, gamma radiation is simply a very high-energy photon and is a form of ionizing radiation. This process is depicted in Figure 33.8.

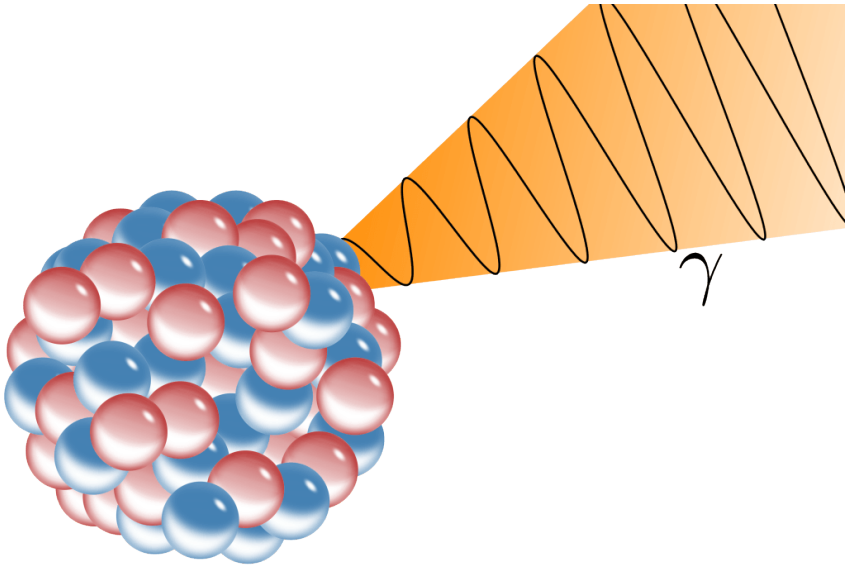


Figure 33.8 – Gamma decay occurs when an excited nucleus decays to a lower energy state and emits a photon in the process. Gamma decay, by Inductiveload, is licensed under Public Domain.

Because gamma radiation is unaffected by the magnetic deflection force, it was shown to have no charge. Because it exhibits characteristics of electromagnetic waves (such as traveling at the speed of light in free space), it was determined to be part of the electromagnetic spectrum.

Gamma radiation can be summarized by the equation:

$$ {}^{A}_{Z}\mathrm{X}^{*} \rightarrow {}^{A}_{Z+1}\mathrm{X} + \gamma, $$

where the asterisk denotes that the unstable nucleus is in an excited state. Note that after the gamma decay, the daughter nucleus is the same isotope since neither the atomic mass number nor the atomic number is unchanged after the decay.

Gamma radiation is very hazardous and also more difficult to shield against than alpha or beta radiation. Gamma rays can be blocked with thick blocks of lead, concrete, or even large amounts of water.

Other decay modes

In addition to the three basic radioactive decays (four if you count β^+ decay), there are other decay mechanisms that can occur. Four additional processes are briefly listed below (although there are other extremely rare modes that are not listed):

1. proton emission (in which a nucleus ejects a single proton), which is an extremely rare process,
2. neutron emission (which occurs in some neutron-rich isotopes as a result of a repulsive effect of the strong nuclear force),
3. electron capture (in which a nuclear proton and an atomic electron merge to form a nuclear neutron with the emission of a neutrino, also called reverse beta decay), and
4. neutron capture (in which a nucleus absorbs a neutron).

All of the decay processes discussed are summarized in the table below.

alpha decay	$^{A}_{Z}\mathrm{X} \rightarrow ^{A-4}_{Z-2}\mathrm{X'} + ^{4}_{2}\mathrm{He}$	$^{226}_{92}\mathrm{Ra} \rightarrow ^{222}_{Z-2}\mathrm{Rn} + ^{4}_{2}\mathrm{He}$
beta-minus decay	$^{A}_{Z}\mathrm{X} \rightarrow ^{A}_{Z+1}\mathrm{X'} + ^{0}_{-1}\mathrm{e}$	$^{234}_{90}\mathrm{Th} \rightarrow ^{234}_{91}\mathrm{Pa} + \mathrm{e}^{-}$
beta-plus decay (positron emission)	$^{A}_{Z}\mathrm{X} \rightarrow ^{A}_{Z-1}\mathrm{X'} + ^{0}_{1}\mathrm{e}$	$^{22}_{11}\mathrm{Na} \rightarrow ^{22}_{10}\mathrm{Ne} + \mathrm{e}^{+}$
gamma decay	$^{A}_{Z}\mathrm{X}^{*} \rightarrow ^{A}_{Z}\mathrm{X} + \gamma$	$^{137}_{56}\mathrm{Ba}^{*} \rightarrow ^{137}_{56}\mathrm{Ba} + \gamma$
proton emission	$^{A}_{Z}\mathrm{X} \rightarrow ^{A-1}_{Z-1}\mathrm{X'} + ^{1}_{1}\mathrm{p}$	$^{53}_{27}\mathrm{Co} \rightarrow ^{52}_{26}\mathrm{Fe} + \mathrm{p}^{+}$
neutron emission	$^{A}_{Z}\mathrm{X} \rightarrow ^{A-1}_{Z}\mathrm{X} + ^{1}_{0}\mathrm{n}$	$^{13}_{4}\mathrm{Ba} \rightarrow ^{12}_{4}\mathrm{Ba} + \mathrm{n}^{0}$
electron capture (reverse beta decay)	$^{A}_{Z}\mathrm{X} + ^{0}_{1}\mathrm{e} \rightarrow ^{A}_{Z-1}\mathrm{X'}$	$^{106}_{47}\mathrm{Ag} + \mathrm{e}^{-} \rightarrow ^{106}_{46}\mathrm{Pd}$
neutron capture	$^{A}_{Z}\mathrm{X} + ^{1}_{0}\mathrm{n} \rightarrow ^{A+1}_{Z}\mathrm{X}$	$^{113}_{48}\mathrm{Cd} + \mathrm{n}^{0} \rightarrow ^{114}_{48}\mathrm{Cd}$

It should be noted that the three beta decay modes described in the table above are mediated by the weak nuclear force. Because the weak nuclear **force** is neither attractive nor repulsive (in that it does not accelerate particles), a better term is the weak nuclear **interaction**. The weak interaction is responsible for changing quarks from one type to another. It is the changing of the quarks' types that converts neutrons into protons and vice versa. Further discussion of this topic is beyond the scope of this textbook.

Half-life

The decay process of unstable elements is characterized by something known as the half-life. This half-life is a description of the amount of time it takes, on average, for half of a particular isotope to decay. Every time a half-life elapses, half of the sample (on average) will have undergone the radioactive decay process, leaving half in the original unstable state. Because the subsequent decay occurs with half of the remaining sample, the percentage of the sample that has not undergone the decay process yet will decrease exponentially over time. This is depicted in Figure 33.9.

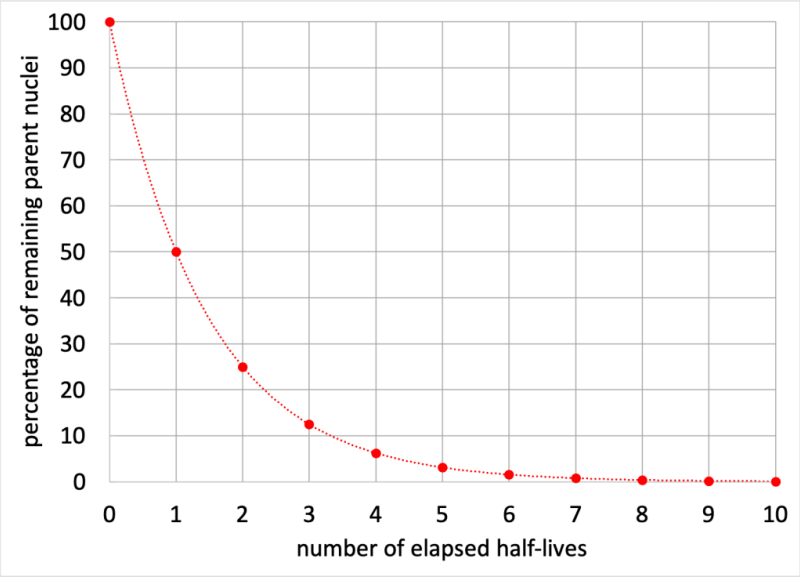


Figure 33.9 – Every time a half-life elapses, 50% of the remaining parent nuclei (on average) will decay. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

A decay process is probabilistic (similar to rolling dice or flipping coins). For example, when rolling a fair single 6-sided die, it cannot be said with any certainty that the next roll will result in any particular number (such as the number 1, for example). What can be said is that there is a one-in-six chance that the next roll will result in any particular number. Even if a die has been rolled a hundred times without getting a 1 (an unlikely but not impossible occurrence), there is still only a one-in-six chance that the next roll comes up 1.

As another example, tossing a coin 100 times will likely result in roughly equal amounts of heads and tails outcomes. However, during this experiment, it is possible that multiple or even many subsequent coin tosses may result in all “heads-up” or all “tails-up.” Each coin toss is an independent event with a 50% chance of obtaining a particular outcome.

Similarly, any particular nucleus of an unstable isotope may decay in a timeframe that is either longer or shorter than the known half-life. In other words, there is no way to predict whether a particular nucleus will decay within a certain time. All that can be said is that within a certain time frame, a certain fraction of unstable parent nuclei will decay (on average).

Every isotope has its own half-life. Xenon-124 has one of the longest known half-lives of 10^{22} years, much, much, much older than the known age of the universe! Scientists have detected the decay of xenon-124, even though it is an extremely rare occurrence.

On the other end of the spectrum, a synthetic isotope of hydrogen, hydrogen-5, has the shortest known half-life of 10^{-23} seconds.

Because the half-life defines the amount of time it takes for half of an unstable isotope to decay, how long would it take 100% of an unstable isotope to decay? Consider an isotope with a mass of 100 grams and half-life of one day. After one day, approximately 50% (50 grams) of the sample will have decayed. After two days, approximately 50% of the remaining unstable sample, or 25 grams, will decay. After three days, an additional 12.5 grams will decay. This means that after three days, approximately 87.5% of the sample will have decayed.

After 7 days, one week, about 99.2% of the sample will have decayed. After two weeks, about 99.99% of the sample will have decayed. When does that value become 100%? The answer is that the decay process is asymptotic. We see that it approaches 100% over a long period of time, but the answer to the question of “when does the very last atom of the isotope decay” is completely probabilistic and cannot be definitively answered.

Unstable isotopes with long half-lives that are present in nuclear waste or contamination can be particularly dangerous. This is because the long half-life indicates that a lot of time will have to pass before the isotope decays to a safe level. Even when the isotope does decay, it may decay to a second unstable isotope. It's possible that it will take days, years, decades, centuries, millennia, or longer for half of some radionuclides to decay to a stable isotope.

Carbon dating

Carbon dating is used to determine how old organic archeological samples are. All organic objects contain carbon. Living plants and animals will naturally replenish this carbon, which includes carbon-12, a stable isotope, and carbon-14, an unstable isotope that undergoes beta decay. After death, the carbon-14 will no longer be replenished and will undergo beta decay with a half-life of 5,730 years. This means that the ratio of carbon-14 relative to carbon-12 in an artifact can be used to calculate its age with reasonable accuracy to about 50,000 years.

Radiation dosage and exposure

The amount of radioactivity that humans encounter can be characterized by the energy we absorb from that radioactivity. This is known as rad: radiation absorbed dose. One rad is equivalent to 0.01 joules per kilogram of tissue.

However, not all radiation is equally dangerous. The unit of rem, roentgen equivalent man, attempts to quantify the dosage of radiation based on the damage it causes to humans. Alpha radiation, for example, may only emit 1 rad, but has a health effect of 10 rems. Beta radiation of 1 rad would have a health effect of 1 rem.

A lethal dose of radiation for a human starts at about 500 rem distributed over the entire body. Such an exposure causes death to 50% of the population within 30 days. Localized dosages of equivalent amounts of radiation may be survivable.

While the concept of radioactivity sounds scary, the truth is that we are surrounded by it. There is a natural background of environmental radiation that we're exposed to in the air, from common rocks and minerals around us, and from cosmic radiation from the Sun and stars. A small amount of radiation exposure is inevitable from our daily lives, and these small exposures pose miniscule risk in our lives. For perspective, a typical annual dose of naturally occurring radiation is about 0.3 rem, a very small amount! Figure 33.10 (source [PDF, 45 kB]) shows the annual average dose of different sources of radiation exposure (industrial and occupational doses are so small that they do not meaningfully show up on the graph). (Download this data [XLSX, 9 kB]) Radiation exposure from medical sources is only relevant for people undergoing these medical procedures. The dosage for a computed tomography (CT) scan, for example, is only received by the patient undergoing the scan.

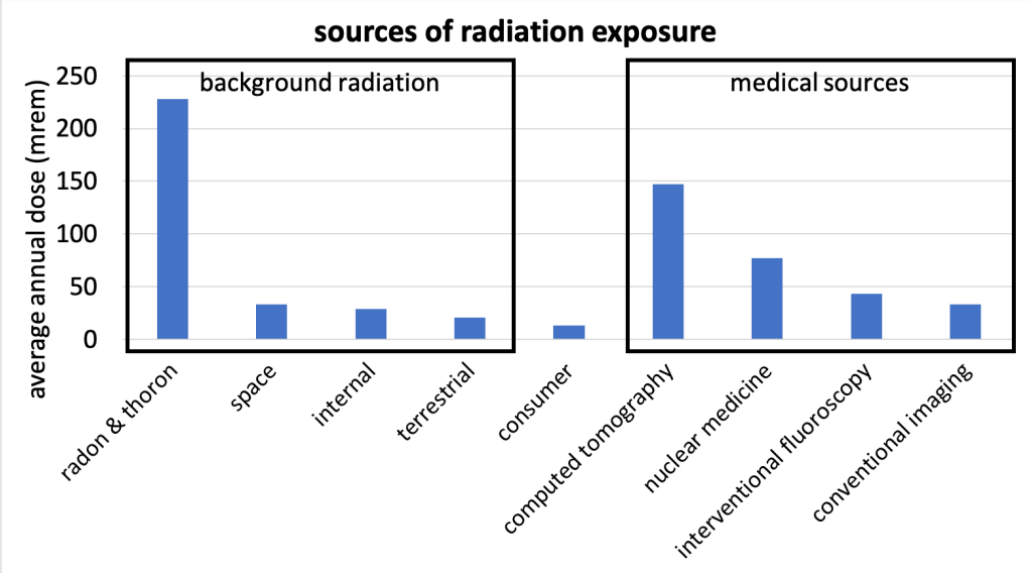


Figure 33.10 – A bar graph of sources of radiation exposure and their corresponding average annual dosage in mrem. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Medical diagnostics such as CT scans and nuclear medicine represent another source of possible radiation in our lives. The idea

is that the benefit of the medical exposure to radiation outweighs the risk of the radiation exposure.

An isotope of radon, radon-222, is a heavy gas that may accumulate in basements that lack proper ventilation. High levels of radon-222 are dangerous, and radon detection and mitigation are both important to avoid unnecessary exposure to its alpha radiation.

There are other sources of environmental radiation. Historical atmospheric nuclear testing occurred between the 1940s and the 1960s. Thankfully, it was banned in 1963, so the dosage we receive from this historical time period is fortunately decreasing every day.

Occupational exposure is also possible. The National Council of Radiation Protection (NCRP) recommends that exposure not exceed 5 rem per year. Due to greater exposure to cosmic radiation at high altitudes, airline pilots and flight crews have limits on their annual flight hours. Other jobs that may encounter exposure to radiation will require employees to wear dosimeters to track their exposure. A collection of dosimeters hanging outside of the Advanced Photon Source at Argonne National Lab is shown in Figure 33.11.



Figure 33.11 – Dosimeters are used to measure exposure to radioactivity in workplaces. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Burning coal, for example in power plants, leads to emission of radioactive thorium and uranium, and is a larger source of environmental exposure to radiation than nuclear power plants. Nuclear accidents (such as Three Mile Island, Chernobyl, and Fukushima) and nuclear waste are other sources of possible radiation exposure.

One of the best ways to reduce your exposure to radiation is to quit smoking if you're a smoker. Cigarettes contain polonium-210, which undergoes alpha decay. When inhaled, it can represent a very large percentage of the annual radiation exposure of a smoker.

Artificial transmutation

The term transmutation refers to the change of one nuclide into a completely different nuclide. The processes of natural transmutation have been discussed: alpha and beta decay. (Recall that gamma decay does not change the isotope.) However, it is also possible to artificially transmute elements. Two particles can be smashed together to generate a new atomic nucleus.

For example, helium-4 and nitrogen-14 can be artificially transmuted to oxygen-17 and hydrogen-1 (a single proton):

$$^{4}_{2}\mathrm{He} + ^{14}_{7}\mathrm{N} \rightarrow ^{17}_{8}\mathrm{O} + ^{1}_{1}\mathrm{H}.$$

Just as with natural transmutation, in the artificial transmutation process, atomic mass number and charge are conserved.

Long ago, alchemy was the study of transmuting elements such as lead into gold. It was hoped that a process could be found that

would allow people to generate the precious metal.

Today, it is possible to use artificial transmutation to create gold. However, it is not remotely feasible economically. This is because the amount of energy that goes into the production of gold is enormous, and the gold that is created in the transmutation process would have to be the one isotope of gold that's stable, otherwise it will naturally transmute into other elements.

Artificial transmutation has been used to generate atoms that don't exist naturally. The so-called superheavy elements on the periodic table that have an atomic number greater than 103 have been created in laboratory settings. These isotopes have such short half-lives that if any natural sources existed (for example, as a result of a supernova), they have long since decayed. The creation and study of these elements leads to advances in physics and chemistry.

Artificial transmutation is also used in radiopharmacology to generate isotopes with short half-lives to use in medical treatment or diagnosis. Some elements such as technetium-99 need to be generated almost immediately before use, as the short half-life would otherwise cause them to decay almost entirely before their use.

Further reading

- Detection of xenon-124 decay – Researchers observed what they believe to be radioactive decay of xenon-124, which has a half-life of 1.8×10^{22} years, one trillion times longer than the current age of the universe.
- What's up with U? – This game from the National Museum of Nuclear Science and History explains the concept of half-life as well as the decay process of uranium-235.
- Radiation dose calculator – This online calculator from the American Nuclear Society estimates your average annual dose of radiation exposure.
- ORAU Museum of Radiation and Radioactivity – The website for the Oak Ridge Associated Universities Museum of Radiation and Radioactivity has a lot of information about radiation, radioactivity, and their history in the United States.
- Fiestaware – This webpage contains information about radioactive Fiesta dinnerware.
- Radioactivity in antiques – This website from the United States Environmental Protection Agency (EPA) describes different antique items that contain radioactive substances and emit alpha, beta, and/or gamma radiation.

Practice questions

Conceptual comprehension

1. Americium-241, which is used in household smoke detectors, is produced from a plutonium isotope. What type of decay process must occur for plutonium to become americium?
2. Consider two 100 g samples, one of nickel-63 (half-life: 100 years), and the other of nickel-56 (half-life: 6 days). Which sample will consist of more undecayed nickel isotopes after any given period of time (say, one week or one month)? Why is that?

Numerical analysis

1. Determine the decay byproduct (the element, number of protons, and number of nucleons) of americium-241, which undergoes alpha decay.
2. Determine the decay byproduct (the element, number of protons, and number of nucleons) of radium-223, which undergoes alpha decay.
3. Determine the decay byproduct (the element, number of protons, and number of nucleons) of carbon-14, which undergoes beta decay.
4. What is the decay mechanism that occurs in each of the following decays:
    1. polonium-218 decays into astatine-218
    2. bismuth-210 decays into thallium-214
    3. aluminum-26 decays into magnesium-26
    4. nitrogen-11 decays into carbon-10
5. Uranium-234 eventually decays into a stable isotope of lead. Which isotope of lead is the possible result: lead-206, lead-207, lead-208, or lead-209?
6. In the previous question, uranium-234 decays into lead through a series of alpha and beta (minus) decays. How many alpha decays and how many beta-minus decays occur in this decay chain?
7. Radium-223 has a half-life of 11.4 days. On average, how much radium-223 would be left in a 100 g sample after 22.8 days?

45.6 days?

8. If a 100 g sample of an unknown radioactive element is left to decay for 8 days and, at that point, only contains 25% of the undecayed element, calculate the half-life of the element.
9. If a flight crew receives 0.06 rem/hr on a long-haul polar route, calculate the maximum hours that the crew can fly in order to keep within the National Council of Radiation Protection recommendation of no more than 5 rem/yr of exposure.

Hands-on experiments

1. You can simulate the radioactive decay process by shaking a box containing a large number of dice or pennies (that represent unstable nuclei) and removing those that come up a certain way (the decay products). For example, if you are shaking a box of pennies, you could remove all the pennies that come up tails after a shake and then continue with the remaining pennies until all of the pennies have "decayed." Each shake represents a specific time interval.
    1. Start by placing a large number of pennies (or other coins) into a box. Securely fasten the lid. Shake the box vigorously for a couple of seconds. Open the box and remove all of the pennies that come up tails and set them aside. With the remaining coins left in the box, repeat the shaking process until you have no pennies left. How many time intervals does it take for the pennies to decay to half of their original value?
    2. Repeat the process using dice instead of pennies. Start by removing all of the dice that come up as 1. How many time intervals does it take for the dice to decay to half of their original value? How does this compare to the pennies?
    3. Repeat the process with dice, removing dice that come up as an odd number. How many time intervals does it take for the dice to decay to half of their original value? How does this compare to the pennies?
    4. Repeat the process with dice, removing dice that come up with a value between 1 and 5. How many time intervals does it take for the dice to decay to half of their original value? How does this compare to the pennies?

This page titled Chapter 33: Nuclear physics and radioactivity is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Chapter 34: Fission and fusion

Summary

This chapter will discuss the energy that can be harnessed by the atomic nucleus in the processes of nuclear fission (the splitting of heavy atomic nuclei into lighter pieces) and fusion (the combining of light atomic nuclei into a heavier piece).

Fission

As previously discussed, radioactive decay occurs when an atomic nucleus is unstable. This can happen due to the presence of so many protons that electromagnetic repulsion becomes stronger than the strong nuclear force that otherwise keeps the nucleus of an atom glued together, or when there are too many neutrons for the strong nuclear force to keep the unstable neutrons from decaying. The discussion that follows focuses on the former.

An atomic nucleus is not a static object. The nucleons (protons and neutrons) in the nucleus vibrate and jiggle around like a liquid drop. This internal motion can cause the nucleus to occasionally stretch into an elongated shape. If this elongation exceeds a critical amount, the electrical repulsion can overcome the strong nuclear force, causing the nucleus to pinch in the middle and then split apart. This is depicted in Figure 34.1. While rare, this does happen naturally, particularly in uranium and elements of greater atomic number. (Elements with atomic number greater than 92 are called transuranic.) This naturally occurring splitting is called spontaneous nuclear fission.

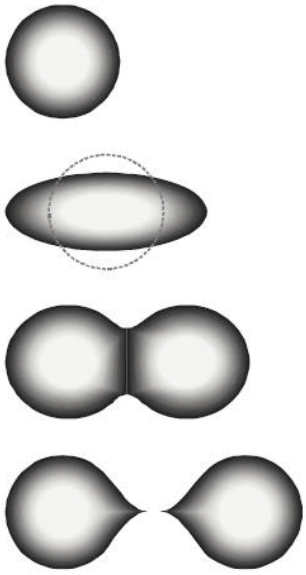


Figure 34.1 – A circular atomic nucleus may become elongated. At this point, the force of electrostatic repulsion becomes stronger than the strong nuclear force and the nucleus will pinch in the middle and split apart. Stages of nuclear fission (liquid drop model), by Hullernuc, is licensed under CC BY-SA 3.0.

If instead we artificially elongate a heavy nucleus, say by bombarding it with a neutron, the nucleus can be induced to elongate beyond the critical point so that electromagnetic repulsion will cause the atom to split in two roughly equal pieces. This process is known as induced nuclear fission.

Nuclear fission was discovered in 1938 by scientists Otto Hahn, Fritz Strassmann, and Lise Meitner when they found that barium isotopes were being produced after bombarding uranium-238 nuclei with neutrons. Lise Meitner and her nephew Otto Frisch came up with the theoretical explanation of what was occurring in these observations.

The process of nuclear fission not only causes the generation of two different nuclei from one original nucleus, it also releases a lot of energy in the process. Where does this energy come from? Consider a nucleus that will undergo fission. We carefully measure its mass before the fission process occurs. After bombarding the nucleus with a neutron, the nucleus fragments into two new nuclei (and some stray neutrons). If we carefully measure the mass of all of the byproducts of the fission process, that mass will actually be smaller than the original mass...by a very small amount.

We've talked about conservation of energy. Interestingly, in addition to other possible forms of energy, matter has energy due to its mass. This mass-energy must be taken into account when considering energy conservation, particularly in nuclear interactions. Mass and energy are related by Albert Einstein's famous equation

$$E = mc^2,$$

where E is the energy that something has by virtue of its mass m, and c is the speed of light. Note that this formula indicates that a tiny amount of mass contains a very large amount of energy (since the speed of light squared is a huge quantity).

The energy liberated in the fission process comes from the mass that is lost during the fission process. This equation basically identifies mass as another form of stored energy. Because the speed of light is such a large number, even a small amount of mass has the potential to release an enormous amount of energy. When Meitner calculated the mass of all the constituents before and after the reaction, it was found that the energy released in the reaction was equal to the energy calculated from the mass difference before and after the reactions. That is,

$$E_{released} = \Delta mc^2.$$

Can all atoms be used to generate energy by way of fission? The short answer is no. Very heavy atoms are capable of releasing energy from fission. Very small atoms are enormously hard to split apart due to their small size and the power of the strong force. If we look at a graph of mass per nucleon in an atom (shown in Figure 34.2), a hydrogen atom has a lot of mass in its one nucleon. On the other end of the spectrum is uranium. In between, with the least mass per nucleon, is iron. Atoms heavier than iron can release energy when they undergo nuclear fission. But the most energy will be generated with the heaviest atoms, and uranium-235 is the source of most nuclear power generated in the United States. (Note that u, the unit used on the y-axis of Figure 34.2, describes mass, and will be defined at the end of this chapter.)

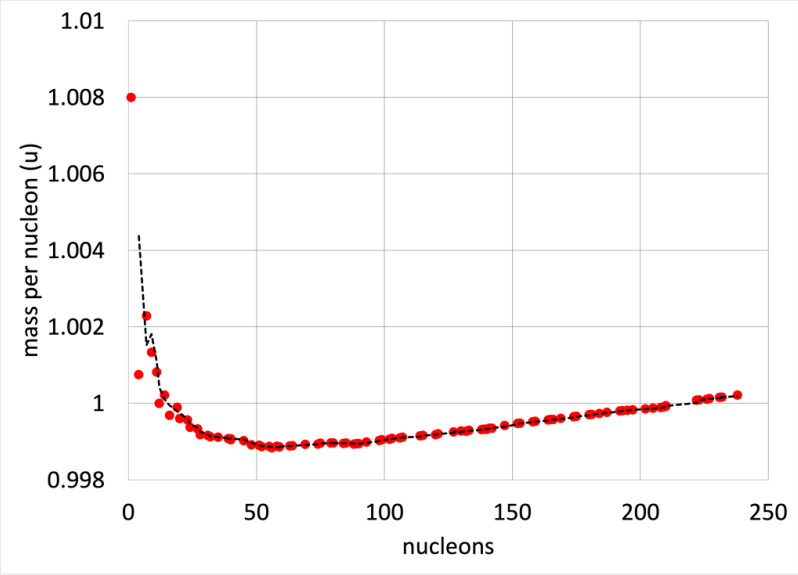


Figure 34.2 – A graph depicting the mass per nucleon of elements between hydrogen and uranium. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Uranium-235 is a fissile isotope of uranium. When a neutron is shot at uranium-235, the uranium-235 can capture the neutron to become uranium-236. The uranium-236 then fissions with one possible set of byproducts being krypton-92, barium-141, and three neutrons:

$$^{235}_{92}\mathrm{U} + ^{1}_{0}\mathrm{n} \rightarrow ^{236}_{92}\mathrm{U} \rightarrow ^{89}_{36}\mathrm{Kr} + ^{144}_{56}\mathrm{Ba} + ^{3}_{0}\mathrm{n}.$$

This process is depicted in Figure 34.3.

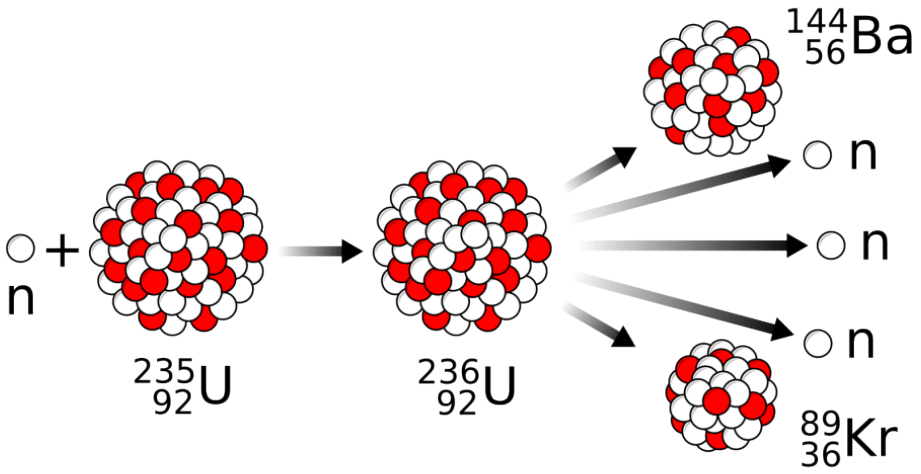


Figure 34.3 – The fission of uranium-235 leads to the generation of an isotope of krypton, barium, and three neutrons. Nuclear fission reaction, by MikeRun, is licensed under CC BY-SA 4.0.

The total number of nucleons is conserved in this process, as is the charge and total energy. As mentioned, while mass is not conserved, mass-energy is.

Usually, the fission process is started by shooting a neutron at an atom to cause the elongation. Why is a neutron used instead of a proton? The neutron, being neutral, will not undergo any electromagnetic repulsion forces that would cause it to deflect away from an atomic nucleus. This is why neutron bombardment is a good method to explore atomic structure.

Some fission reactions liberate even more neutrons. If enough fissionable material is present, the liberation of new neutrons after each fission event can start what is known as a chain reaction: a self-sustaining reaction that doesn't require any intervention to keep it going.

When a large enough quantity, known as a critical mass, of fissionable material is present, the energy is released so quickly that an explosion can result, releasing a massive amount of energy in a very short time. This is the principle behind nuclear weapons.

Fissionable vs. fissile nuclei

All unstable nuclei that will undergo fission after being bombarded with a suitable neutron are known as fissionable. However, it's possible that some nuclei require neutrons with a very large amount of energy before they will undergo fission. Neutrons with a lot of energy travel very fast, and their de Broglie wavelength is therefore quite small. The de Broglie wavelength is qualitatively related to the likelihood of the neutron being captured by a nucleus. The smaller the wavelength, the lower the odds of being captured by a nucleus. This means that fissionable nuclei that require high-energy neutrons do not necessarily undergo the fission process in an efficient manner.

A fissile nucleus is an isotope that can undergo fission after being bombarded with slow neutrons (known as "thermal" neutrons). These neutrons, because they move slowly, have a larger de Broglie wavelength and therefore have a higher chance of interacting with fissile nuclei. A fissile isotope does not require as much material to be present to sustain a nuclear chain reaction. (Fissile isotopes are a subset of fissionable isotopes. All fissile nuclei are fissionable, but not all fissionable nuclei are fissile. For example both uranium-235 and uranium-238 are fissionable. However, while uranium-235 is fissile, uranium-238 is not.)

Chain reactions

A chain reaction is a self-sustaining fission process. For instance, after the fission process, a uranium-235 nucleus can liberate two or three new neutrons depending on the daughter nuclei. If a sufficient amount of uranium-235 is present, the liberated new neutrons will cause more of the atoms to undergo fission, which will release more neutrons, create more fission, and so on. This is shown in Figure 34.4.

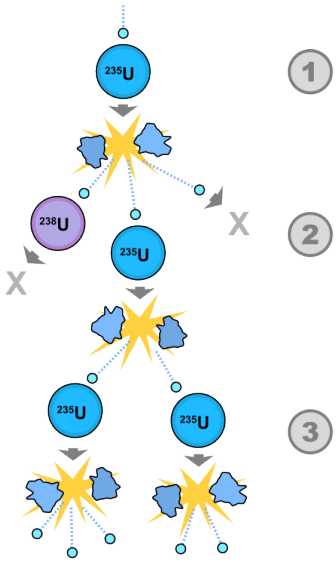


Figure 34.4 – In a chain reaction, enough neutrons are generated from each fission process to create a self-sustatining fission reaction. Fission chain reaction, by Fastfission, is licensed under Public Domain.

Although two or three neutrons are liberated in each fission, it is not necessarily true that each single fission process will produce two or three more. It's possible some of the excess neutrons may escape the material, or be absorbed by a different nucleus, such as those in a control rod, to keep a nuclear reaction from generating too much energy.

The number of reactions that are provoked by a single fission event relates to criticality. A subcritical process is one where, on average, less than one new fission reaction occurs after a single fission event. This would not lead to a self-sustaining nuclear reaction. Over time the reactions will stop on their own.

In a critical reaction, each fission process generates, on average, one additional fission event. A critical reaction is self-sustaining and will continue until all of the fissionable material is consumed. An analogy would be a line-up of dominoes. When the first domino is knocked over, it knocks over a second, which knocks over a third, and so on. This process continues until there are no more dominoes to knock over. This is shown in the video below.

In a supercritical reaction, each fission process generates, on average, more than one additional fission event. If one fission event provokes two more, those two will provoke four, which will provoke eight, and so on. A supercritical reaction grows exponentially. That means the number of fission events becomes extremely large extremely quickly.

A variable known as the effective neutron multiplication factor (which uses the symbol k) is used to represent the average number of neutrons that are generated in a fission reaction that go on to provoke new fission reactions. When k is less than one, a reaction is subcritical and will eventually die out. When k is equal to one, a reaction is critical (each fission event generates, on average, one neutron that will go on to initiate a new fission event). This is the desired situation in a nuclear reactor. When k is greater than one, a reaction is supercritical. Nuclear weapons are capable of releasing enormous amounts of energy because they are based on supercritical reactions.

If we look at a graph of the number of fission events that occur after each iteration (shown in Figure 34.5), we need to use a logarithmic scale in order to even compare critical and supercritical reactions. Assuming that one fission reaction started a chain reaction, after ten iterations, a critical reaction will have one fission event occurring; a supercritical reaction in which each iteration generates two new events (labeled k=2 in the graph) will have 512 fission events occurring; and a supercritical reaction that generates three new events with each iteration (labeled k=3 in the graph) will have nearly 20,000 fission events occurring!

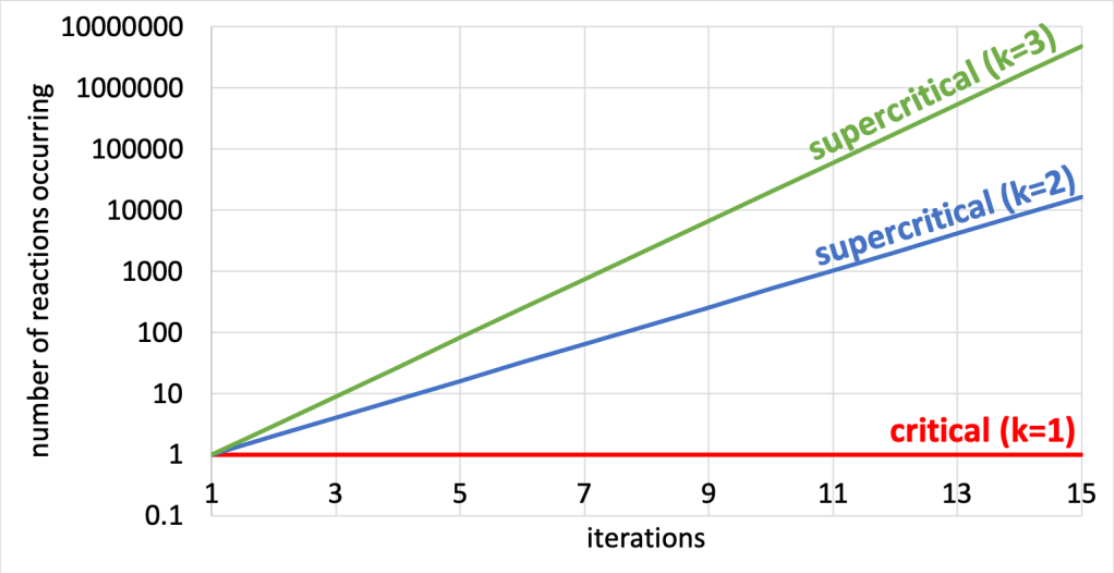


Figure 34.5 – A critical chain reaction will always have one reaction occurring at a time. Supercritical reactions create an exponentially increasing number of reactions over time. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Supercritical reactions are not tenable in nuclear power generation. To control the number of neutrons present in a chain reaction, control rods are used to absorb excess neutrons and keep a reaction at the critical stage. The control rod material (for example, cadmium or boron) is capable of absorbing neutrons without undergoing fission itself.

Critical mass

Because the liberated neutrons travel some average distance before initiating another fission event, there is a lower limit to how small a piece of fissionable material can be in order to maintain a chain reaction. A piece that is smaller than what is called the critical mass will allow liberated neutrons to reach the surface of the material and escape before they can trigger another fission event. (Recall the scaling effects of surface area and volume discussed earlier in this textbook: as the size of an object is scaled up, the volume grows faster than the surface area.) With a piece larger than the critical mass, the liberated neutrons will be able to initiate another fission event (due to the much greater abundance of nuclei in the volume) before they can escape to the surface.

This is the primary principle behind a nuclear fission bomb. In such a device two or more subcritical pieces (that is, pieces whose masses are less than the critical mass) are driven together using conventional explosives. When the pieces quickly come together, their combined masses form a single piece whose mass is well above the critical mass. The liberated neutrons can no longer escape to the surface of the now supercritical mass, causing the reaction to grow exponentially and resulting in a devastating explosion.

The critical mass depends on other factors such as shape and purity. The critical mass for pure uranium-235 is about 50 kg. In spherical form, this would be a ball roughly 17 cm in diameter.

Nuclear power

Natural sources of uranium consist primarily of the isotope uranium-238, which is fissionable, but not fissile. Samples of uranium ore will naturally contain some uranium-235, which is fissile. However, today less than 1% of natural uranium ore contains uranium-235. To sustain a critical chain reaction, a greater concentration of uranium-235 is required. Enrichment is a process by which the proportion of uranium-235 is increased.

Most nuclear power plants use enriched uranium (typically about 4% on average) as a fuel to create a critical chain reaction. A moderator is used to slow the neutrons released in a fission event until they travel at the right speed to cause fission in uranium-235. As mentioned earlier, control rods are used to absorb excess neutrons and prevent a super-critical reaction from occurring. Pressurized water runs through the reactor and heats up as it absorbs the energy generated by the fission reaction. In nuclear power plants in the United States, water also acts as the moderator.

The pressurized water from the reactor is sent through a heat exchanger, in which the water from the reactor transfers the heat to water in a separate line. The water in the second line boils, creating steam that spins a turbine, causing a relative rotation of a magnetic field and coil of wire. As previously discussed, this generates electricity. While the fuel is different in all turbine-powered generators, the actual mechanics of the power generation is the same for nuclear generators and fossil-fuel powered generators.

The major benefit of a nuclear reaction over fossil fuel is that there is no combustion during the actual fission process, and therefore there is no emission of greenhouse gases. In addition, uranium is a very energy dense material and per kilogram will generate much more energy than equivalent amounts of coal, oil, or gas.

Naturally, there are drawbacks to nuclear energy. While minimal, the nuclear reaction process does release small amounts of radioactivity into the environment. The uranium enrichment process can be used to create nuclear weapons. The byproducts of the fission process are themselves radioactive and have long half-lives. The safe storage of these radioactive wastes is an ongoing issue in the United States. Finally, there is always the risk of disaster in any nuclear power plant. Three Mile Island in the United States, Chernobyl in Ukraine, and Fukushima in Japan are all examples of what can happen if things go wrong. (A sidebar describing these three incidents is provided at the end of this chapter.)

Fusion

Fission is not the only way to generate energy from atomic nuclei. You may have noticed from Figure 34.2 that light elements contain a lot of mass per nucleon. If we smash hydrogen nuclei together, the result, helium, contains less mass per nucleon. This means we would generate energy in the process of binding them together. This smashing of nuclei together to form a heavier nucleus is known as nuclear fusion. All nuclei lighter than iron can release energy in the fusion process, just not with the same efficiency as the fusion of hydrogen.

Fusion is perhaps the most prevalent source of energy in our lives. That's because the Sun generates energy using fusion, and that is where the vast majority of our energy originates! Our Sun contains a massive amount of hydrogen and converts that hydrogen to helium in the fusion process. This process is depicted in Figure 34.6. Our Sun has been generating energy in this manner for 4.6 billion years, and thankfully for us has enough fuel to keep going for about 5 billion more.

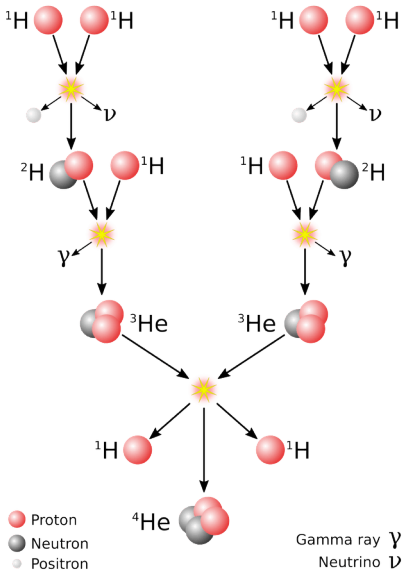


Figure 34.6 – The sun converts hydrogen into helium by fusing hydrogen nuclei together until a helium nucleus is generated. Fusion in the Sun, by Sarang, is licensed under Public Domain.

Fusion is an effective process for generating energy. Why don't we use it on Earth? The problem is that hydrogen nuclei will experience huge amounts of electromagnetic repulsion when we try to get them to come together. To overcome this, we need to get hydrogen nuclei to come together at extraordinarily high speeds. That means extremely high temperatures (such as those found in the core of our Sun and in other stars) need to be achieved in order to get the nuclei moving fast enough.

Prior to the development of fission bombs (described earlier in this chapter), attaining the extremely high temperatures necessary to overcome the electrostatic repulsion of the hydrogen nuclei and ignite fusion was not achievable. However, the energy released

in fission bombs can produce such temperatures. In 1952, the first fusion bomb was detonated in the Marshall Islands using atomic bombs as the “triggers” to detonate the fusion bomb.

While such devices demonstrate the destructive application of this knowledge, the goal of controlling fusion for use as a source of clean energy production is the other side of the coin. Harnessing the nuclear fusion of hydrogen as an energy source would be one of the pinnacles of human achievement. The fuel consists of the most abundant element in the universe (hydrogen) and the waste product is what is commonly used to fill balloons (helium). However, the quest towards achieving controlled fusion has been an arduous journey. The technological obstacles that need to be overcome are still formidable. Recent approaches focus on one of two basic schemes: magnetic confinement (in which the hot fusion plasma is held in a magnetic field in a toroidal [donut-shaped] chamber) and inertial containment (in which pellets of hydrogen fuel are bombarded by high-powered lasers from multiple directions to compress and heat the fuel) to initiate the fusion reaction.

While some recent preliminary evidence indicates that it may be possible to break even on the energy required to speed these hydrogen atoms up to the point of fusion, versus the amount of energy they produce, we are far from creating a sustained fusion reaction, let alone getting enough energy from the fusion process to use it to light up and heat our homes.

Energy in nuclear reactions

We have found that energy is released when splitting a very heavy nucleus into smaller ones (nuclear fission) and when combining very light nuclei into a heavy one (nuclear fusion). This energy comes from the conversion of mass energy according to Einstein’s famous equation,

$$E_{released} = |\Delta m|c^2,$$

where Δm is the difference in mass between the initial reactants and the final products ($m_{final} - m_{initial}$.) Plugging in this definition of Δm, the equation becomes

$$E_{released} = |(m_{final} - m_{initial})|c^2.$$

When solving this equation with numerical values, it’s helpful to use convenient units of measure (units that are neither extremely small nor extremely large). Just as we do not measure apples and baseballs in units of Earth masses, physicists generally do not measure atomic nuclei in kilograms. The kg unit is inconveniently large when discussing the mass of atomic-scale objects. For example, the mass of a proton is 1.673×10^{-27} kg. Similarly, the familiar energy unit, the joule, is also inconveniently large when dealing with atomic-scale objects. The mass energy of a proton is

$$\begin{aligned} E &= mc^2 \\ &= (1.67 \times 10^{-27}\ \text{kg})(3.00 \times 10^{8}\ \text{m/s})^2 \\ &= 1.51 \times 10^{-10}\ \text{J}.. \end{aligned}$$

To avoid cumbersome powers of ten, more convenient units are used. For mass, we use the unified atomic mass unit (abbreviated amu or often just u). The unified atomic mass unit is defined to be 1/12 the mass of a carbon-12 atom. (For comparison, $1\ \text{u} = 1.66054 \times 10^{-27}$ kg.) This results in protons having a mass of 1.007276 u. The mass of a neutron is 1.008665 u.

Similarly, the energy unit generally used in atomic and nuclear physics is the electron-volt (symbol: eV), defined as the amount of energy acquired by an electron when it is accelerated through a potential difference of one volt. Nuclear energies range in the millions of eV, so the convenient unit is the MeV where $1\ \text{MeV} = 10^6\ \text{eV}$.

1 u has a mass energy equal to 931.5 MeV. This is derived below, using the mass of 1 u in kilograms, the speed of light, and the conversion between joules and electron-volts.

$$\begin{aligned} E &= (1.66054 \times 10^{-27}\ \text{kg})(2.99792 \times 10^{8}\ \text{m/s})^2 \\ &= 1.49241 \times 10^{-10}\ \text{J} \\ &= \frac{1.49241 \times 10^{-10}\ \text{J}}{1.60218 \times 10^{-19}\ \text{J/eV}} \\ &= 9.315 \times 10^{8}\ \text{eV} \\ &= 931.5\ \text{MeV}. \end{aligned}$$

This provides a convenient conversion factor for the square of the speed of light: $c^2 = 931.5\ \text{MeV/u}$.

For example, a uranium-235 nucleus can absorb a neutron to become uranium-236, which then undergoes fission. One possible outcome of the splitting is krypton-92, barium-141, and three neutrons:

$$^{236}\mathrm{U} \rightarrow ^{92}\mathrm{Kr} + ^{141}\mathrm{Ba} + 3\mathrm{n}.$$

To calculate the energy released in this fission reaction, we must know the masses of each particle involved in the reaction. These are listed in the table below.

uranium-236	236.045568 u
krypton-92	91.926152 u
barium-141	140.91440 u
neutron	1.008665 u

The initial mass ($m_{initial}$) is 236.045568 u, which is the mass of the original nucleus. The final mass is the sum of the masses of each byproduct

$$\begin{aligned} m_{final} &= 91.926152\ \mathrm{u} + 140.91440\ \mathrm{u} + 3(1.008665\ \mathrm{u}) \\ &= 235.86655\ \mathrm{u}. \end{aligned}$$

Note that the final mass is less than the initial mass, resulting in a negative Δm value. The mass isn't truly "missing," but was instead converted into the energy that was released in the fission process. That released energy can be calculated as

$$\begin{aligned} E_{released} &= |(m_{final} - m_{initial})|c^2 \\ &= |235.86655\ \mathrm{u} - 236.045568\ \mathrm{u}|(931.5\ \mathrm{MeV/u}) \\ &= (0.179018\ \mathrm{u})(931.5\ \mathrm{MeV/u}) \\ &= 166.8\ \mathrm{MeV}. \end{aligned}$$

As a second example, consider the fusion reaction in which hydrogen-2 (also known as deuterium) combines with hydrogen-3 (also known as tritium) to form helium-4 and a single neutron:

$$^{2}\mathrm{H} + ^{3}\mathrm{H} \rightarrow ^{4}\mathrm{He} + \mathrm{n}.$$

The mass of each particle is given in the table below.

hydrogen-2	2.013553 u

hydrogen-3	3.015501 u
helium-4	4.001506 u
neutron	1.008665 u

The difference in mass between the initial reactants and final products can be calculated

$$\begin{aligned} m_{final} - m_{initial} &= (4.001506 \text{ u} + 1.008665 \text{ u}) - (2.013553 \text{ u} + 3.015501 \text{ u}) \\ &= -0.018883 \text{ u}. \end{aligned}$$

The energy released in this fusion reaction is

$$\begin{aligned} E_{released} &= |(m_{final} - m_{initial})|c^2 \\ &= |235.86655 \text{ u} - 236.045568 \text{ u}|(931.5 \text{ MeV/u}) \\ &= (0.179018 \text{ u})(931.5 \text{ MeV/u}) \\ &= 166.8 \text{ MeV}. \end{aligned}$$

In short, to determine whether energy will be released in a nuclear reaction, compare the total mass of the initial reactants to the total mass of the final products. If $m_{final} < m_{initial}$, then energy is released and the reaction can occur spontaneously. Using unified atomic mass units (u) for the masses, multiply the mass difference by 931.5 MeV/u to calculate the energy released in MeV.

Binding energy

By carefully measuring the mass of any atomic nucleus (except hydrogen-1), it is found that the composite nucleus contains less mass than the sum of the masses of the individual nucleons that comprise it. In other words, the composite nucleus is less than the sum of its parts! How can this be?

The fact that protons and neutrons are bound together in the nucleus implies that energy is required to disassemble a nucleus into the individual nucleons that comprise it. The amount of energy that some external agent must supply to accomplish this task shows up as extra mass found in the individual protons and neutrons. This energy, known as the nuclear binding energy, describes the energy required to dissociate the nucleus into its constituent nucleons.

To calculate the nuclear binding energy of a particular nucleus, the mass of the nucleus as well as the number of protons and neutrons in that nucleus are required. As an example, we can calculate the binding energy of a helium-4 nucleus. Helium-4 has a mass of 4.001506 u and consists of two protons and two neutrons. The mass of two individual protons and two individual neutrons is

$$\begin{aligned} 2m_p + 2m_n &= 2(1.007276 \text{ u})(1.008665 \text{ u}) \\ &= 4.031882 \text{ u}. \end{aligned}$$

Note that the four individual nucleons have more mass (0.030376 u more mass) than the helium-4 nucleus itself. Multiplying this mass difference by 931.5 MeV/u yields 28.30 MeV. This means that an external agent must supply 28.30 MeV of energy to dissociate the helium-4 nucleus into the two protons and two neutrons that formed it.

Of greater importance is the binding energy per nucleon. This is defined as the total nuclear binding energy divided by the total number of nucleons. The binding energy per nucleon in helium-4 is

$$\frac{28.30 \text{ MeV}}{4} = 7.075 \text{ MeV}.$$

For a uranium-236 nucleus (which consists of 92 protons and 144 neutrons), a similar calculation shows that the nucleus is lighter than the 236 individual nucleons by 1.870144 u. This gives a nuclear binding energy of 1742.039 MeV. Dividing this nuclear binding energy by the 236 nucleons gives 7.382 MeV per nucleon.

The isotopes with the highest nuclear binding energy per nucleon are nickel-62 and iron-56, both of which have a value of 8.79 MeV per nucleon.

Figure 34.7 shows the average binding energy per nucleon for common isotopes plotted as a function of the number of nucleons. This is consistent with the graph shown in Figure 34.2, which shows that iron-56 has the lowest mass per nucleon. (It should be pointed out that nickel-62 has a nuclear binding energy per nucleon nearly equal to that of iron-56.) It takes more energy to dissociate an iron-56 nucleus (or a nickel-62 nucleus) into its individual nucleons than any other nucleus. This larger energy difference leads to the larger mass difference and explains why the two graphs are the same shape, just upside down.

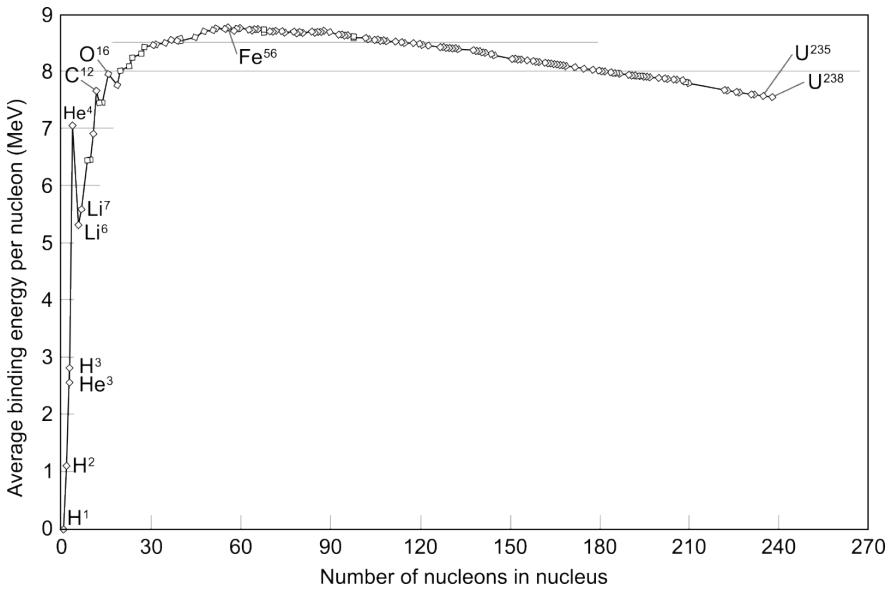


Figure 34.7 – A graph depicting the average binding energy per nucleon of common isotopes between hydrogen and uranium. Binding energy curve – common isotopes, by Fastfission, is licensed under Public Domain.

Isotopes that are lighter than iron-56 can cause a release of energy in the fusion process. Because the mass of individual nucleons is greater than the mass of these isotopes (except for hydrogen-1, which is composed of a single proton), the change in mass when binding the nucleons together into an isotope will be negative, leading to a release in energy based on $E = mc^2$. The fusion of isotopes with more nucleons than iron-56 will have a positive change in mass, which means that fusion of heavy elements can only occur if energy is input into the process.

On the right side of the graph, fission will cause a release of energy for isotopes that have more nucleons than iron-56 (and would require an input of energy for isotopes that have fewer nucleons than iron-56).

Stars shine by means of nuclear fusion in their cores. For relatively massive stars, the nuclear fuel is spent once the lighter nuclei have fused into iron. At that point, the star can no longer generate energy by nuclear fusion and the star collapses. The collapsing outer layers of the star bounce off the star's dense core. This rebound results in a spectacular explosion called a supernova. It is in these explosions that the elements with atomic numbers greater than that of iron are formed. The energy of the explosion is what creates the heavy elements. All of the elements that have atomic numbers greater than those of iron and nickel were formed in the supernova explosions of previous generations of stars. These materials are then incorporated into the next generation of stars and their planetary systems. We are literally made of stardust.

Nuclear Incidents

Three Mile Island

In 1979, a nuclear power station near Harrisburg, Pennsylvania, suffered a partial meltdown. The exact cause for this is complicated, as it was a series of many seemingly minor issues that led to the near-disaster.

First, an emergency water valve was left closed after maintenance had been performed on the power plant. This valve would allow water to flow into the cooling system that is part of the generator itself and that is part of the non-radioactive portion of the power plant (this water serves two purposes: it keeps the reactor cool and after heating up in this process, generates electricity by converting to steam and spinning a turbine). At the same time that this valve was closed, maintenance was being performed on the regular water lines feeding into the steam turbine. Minerals in hard water can cause plumbing to clog, and clogs of this nature were being flushed out. A clog made its way through a stuck valve and caused the water flow into the system to shut off. In an event such as this, the emergency water system should turn on to keep pressure and water flow levels normal; however, due to the valve that was closed during maintenance, this was unable to occur.

Because of the lack of cooling water reaching the reactor, pressure and temperature in the reactor portion of the power plant started to increase. This part of the plant has water in direct contact with the fuel. This water (and steam, which is also present in this type of reactor) becomes slightly radioactive. Fortunately, control rods were automatically inserted to shut down the reaction due to increases in temperature and pressure. However, a nuclear reactor does continue to generate heat even after being shut off. To cool this so-called decay heat, water is required. Because of rising pressure levels in the reactor, pressure relief valves opened to prevent a dangerous buildup of steam. Unfortunately, one of the relief valves did not close after relieving the pressure back to safe levels. To add to the issue, the indicator light did not actually indicate that the relief valve was open, so operators were unaware of this problem. This open valve caused cooling water to continue flowing out of the reactor.

Meanwhile, plant operators were concerned about keeping the correct ratio of water and steam in the reactor. Too much of either one could cause issues during normal operation of this type of power plant. Operators added some water to attempt to alleviate the issue of falling water levels, but were hesitant of accidentally adding too much. This would not have been a problem if water was not continuing to flow out of the system through the open relief valve. In addition, steam was now starting to form adjacent to the fuel itself, which was extremely dangerous as now the decay heat from the nuclear fuel could not be adequately cooled down. This caused a positive feedback loop (more steam means less cooling, which means there's more heat, which boils more of the remaining water into steam, which means less cooling, etc.).

Fortunately, the relief valve was eventually closed, plant personnel were evacuated, and water was eventually added back to the reactor, reducing steam levels. Unfortunately, a small hydrogen explosion occurred (which, while it did not cause a lot of damage, complicated clean-up efforts) and the contaminated water and steam that had vented through the relief valve caused the relief tank to overfill, and vented into the atmosphere.

While this event could have been much worse, it remains unclear if the Three Mile Island incident caused an increase in cancer rates in the surrounding area.

Chernobyl

The Chernobyl Nuclear Power Plant, situated north of Kiev in what is now Ukraine (but at the time was part of the Soviet Union), was a power plant built in the late 1970s. The type of reactor used in this power plant is known as an RBMK reactor, which among other design flaws does not include a containment structure around the reactor. This is due to the large size of the reactor; it would take an enormously large amount of space, lots of material, and lots of money to build a suitable containment environment.

In April 1986, operators were attempting to conduct a test of a possible safety feature. The idea was to determine if, while running down the reactor, the momentum of the steam turbine would be able to power the plant's safety features for the amount of time required to get backup diesel generators up and running. To conduct the test, the reactor needed to be powered down to a very low level (to simulate a shutdown). Operators were under a significant amount of pressure to complete the test quickly. At the same time, the test could not be conducted during the day while the day shift was operating the plant due to the fact that the power being generated from the plant was needed in the electrical system; reducing the output of the plant may have impacted the electrical needs of the surroundings. This delayed the test until closer to midnight.

The day shift initially was able to reduce power to about half of normal output. At the same time, they also shut down emergency cooling systems. While turning off the emergency cooling system had no direct effect on what was to come later, it was a sign of the lax safety culture in existence in the power plant to turn off (and keep off) an emergency backup system.

When the plant operators were finally allowed to reduce power, power levels were reduced to lower levels than anticipated. This caused an issue known as “xenon poisoning” where lots of xenon-135, a byproduct of the fission process, is generated. Because the power levels were low, this xenon-135 wasn’t being bombarded by as many neutrons to convert it to xenon-136. Xenon-135 is a great absorber of neutrons (xenon-136 is not), and acts to slow the reaction further. Because xenon-135 has a relatively long half-life, it takes a long time for the effects of xenon poisoning to stop. (If the fission reaction is occurring at high enough levels, the abundance of neutrons converts the xenon-135 to xenon-136, causing this to be a non-issue.) Operators removed many of the control rods in the reactor to make up for this xenon poisoning and to keep the reaction critical.

Meanwhile, a lot of cooling water was flowing into the reactor, more so than was needed for the lower power output of the plant. This caused low levels of steam and pressure in the reactor, and also acted as a further neutron absorber (water is a good absorber of neutrons). More control rods were removed to keep the reaction critical as a response to the excess water levels. Some of the cooling water was being pumped in by pumps connected to the steam generator as part of the safety test. (The question was: if the reactor shuts down, will the momentum of the steam generator produce enough power to keep cooling water flowing before the diesel backups start powering the pumps?)

When the test finally started, the valve to release steam into the turbine and generator was closed. This caused steam pressure to build up in the reactor (if it could not go into the turbine, it instead would stay in the reactor area). Meanwhile, the cooling water pumps powered by the generator (which was no longer producing much, if any, power, as expected) stopped flowing. This caused an increase in temperatures, causing much of the cooling water in the reactor to convert to steam. Steam does not absorb neutrons as well as liquid water does due to its low density.

At this point, with most of the control rods having been removed, and water converting to steam, there was now an increase in neutrons and the reaction levels increased. This generated more heat and steam, which led to more neutrons and a higher reaction level. This process continued in a positive feedback loop.

An emergency shutdown button was activated, which caused control rods to be inserted back into the reactor. Unfortunately, the design of the RBMK reactor control rods contained a huge flaw: the ends of the control rods that would move first into the reactor as they were lowered down were graphite, a moderator. Instead of slowing down the reaction, this caused an increase in the reaction!

Eventually, the pressure in the reactor became so great that there was a steam explosion. Because there was no containment structure, radioactive debris and fuel were ejected all around the plant. This caused many fires and severed coolant lines. (Although at this point cooling water would not have stopped the damage.) The fire spread radioactive particles throughout not only the plant, but also throughout the USSR and Europe. Many plant operators, firefighters, and rescue workers were killed either directly by the explosion, or later by their exposure to extreme levels of radioactive nuclei.

A steel and concrete structure known as a “sarcophagus” was built to reduce the spread of further contamination, but about 2,600 km^2 of the surrounding area has been declared an “exclusion zone” to prevent human inhabitants from residing in an extremely radioactive environment.

Fukushima

The Fukushima nuclear accident occurred on March 11, 2011 as a result of the Great East Japan Earthquake. While the earthquake and subsequent tsunami caused significant damage to the nuclear power plant, the design of the plant itself contributed greatly to the severity of the accident.

The Fukushima Daiichi Nuclear Power Plant (the first of two nuclear power plants in Fukushima) was built along the coastline of Japan. The proximity to the Pacific Ocean ensured that the reactor would have access to lots of cooling water. When construction began in the late 1960s, a cliff that was originally present on the coastline at the site was removed. This allowed the power plant to be more easily anchored to bedrock (an important consideration in an earthquake-prone area), and also enabled easier pumping of cooling ocean water into the plant. (If the plant had been built on top of a cliff, pumps would have to overcome gravity to pump the water significantly above sea level. This would essentially make the power plant less efficient, as some of the power generated by the plant would be spent pumping water instead of going directly to the power grid.)

Backup diesel generators (used in case of a power outage to keep backup and safety systems online) were installed in the lowest level of the plant. This may have been an OK decision for locations significantly above sea level, but was a very problematic decision in the construction of the Fukushima Daiichi plant.

In general, the power plant was designed to withstand probable, predictable geological events, not a worst-case scenario. There is always a tradeoff between extremely safe design (designing for a worst-case scenario) and keeping a design cost- and power-efficient (designing for probable scenarios). Unfortunately, the design of this reactor would contribute to its downfall.

On March 11, 2011 the Great East Japan Earthquake registered at a 9.0 on the Richter scale. This earthquake knocked out power to the reactor. Thankfully, the diesel backup generators came online to power safety systems. All of the reactors at the plant were automatically shut down. However, fission reactors do continue to produce heat (known as decay heat because it is generated by byproducts of radioactive decay) even after the fusion reaction is shut down.

A few minutes after the earthquake hit the coast, a 13-meter-tall tsunami hit the power plant. The plant had been modified at that point to withstand 5.7 meters, which was insufficient to protect from the effect of the gigantic waves from the tsunami. The plant was flooded, cooling lines were severed, and the diesel generators were flooded out. Now all of the safety systems went offline. Fresh cooling water could not be pumped into the reactors, and the cooling water that was already in the reactors could not be pumped out after heating up. The core temperatures of many units became so high that the fuel melted through the containment structures (this is known as a meltdown). In addition, hydrogen gases built up and caused explosions.

Radioactive byproducts were released into the surrounding areas, and many thousands of people living near the power plant were evacuated. More recently (2023), there has been public outcry over the release of radioactive cooling water back into the Pacific Ocean.

Further reading

- Chernobyl exclusion zone. (2024, March 2). In *Wikipedia*. en.wikipedia.org/w/index.php?title=Chernobyl_exclusion_zone&oldid=1211408381
- Distinction between Fissionable, Fissile and Fertile | nuclear-power.com. (n.d.). Nuclear Power. https://www.nuclear-power.com/glossary/distinction-between-fissionable-fissile-and-fertile/
- Filburn, T., & Bullard, S. (2016). *Three Mile Island, Chernobyl and Fukushima: Curse of the Nuclear Genie*. Cham Springer International Publishing.
- Fukushima nuclear accident. (2024, March 21). In *Wikipedia*. en.wikipedia.org/w/index.php?title=Fukushima_nuclear_accident&oldid=1214772218
- Three Mile Island accident. (2024, March 20). In *Wikipedia*. en.wikipedia.org/w/index.php?title=Three_Mile_Island_accident&oldid=1214713366
- United States Nuclear Regulatory Commission: Backgrounder on NRC Response to Lessons Learned from Fukushima. (n.d.). Www.nrc.gov. https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/japan-events.html
- United States Nuclear Regulatory Commission. (2018, June 21). NRC: Backgrounder on the Three Mile Island Accident. Nrc.gov; United States Nuclear Regulatory Commission. https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html
- United States Nuclear Regulatory Commission. (2022, March 1). NRC: Backgrounder on Chernobyl Nuclear Power Plant Accident. Nrc.gov; United States Nuclear Regulatory Commission. https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/chernobyl-bg.html

Further reading

- United States Nuclear Regulatory Commission – The US NRC is tasked with regulating the safety of nuclear power plants in the United States. Their website contains a lot of information about nuclear power and related issues.
- What are Small Modular Reactors (SMRs) – This website explains the physics behind small modular reactors, which can produce a lot of energy in a relatively small size.
- Advanced reactor systems – This website explains some new nuclear reactor technologies (known as Gen 4) that may be able to fuel our power needs into the next few decades.

Practice questions

Conceptual comprehension

1. In a commercial nuclear fission reactor that generates electricity, what value for the neutron multiplication number, k, is most desirable? Why?
2. A uranium-235 nucleus absorbs a neutron to become uranium-236. One possible outcome of the resulting fission reaction is cesium-137, rubidium-97, and some number of individual neutrons. How many neutrons would be generated in this reaction?
3. If instead the uranium-236 in the previous question fissions into tellurium-137, an isotope of zirconium, and two neutrons, what is the atomic mass number of the zirconium isotope involved in this reaction?
4. How is it possible that energy can be released by splitting heavy nuclei into lighter ones and by fusing light nuclei into heavier ones?

Numerical analysis

1. Consider a supercritical chain reaction in which each fission event induces two successive events. (That is, the neutron multiplication number, k, is equal to two.) How many fission events will take place during the 6th cycle? After the 6th cycle, how many total fissions will have taken place since the start of the reaction?
2. In a critical fission reaction (k=1), each fission event goes on to initiate only one successive event. If each successive event takes place with an average time interval of 10 nanoseconds, how long will it take for 1023 fission events to occur?
3. If instead, the reaction described in the previous questions was supercritical with a neutron multiplication number k=2, how long would it take for 1023 fission events to occur? How many times faster is this reaction compared to the critical reaction? (Hint: First determine how many cycles occur for a total of 1023 fission events.)
4. The power output of the Sun is $3.86 \times 10^{26}\ \mathrm{W}$. Calculate the rate at which the sun is converting mass into energy. Give your answer in kg/s. (Hint: Recall that 1 W = 1 J/s. Then use Einstein's famous equation.)
5. How much energy is released when a plutonium-244 nucleus alpha decays into neptunium-240? The required masses are given in the table below.

plutonium-244	244.064204 u
neptunium-240	240.056162 u
helium-4	4.001506 u

6. Consider the following fission reaction: $\mathrm{n} + {}^{235}\mathrm{U} \rightarrow {}^{144}\mathrm{Xe} + {}^{90}\mathrm{Sr} + 2\mathrm{n}$. How much energy is released in this reaction? The required masses are given in the table below.

neutron	1.008665 u

uranium-235	235.043930 u
xenon-144	143.93851 u
strontium-90	89.907738 u

7. Consider the nuclear reactions listed below. How much energy is released in each reaction? The mass of helium-3 is 3.016029 u. Refer to earlier tables for the masses of the other particles.
   1. $^{2}\mathrm{H} + ^{2}\mathrm{H} \rightarrow ^{3}\mathrm{H} + ^{1}\mathrm{H}$
   2. $^{3}\mathrm{He} + ^{3}\mathrm{He} \rightarrow ^{4}\mathrm{He} + ^{1}\mathrm{H} + ^{1}\mathrm{H}$
8. Calculate how much energy is needed to disassemble a nickel-62 nucleus into its individual protons and neutrons.
9. Calculate the average mass per nucleon of a carbon-12 nucleus.
10. Calculate the energy required to dissociate 12 micrograms of carbon-12 into individual protons and neutrons. If the burning of one gallon of gasoline releases 140 MJ of energy, calculate how many gallons of gasoline are potentially needed to dissociate those 12 micrograms of carbon-12 atoms.

Hands-on experiments

1. If you have a set of dominoes, carefully arrange them so that one domino will knock over a second, which will knock over a third, and so on. This will represent a critical chain reaction. If you have a stopwatch, knock over the first domino while pressing the start button on your stopwatch. Stop the stopwatch when the last domino falls. How long did the reaction take? Divide by the number of dominoes minus one (the first knock that you initiated) to determine how long it took for each individual reaction event to occur. If you can, try setting up the dominoes such that one will knock over two, and each of those dominoes knock over two more. This represents a supercritical (k=2) chain reaction. How long did that reaction take to knock over the same number of dominoes? Did you encounter any difficulties in setting up the supercritical reaction?

This page titled Chapter 34: Fission and fusion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

APPENDICES

Appendix A: Vectors and vector arithmetic
Appendix B: Units and unit conversions
Appendix C: Answers to selected questions

This page titled Appendices is shared under a license and was authored, remixed, and/or curated by LibreTexts.

Appendix A: Vectors and vector arithmetic

A vector is a quantity that requires two or more numbers (with units) to be fully described and that obeys certain rules of combination. For our purposes, it will be sufficient to say that a vector quantity is one that requires both a magnitude (strength) and a direction for a complete description. As we saw in our discussion of forces in this textbook, not only can different strengths of forces be applied to an object, they can be applied in different directions.

A quantity that does not contain direction is said to be a scalar quantity. Mass is a scalar. It has a magnitude but no direction. It wouldn't make any sense for a box to have a mass of 20 kg down, or up, or left, or right. The box would just have 20 kg of mass.

Vector quantities, in this textbook, will be denoted using an arrow over the symbol. For example: $\vec{F}$. Other textbooks may use a bold font to denote a vector quantity.

Symbolic representation of vectors

$\vec{F}$

$\boldsymbol{F}$

Drawing vectors

A vector is denoted graphically by drawing it as an arrow. The length of the arrow tells us about the magnitude of the vector. The positioning of the arrowhead tells us the direction of the vector. For example, you can draw an arrow to demonstrate a vector with a magnitude of 10 units pointing down by a certain length. A vector with a magnitude of 20 units pointing down would be twice as long.

Because a vector can be used to represent any direction in three-dimensional space, it can be tricky to discuss the direction of a vector. In this textbook, we'll mostly limit the discussion of vectors to two-dimensional space. In two dimensions, a vector can be pointed in any of 360 degrees.

Vector direction

One way that we can discuss the direction of a vector is used for vectors that point completely vertically or completely horizontally. For a vertically directed vector, a plus sign means that a vector points up, and a minus sign means that a vector points down. For a horizontal vector, a plus sign means that a vector points to the right, and a minus sign means that a vector points to the left. This is the same as the signs used in the Cartesian coordinate system.

Alternatively, we could use a cardinal direction such as north, south, east, or west to define the direction of a vector. This is not always relevant, but can be a nice way to talk about direction when discussing changes in location, for example.

Finally, we could use angle relative to a reference direction to discuss direction. It is customary to use the positive x-axis as the reference direction. This is how direction is used for vectors in many physics classes. Although we will discuss angles at times, we will not be making use of trigonometry in this textbook.

up	north	90 degrees

down	south	270 degrees
left	west	180 degrees
right	east	0 degrees

Vector components and magnitude

Any vector, regardless of its direction, can be broken down into components. Those components can be used to calculate the magnitude (length) of the vector. The magnitude of a vector will always be a positive number (length cannot be negative).

The Pythagorean theorem states that

$$a^2 + b^2 = c^2,$$

where a and b are the lengths of two sides of a right triangle, and c is the hypotenuse of that triangle, as shown in Figure A.1.

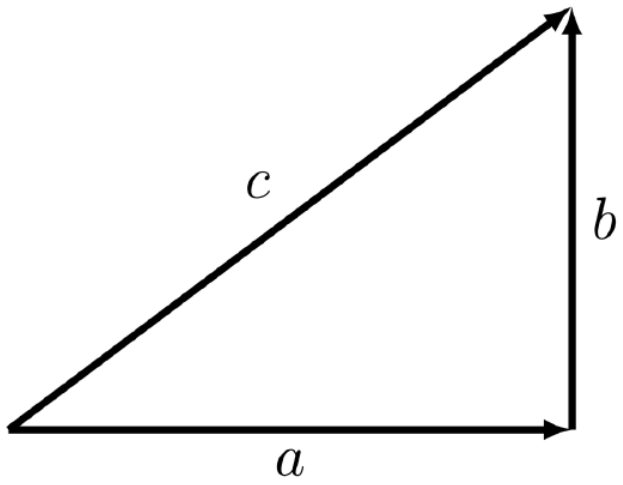


Figure A.1 – The Pythagorean theorem can be used to calculate the length of the hypotenuse of a right triangle. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

By splitting a vector into horizontal and vertical components, we have found that right triangle. Once both the horizontal and vertical coordinates of a vector are known, the Pythagorean theorem is used to calculate the magnitude of the vector, using the equation

$$\Delta\vec{x}^2 + \Delta\vec{y}^2 = l^2,$$

where $\Delta\vec{x}$ is the horizontal component of the vector, $\Delta\vec{y}$ is the vertical component of the vector, and l is the magnitude of the vector.

On the other hand, if the magnitude and direction of a vector are known, the horizontal and vertical components can be found graphically by projecting the shadow of the vector on the x-axis and y-axis, respectively (as shown in Figure A.2). To do this

mathematically generally requires the use of trigonometry and thus will not be discussed further.

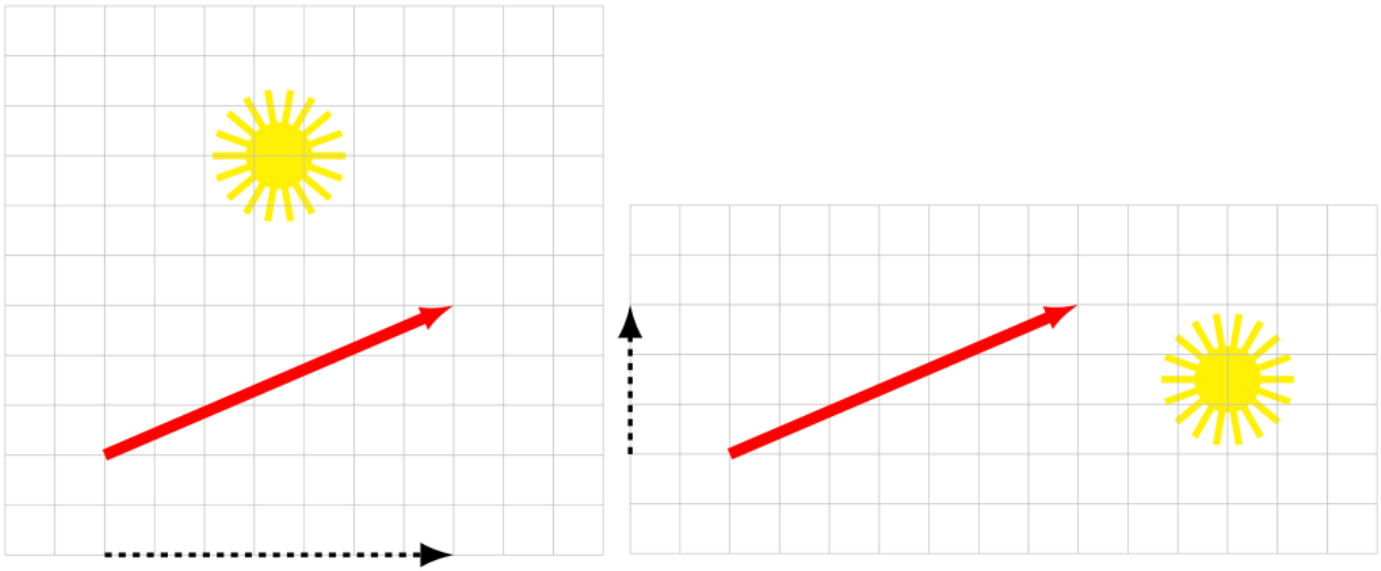


Figure A.2 – To determine the components of a vector, the shadow cast on the x-axis (left) or y-axis (right) can be used. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Horizontal vector

The vector shown in Figure A.3 has a starting position of (4,0) and an ending position of (0,0). The arrow will point from the start to the end. This is a horizontal line pointing toward the left. Because the vector is purely horizontal, there is no vertical component associated with it.

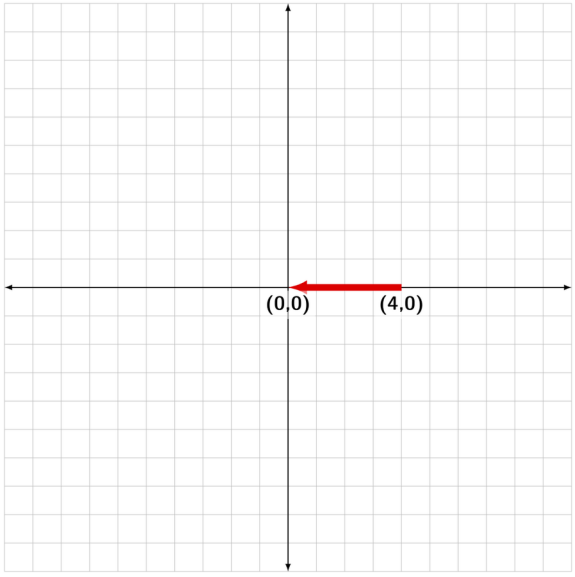


Figure A.3 – A horizontal vector with starting coordinate of (4,0) and ending coordinate of (0,0) drawn on a Cartesian axis. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The horizontal component of the vector is equal to the final x position of the vector minus the initial x position of the vector:

$$\Delta\vec{x} = x_{final} - x_{initial}.$$

The vector in Figure A.3 has an x component of

$$\Delta\vec{x} = 0 - 4 = -4.$$

The negative sign of the component indicates that the direction of the vector points to the left (which is defined as negative using the Cartesian coordinate system). The magnitude (length) of the vector is four.

Vertical vector

The vector shown in Figure A.4 has a starting position of (-3,-2) and an ending position of (-3,5). The arrow points from start to finish. This is a vertical line pointing upward. This vector is completely vertical and has no horizontal component to it.

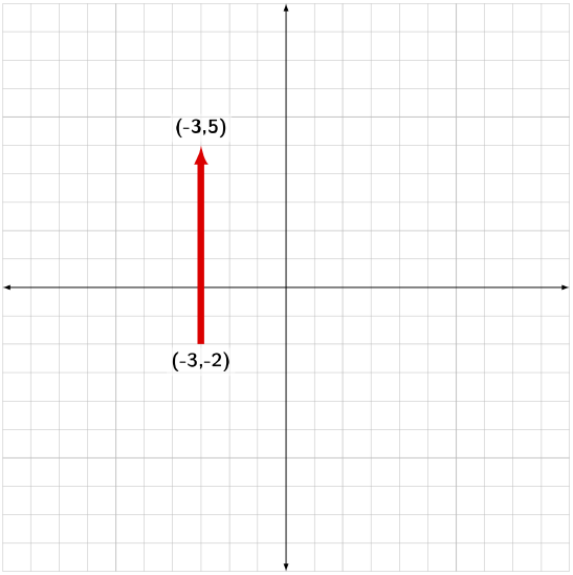


Figure A.4 – A vertical vector with starting coordinate of (-3,-2) and ending coordinate of (-3,5) drawn on a Cartesian axis. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The vertical component of the vector is equal to the final y position of the vector minus the initial y position of the vector

$$\Delta\vec{y} = y_{final} - y_{initial}.$$

The vector in Figure A.4 has a y component of

$$\Delta\vec{y} = 5 - (-2) = 7.$$

The y component of the vector is therefore equal to positive seven. The positive sign indicates that it points up. The magnitude of the vector is seven, which describes the vector's length.

Two-dimensional vector

The next example, shown in Figure A.5, has a starting location of (-4,-1) and an ending location of (6,8). The arrow points from start to finish. This vector has both horizontal and vertical components.

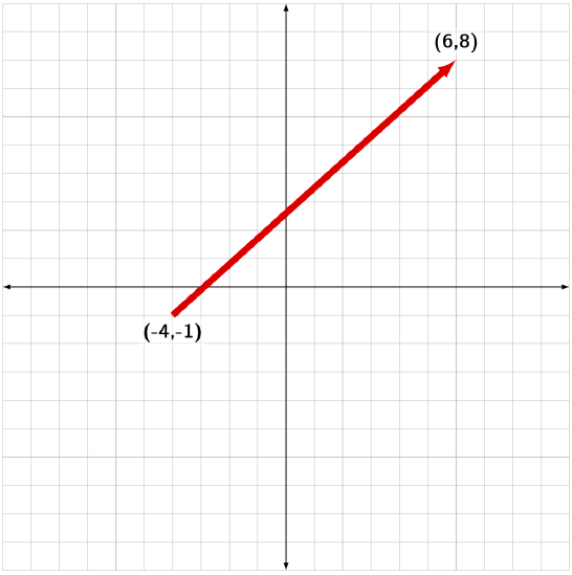


Figure A.5 – A two-dimensional vector with starting coordinate of (-4,-1) and ending coordinate of (6,8) drawn on a Cartesian axis. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The horizontal component defines the horizontal motion of this vector, ignoring the vertical parts. The horizontal component is

$$\Delta\vec{x} = 6 - (-4) = 10.$$

The vertical component defines the vertical motion of the vector, ignoring the horizontal parts. The vertical component is

$$\Delta\vec{y} = 8 - (-1) = 9.$$

Now that we know the value of each component, we can use the Pythagorean theorem to calculate the magnitude of the vector itself. In this example, the magnitude is

$$\sqrt{10^2 + 9^2} = 13.45.$$

Adding vectors

What if we want to add two or more vectors together? Sometimes we will need to do this, especially when calculating the sum of all forces acting on an object to determine whether or not it is in equilibrium. There are two methods for adding vectors together: the graphical method and the algebraic method. Either one can be used. Both will give the same result. Try both of them and use the one that feels more comfortable to you.

Graphical method

To add vectors graphically, draw one of the vectors (typically starting at the origin). Then, at the tip of that vector, draw the second vector. If there are more than two vectors, each will be connected together tail to tip, tail to tip. The order in which you draw the vectors does not matter. The sum of the vectors is the straight line from the start of the first vector to the end of the last vector.

Consider two vectors with components of (4,9) and (-6,-5). Each individual vector is drawn tail to tip, and the overall sum vector is drawn from the origin to the end of the last vector. This is represented in Figure A.6.

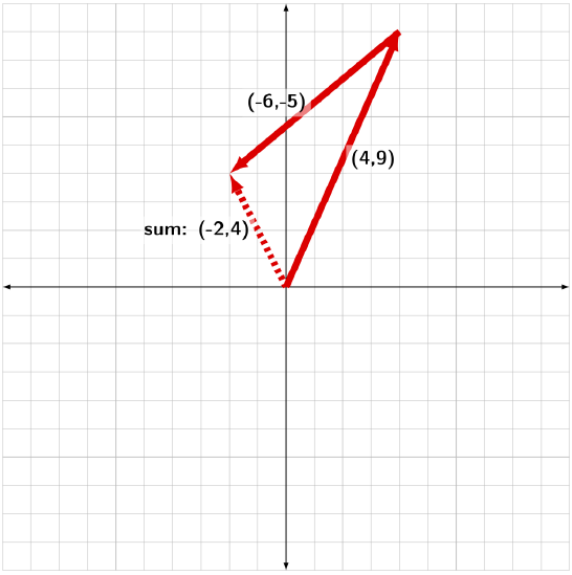


Figure A.6 – Two vectors (4,9) and (-6, -5) are added using the graphical method. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

In this example, the sum vector has a horizontal component of negative two and a vertical component of positive four: (-2, 4). The Pythagorean theorem can be used to calculate the magnitude of the sum of the vectors:

$$\sqrt{(-2)^2 + (4)^2} = 4.47.$$

Algebraic method

Alternatively, we can add the two vectors together algebraically. To add the vectors algebraically, write out each vector as a set of components. The first vector is (4,9), and the second vector is (-6,-5). Add all of the x components together.

$$\begin{aligned} x &= 4 + (-6) \\ &= -2 \end{aligned}$$

Then add all of the y components together.

$$\begin{aligned} y &= 9 + (-5) \\ &= 4 \end{aligned}$$

Therefore, the sum of the vectors is (-2, 4). Note that we should, and did, get the same result using both methods (graphical and algebraic). If there are more than two vectors, simply add all of the x components together to find the x component of the sum. Then add all of the y components together to find the y component of the sum.

Vector properties

There are certain important properties that apply when doing arithmetic with vector quantities.

Commutativity of addition

When adding vectors, addition is a commutative property. That is,

$$\vec{a} + \vec{b} = \vec{b} + \vec{a}.$$

The order in which we add vectors does not matter! Consider two vectors, $\vec{a} = (3, -5)$ and $\vec{b} = (-10, 8)$. Figure A.7 demonstrates the addition of $\vec{a} + \vec{b}$ which has a sum of (-7,3).

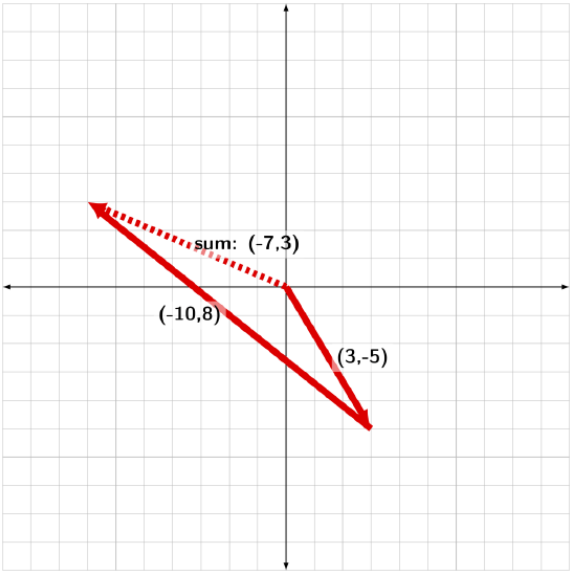


Figure A.7 – Vector A (3,-5) is added to vector B (-10,8) using the graphical method. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Figure A.8 demonstrates the addition of $\vec{b} + \vec{a}$ which also has a sum of (-7,3). This shows that the order of addition does not matter. This is not only true for two vectors: when adding any number of vectors together, the order does not matter!

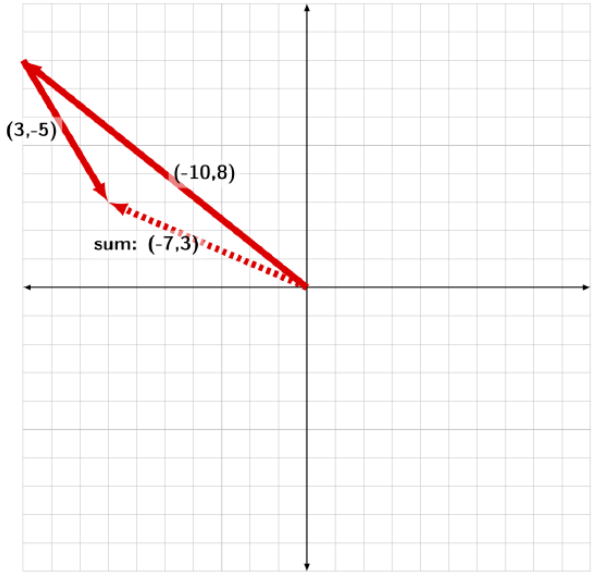


Figure A.8 – Vector B (-10,8) is added to vector A (3,-5) using the graphical method. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Associativity of addition

Vectors obey the associative property for addition. That is,

$$(\vec{A} + \vec{B}) + \vec{C} = \vec{A} + (\vec{B} + \vec{C}).$$

This property is demonstrated in Figure A.9, where vectors $\vec{A}$ (5,5), $\vec{B}$ (3,-2), and $\vec{C}$ (0,-6) are added in both orders with the same result (8,-3).

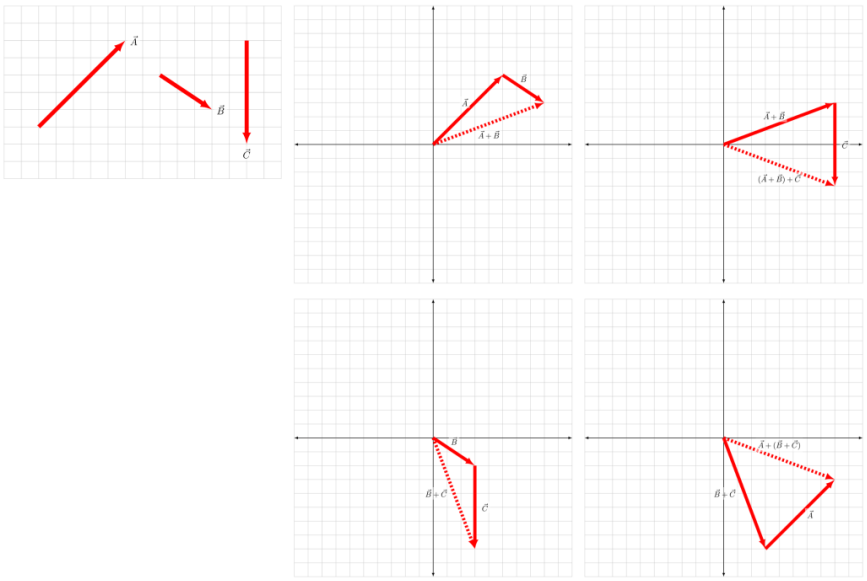


Figure A.9 – The associative law states that (A+B)+C = A+(B+C), which is verified in this image. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Scalar multiplication

It is possible to multiply a vector by a scalar. That is,

$$c(x, y) = (cx, cy).$$

Consider the example of $\vec{a} = (-4, -1)$. When multiplied by (for example) positive two, the result is $2\vec{a} = (-8, -2)$. It is also possible to multiply the vector by (for example) negative one half. That result is $-\frac{1}{2}\vec{a} = (2, 0.5)$. These results are demonstrated graphically in Figure A.10.

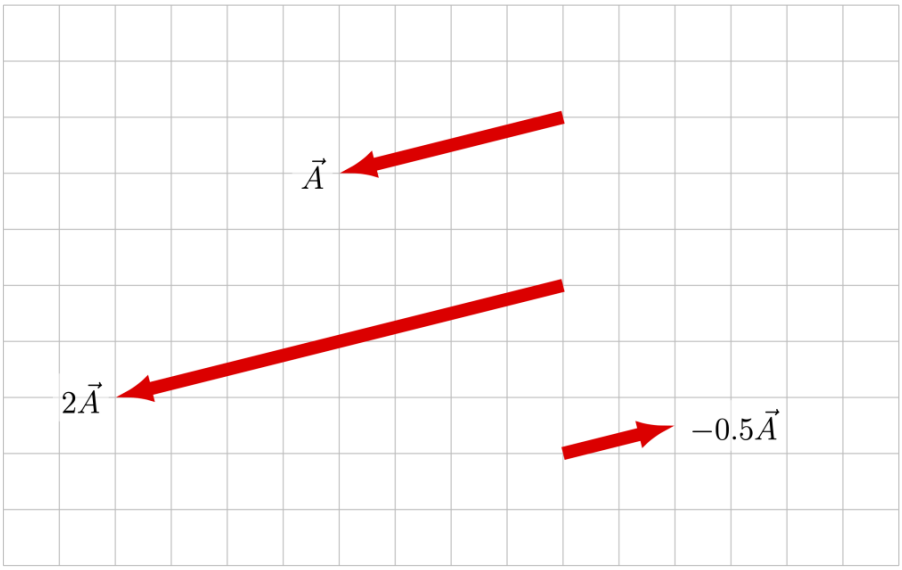


Figure A.10 – A vector can be multiplied by any scalar quantity. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Because it is possible to multiply a vector by a negative scalar value, we can think of subtraction as addition performed after scalar multiplication. In other words,

$$\vec{a} - \vec{b} = \vec{a} + (-1)\vec{b}.$$

Further reading

- PhET vector addition simulation – Use this simulation to practice adding vectors together.
- Vector addition: does order matter – Play with this simulation to see for yourself that the order in which you add vectors does not matter.

Practice questions

Conceptual comprehension

1. Recall that a vector quantity is one in which two or more numbers with units are needed for a complete description. For example, a complete description of a location on a two-dimensional grid can be provided by the x-component and the y-component of the vector. What are two different numbers that could be used to completely describe the same vector?

Numerical analysis

1. Vector $\vec{A} = (3, 2)$. Calculate the magnitude of this vector.
2. Vector $\vec{B} = (-1, 4)$. Calculate the magnitude of this vector.
3. Vector $\vec{A} = (3, -2)$, and vector $\vec{B} = (-1, 4)$. Calculate the components of the resultant vector $\vec{C} = \vec{A} + \vec{B}$. Then, calculate the magnitude of the resultant vector $\vec{C}$.
4. Vector $\vec{X} = (5, 7)$. Calculate the magnitude of this vector.
5. Vector $\vec{Y} = (2, -3)$. Calculate the magnitude of this vector.
6. Vector $\vec{X} = (5, 7)$, and vector $\vec{Y} = (2, -3)$. Calculate the components of vector $\vec{Z} = \vec{X} - \vec{Y}$. Then, calculate the magnitude of vector $\vec{Z}$.

Hands-on experiments

1. If you have access to safe outdoor space, walk, run, or roll to a particular spot (feel free to mark it with an object) that you will call the origin. Then move 5-10 meters (the exact distance does not matter as long as you try to be consistent) east (or west), followed by moving 5-10 meters north (or south). Consider your new location with respect to the origin. What is your displacement? If you start by moving north/south, and then east/west, does that affect your displacement?
2. If you regularly commute to work, school, or another common location, look at a map and consider all of the different routes you can take between your starting and ending locations. Which has the shortest path? Which has the longest path? Does the choice of path affect the displacement? Along your commute, consider your instantaneous speed and velocity. After your commute, consider your average speed and velocity.

This page titled Appendix A: Vectors and vector arithmetic is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Appendix B: Units and unit conversions

When describing a physical property, a unit of measure is typically required. A unit gives a sense of perspective or scale relating to the magnitude of the value. To give an example: say you want to determine what to wear outside. You ask for the temperature and are told that the temperature is 10. Is that hot, cold, or somewhere in between? Without a unit to tell you the value of that number, it is meaningless. It is likely to be either 10 Celsius or 10 Fahrenheit (the most common units used to describe temperature), but those two values would require very different wardrobes when going outside!

In fact, many scientific units are given precise definitions that have to do with fundamental properties of our universe. This is to remove any extraneous errors that could result due to differences in definitions or in the measurement techniques used to determine the exact "value" of each unit. We could also consider that our definitions of units could be used by people (or hypothetical aliens from other planets) who don't know what a term such as "gram" or "kilometer" means, but could deduce those values from an understanding of the fundamental properties of the universe.

It is important to note that two quantities with different units cannot be combined together using addition or subtraction. (A common phrase used by physics professors is that "you can't add apples to oranges," as both quantities are fundamentally different.) For example, it would be meaningless to add 5 meters to 8 kilograms.

It is possible to add (or subtract) two quantities that measure the same property (such as length), however, the units have to be the same. For example, adding 5 meters to 8 kilometers requires that one of the two units be converted to the other. (Either convert 5 meters to kilometers, and add to 8, or convert 8 kilometers to meters, and add to 5.)

Unit conventions

Historically, many different units have been developed to define quantities such as length, weight (force), and time. Today, while there are regional differences in the units used for many properties (speed, length, and mass), there are other properties that have units which are relatively standard across the globe. For example, modern units used with electricity such as the ampere, coulomb, and volt are part of the SI system of units and are used internationally, even in the United States.

United States customary units

The United States is one of a very small number of countries on the planet that does not primarily use a metric system of units. Many of the units used in the US are derived from British imperial units (the two unit systems are very similar although there are notable differences, including the definition of a fluid ounce used to measure volume).

Commonly used US customary units include:

- inch, foot, and mile (distance);
- teaspoon, tablespoon, cup, pint, quart, and gallon (fluid volume);
- ounce, pound, and ton (weight); and
- degree Fahrenheit (temperature).

Metric system

A metric system is simply a system of units that uses powers of ten to define units. United States customary units are not considered metric. This is because, for example, there are 12 inches in a foot, 3 feet in a yard, and 1760 yards in a mile. None of these values are powers of ten. However, there are 1000 millimeters in a meter, 1000 meters in a kilometer, and so on. These are all based on powers of ten. The SI system of units (described below) is considered a metric system, as it has units that use powers of ten.

The international standard of units (SI)

The international standard of units (SI) is the modern system of units, and is considered a metric system as it is composed of

units that are based on powers of ten. This system of units uses base units (described below) that are explained by physical properties of our universe (such as the speed of light in a vacuum, a value that does not change regardless of one's position or motion in the universe) to derive units for other physical properties. The SI system of units is the established standard used in scientific fields (even in the United States).

SI base units

There are seven units in the SI system that can be used to derive every other unit in the system. These seven units are known as SI base units. The SI base units are defined in the table below.

s	second	time
m	meter	length
kg	kilogram	mass
A	ampere	electrical current
K	kelvin	temperature
mol	mole	quantity of a substance
cd	candela	luminous intensity

Each of these base units has a scientific explanation that uses fundamental properties of our universe to define it. For example, the base unit of second is defined with respect to the frequency of light emitted in a particular electron transition in a cesium-133 atom. Regardless of the person who measures that frequency, an identical description of the unit of second should result.

Historically, many units were justified by a physical item. For example, there was an object formulated to have a mass of exactly one kilogram, used as a reference to calibrate precise scientific measurement tools. "The" kilogram (along with several reference copies) was preserved in a vacuum and only cleaned using a carefully defined procedure to avoid either adding or sloughing off atoms and changing the mass of the reference. (As the unit of kilogram was defined by this reference kilogram, any change in its mass would have changed the very definition of what a kilogram was!) However, because a reference item is inherently prone to changing over time, and because it must be so carefully preserved and maintained, it is less useful than a scientific justification of the unit of mass. In 2019, a scientific justification of the unit of kilogram was formulated based on the value of Planck's constant.

SI derived units

All other units can be derived from base units and are therefore called derived units. For example, the unit used for velocity (meters per second) is derived from the base units of meter and second: m/s.

The joule is a unit of energy and work, and is derived from the base units of kilograms, meters, and seconds: $kg \cdot m^2/s^2$. The coulomb is a unit of electrical charge, and is derived from the base units of ampere and second: A·s (second times ampere).

Quantities that are derived from derived units can also have their base units determined. For example, a volt is defined as a joule divided by a coulomb. Because we know the base units of joule and coulomb, we can derive the base units of a volt:

$$\begin{aligned} \mathrm{V} &= \frac{\mathrm{J}}{\mathrm{C}} \\ &= \frac{\frac{\mathrm{kgm}^2}{\mathrm{s}^2}}{\mathrm{sA}} \\ &= \frac{\mathrm{kgm}^2}{\mathrm{s}^3\mathrm{A}}. \end{aligned}$$

SI prefixes

At times, there are quantities that are either extremely small or extremely large. Instead of using scientific notation to deal with these quantities, we can instead use prefixes. The SI prefixes have been standardized for use with the International System of Units (SI). These prefixes can be used with any type of SI unit: meter, volt, second, gram, and so on. Common SI prefixes that are used in this textbook are defined in the table below.

atto	a	10^{-18}	am (attometers)
femto	f	10^{-15}	fm (femtometers)
pico	p	10^{-12}	pm (picometers)
nano	n	10^{-9}	nm (nanometers)
micro	μ	10^{-6}	μm (micrometers)
milli	m	10^{-3}	mm (millimeters)
		10^{0}	m (meters)
kilo	k	10^{3}	km (kilometers)
mega	M	10^{6}	Mm (megameters)

giga	G	10^9	Gm (gigameters)
tera	T	10^{12}	Tm (terameters)
peta	P	10^{15}	Pm (petameters)
exa	E	10^{18}	Em (exameters)

How to convert units

When performing numerical calculations in physics, we frequently need to convert a quantity from one type of unit to another (for example, we may need to convert from cm to m).

To perform a conversion, we rely on the algebraic rule that any value can be multiplied by one without changing its value. If the conversion between units is known, the ratio of one unit to the other will be equal to one. The following examples will demonstrate this.

To convert 3.65 m (meters) to cm (centimeters), the conversion factor between m and cm must be known. There are 100 cm in 1 m. Because we want to convert m to cm, we need to multiply 3.65 m by the ratio containing the unit cm in the numerator and the unit m in the denominator:

$$(3.65\text{ m})\left(\frac{100\text{ cm}}{1\text{ m}}\right).$$

The units of m will cancel out in this equation, resulting in a final unit of cm. The result is

$$(3.65)(100\text{ cm}) = 365\text{ cm}.$$

Unit conversions may require more than one conversion factor. Consider the specific heat capacity of water, which is 1 cal/(gC^o). To determine the specific heat capacity with units of J/(kgK), we must convert between cal (calories) and J (joules), between g (grams) and kg (kilograms), and between C^o (Celsius degree) and K (kelvin). To do this, we can simply multiply by all three ratios, ensuring that the ratios are configured so that units of cal, g, and C^o will cancel out. This will yield

$$\left(1\ \frac{\text{cal}}{\text{gC}^\circ}\right)\left(\frac{4.184\text{ J}}{1\text{ cal}}\right)\left(\frac{1000\text{ g}}{1\text{ kg}}\right)\left(\frac{1\text{ C}^\circ}{1\text{ K}}\right).$$

The result will be

$$(1)(4.184\text{ J})\left(\frac{1000}{\text{kg}}\right)\left(\frac{1}{\text{K}}\right) = 4184\ \frac{\text{J}}{\text{kgK}}.$$

If a unit is raised to a power (squared or cubed, for example), then a conversion can be raised to the same power. For example, to convert the density of water (which is 1 g/cm^3) into units of kg/m^3, convert g to kg, and then cube the conversion between cm and m:

$$\left(1\ \frac{\text{g}}{\text{cm}^3}\right)\left(\frac{1\text{ kg}}{1000\text{ g}}\right)\left(\frac{100\text{ cm}}{1\text{ m}}\right)^3 = \left(1\ \frac{\text{g}}{\text{cm}^3}\right)\left(\frac{1\text{ kg}}{1000\text{ g}}\right)\left(\frac{1000000\text{ cm}^3}{1\text{ m}^3}\right).$$

The result is

$$(1)\left(\frac{1\ \text{kg}}{1000}\right)\left(\frac{1000000}{\text{m}^3}\right) = 1000\ \frac{\text{kg}}{\text{m}^3}.$$

Common unit conversions

Length

1 inch = 2.54 cm

1 foot = 0.3048 m

1 m = 39.37 in

1 mi = 1.609 km

Volume

1 L = 10^{-3} m^3

Time

1 year = 365.25 days

1 day = 24 hours

1 hour = 60 minutes

1 minute = 60 seconds

Mass

1 kg = 1000 g

1 kg has a weight of 2.205 lb on Earth

Force

1 lb = 4.448 N

Pressure

1 Pa = 1 N/m^2

1 atm = 1.01 x 10^5 Pa

1 lb/in^2 (psi) = 6895 Pa

Energy and power

1 cal = 4.184 J

1 kWhr = 3.6 x 10^6 J

1 eV = 1.6 x 10^{-19} J

1 hp = 746 W

Temperature

1 C^{o} (change of one degree C) = 1.8 F^{o} (change of 1.8 degrees F)

1 F^{o} (change of one degree F) = 5/9 C^{o} (change of 5/9 degree C)

Speed

1 m/s = 3.60 km/hr (kph) = 2.24 mi/hr (mph)

1 km/hr (kph) = 0.621 mi/hr

Physical constants

speed of light in vacuum	$3.00 \times 10^{8}\ \mathrm{m/s}$
speed of sound at 20 °C	$343\ \mathrm{m/s}$
acceleration due to gravity	$9.81\ \mathrm{m/s^2}$
average distance between the Earth and Sun	$1.50 \times 10^{11}\ \mathrm{m}$
average distance between the Earth and Moon	$3.84 \times 10^{8}\ \mathrm{m}$
average radius of the Earth	$6.37 \times 10^{6}\ \mathrm{m}$
solar radius	$6.96 \times 10^{8}\ \mathrm{m}$
average radius of the Moon	$1.74 \times 10^{6}\ \mathrm{m}$
mass of the Earth	$5.97 \times 10^{24}\ \mathrm{kg}$

mass of the Sun	$1.99 \times 10^{30}\ \text{kg}$
mass of the Moon	$7.34 \times 10^{22}\ \text{kg}$
mass of a proton	$1.67 \times 10^{-27}\ \text{kg}$
mass of a neutron	$1.68 \times 10^{-27}\ \text{kg}$
mass of an electron	$9.11 \times 10^{-31}\ \text{kg}$
charge of a proton	$+1.60 \times 10^{-19}\ \text{C}$
charge of an electron	$-1.60 \times 10^{-19}\ \text{C}$
universal gravitational constant	$6.67 \times 10^{-11}\ \text{Nm}^2/\text{kg}^2$
Coulomb's constant	$8.99 \times 10^{9}\ \text{Nm}^2/\text{C}^2$
Planck's constant	$6.63 \times 10^{-34}\ \text{Js}$
Avogadro's number	$6.02 \times 10^{23}\ \text{mol}^{-1}$

Further reading

- SI Units – This website from NIST gives an overview of the international system of units (SI).

Practice questions

Conceptual comprehension

1. Is it possible to add a distance in feet to a distance in miles? If so, how? If not, why not?
2. Is it possible to add a Celsius temperature to a Fahrenheit temperature? If so, how? If not, why not?
3. Is it possible to multiply quantities of different units and get a meaningful result? If so, give an example. If not, why not?

Numerical analysis

1. A car travels at a speed of 72 km/hr. Convert this speed to meters per second.
2. A block has a mass of 500 grams. Express this mass in kilograms.
3. An object is accelerated at 20 m/s^2. Calculate the acceleration in terms of km/hr^2.
4. A length of wire measures 25 centimeters. Convert this length to meters.

This page titled Appendix B: Units and unit conversions is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

Appendix C: Answers to selected questions

Chapter 1

Conceptual comprehension

1. A scientific theory is an explanation of how the universe works. It is supported by multiple, repeated evidence. (In other words, it is not merely an untested idea.) On the other hand, a scientific law is a statement describing the relationship between observed or measured quantities. Most of the time, a scientific law is expressed as one or more equations.

Chapter 2

Numerical analysis

1. The scenario is depicted in Figure C.1.

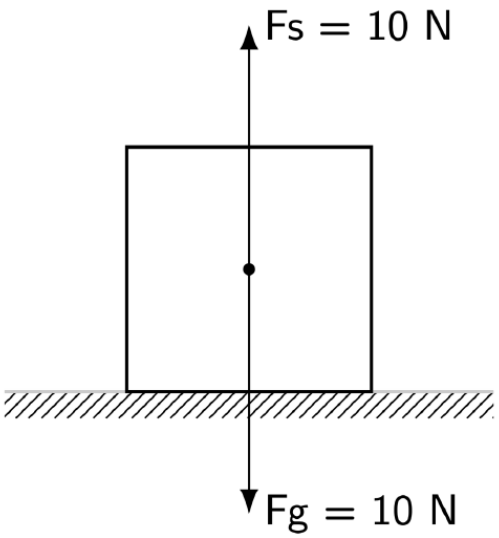


Figure C.1 – A block has a gravitational force (weight) of 10 N and a support force of 10 N. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The magnitude of the net force acting on the box is

$$\begin{aligned} |\vec{F}_{net}| &= |\vec{F}_s + \vec{F}_g| \\ &= |10\ \text{N} + (-10\ \text{N})| \\ &= 0\ \text{N}. \end{aligned}$$

Because the net force acting on the box is 0 N, the box is in mechanical equilibrium.

2. The scenario is depicted in Figure C.2.

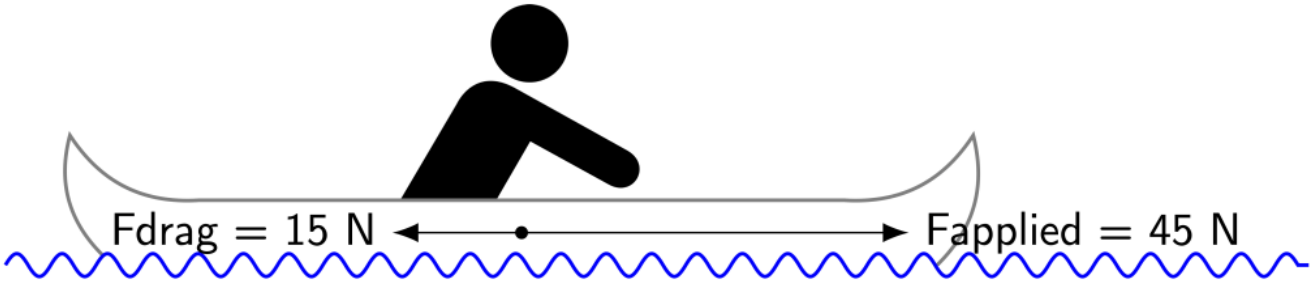


Figure C.2 – An applied force of 45 N acts on the boat in the forward direction. A drag force of 15 N acts on the boat in the opposite direction. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The magnitude of the net force acting on the boat is

$$\begin{aligned} |\vec{F}_{net}| &= |\vec{F}_{applied} + \vec{F}_{drag}| \\ &= |45\text{ N} + (-15\text{ N})| \\ &= 30\text{ N}. \end{aligned}$$

Because the net force acting in the boat is not equal to zero, the boat is not in mechanical equilibrium.

3. The scenario is depicted in Figure C.3.

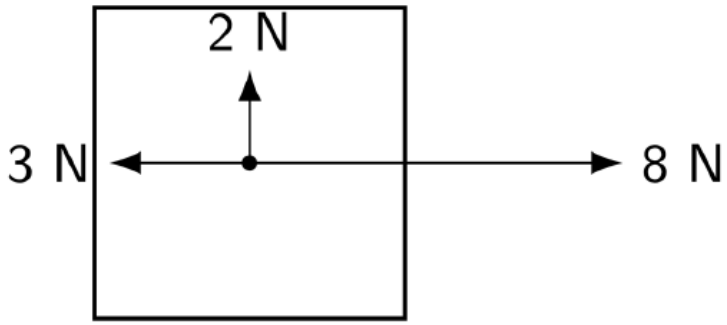


Figure C.3 – A horizontal force of 8 N acts on a block pointing to the right. A horizontal force of 3 N acts on a block pointing to the left. A vertical force of 2 N acts on a block pointing up. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The overall horizontal component of the force acting on the block is

$$\begin{aligned} \Sigma\vec{F}_x &= 8\text{ N} + (-3\text{ N}) \\ &= 5\text{ N}. \end{aligned}$$

The overall vertical component of the force acting on the block is

$$\Sigma\vec{F}_y = 2\text{ N}.$$

The magnitude of the net force is

$$\begin{aligned} \vec{F}_{net}^2 &= \vec{F}_x^2 + \vec{F}_y^2 \\ \vec{F}_{net} &= \sqrt{\vec{F}_x^2 + \vec{F}_y^2} \\ &= \sqrt{(5\text{ N})^2 + (2\text{ N})^2} \\ &= \sqrt{25\text{ N}^2 + 4\text{ N}^2} \\ &= \sqrt{29\text{ N}^2} \\ &= 5.385\text{ N}. \end{aligned}$$

4. The scenario is depicted in Figure C.4. (Note that the force arrows are not to scale with each other in this figure.)

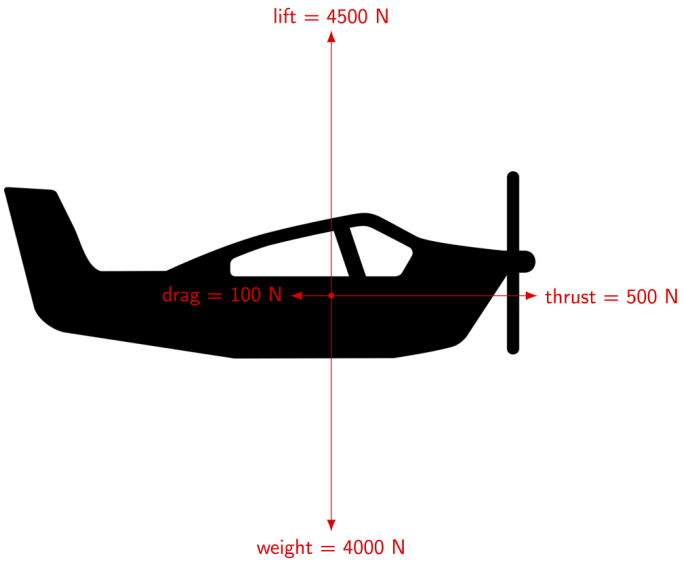


Figure C.4 – An airplane has a lift force of 4500 N, a weight force of 4000 N, a thrust force of 500 N, and a drag force of 100 N. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The overall horizontal component of the force acting on the airplane is

$$\begin{aligned}\Sigma\vec{F}_x &= 500\text{ N} + (-100\text{ N}) \\ &= 400\text{ N}.\end{aligned}$$

The overall vertical component of the force acting on the block is

$$\begin{aligned}\Sigma\vec{F}_y &= 4500\text{ N} + (-4000\text{ N}) \\ &= 500\text{ N}.\end{aligned}$$

The magnitude of the net force is

$$\begin{aligned}\vec{F}_{net}^2 &= \vec{F}_x^2 + \vec{F}_y^2 \\ \vec{F}_{net} &= \sqrt{\vec{F}_x^2 + \vec{F}_y^2} \\ &= \sqrt{(400\text{ N})^2 + (500\text{ N})^2} \\ &= \sqrt{160000\text{ N}^2 + 250000\text{ N}^2} \\ &= \sqrt{410000\text{ N}^2} \\ &= 640.312\text{ N}.\end{aligned}$$

Chapter 3

Conceptual comprehension

1. Draw out the piece of string in a straight line from the starting position to the ending position. Cut the string to that length. Then, using the distance scale on the map, determine how the length of the string relates to the straight-line distance in miles or kilometers. To determine direction, note the overall direction that the car traveled in relation to the compass rose.

Numerical analysis

1. The scenario is depicted in Figure C.5.

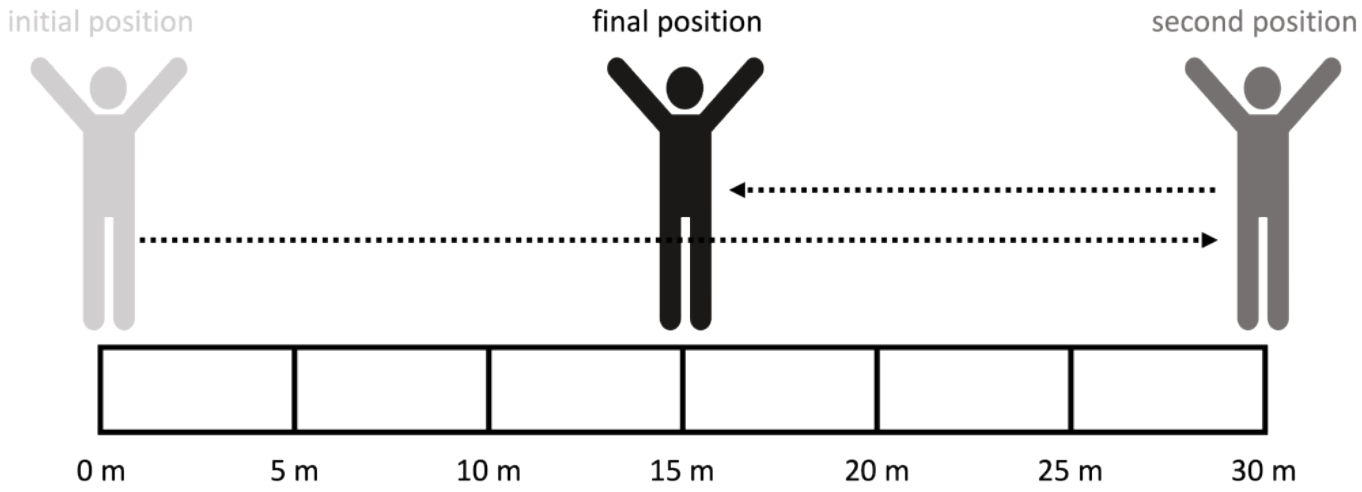


Figure C.5 – A person moves 30 m to the right, and then moves 15 m to the left of that position. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Path is the overall distance traveled by the person, which is

$$\begin{aligned} \text{path} &= 30\ \text{m} + 15\ \text{m} \\ &= 45\ \text{m}. \end{aligned}$$

Displacement is the change in position of the person, which is

$$\begin{aligned} \text{displacement} &= x_{final} - x_{initial} \\ &= 15\ \text{m} - 0\ \text{m} \\ &= 15\ \text{m}. \end{aligned}$$

2. The scenario is depicted in Figure C.6.

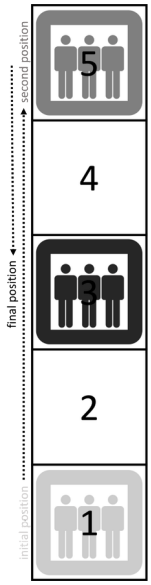


Figure C.6 – An elevator moves from the ground floor to the fifth floor, and then to the third floor of a building. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Path is the overall distance traveled by the elevator, which is

$$\begin{aligned} \text{path} &= 4\ \text{floors} + 2\ \text{floors} \\ &= 6\ \text{floors}. \end{aligned}$$

Displacement is the change in position of the elevator, which is

$$\begin{aligned} \text{displacement} &= y_{final} - y_{initial} \\ &= 3 \text{ floors} - 1 \text{ floor} \\ &= 2 \text{ floors}. \end{aligned}$$

3. The scenario is depicted in Figure C.7.

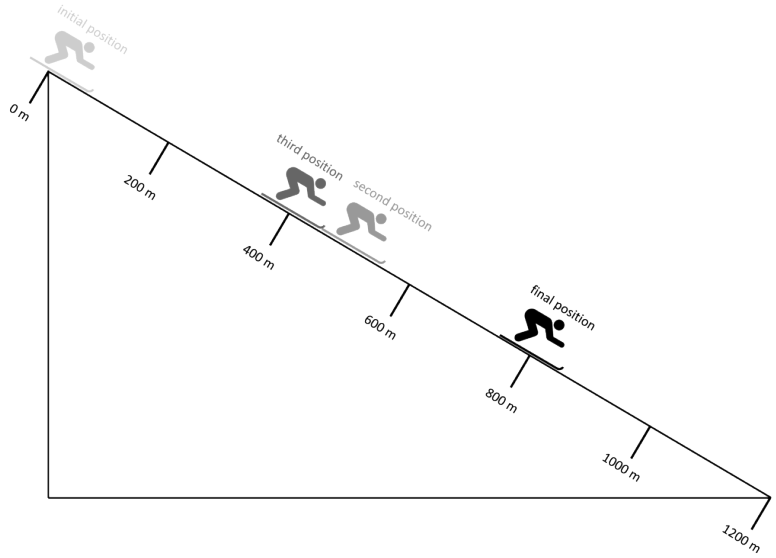


Figure C.7 – A skier skis 500 m downhill, climbs 100 m uphill, then skis 400 m downhill. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

Path is the overall distance traveled by the skier, which is

$$\begin{aligned} \text{path} &= 500 \text{ m} + 100 \text{ m} + 400 \text{ m} \\ &= 1000 \text{ m}. \end{aligned}$$

Displacement is the change in position of the skier, which is

$$\begin{aligned} \text{displacement} &= d_{final} - d_{initial} \\ &= 800 \text{ m} - 0 \text{ m} \\ &= 800 \text{ m}. \end{aligned}$$

4. Path is the overall distance traveled by the boat, which is

$$\begin{aligned} \text{path} &= 10 \text{ km} + 10 \text{ km} \\ &= 20 \text{ km}. \end{aligned}$$

Displacement is the change in position of the boat, which is 0 km, as the boat's final location is the same as its initial location.

5. Average speed is equal to distance (path) divided by time.

$$\begin{aligned} \text{speed} &= \left(\frac{200 \text{ km}}{2 \text{ hr}}\right)\left(\frac{1000 \text{ m}}{1 \text{ km}}\right)\left(\frac{1 \text{ hr}}{60 \text{ min}}\right)\left(\frac{1 \text{ min}}{60 \text{ s}}\right) \\ &= \left(\frac{200000 \text{ m}}{7200 \text{ s}}\right) \\ &= 27.8 \text{ m/s}. \end{aligned}$$

Velocity includes direction. The average velocity of the car is 27.778 m/s west.

6. The path the bird travels as it flies north is 2 km. The path the bird travels as it flies south is 2 km. The overall path of the bird is 4 km. The overall displacement of the bird is 0 km (initial and final locations are the same).
    1. The instantaneous speed of the bird as it flies north is

$$\begin{aligned}\text{speed} &= \left(\frac{2\text{ km}}{20\text{ min}}\right)\left(\frac{1000\text{ m}}{1\text{ km}}\right)\left(\frac{1\text{ min}}{60\text{ s}}\right)\\ &= \left(\frac{2000\text{ m}}{1200\text{ s}}\right)\\ &= 1.67\text{ m/s}.\end{aligned}$$

2. The instantaneous speed of the bird as it flies south is

$$\begin{aligned}\text{speed} &= \left(\frac{2\text{ km}}{15\text{ min}}\right)\left(\frac{1000\text{ m}}{1\text{ km}}\right)\left(\frac{1\text{ min}}{60\text{ s}}\right)\\ &= \left(\frac{2000\text{ m}}{900\text{ s}}\right)\\ &= 2.22\text{ m/s}.\end{aligned}$$

3. The average speed of the bird over its entire journey is

$$\begin{aligned}\text{speed} &= \left(\frac{4\text{ km}}{35\text{ min}}\right)\left(\frac{1000\text{ m}}{1\text{ km}}\right)\left(\frac{1\text{ min}}{60\text{ s}}\right)\\ &= \left(\frac{4000\text{ m}}{2100\text{ s}}\right)\\ &= 1.90\text{ m/s}.\end{aligned}$$

4. The average velocity for the bird over its entire journey is 0 m/s, as the overall displacement of the bird is 0 m.

7. The position vs. time graph is depicted in Figure C.8.

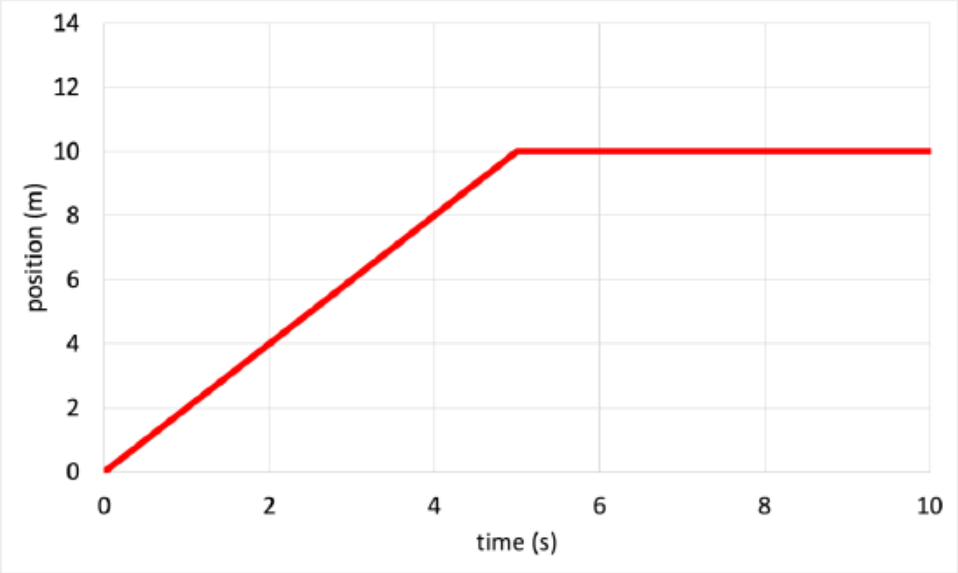


Figure C.8 – Position vs. time graph corresponding to an object moving with a steady pace in the positive direction, then stopping and staying still for several moments. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The velocity vs. time graph is depicted in Figure C.9.

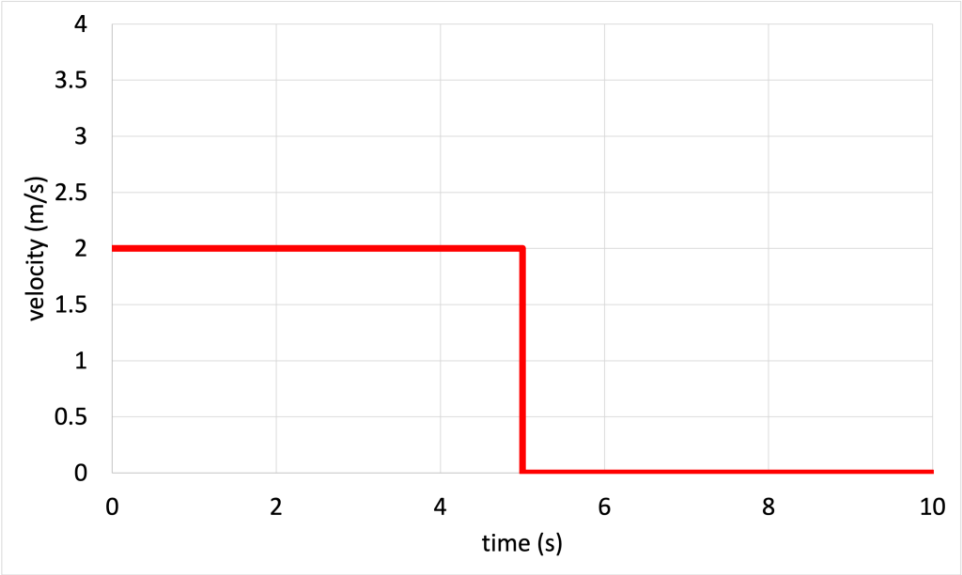


Figure C.9 – Velocity vs. time graph corresponding to an object moving with a steady pace in the positive direction, then stopping and staying still for several moments. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

8. The position vs. time graph is depicted in Figure C.10.

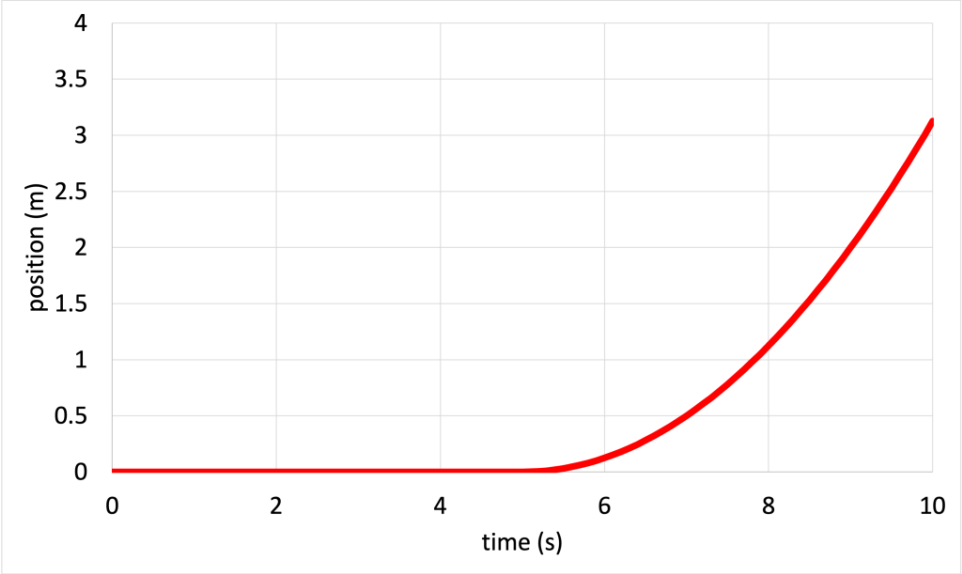


Figure C.10 – Position vs. time graph corresponding to an object that starts at rest for several moments, then moves at a constantly increasing speed for the next several moments. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The velocity vs. time graph is depicted in Figure C.11.

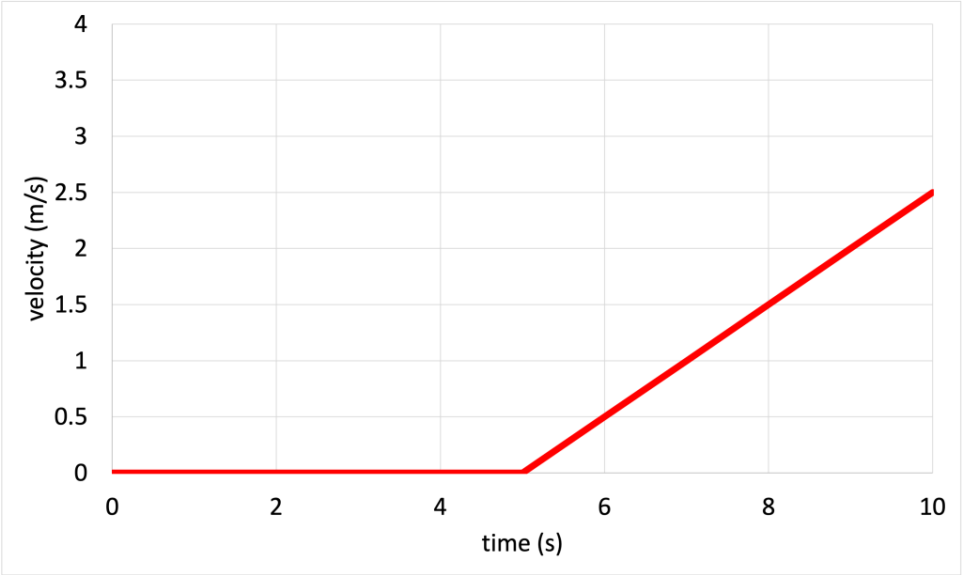


Figure C.11 – Velocity vs. time graph corresponding to an object that starts at rest for several moments, then moves at a constantly increasing speed for the next several moments. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

9. The position vs. time graph is depicted in Figure C.12.

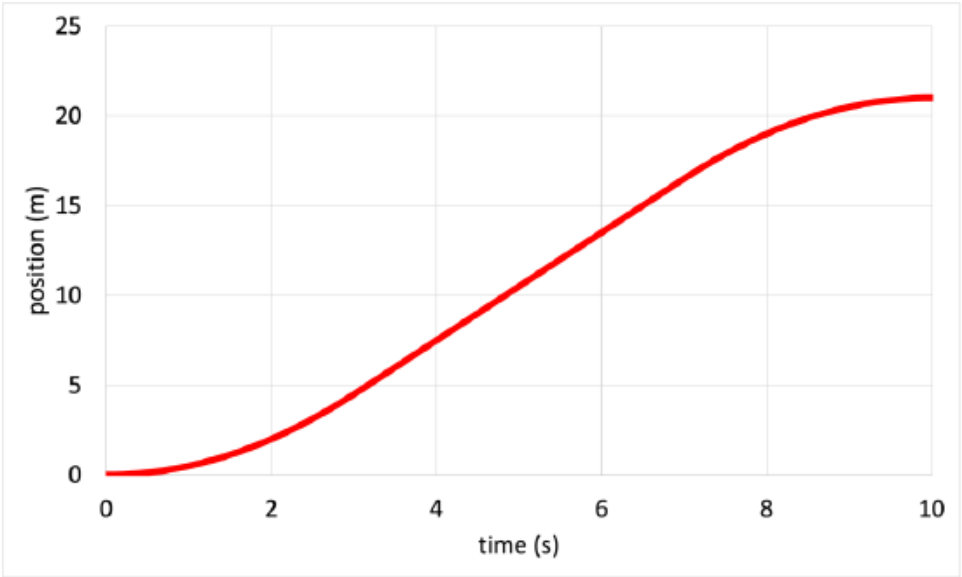


Figure C.12 – Position vs. time graph corresponding to an object that speeds up at an increasing rate, travels at a constant rate for a few moments, and then slows down until it stops. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

The velocity vs. time graph is depicted in Figure C.13.

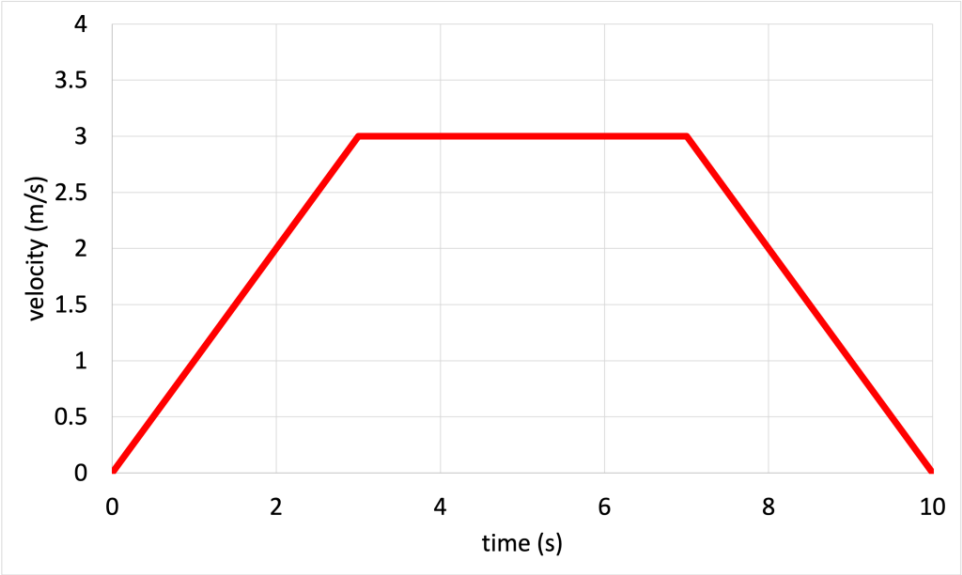


Figure C.13 – Velocity vs. time graph corresponding to an object that speeds up at an increasing rate, travels at a constant rate for a few moments, and then slows down until it stops. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

10. For the first 3 seconds, speed up at an increasing rate in the positive direction. For the next 4 seconds, stop and stand still. Then, for the last 3 seconds, turn around and move back to the starting location at a constant speed.
11. For the first 2 seconds, stop and stand still. For the next 3 seconds, speed up at a constant rate. For the next 2 seconds, continue to move at that constant fast pace. For the next second, slow down at a constant rate until stopped. Then, stand still for the final 2 seconds.
12. The velocity of the train is

$$\begin{aligned} |v| &= \left|\frac{d}{t}\right| \\ &= \left|\frac{500\text{ m}}{20\text{ s}}\right| \\ &= 25\text{ m/s}. \end{aligned}$$

13. The final velocity of the car is

$$\begin{aligned} v_f &= v_i + at \\ &= 0\text{ m/s} + (2\text{ m/s}^2)(5\text{ s}) \\ &= 10\text{ m/s}. \end{aligned}$$

The distance traveled by the car is

$$\begin{aligned} d &= \frac{1}{2}at^2 + v_i t \\ &= \frac{1}{2}(2\text{ m/s}^2)(5\text{ s})^2 + (0\text{ m/s})(5\text{ s}) \\ &= \frac{1}{2}(2\text{ m/s}^2)(25\text{ s}^2) \\ &= 25\text{ m}. \end{aligned}$$

14. First, calculate how long the object will be in the air between the start of its motion and the top of the trajectory. The instantaneous speed at the top of the trajectory is 0 m/s.

\[

$$
\begin{aligned}
v_f &= v_i + at \\
t &= \frac{v_f - v_i}{a} \\
&= \frac{0 \text{ m/s} - 20 \text{ m/s}}{-9.8 \text{ m/s}^2} \\
&= \frac{-20 \text{ m/s}}{-9.8 \text{ m/s}^2} \\
&= 2.04 \text{ s} \\
d &= \frac{1}{2}at^2 + v_i t \\
&= \frac{1}{2}at^2 \\
t &= \sqrt{\frac{2d}{a}} \\
&= \sqrt{\frac{2(-40 \text{ m})}{-9.8 \text{ m/s}^2}} \\
&= \sqrt{\frac{-80 \text{ m}}{-9.8 \text{ m/s}^2}} \\
&= \sqrt{8.16 \text{ s}^2} \\
&= 2.86 \text{ s}
\end{aligned}
$$

Now, use the distance equation to calculate the maximum height of the projectile.

$$
\begin{aligned}
d &= \frac{1}{2}at^2 + v_i t \\
&= \frac{1}{2}(-9.8 \text{ m/s}^2)(2.04 \text{ s})^2 + (20 \text{ m/s})(2.04 \text{ s}) \\
&= \frac{1}{2}(-9.8 \text{ m/s}^2)(4.16 \text{ s}^2) + 40.8 \text{ m} \\
&= -20.4 \text{ m} + 40.8 \text{ m} \\
&= 20.4 \text{ m}
\end{aligned}
$$

15. First, use the distance equation to calculate the time the object is in free fall. If it starts at 0 m and moves down 40 m, its displacement is -40 m. The initial velocity can be assumed to be 0 m/s.

$$
\begin{aligned}
d &= \frac{1}{2}at^2 + v_i t \\
&= \frac{1}{2}at^2 \\
t &= \sqrt{\frac{2d}{a}} \\
&= \sqrt{\frac{2(-40 \text{ m})}{-9.8 \text{ m/s}^2}} \\
&= \sqrt{\frac{-80 \text{ m}}{-9.8 \text{ m/s}^2}} \\
&= \sqrt{8.16 \text{ s}^2} \\
&= 2.86 \text{ s}
\end{aligned}
$$

Now, use the velocity equation to calculate the final speed of the object before it hits the ground.

$$
\begin{aligned}
v_f &= v_i + at \\
&= 0 \text{ m/s} + (-9.8 \text{ m/s}^2)(2.86 \text{ s}) \\
&= -28.0 \text{ m/s}
\end{aligned}
$$

Chapter 4

Conceptual comprehension

1. The BB will continue traveling in a straight-line after leaving the plastic tubing. It is only capable of traveling in a circular path when passing through the tubing because the tubing supplies a support force that changes the direction of the BB. This is shown in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Numerical analysis

1. The net force acting on the object is

$$\begin{aligned}\vec{F}_{net} &= m\vec{a}\\ &= (5\text{ kg})(2\text{ m/s}^2)\\ &= 10\text{ N}.\end{aligned}$$

2. The acceleration that will be created is

$$\begin{aligned}\vec{a} &= \frac{\vec{F}_{net}}{m}\\ &= \left(\frac{0.2\text{ N}}{10\text{ g}}\right)\left(\frac{1000\text{ g}}{1\text{ kg}}\right)\\ &= 20\text{ m/s}^2.\end{aligned}$$

3. The net force acting on the bicycle is 0 N, as the bicycle is moving at a constant velocity (acceleration is 0 m/s^2).
4. The net force acting on the spacecraft is

$$\begin{aligned}\vec{F}_{net}^2 &= \vec{F}_x^2 + \vec{F}_y^2\\ \vec{F}_{net} &= \sqrt{\vec{F}_x^2 + \vec{F}_y^2}\\ &= \sqrt{(2000\text{ N})^2 + (1000\text{ N})^2}\\ &= \sqrt{4000000\text{ N}^2 + 1000000\text{ N}^2}\\ &= \sqrt{5000000\text{ N}^2}\\ &= 2236.068\text{ N}.\end{aligned}$$

The acceleration of the spacecraft is equal to

$$\begin{aligned}\vec{a} &= \frac{\vec{F}_{net}}{m}\\ &= \frac{2240\text{ N}}{2000000\text{ kg}}\\ &= 0.00112\text{ m/s}^2.\end{aligned}$$

5. The gravitational force acting on the object is

$$\begin{aligned}\vec{F}_g &= m\vec{g} \\ &= (10\text{ kg})(-9.8\text{ m/s}^2) \\ &= -98\text{ N}.\end{aligned}$$

6. The gravitational force acting on the person on the Moon is

$$\begin{aligned}\vec{F}_g &= m\vec{g} \\ &= (70\text{ kg})(-1.6\text{ m/s}^2) \\ &= -112\text{ N}.\end{aligned}$$

Chapter 5

Conceptual comprehension

2. Both the mosquito and the sports car experience identical magnitudes of force during the impact. This is because the forces are action-reaction pairs, which are equal in magnitude and opposite in direction.

Numerical analysis

1. —
    1. Because the force that the rock exerts on the person during the kick and the force with which the person kicks the rock are action-reaction pairs, the magnitude of the force that the rock exerts on the person during the kick is 10 N.
    2. The acceleration of the person is

$$\begin{aligned}\vec{a} &= \frac{\vec{F}_{net}}{m} \\ &= \frac{10\text{ N}}{80\text{ kg}} \\ &= 0.125\text{ m/s}^2.\end{aligned}$$

    3. The acceleration of the rock is

$$\begin{aligned}\vec{a} &= \frac{\vec{F}_{net}}{m} \\ &= \frac{10\text{ N}}{5\text{ kg}} \\ &= 2\text{ m/s}^2.\end{aligned}$$

2. —
    1. Because the force that the Earth exerts on the rocket is and the force that the rocket exerts on the Earth are action-reaction pairs, the magnitude of the force that the Earth exerts on the rocket is 350,000 N.
    2. The acceleration of the rocket is

$$\begin{aligned}\vec{a} &= \frac{\vec{F}_{net}}{m} \\ &= \frac{350000\text{ N}}{3000000\text{ kg}} \\ &= 0.117\text{ m/s}^2.\end{aligned}$$

    3. The acceleration of the Earth is

$$\begin{aligned}\vec{a} &= \frac{\vec{F}_{net}}{m} \\ &= \frac{350000\text{ N}}{5.97\times10^{24}\text{ kg}} \\ &= 5.86\times10^{-20}\text{ m/s}^2.\end{aligned}$$

Chapter 6

Conceptual comprehension

1. The unit of Ns is equal to kgm/s, as shown below.

$$\begin{aligned}\text{Ns} &= \\ &= \frac{\text{kgm}}{\text{s}^2}s \\ &= \frac{\text{kgm}}{\text{s}}\end{aligned}$$

Numerical analysis

1. The momentum of the car is equal to

$$\begin{aligned}\vec{p} &= m\vec{v} \\ &= (1200\text{ kg})(20\text{ m/s}) \\ &= 24000\text{ kgm/s}.\end{aligned}$$

2. The change in momentum of the soccer ball is equal to

$$\begin{aligned}\Delta\vec{p} &= \Delta(m\vec{v}) \\ &= m_{final}\vec{v}_{final} - m_{initial}\vec{v}_{initial} \\ &= (0.45\text{ kg})(0\text{ m/s}) - (0.45\text{ kg})(10\text{ m/s}) \\ &= -4.5\text{ kgm/s}.\end{aligned}$$

3. The impulse of the tennis ball is equal to

$$\begin{aligned}\Delta\vec{p} &= \Delta(m\vec{v}) \\ &= m_{final}\vec{v}_{final} - m_{initial}\vec{v}_{initial} \\ &= (0.05\text{ kg})(-10\text{ m/s}) - (0.05\text{ kg})(15\text{ m/s}) \\ &= -0.5\text{ kgm/s} - 0.75\text{ kgm/s} \\ &= -1.25\text{ kgm/s}.\end{aligned}$$

4. —
    1. The initial momentum of the system is

$$\begin{aligned}\vec{p}_{initial} &= m_A\vec{v}_{A,initial} + m_B\vec{v}_{B,initial} \\ &= (1200\text{ kg})(15\text{ m/s}) + (800\text{ kg})(-10\text{ m/s}) \\ &= 18000\text{ kgm/s} - 8000\text{ kgm/s} \\ &= 10000\text{ kgm/s}.\end{aligned}$$

    2. The final momentum of the system is also equal to 10,000 kgm/s, due to the law of conservation of momentum.
    3. The final velocity of both cars is equal to

$$\begin{aligned}\vec{p}_{final} &= m_{final}\vec{v}_{final} \\ \vec{v}_{final} &= \frac{\vec{p}_{final}}{m_{final}} \\ &= \frac{10000\text{ kgm/s}}{1200\text{ kg} + 800\text{ kg}} \\ &= \frac{10000\text{ kgm/s}}{2000\text{ kg}} \\ &= 5\text{ m/s}.\end{aligned}$$

    4. This collision is perfectly inelastic because both cars travel together after the collision. It can also be shown that the collision is inelastic by comparing the initial and final kinetic energy of the system. Because the final KE is not equal to the initial KE, the collision must have been inelastic.

$$
\begin{aligned}
KE_{initial} &= \frac{1}{2}m_A v_{A,initial}^2 + \frac{1}{2}m_B v_{B,initial}^2 \\
&= \frac{1}{2}(1200\text{ kg})(15\text{ m/s})^2 + \frac{1}{2}(800\text{ kg})(-10\text{ m/s})^2 \\
&= \frac{1}{2}(1200\text{ kg})(225\text{ m}^2/\text{s}^2) + \frac{1}{2}(800\text{ kg})(100\text{ m}^2/\text{s}^2) \\
&= 135000\text{ J} + 40000\text{ J} \\
&= 175000\text{ J} \\
KE_{final} &= \frac{1}{2}m_A v_{A,final}^2 + \frac{1}{2}m_B v_{B,final}^2 \\
&= \frac{1}{2}(1200\text{ kg})(5\text{ m/s})^2 + \frac{1}{2}(800\text{ kg})(5\text{ m/s})^2 \\
&= \frac{1}{2}(1200\text{ kg})(25\text{ m}^2/\text{s}^2) + \frac{1}{2}(800\text{ kg})(25\text{ m}^2/\text{s}^2) \\
&= 15000\text{ J} + 10000\text{ J} \\
&= 25000\text{ J}
\end{aligned}
$$

5. —
   1. The initial momentum of the system is

$$
\begin{aligned}
\vec{p}_{initial} &= m_A\vec{v}_{A,initial} + m_B\vec{v}_{B,initial} \\
&= (0.2\text{ kg})(0\text{ m/s}) + (0.3\text{ kg})(5\text{ m/s}) \\
&= 0\text{ kgm/s} + 1.5\text{ kgm/s} \\
&= 1.5\text{ kgm/s}.
\end{aligned}
$$

   2. The final momentum of the system is also equal to 1.5 kgm/s, due to the law of conservation of momentum.
   3. The final velocity of both hockey pucks is equal to

$$
\begin{aligned}
\vec{p}_{final} &= m_{final}\vec{v}_{final} \\
\vec{v}_{final} &= \frac{\vec{p}_{final}}{m_{final}} \\
&= \frac{1.5\text{ kgm/s}}{0.2\text{ kg} + 0.3\text{ kg}} \\
&= \frac{1.5\text{ kgm/s}}{0.5\text{ kg}} \\
&= 3\text{ m/s}.
\end{aligned}
$$

   4. This collision is perfectly inelastic because both hockey pucks travel together after the collision. It can also be shown that the collision is inelastic by comparing the initial and final kinetic energy of the system. Because the final KE is not equal to the initial KE, the collision must have been inelastic.

$$
\begin{aligned}
KE_{initial} &= \frac{1}{2}m_A v_{A,initial}^2 + \frac{1}{2}m_B v_{B,initial}^2 \\
&= \frac{1}{2}(0.2\text{ kg})(0\text{ m/s})^2 + \frac{1}{2}(0.3\text{ kg})(5\text{ m/s})^2 \\
&= 0\text{ J} + \frac{1}{2}(0.3\text{ kg})(25\text{ m}^2/\text{s}^2) \\
&= 0\text{ J} + 3.75\text{ J} \\
&= 3.75\text{ J} \\
KE_{final} &= \frac{1}{2}m_A v_{A,final}^2 + \frac{1}{2}m_B v_{B,final}^2 \\
&= \frac{1}{2}(0.2\text{ kg})(3\text{ m/s})^2 + \frac{1}{2}(0.3\text{ kg})(3\text{ m/s})^2 \\
&= \frac{1}{2}(0.2\text{ kg})(9\text{ m}^2/\text{s}^2) + \frac{1}{2}(0.3\text{ kg})(9\text{ m}^2/\text{s}^2) \\
&= 0.90\text{ J} + 1.35\text{ J} \\
&= 2.25\text{ J}
\end{aligned}
$$

6. —
   1. The initial momentum of the system is equal to

$$\begin{aligned}\vec{p}_{initial} &= m_1\vec{v}_{1,initial} + m_2\vec{v}_{2,initial}\\ &= (1200\text{ kg})(20\text{ m/s}) + (900\text{ kg})(-10\text{ m/s})\\ &= 24000\text{ kgm/s} - 900\text{ kgm/s}\\ &= 15000\text{ kgm/s}.\end{aligned}$$

2. The final momentum of the system is also equal to 15000 kgm/s, due to the law of conservation of momentum.
3. The final velocity of car 2 is equal to

$$\begin{aligned}\vec{p}_{final} &= m_1\vec{v}_{1,final} + m_2\vec{v}_{2,final}\\ \vec{v}_{2,final} &= \frac{\vec{p}_{final} - m_1\vec{v}_{1,final}}{m_2}\\ &= \frac{15000\text{ kgm/s} - (1200\text{ kg})(-6\text{ m/s})}{900\text{ kg}}\\ &= \frac{15000\text{ kgm/s} - (-7200\text{ kgm/s})}{900\text{ kg}}\\ &= \frac{22200\text{ kgm/s}}{900\text{ kg}}\\ &= 24.7\text{ m/s}.\end{aligned}$$

7. The magnitude of the force of the collision is equal to

$$\begin{aligned}m\Delta\vec{v} &= \vec{F}\Delta t\\ |\vec{F}| &= \left|\frac{m\Delta\vec{v}}{\Delta t}\right|\\ &= \left|\frac{m(\vec{v}_{final} - \vec{v}_{initial})}{\Delta t}\right|\\ &= \left|\frac{(1500\text{ kg})(0\text{ m/s} - 40\text{ m/s})}{0.5\text{ s}}\right|\\ &= \left|\frac{(1500\text{ kg})(-40\text{ m/s})}{0.5\text{ s}}\right|\\ &= \left|\frac{-60000\text{ kgm/s}}{0.5\text{ s}}\right|\\ &= |-120000\text{ N}|\\ &= 120000\text{ N}.\end{aligned}$$

8. The magnitude of the force of the collision is equal to

$$\begin{aligned}m\Delta\vec{v} &= \vec{F}\Delta t\\ |\vec{F}| &= \left|\frac{m\Delta\vec{v}}{\Delta t}\right|\\ &= \left|\frac{m(\vec{v}_{final} - \vec{v}_{initial})}{\Delta t}\right|\\ &= \left|\frac{(1500\text{ kg})(0\text{ m/s} - 40\text{ m/s})}{4\text{ s}}\right|\\ &= \left|\frac{(1500\text{ kg})(-40\text{ m/s})}{4\text{ s}}\right|\\ &= \left|\frac{-60000\text{ kgm/s}}{4\text{ s}}\right|\\ &= |-15000\text{ N}|\\ &= 15000\text{ N}.\end{aligned}$$

Chapter 7

Numerical analysis

1. The gravitational potential energy of the object is equal to

$$GPE = mgh \\ = (10\ \text{kg})(9.8\ \text{m/s}^2)(5\ \text{m}) \\ = 490\ \text{J}. \tag{C.1}$$

2. The gravitational potential energy of the object is equal to

$$GPE = mgh \\ = (15\ \text{kg})(9.8\ \text{m/s}^2)(20\ \text{m}) \\ = 2940\ \text{J}.$$

3. The gravitational potential energy of the spaceship is equal to

$$GPE = mgh \\ = (15000\ \text{kg})(1.6\ \text{m/s}^2)(50\ \text{m}) \\ = 1200000\ \text{J}.$$

4. The kinetic energy of the object is equal to

$$KE = \frac{1}{2}mv^2 \\ = \frac{1}{2}(12\ \text{kg})(20\ \text{m/s})^2 \\ = \frac{1}{2}(12\ \text{kg})(400\ \text{m}^2/\text{s}^2) \\ = 2400\ \text{J}.$$

5. The kinetic energy of the object is equal to

$$KE = \frac{1}{2}mv^2 \\ = \frac{1}{2}(3.2\ \text{kg})(-8.5\ \text{m/s})^2 \\ = \frac{1}{2}(3.2\ \text{kg})(72\ \text{m}^2/\text{s}^2) \\ = 115\ \text{J}.$$

6. The mechanical energy will be conserved throughout the motion of the object. Therefore the mechanical energy at the start is equal to the mechanical energy when the object is 5 m above the ground. The mechanical energy can be easily calculated at the start of the motion, and is equal to

$$ME = GPE + KE \\ = mgh + \frac{1}{2}mv^2 \\ = (2\ \text{kg})(9.8\ \text{m/s}^2)(10\ \text{m}) + \frac{1}{2}(2\ \text{kg})(0\ \text{m/s}) \\ = 196\ \text{J} + 0\ \text{J} \\ = 196\ \text{J}.$$

The KE at the starting position is equal to zero, because the object is at rest at that point. When the object is 5 m above the ground, the KE is equal to

$$\begin{aligned} ME &= GPE + KE \\ KE &= ME - GPE \\ &= ME - mgh \\ &= 196 \text{ J} - (2 \text{ kg})(9.8 \text{ m/s}^2)(5 \text{ m}) \\ &= 196 \text{ J} - 98 \text{ J} \\ &= 98 \text{ J}. \end{aligned}$$

7. The total mechanical energy will be conserved throughout the object's motion. This can be used to calculate the final velocity of the object.

$$\begin{aligned} ME_{initial} &= ME_{final} \\ GPE_{initial} + KE_{initial} &= GPE_{final} + KE_{final} \\ mgh_{initial} + \frac{1}{2}mv_{initial}^2 &= mgh_{final} + \frac{1}{2}mv_{final}^2 \end{aligned}$$

The initial velocity of the object is 0 m/s, and the final height of the object (at the instant before it hits the ground) is 0 m. This means that the $\frac{1}{2}mv_{initial}^2$ and mgh_{final} terms will both be equal to 0.

$$\begin{aligned} ME_{initial} &= ME_{final} \\ GPE_{initial} + KE_{initial} &= GPE_{final} + KE_{final} \\ mgh_{initial} + \frac{1}{2}mv_{initial}^2 &= mgh_{final} + \frac{1}{2}mv_{final}^2 \end{aligned}$$

8. The total mechanical energy will be conserved throughout the object's motion. This can be used to calculate the final velocity of the object. As with the previous question, the initial velocity is 0 m/s and the final height is 0 m. This simplifies the equation that can be used to solve for the final velocity of the object.

$$\begin{aligned} v_{final} &= \sqrt{2gh_{initial}} \\ &= \sqrt{2(9.8 \text{ m/s}^2)(8 \text{ m})} \\ &= \sqrt{156.8 \text{ m}^2/\text{s}^2} \\ &= 12.5 \text{ m/s} \end{aligned}$$

9. The total mechanical energy will be conserved throughout the pendulum's motion. This can be used to calculate the final velocity of the pendulum. As with the previous two questions, the velocity at the top of the swing is 0 m/s and the height at the bottom of the swing is 0 m. This simplifies the equation that can be used to solve for the final velocity of the pendulum.

$$\begin{aligned} v_{final} &= \sqrt{2gh_{initial}} \\ &= \sqrt{2(9.8 \text{ m/s}^2)(2 \text{ m})} \\ &= \sqrt{39.2 \text{ m}^2/\text{s}^2} \\ &= 6.26 \text{ m/s} \end{aligned}$$

10. The work done by the car engine is equal to

$$\begin{aligned} W &= F_{//}d \\ &= (5000 \text{ N})(100 \text{ m}) \\ &= 500000 \text{ J}. \end{aligned}$$

11. The work done by the person on the sled is equal to

$$\begin{aligned} W &= F_{//}d \\ &= (33 \text{ N})(10 \text{ m}) \\ &= 330 \text{ J}. \end{aligned}$$

12. The power output of the pulley system is equal to

$$\begin{aligned} P &= \frac{W}{t} \\ &= \frac{F_{//}d}{t} \\ &= \frac{mgd}{t} \\ &= \frac{(100\text{ kg})(9.8\text{ m/s}^2)(5\text{ m})}{10\text{ s}} \\ &= \frac{4900\text{ J}}{10\text{ s}} \\ &= 490\text{ W}. \end{aligned}$$

13. The power applied to lift the object is equal to

$$\begin{aligned} P &= \frac{W}{t} \\ &= \frac{F_{//}d}{t} \\ &= \frac{mgd}{t} \\ &= \frac{(500\text{ kg})(9.8\text{ m/s}^2)(4\text{ m})}{20\text{ s}} \\ &= \frac{19600\text{ J}}{20\text{ s}} \\ &= 980\text{ W}. \end{aligned}$$

14. Use conservation of work to solve for mechanical advantage. The mechanical advantage of the lever is equal to

$$\begin{aligned} W_{in} &= W_{out} \\ F_{in}d_{in} &= F_{out}d_{out} \\ \frac{F_{out}}{F_{in}} &= \frac{d_{in}}{d_{out}} \\ MA = \frac{F_{out}}{F_{in}} &= \frac{d_{in}}{d_{out}} \\ &= \frac{50\text{ cm}}{10\text{ cm}} \\ &= 5. \end{aligned}$$

15. —
    1. Use conservation of work to solve for the distance. The distance the object was moved upward is equal to

$$\begin{aligned} W_{in} &= W_{out} \\ F_{in}d_{in} &= F_{out}d_{out} \\ d_{out} &= \frac{F_{in}d_{in}}{F_{out}} \\ &= \frac{(80\text{ N})(5\text{ m})}{100\text{ N}} \\ &= \frac{400\text{ Nm}}{100\text{ N}} \\ &= 4\text{ m}. \end{aligned}$$

    2. The mechanical advantage of the inclined plane is equal to

$$\begin{aligned} MA &= \frac{F_{out}}{F_{in}} \\ &= \frac{100\text{ N}}{80\text{ N}} \\ &= 1.25. \end{aligned}$$

16. The mechanical advantage of the pulley system is equal to 2.
17. —
    1. The mechanical advantage of the lever is equal to

$$\begin{aligned} MA &= \frac{F_{out}}{F_{in}} \\ &= \frac{80\ \text{N}}{40\ \text{N}} \\ &= 2. \end{aligned}$$

2. The work input to the simple machine is equal to

$$\begin{aligned} W_{in} &= F_{in} d_{in} \\ &= (40\ \text{N})(50\ \text{cm}) \left(\frac{1\ \text{m}}{100\ \text{cm}} \right) \\ &= 20\ \text{J}. \end{aligned}$$

3. The workout output of the simple machine is equal to

$$\begin{aligned} W_{out} &= F_{out} d_{out} \\ &= (80\ \text{N})(20\ \text{cm}) \left(\frac{1\ \text{m}}{100\ \text{cm}} \right) \\ &= 16\ \text{J}. \end{aligned}$$

4. The efficiency of the lever is equal to

$$\begin{aligned} e &= \frac{W_{out}}{W_{in}} \\ &= \frac{16\ \text{J}}{20\ \text{J}} \\ &= 0.8 \\ &= 80\%. \end{aligned}$$

18. —
    1. The work input to the pulley system is equal to

$$\begin{aligned} W_{in} &= F_{in} d_{in} \\ &= (60\ \text{N})(10\ \text{m}) \\ &= 600\ \text{J}. \end{aligned}$$

    2. The efficiency of the pulley system is equal to

$$\begin{aligned} e &= \frac{W_{out}}{W_{in}} \\ &= \frac{480\ \text{J}}{600\ \text{J}} \\ &= 0.8 \\ &= 80\%. \end{aligned}$$

Chapter 8

Conceptual comprehension

1. Torque is equal to perpendicular force times distance from the axis of rotation. By maximizing the distance between the applied force (where the doorknob is positioned) and the axis of rotation (which is where the hinges are positioned), torque is maximized, which makes it easier to open the door.

Numerical analysis

1. The rotational speed of the fan is equal to...
    1. ...in rotations per minute:

$$\begin{aligned}\text{rotational speed} &= \frac{120 \text{ rotations}}{1 \text{ min}} \\ &= 120 \text{ RPM}.\end{aligned}$$

2. ...in rotations per second:

$$\begin{aligned}\text{rotational speed} &= \frac{120 \text{ rotations}}{\text{min}} \left(\frac{1 \text{ min}}{60 \text{ s}}\right) \\ &= 2 \text{ rotations/s}.\end{aligned}$$

3. ...in degrees per second:

$$\begin{aligned}\text{rotational speed} &= \frac{2 \text{ rotations}}{\text{s}} \left(\frac{360^\circ}{1 \text{ rotation}}\right) \\ &= 720^\circ/\text{s}.\end{aligned}$$

4. ...in radians per second:

$$\begin{aligned}\omega &= \frac{2 \text{ rotations}}{\text{s}} \left(\frac{2\pi \text{ rad}}{1 \text{ rotation}}\right) \\ &= 12.6 \text{ rad/s}.\end{aligned}$$

2. The rotational speed of a long-playing vinyl record is equal to

$$\begin{aligned}\omega &= \left(33\frac{1}{3} \text{ rotations/min}\right) \left(\frac{1 \text{ min}}{60 \text{ s}}\right) \left(\frac{2\pi \text{ rad}}{1 \text{ rotation}}\right) \\ &= 3.49 \text{ rad/s}.\end{aligned}$$

The tangential velocity at the outer rim of the record is equal to

$$\begin{aligned}v_t &= \omega r \\ &= \omega \frac{d}{2} \\ &= (3.49 \text{ rad/s}) \left(\frac{25 \text{ cm}}{2}\right) \left(\frac{1 \text{ m}}{100 \text{ cm}}\right) \\ &= 0.436 \text{ m/s}.\end{aligned}$$

3. The rotational speed of the bicycle wheel is equal to

$$\begin{aligned}\omega &= (60 \text{ rotations/min}) \left(\frac{1 \text{ min}}{60 \text{ s}}\right) \left(\frac{2\pi \text{ rad}}{1 \text{ rotation}}\right) \\ &= 6.28 \text{ rad/s}.\end{aligned}$$

The tangential velocity at the outer rim of the wheel is equal to

$$\begin{aligned}v_t &= \omega r \\ &= (6.28 \text{ rad/s})(0.35 \text{ m}) \\ &= 2.2 \text{ m/s}.\end{aligned}$$

4. When played from the outer rim, the rotational speed of the CD is equal to

$$\begin{aligned}v_t &= \omega r \\ &= \omega \frac{d}{2} \\ \omega &= \frac{2v_t}{d} \\ &= \frac{2(1.2 \text{ m/s})}{(12 \text{ cm}) \left(\frac{1 \text{ m}}{100 \text{ cm}}\right)} \\ &= \frac{2.4 \text{ m/s}}{0.12 \text{ m}} \\ &= 20 \text{ rad/s}.\end{aligned}$$

When played from a point halfway between the center and the outer rim, the rotational speed of the CD is equal to

$$\begin{aligned} v_t &= \omega r \\ &= \omega \frac{d}{2} \\ \omega &= \frac{2v_t}{d} \\ &= \frac{2(1.2 \text{ m/s})}{(6 \text{ cm})\left(\frac{1 \text{ m}}{100 \text{ cm}}\right)} \\ &= \frac{2.4 \text{ m/s}}{0.06 \text{ m}} \\ &= 40 \text{ rad/s}. \end{aligned}$$

5. The torque applied to the bolt is equal to

$$\begin{aligned} \tau &= Fd \\ &= (40 \text{ N})(0.2 \text{ m}) \\ &= 8 \text{ Nm}. \end{aligned}$$

6. To balance the see-saw, the person on the other end should sit at a distance equal to

$$\begin{aligned} \tau_{left} &= \tau_{right} \\ F_{left} d_{left} &= F_{right} d_{right} \\ d_{left} &= \frac{F_{right} d_{right}}{F_{left}} \\ &= \frac{m_{right} g d_{right}}{m_{left} g} \\ &= \frac{m_{right} d_{right}}{m_{left}} \\ &= \frac{(60 \text{ kg})(2 \text{ m})}{80 \text{ kg}} \\ &= \frac{120 \text{ kgm}}{80 \text{ kg}} \\ &= 1.5 \text{ m}. \end{aligned}$$

7. The torque applied to the wrench is equal to

$$\begin{aligned} \tau &= F_{\perp} d \\ &= (12 \text{ N})(0.4 \text{ m}) \\ &= 4.8 \text{ Nm}. \end{aligned}$$

8. —
    1. The centripetal force is equal to

$$\begin{aligned} F_c &= \frac{mv^2}{r} \\ &= \frac{(1200 \text{ kg})(20 \text{ m/s})^2}{50 \text{ m}} \\ &= \frac{(1200 \text{ kg})(400 \text{ m}^2/\text{s}^2)}{50 \text{ m}} \\ &= \frac{480000 \text{ kgm}^2/\text{s}^2}{50 \text{ m}} \\ &= 9600 \text{ N}. \end{aligned}$$

    2. The centripetal acceleration is equal to

$$\begin{aligned} a_c &= \frac{v^2}{r} \\ &= \frac{(20\ \text{m/s})^2}{50\ \text{m}} \\ &= \frac{400\ \text{m}^2/\text{s}^2}{50\ \text{m}} \\ &= 8\ \text{m/s}^2. \end{aligned}$$

9. —
    1. The centripetal force acting on the ball will be equal to 10 N. The centripetal force is supplied by (and is equal to) the tension in the string.
    2. The centripetal acceleration of the ball is equal to

$$\begin{aligned} a_c &= \frac{F_c}{m} \\ &= \frac{10\ \text{N}}{0.5\ \text{kg}} \\ &= 20\ \text{m/s}^2. \end{aligned}$$

Chapter 9

Numerical analysis

1. First, calculate the mass of the person, which is equal to

$$\begin{aligned} \vec{F}_{gravity} &= m\vec{g} \\ m &= \left| \frac{\vec{F}_{gravity}}{\vec{g}} \right| \\ &= \left| \frac{-600\ \text{N}}{-9.8\ \text{m/s}^2} \right| \\ &= 61.2\ \text{kg}. \end{aligned}$$

The apparent weight of the person while the elevator is accelerating upward is equal to

$$\begin{aligned} \vec{F}_{net} = m\vec{a} &= \vec{F}_{support} + \vec{F}_{gravity} \\ \vec{F}_{support} &= m\vec{a} - \vec{F}_{gravity} \\ &= (61.2\ \text{kg})(2\ \text{m/s}^2) - (-600\ \text{N}) \\ &= 122\ \text{N} + 600\ \text{N} \\ &= 722\ \text{N}. \end{aligned}$$

2. The apparent weight at the top of the loop is 0 N. This can be deduced from the fact that the passengers experience a sensation of weightlessness.
3. The gravitational force is equal to

$$\begin{aligned} F &= G\frac{m_1 m_2}{d^2} \\ &= (6.67 \times 10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{(50\ \text{kg})(50\ \text{kg})}{(10\ \text{m})^2} \\ &= (6.67 \times 10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{2500\ \text{kg}^2}{100\ \text{m}^2} \\ &= (6.67 \times 10^{-11}\ \text{Nm}^2/\text{kg}^2)(25\ \text{kg}^2/\text{m}^2) \\ &= 1.67 \times 10^{-9}\ \text{N}. \end{aligned}$$

4. The gravitational force is equal to

$$\begin{aligned}
F &= G\frac{m_1 m_2}{d^2} \\
&= G\frac{m_1 m_2}{(r_{earth}+h)^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{(5.97\times10^{24}\ \text{kg})(200000\ \text{kg})}{(6.37\times10^{6}\ \text{m}+500\ \text{km}\,(\frac{1000\ \text{m}}{1\ \text{km}}))^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{1.19\times10^{30}\ \text{kg}^2}{(6.37\times10^{6}\ \text{m}+500000\ \text{m})^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{1.19\times10^{30}\ \text{kg}^2}{(6.87\times10^{6}\ \text{m})^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{1.19\times10^{30}\ \text{kg}^2}{4.72\times10^{13}\ \text{m}^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)(2.52\times10^{16}\ \text{kg}^2/\text{m}^2) \\
&= 1.68\times10^{6}\ \text{N}.
\end{aligned}$$

5. The gravitational force is equal to

$$\begin{aligned}
F &= G\frac{m_1 m_2}{d^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{(1.99\times10^{30}\ \text{kg})(1\times10^{26}\ \text{kg})}{(4.5\times10^{9}\ \text{km}\,(\frac{1000\ \text{m}}{1\ \text{km}}))^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{(1.99\times10^{30}\ \text{kg})(1\times10^{26}\ \text{kg})}{(4.5\times10^{12}\ \text{m})^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{1.99\times10^{56}\ \text{kg}^2}{2.03\times10^{25}\ \text{m}^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)(9.80\times10^{30}\ \text{kg}^2/\text{m}^2) \\
&= 6.54\ \text{N}.
\end{aligned}$$

6. The magnitude of the gravitational acceleration on the surface of Jupiter is equal to

$$\begin{aligned}
g &= G\frac{m}{r^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{2\times10^{27}\ \text{kg}}{(70000\ \text{km}\,(\frac{1000\ \text{m}}{1\ \text{km}}))^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{2\times10^{27}\ \text{kg}}{(7\times10^{7}\ \text{m})^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{2\times10^{27}\ \text{kg}}{4.9\times10^{15}\ \text{m}^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)(4.08\times10^{11}\ \text{kg}/\text{m}^2) \\
&= 27.2\ \text{m}/\text{s}^2.
\end{aligned}$$

7. The magnitude of the gravitational acceleration on the surface of Venus is equal to

$$\begin{aligned}
g &= G\frac{m}{r^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{5\times10^{24}\ \text{kg}}{(6000\ \text{km}\,(\frac{1000\ \text{m}}{1\ \text{km}}))^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{5\times10^{24}\ \text{kg}}{(6000000\ \text{m})^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{5\times10^{24}\ \text{kg}}{3.6\times10^{13}\ \text{m}^2} \\
&= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)(1.39\times10^{11}\ \text{kg}/\text{m}^2) \\
&= 9.27\ \text{m}/\text{s}^2.
\end{aligned}$$

8. The magnitude of the gravitational acceleration on the surface of Titan is equal to

$$\begin{aligned} g &= G\frac{m}{r^2} \\ &= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{1.3\times10^{23}\ \text{kg}}{(2600\ \text{km}\,(\frac{1000\ \text{m}}{1\ \text{km}}))^2} \\ &= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{1.3\times10^{23}\ \text{kg}}{(2600000\ \text{m})^2} \\ &= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{1.3\times10^{23}\ \text{kg}}{6.76\times10^{12}\ \text{m}^2} \\ &= (6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)(1.92\times10^{10}\ \text{kg/m}^2) \\ &= 1.28\ \text{m/s}^2. \end{aligned}$$

9. The velocity of the satellite is equal to

$$\begin{aligned} F_c &= F_g \\ \frac{m_{satellite}v^2}{r} &= G\frac{m_{satellite}m_{Earth}}{r^2} \\ v^2 &= G\frac{m_{Earth}}{r} \\ v &= \sqrt{G\frac{m_{Earth}}{r}} \\ &= \sqrt{(6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{6\times10^{24}\ \text{kg}}{6.37\times10^{6}\ \text{m}+400\ \text{km}\,(\frac{1000\ \text{m}}{1\ \text{km}})}} \\ &= \sqrt{(6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{6\times10^{24}\ \text{kg}}{6.37\times10^{6}\ \text{m}+400000\ \text{m}}} \\ &= \sqrt{(6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)\frac{6\times10^{24}\ \text{kg}}{6.77\times10^{6}\ \text{m}}} \\ &= \sqrt{(6.67\times10^{-11}\ \text{Nm}^2/\text{kg}^2)(8.86\times10^{17}\ \text{kg/m})} \\ &= \sqrt{5.91\times10^{7}\ \text{m}^2/\text{s}^2} \\ &= 7690\ \text{m/s}. \end{aligned}$$

Chapter 10

Numerical analysis

1. —
    1. The time it takes the ball to reach its maximum height is equal to

$$\begin{aligned} v_f &= v_i + at \\ t &= \frac{v_f - v_i}{a} \\ &= \frac{0\ \text{m/s} - 10\ \text{m/s}}{-9.8\ \text{m/s}^2} \\ &= \frac{-10\ \text{m/s}}{-9.8\ \text{m/s}^2} \\ &= 1.02\ \text{s}. \end{aligned}$$

    2. The maximum height of the ball is equal to

$$\begin{aligned}|d| &= \left|\frac{1}{2}at^2 + v_i t\right| \\ &= \left|\frac{1}{2}(-9.8\text{ m/s}^2)(1.02\text{ s})^2 + (10\text{ m/s})(1.02\text{ s})\right| \\ &= \left|\frac{1}{2}(-9.8\text{ m/s}^2)(1.04\text{ s}^2) + (10.2\text{ m})\right| \\ &= |-5.1\text{ m} + 10.2\text{ m}| \\ &= 5.1\text{ m}.\end{aligned}$$

2. —
    1. First, calculate the amount of time it takes the ball to travel before hitting the ground. This is limited by the vertical direction of travel. The initial component of the velocity in the vertical direction can be assumed to be zero.

$$\begin{aligned}y &= \frac{1}{2}at^2 + v_{i,y}t \\ &= \frac{1}{2}at^2 \\ t^2 &= \frac{2y}{a} \\ t &= \sqrt{\frac{2y}{a}} \\ &= \sqrt{\frac{2(-2\text{ m})}{-9.8\text{ m/s}^2}} \\ &= \sqrt{\frac{-4\text{ m}}{-9.8\text{ m/s}^2}} \\ &= \sqrt{0.408\text{ s}^2} \\ &= 0.639\text{ s}.\end{aligned}$$

The distance that the ball travels horizontally before hitting the ground is equal to

$$\begin{aligned}v_x &= \frac{x}{t} \\ x &= v_x t \\ &= (15\text{ m/s})(0.639\text{ s}) \\ &= 9.59\text{ m}.\end{aligned}$$

    2. The vertical speed of the ball immediately before impact is equal to

$$\begin{aligned}|v_{f,y}| &= |v_{i,y} + at| \\ &= |0\text{ m/s} + (-9.8\text{ m/s}^2)(0.639\text{ s})| \\ &= |-6.26\text{ m/s}| \\ &= 6.26\text{ m/s}.\end{aligned}$$

    3. The final horizontal speed of the ball is equal to the initial horizontal speed of the ball (15 m/s) because there is no force acting on the ball in the horizontal direction while it is in motion.
    4. The total magnitude of the ball's velocity immediately before impact is equal to

$$\begin{aligned}v_{total}^2 &= v_{f,y}^2 + v_x^2 \\ v_{total} &= \sqrt{v_{f,y}^2 + v_x^2} \\ &= \sqrt{(6.26\text{ m/s})^2 + (15\text{ m/s})^2} \\ &= \sqrt{39.2\text{ m}^2/\text{s}^2 + 225\text{ m}^2/\text{s}^2} \\ &= \sqrt{264.2\text{ m}^2/\text{s}^2} \\ &= 16.3\text{ m/s}.\end{aligned}$$

3. —
    1. The distance the ball travels horizontally before hitting the ground is equal to

$$\begin{aligned} v_x &= \frac{x}{t} \\ x &= v_x t \\ &= (20\text{ m/s})(4\text{ s}) \\ &= 80\text{ m}. \end{aligned}$$

2. The distance the ball travels vertically during that time is equal to

$$\begin{aligned} |y| &= \left|\frac{1}{2}at^2 + v_{i,y}t\right| \\ &= \left|\frac{1}{2}at^2\right| \\ &= \left|\frac{1}{2}(-9.8\text{ m/s}^2)(4\text{ s})^2\right| \\ &= \left|\frac{1}{2}(-9.8\text{ m/s}^2)(16\text{ s}^2)\right| \\ &= |-78.4\text{ m}| \\ &= 78.4\text{ m}. \end{aligned}$$

3. The vertical speed of the ball immediately before impact is equal to

$$\begin{aligned} |v_{f,y}| &= |v_{i,y} + at| \\ &= |0\text{ m/s} + (-9.8\text{ m/s}^2)(4\text{ s})| \\ &= |-39.2\text{ m/s}| \\ &= 39.2\text{ m/s}. \end{aligned}$$

4. The horizontal speed of the ball immediately before impact is equal to the initial horizontal speed, 20 m/s. This is because there is no force acting on the ball in the horizontal direction while it is in motion.
5. The total magnitude of the ball's velocity immediately before impact is equal to

$$\begin{aligned} v_{total}^2 &= v_{f,y}^2 + v_x^2 \\ v_{total} &= \sqrt{v_{f,y}^2 + v_x^2} \\ &= \sqrt{(39.2\text{ m/s})^2 + (20\text{ m/s})^2} \\ &= \sqrt{1540\text{ m}^2/\text{s}^2 + 400\text{ m}^2/\text{s}^2} \\ &= \sqrt{1940\text{ m}^2/\text{s}^2} \\ &= 44.0\text{ m/s}. \end{aligned}$$

4. The height of the bridge above the water's surface is equal to

$$\begin{aligned} |y| &= \left|\frac{1}{2}at^2 + v_i t\right| \\ &= \left|\frac{1}{2}at^2\right| \\ &= \left|\frac{1}{2}(-9.8\text{ m/s}^2)(3\text{ s})^2\right| \\ &= \left|\frac{1}{2}(-9.8\text{ m/s}^2)(9\text{ s}^2)\right| \\ &= |-44.1\text{ m}| \\ &= 44.1\text{ m}. \end{aligned}$$

5. —
   1. The time it takes for the rocket to hit the ground is equal to

$$\begin{aligned} v_{f,y} &= v_{i,y} + at \\ t &= \frac{v_{f,y} - v_{i,y}}{a} \\ &= \frac{-30 \text{ m/s} - 30 \text{ m/s}}{-9.8 \text{ m/s}^2} \\ &= \frac{-60 \text{ m/s}}{-9.8 \text{ m/s}^2} \\ &= 6.12 \text{ s}. \end{aligned}$$

2. The horizontal distance traveled by the rocket is equal to

$$\begin{aligned} v_x &= \frac{x}{t} \\ x &= v_x t \\ &= (40 \text{ m/s})(6.12 \text{ s}) \\ &= 245 \text{ m}. \end{aligned}$$

6. The kinetic energy of the satellite is equal to

$$\begin{aligned} KE &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}(500 \text{ kg})(7615 \text{ m/s})^2 \\ &= \frac{1}{2}(500 \text{ kg})(5.8 \times 10^7 \text{ m}^2/\text{s}^2) \\ &= 1.45 \times 10^{10} \text{ J}. \end{aligned}$$

The gravitational potential energy of the satellite is equal to

$$\begin{aligned} GPE &= -G\frac{m_1 m_2}{r} \\ &= -G\frac{m_{satellite} m_{Earth}}{r_{Earth} + \text{altitude}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{(500 \text{ kg})(5.97 \times 10^{24} \text{ kg}}{6.37 \times 10^6 \text{ m} + (500 \text{ km})\left(\frac{1000 \text{ m}}{1 \text{ km}}\right)} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{2.99 \times 10^{27} \text{ kg}^2}{6.37 \times 10^6 \text{ m} + 500000 \text{ m}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{2.99 \times 10^{27} \text{ kg}^2}{6.87 \times 10^6 \text{ m}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)(4.35 \times 10^{20} \text{ kg}^2/\text{m}) \\ &= -2.90 \times 10^{10} \text{ J}. \end{aligned}$$

The total mechanical energy of the satellite is equal to

$$\begin{aligned} ME &= KE + GPE \\ &= 1.45 \times 10^{10} \text{ J} + 2.45 \times 10^9 \text{ J} \\ &= 1.695 \times 10^{10} \text{ J}. \end{aligned}$$

7. First, it is necessary to calculate the velocity of the satellite. This is equal to the speed of rotation of the Earth around its axis, which is equal to

$$\begin{aligned} v &= \frac{d}{t} \\ &= \frac{2\pi r}{t} \\ &= \frac{2\pi\left(r_{Earth} + \text{altitude}\right)}{t} \\ &= \frac{2\pi(6.37 \times 10^6 \text{ m} + 35786 \text{ km}\left(\frac{1000 \text{ m}}{1 \text{ km}}\right))}{(1 \text{ day})\left(\frac{24 \text{ hr}}{1 \text{ day}}\right)\left(\frac{60 \text{ min}}{1 \text{ hr}}\right)\left(\frac{60 \text{ s}}{1 \text{ min}}\right)} \\ &= \frac{2.65 \times 10^8 \text{ m}}{86400 \text{ s}} \\ &= 3070 \text{ m/s}. \end{aligned}$$

The kinetic energy of the satellite is therefore equal to

$$\begin{aligned} KE &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}(1000 \text{ kg})(3070 \text{ m/s})^2 \\ &= \frac{1}{2}(1000 \text{ kg})(9.42 \times 10^6 \text{ m}^2/\text{s}^2) \\ &= 4.7 \times 10^9 \text{ J}. \end{aligned}$$

The gravitational potential energy of the satellite is equal to

$$\begin{aligned} GPE &= -G\frac{m_1 m_2}{r} \\ &= -G\frac{m_{satellite} m_{Earth}}{r_{Earth} + \text{altitude}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{(1000 \text{ kg})(5.97 \times 10^{24} \text{ kg}}{6.37 \times 10^6 \text{ m} + (35786 \text{ km})\left(\frac{1000 \text{ m}}{1 \text{ km}}\right)} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{5.97 \times 10^{27} \text{ kg}^2}{6.37 \times 10^6 \text{ m} + 3.58 \times 10^7 \text{ m}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{5.97 \times 10^{27} \text{ kg}^2}{4.22 \times 10^7 \text{ m}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)(1.41 \times 10^{20} \text{ kg}^2/\text{m}) \\ &= -9.4 \times 10^9 \text{ J}. \end{aligned}$$

The total mechanical energy of the satellite is equal to

$$\begin{aligned} ME &= KE + GPE \\ &= 1.1 \times 10^8 \text{ J} + 3.5 \times 10^{11} \text{ J} \\ &= 3.501 \times 10^{11} \text{ J}. \end{aligned}$$

8. —
    1. The gravitational potential energy of the satellite at an altitude of 200 km is equal to

$$\begin{aligned} GPE &= -G\frac{m_1 m_2}{r} \\ &= -G\frac{m_{satellite} m_{Earth}}{r_{Earth} + \text{altitude}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{(2500 \text{ kg})(5.97 \times 10^{24} \text{ kg}}{6.37 \times 10^6 \text{ m} + (200 \text{ km})\left(\frac{1000 \text{ m}}{1 \text{ km}}\right)} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{1.49 \times 10^{28} \text{ kg}^2}{6.37 \times 10^6 \text{ m} + 200000 \text{ m}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)\frac{1.49 \times 10^{28} \text{ kg}^2}{6.57 \times 10^6 \text{ m}} \\ &= -(6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2)(2.27 \times 10^{21} \text{ kg}^2/\text{m}) \\ &= -1.51 \times 10^{11} \text{ J}. \end{aligned}$$

The gravitational potential energy when the satellite has twice the altitude (400 km), will be twice the value of the GPE at 200

km, therefore it will be 9.8×10^9 J.

2. The change in kinetic energy of the satellite is equal to

$$\begin{aligned}\Delta KE &= -\Delta GPE \\ &= -(GPE_{400km} - GPE_{200km}) \\ &= -(9.8 \times 10^9 \text{ J} - 4.9 \times 10^9 \text{ J}) \\ &= -4.9 \times 10^9 \text{ J}.\end{aligned}$$

3. Because mechanical energy is conserved, the change in mechanical energy of the satellite is equal to 0.

Chapter 11

Conceptual comprehension

3. Information about each nucleus is given in the table below.

nucleus	number of protons (Z)	atomic mass number (A)	number of neutrons (N)
^{18}Fe	9	18	9
^{20}Na	11	20	9
^{21}Mg	12	21	9
^{23}Si	14	23	9

All four nuclei contain the same number of neutrons.

Numerical analysis

1. —
    1. Hydrogen contains 1 proton.
    2. Helium contains 2 protons.
    3. Carbon contains 6 protons.
    4. Oxygen contains 8 protons.
    5. Iron contains 26 protons.
2. —
    1. The H_2O molecule contains $(2)(1) + (1)(8) = 10$ protons.
    2. The CH_4 molecule contains $(1)(6) + (4)(1) = 10$ protons.
    3. The N_2 molecule contains $(2)(7) = 14$ protons.
    4. The CO_2 molecule contains $(1)(6) + (2)(8) = 22$ protons.
    5. The $C_6H_{12}O_6$ molecule contains $(6)(6) + (12)(1) + (6)(8) = 96$ protons.
3. —
    1. Carbon-12 has 6 protons. Therefore, the number of neutrons is equal to

$$\begin{aligned}N &= A - Z \\ &= 12 - 6 \\ &= 6.\end{aligned}$$

    2. Helium-4 has 2 protons. Therefore, the number of neutrons is equal to

$$\begin{aligned}N &= A - Z \\ &= 4 - 2 \\ &= 6.\end{aligned}$$

3. Oxygen-16 has 8 protons. Therefore, the number of neutrons is equal to

$$\begin{aligned} N &= A - Z \\ &= 16 - 8 \\ &= 8. \end{aligned}$$

4. Silicon-28 has 14 protons. Therefore, the number of neutrons is equal to

$$\begin{aligned} N &= A - Z \\ &= 28 - 14 \\ &= 14. \end{aligned}$$

5. Uranium-235 has 92 protons. Therefore, the number of neutrons is equal to

$$\begin{aligned} N &= A - Z \\ &= 235 - 92 \\ &= 143. \end{aligned}$$

4. —
   1. $^{23}\mathrm{Na}^{+}$ has 11 protons and a charge of +1. Therefore, the number of neutrons is equal to

$$\begin{aligned} N &= A - Z \\ &= 23 - 11 \\ &= 12. \end{aligned}$$

The number of electrons is equal to

$$\begin{aligned} N_e &= Z - \text{charge} \\ &= 11 - 1 \\ &= 10. \end{aligned}$$

   2. $^{36}\mathrm{Cl}^{-}$ has 17 protons and a charge of -1. Therefore, the number of neutrons is equal to

$$\begin{aligned} N &= A - Z \\ &= 36 - 17 \\ &= 19. \end{aligned}$$

The number of electrons is equal to

$$\begin{aligned} N_e &= Z - \text{charge} \\ &= 36 - (-1) \\ &= 37. \end{aligned}$$

   3. $^{24}\mathrm{Mg}^{2+}$ has 12 protons and a charge of +2. Therefore, the number of neutrons is equal to

$$\begin{aligned} N &= A - Z \\ &= 24 - 12 \\ &= 12. \end{aligned}$$

The number of electrons is equal to

$$\begin{aligned} N_e &= Z - \text{charge} \\ &= 12 - 2 \\ &= 10. \end{aligned}$$

   4. $^{80}\mathrm{Br}^{-}$ has 35 protons and a charge of -1. Therefore, the number of neutrons is equal to

$$\begin{aligned} N &= A - Z \\ &= 80 - 35 \\ &= 45. \end{aligned}$$

The number of electrons is equal to

$$\begin{aligned} N_e &= Z - \text{charge} \\ &= 80 - (-1) \\ &= 81. \end{aligned}$$

Chapter 12

Numerical analysis

1. The density of the object is equal to

$$\begin{aligned} D &= \frac{m}{V} \\ &= \frac{50\ \text{g}}{20000\ \text{cm}^3} \\ &= 0.0025\ \text{g/cm}^3. \end{aligned}$$

Using a unit of kg/m^3, the volume is equal to

$$\begin{aligned} D &= \frac{m}{V} \\ &= \frac{0.05\ \text{kg}}{0.02\ \text{m}^3} \\ &= 2.5\ \text{kg/m}^3. \end{aligned}$$

2. The density of the stone is equal to

$$\begin{aligned} D &= \frac{m}{V} \\ &= \frac{750\ \text{g}}{400\ \text{cm}^3} \\ &= 1.88\ \text{g/cm}^3. \end{aligned}$$

Using a unit of kg/m^3, the volume is equal to

$$\begin{aligned} D &= \frac{m}{V} \\ &= \frac{0.75\ \text{kg}}{0.0004\ \text{m}^3} \\ &= 1880\ \text{kg/m}^3. \end{aligned}$$

3. The density of the wooden cube is equal to

$$\begin{aligned} D &= \frac{m}{V} \\ &= \frac{m}{l^3} \\ &= \frac{20\ \text{g}}{(3.3\ \text{cm})^3} \\ &= \frac{20\ \text{g}}{35.9\ \text{cm}^3} \\ &= 0.557\ \text{g/cm}^3. \end{aligned}$$

Using a unit of kg/m^3, the volume is equal to

$$\begin{aligned} D &= \frac{m}{V} \\ &= \frac{m}{l^3} \\ &= \frac{0.02 \text{ kg}}{(35.9 \text{ cm}^3)\left(\frac{1 \text{ m}}{100 \text{ cm}}\right)^3} \\ &= \frac{0.02 \text{ kg}}{(35.9 \text{ cm}^3)(1 \times 10^{-6} \text{ m}^3/\text{cm}^3)} \\ &= \frac{0.02 \text{ kg}}{3.59 \times 10^{-5} \text{ m}^3} \\ &= 557 \text{ kg/m}^3. \end{aligned}$$

4. The distance that the spring will stretch is equal to

$$\begin{aligned} F &= k\Delta x \\ \Delta x &= \frac{F}{k} \\ &= \frac{50 \text{ N}}{200 \text{ N/m}} \\ &= 0.25 \text{ m}. \end{aligned}$$

5. The force applied to the spring must be equal to

$$\begin{aligned} F &= k\Delta x \\ &= (300 \text{ N/m})(0.1 \text{ m}) \\ &= 30 \text{ N}. \end{aligned}$$

6. The spring constant of the spring is equal to

$$\begin{aligned} F &= k\Delta x \\ k &= \frac{F}{\Delta x} \\ &= \frac{25 \text{ N}}{0.05 \text{ m}} \\ &= 500 \text{ N/m}. \end{aligned}$$

7. Each spring will stretch a distance that is equal to

$$\begin{aligned} F &= k\Delta x \\ \Delta x &= \frac{F}{k} \\ &= \frac{5 \text{ N}}{30 \text{ N/m}} \\ &= 0.167 \text{ m}. \end{aligned}$$

Therefore, the total distance that both springs stretch will be equal to

$$\begin{aligned} \Delta x_{total} &= \Delta x_1 + \Delta x_2 \\ &= 0.167 \text{ m} + 0.167 \text{ m} \\ &= 0.333 \text{ m}. \end{aligned}$$

8. Each spring only experiences 2.5 N of force. Therefore, the distance both springs will stretch is equal to

$$\begin{aligned} F &= k\Delta x \\ \Delta x &= \frac{F}{k} \\ &= \frac{2.5 \text{ N}}{30 \text{ N/m}} \\ &= 0.083 \text{ m}. \end{aligned}$$

9. —
    1. The surface area of the cube is equal to

$$\begin{aligned} SA &= 6 \times l^2 \\ &= (6)(12\text{ cm})^2 \\ &= (6)(144\text{ cm}^2) \\ &= 864\text{ cm}^2. \end{aligned}$$

2. The volume of the cube is equal to

$$\begin{aligned} V &= l^3 \\ &= (12\text{ cm})^3 \\ &= 1728\text{ cm}^3. \end{aligned}$$

3. The surface area to volume ratio of the cube is equal to

$$\begin{aligned} \frac{SA}{V} &= \frac{864\text{ cm}^2}{1728\text{ cm}^3} \\ &= 0.5\text{ cm}^{-1}. \end{aligned}$$

Chapter 13

Conceptual comprehension

1. Buoyant force is directly proportional to both the density of the fluid and the volume of displaced fluid. Both spheres are in the same fluid, so this variable will not play a role in which sphere (if any) experiences a greater buoyant force. The sphere that sinks displaces more fluid than the sphere that floats. Therefore, the sphere that sinks experiences the greater buoyant force.

Numerical analysis

1. The pressure exerted on the floor by the box is equal to

$$\begin{aligned} P &= \frac{F}{A} \\ &= \frac{mg}{lw} \\ &= \frac{(30\text{ kg})(9.8\text{ m/s}^2)}{(0.5\text{ m})(0.5\text{ m})} \\ &= \frac{294\text{ N}}{0.25\text{ m}^2} \\ &= 1180\text{ Pa}. \end{aligned}$$

2. —
    1. The pressure exerted by the book on the table is equal to

$$\begin{aligned} P &= \frac{F}{A} \\ &= \frac{F}{td} \\ &= \frac{22\text{ N}}{(4.5\text{ cm})\left(\frac{1\text{ m}}{100\text{ cm}}\right)(20.5\text{ cm})\left(\frac{1\text{ m}}{100\text{ cm}}\right)} \\ &= \frac{22\text{ N}}{(0.045\text{ m})(0.205\text{ m})} \\ &= \frac{22\text{ N}}{0.009225\text{ m}^2} \\ &= 2384.8\text{ Pa}. \end{aligned}$$

    2. The pressure exerted by the book on the table is equal to

$$\begin{aligned} P &= \frac{F}{A} \\ &= \frac{F}{dh} \\ &= \frac{22\text{ N}}{(20.5\text{ cm})\left(\frac{1\text{ m}}{100\text{ cm}}\right)(26\text{ cm})\left(\frac{1\text{ m}}{100\text{ cm}}\right)} \\ &= \frac{22\text{ N}}{(0.205\text{ m})(0.26\text{ m})} \\ &= \frac{22\text{ N}}{0.0533\text{ m}^2} \\ &= 412.758\text{ Pa.} \end{aligned}$$

3. —
    1. The pressure exerted by the person on the floor while barefoot is equal to

$$\begin{aligned} P &= \frac{F}{A} \\ &= \frac{550\text{ N}}{(2)(160\text{ cm}^2)\left(\frac{1\text{ m}}{100\text{ cm}}\right)^2} \\ &= \frac{550\text{ N}}{(320\text{ cm}^2)\left(\frac{1\text{ m}}{100\text{ cm}}\right)^2} \\ &= \frac{550\text{ N}}{(320\text{ cm}^2)(0.0001\text{ m}^2/\text{cm}^2)} \\ &= \frac{550\text{ N}}{0.032\text{ m}^2} \\ &= 17200\text{ Pa.} \end{aligned}$$

    2. The pressure exerted by the person on the floor while wearing high heels is equal to

$$\begin{aligned} P &= \frac{F}{A} \\ &= \frac{550\text{ N}}{(2)(1\text{ cm}^2)\left(\frac{1\text{ m}}{100\text{ cm}}\right)^2} \\ &= \frac{550\text{ N}}{(2\text{ cm}^2)(0.0001\text{ m}^2/\text{cm}^2)} \\ &= \frac{550\text{ N}}{0.0002\text{ m}^2} \\ &= 2.75 \times 10^6\text{ Pa.} \end{aligned}$$

4. The total pressure exerted by all four tires on the road is equal to

$$\begin{aligned} P &= \frac{F}{A} \\ &= \frac{mg}{A} \\ &= \frac{(1200\text{ kg})(9.8\text{ m/s}^2)}{4(0.02\text{ m}^2)} \\ &= \frac{11760\text{ N}}{0.08\text{ m}^2} \\ &= 147000\text{ Pa.} \end{aligned}$$

5. The pressure at the bottom of the pool due to the weight of the water is equal to

$$\begin{aligned} P &= Dgdepth \\ &= (1000\text{ kg/m}^3)(9.8\text{ m/s}^2)(2\text{ m}) \\ &= 19600\text{ Pa.} \end{aligned}$$

6. The pressure on the submarine due to the weight of the water above it is equal to

$$\begin{aligned} P &= Dgdepth \\ &= (1025\ \text{kg/m}^3)(9.8\ \text{m/s}^2)(30\ \text{m}) \\ &= 301000\ \text{Pa}. \end{aligned}$$

7. The buoyant force acting on the cube is equal to

$$\begin{aligned} F_b &= D_{fluid} V_{displaced} g \\ &= (1000\ \text{kg/m}^3)(0.02\ \text{m}^3)(9.8\ \text{m/s}^2) \\ &= 196\ \text{N}. \end{aligned}$$

8. The buoyant force acting on the wooden block is equal to

$$\begin{aligned} F_b &= D_{fluid} V_{displaced} g \\ &= (900\ \text{kg/m}^3)(0.05\ \text{m}^3)(9.8\ \text{m/s}^2) \\ &= 441\ \text{N}. \end{aligned}$$

9. —
    1. The weight of the ship is equal to

$$\begin{aligned} W &= mg \\ &= (400000\ \text{kg})(9.8\ \text{m/s}^2) \\ &= 3.92 \times 10^6\ \text{N}. \end{aligned}$$

    2. The weight of the liquid that must be displaced to cause the ship to float is equal to the ship's weight, which is 3.92×10^6 N.
    3. The volume of displaced freshwater is equal to

$$\begin{aligned} F_b &= D_{fluid} V_{displaced} g \\ V_{displaced} &= \frac{F_b}{D_{fluid} g} \\ &= \frac{3.92 \times 10^6\ \text{N}}{(1000\ \text{kg/m}^3)(9.8\ \text{m/s}^2)} \\ &= \frac{3.92 \times 10^6\ \text{N}}{9800\ \text{N/m}^3} \\ &= 400\ \text{m}^3. \end{aligned}$$

    4. The volume of displaced seawater is equal to

$$\begin{aligned} F_b &= D_{fluid} V_{displaced} g \\ V_{displaced} &= \frac{F_b}{D_{fluid} g} \\ &= \frac{3.92 \times 10^6\ \text{N}}{(1020\ \text{kg/m}^3)(9.8\ \text{m/s}^2)} \\ &= \frac{3.92 \times 10^6\ \text{N}}{10000\ \text{N/m}^3} \\ &= 392\ \text{m}^3. \end{aligned}$$

10. The pressure exerted on the fluid is equal to

$$\begin{aligned} P &= \frac{F}{A} \\ &= \frac{200\ \text{N}}{0.02\ \text{m}^2} \\ &= 10000\ \text{Pa}. \end{aligned}$$

11. The force exerted by the large piston is equal to

$$\begin{aligned}\frac{\Delta F_{large}}{A_{large}} &= \frac{\Delta F_{small}}{A_{small}} \\ \Delta F_{large} &= \left(\frac{\Delta F_{small}}{A_{small}}\right)(A_{large}) \\ &= \left(\frac{500\text{ N}}{0.005\text{ m}^2}\right)(0.05\text{ m}^2) \\ &= (100000\text{ N/m}^2)(0.05\text{ m}^2) \\ &= 5000\text{ N}.\end{aligned}$$

12. The force exerted by the large piston on the car is equal to

$$\begin{aligned}\frac{\Delta F_{large}}{A_{large}} &= \frac{\Delta F_{small}}{A_{small}} \\ \Delta F_{large} &= \left(\frac{\Delta F_{small}}{A_{small}}\right)(A_{large}) \\ &= \left(\frac{10000\text{ N}}{0.02\text{ m}^2}\right)(0.1\text{ m}^2) \\ &= (500000\text{ N/m}^2)(0.1\text{ m}^2) \\ &= 50000\text{ N}.\end{aligned}$$

Chapter 14

Numerical analysis

1. The new volume of the gas is equal to

$$\begin{aligned}P_1V_1 &= P_2V_2 \\ V_2 &= \frac{P_1V_1}{P_2} \\ &= \frac{(3\text{ atm})(4\text{ L})}{1.5\text{ atm}} \\ &= \frac{12\text{ atmL}}{1.5\text{ atm}} \\ &= 8\text{ L}.\end{aligned}$$

2. The new volume of the gas is equal to

$$\begin{aligned}P_1V_1 &= P_2V_2 \\ V_2 &= \frac{P_1V_1}{P_2} \\ &= \frac{(200\text{ kPa})(10\text{ L})}{400\text{ kPa}} \\ &= \frac{2000\text{ kPaL}}{400\text{ kPa}} \\ &= 5\text{ L}.\end{aligned}$$

3. The pressure of the gas is equal to

$$\begin{aligned}P_1V_1 &= P_2V_2 \\ P_2 &= \frac{P_1V_1}{V_2} \\ &= \frac{(4\text{ atm})(8\text{ L})}{2\text{ L}} \\ &= \frac{32\text{ atmL}}{2\text{ L}} \\ &= 16\text{ atm}.\end{aligned}$$

4. The final volume of the weather balloon is equal to

$$\begin{aligned} P_1V_1 &= P_2V_2 \\ V_2 &= \frac{P_1V_1}{P_2} \\ &= \frac{(100\text{ kPa})(10\text{ m}^2)}{30\text{ kPa}} \\ &= \frac{1000\text{ kPam}^3}{30\text{ kPa}} \\ &= 33.333\text{ m}^3. \end{aligned}$$

5. The buoyant force acting on the helium balloon is equal to

$$\begin{aligned} F_b &= D_{fluid}V_{displaced}g \\ &= (1.2\text{ kg/m}^3)(5\text{ m}^3)(9.8\text{ m/s}^2) \\ &= 58.8\text{ N}. \end{aligned}$$

Chapter 15

Numerical analysis

1. —
    1. The temperature is equal to

$$\begin{aligned} T_F &= 1.8T_C + 32 \\ &= (1.8)(20) + 32 \\ &= 36\text{ }^\circ\text{F} + 32\text{ }^\circ\text{F} \\ &= 68\text{ }^\circ\text{F}. \end{aligned}$$

    2. The temperature is equal to

$$\begin{aligned} T_K &= T_C + 273 \\ &= 20\text{ K} + 273\text{ K} \\ &= 293\text{ K}. \end{aligned}$$

2. —
    1. The temperature is equal to

$$\begin{aligned} T_C &= (T_F - 32)\left(\frac{5}{9}\right) \\ &= (75 - 32)\left(\frac{5}{9}\right) \\ &= (43)\left(\frac{5}{9}\right) \\ &= 23.889\text{ }^\circ\text{C}. \end{aligned}$$

    2. The temperature is equal to

$$\begin{aligned} T_K &= T_C + 273 \\ &= 23.899\text{ K} + 273\text{ K} \\ &= 296.899\text{ K}. \end{aligned}$$

3. —
    1. The temperature is equal to

$$\begin{aligned} T_C &= T_K - 273 \\ &= 318\text{ }^\circ\text{C} - 273\text{ }^\circ\text{C} \\ &= 45\text{ }^\circ\text{C}. \end{aligned}$$

    2. The temperature is equal to

$$\begin{aligned} T_F &= 1.8T_C + 32 \\ &= (1.8)(45) + 32 \\ &= 81\ ^\circ\text{F} + 32\ ^\circ\text{F} \\ &= 113\ ^\circ\text{F}. \end{aligned}$$

4. The amount of heat required to change the temperature of the water is equal to

$$\begin{aligned} Q &= mc\Delta T \\ &= mc(T_{final} - T_{initial}) \\ &= (500\text{ g})(1\text{ cal/(gC}^\circ))(45\ ^\circ\text{C} - 25\ ^\circ\text{C}) \\ &= (500\text{ g})(1\text{ cal/(gC}^\circ))(20\text{ C}^\circ) \\ &= 10000\text{ cal}. \end{aligned}$$

5. The amount of heat required to change the temperature of the aluminum is equal to

$$\begin{aligned} Q &= mc\Delta T \\ &= mc(T_{final} - T_{initial}) \\ &= (250\text{ g})(0.215\text{ cal/(gC}^\circ))(80\ ^\circ\text{C} - 20\ ^\circ\text{C}) \\ &= (250\text{ g})(0.215\text{ cal/(gC}^\circ))(60\text{ C}^\circ) \\ &= 3230\text{ cal}. \end{aligned}$$

6. The amount of heat required to change the temperature of the lead is equal to

$$\begin{aligned} Q &= mc\Delta T \\ &= mc(T_{final} - T_{initial}) \\ &= (2\text{ kg})\left(\frac{1000\text{ g}}{1\text{ kg}}\right)(0.031\text{ cal/(gC}^\circ))(200\ ^\circ\text{C} - 100\ ^\circ\text{C}) \\ &= (2000\text{ g})(0.031\text{ cal/(gC}^\circ))(100\text{ C}^\circ) \\ &= 6200\text{ cal}. \end{aligned}$$

7. The amount of heat required to heat the water and the copper kettle is equal to

$$\begin{aligned} Q &= m_{water}c_{water}\Delta T_{water} + m_{kettle}c_{kettle}\Delta T_{kettle} \\ &= m_{water}c_{water}(T_{water,final} - T_{water,initial}) \\ &+ m_{kettle}c_{kettle}(T_{kettle,final} - T_{kettle,initial}) \\ &= (800\text{ g})(1\text{ cal/(gC}^\circ))(100\ ^\circ\text{C} - 20\ ^\circ\text{C}) \\ &+ (500\text{ g})(0.092\text{ cal/(gC}^\circ))(100\ ^\circ\text{C} - 20\ ^\circ\text{C}) \\ &= (800\text{ g})(1\text{ cal/(gC}^\circ))(80\text{ C}^\circ) + (500\text{ g})(0.092\text{ cal/(gC}^\circ))(80\text{ C}^\circ) \\ &= 64000\text{ cal} + 3680\text{ cal} \\ &= 67700\text{ cal}. \end{aligned}$$

8. Because both substances have equal masses and have the same specific heat capacity, the final temperature will be an average of each temperature, which is equal to

$$\begin{aligned} T_f &= \frac{T_{i,h} + T_{i,c}}{2} \\ &= \frac{20\ ^\circ\text{C} + 80\ ^\circ\text{C}}{2} \\ &= \frac{100\ ^\circ\text{C}}{2} \\ &= 50\ ^\circ\text{C}. \end{aligned}$$

9. The final temperature is equal to

$$
\begin{aligned}
0 &= Q_h + Q_c \\
&= m_h c_h \Delta T_h + m_c c_c \Delta T_c \\
&= m_h c_h (T_f - T_{i,h}) + m_c c_c (T_f - T_{i,c}) \\
&= m_h c_h T_f - m_h c_h T_{i,h} + m_c c_c T_f - m_c c_c T_{i,c} \\
m_h c_h T_f + m_c c_c T_f &= m_h c_h T_{i,h} + m_c c_c T_{i,c} \\
T_f (m_h c_h + m_c c_c) &= m_h c_h T_{i,h} + m_c c_c T_{i,c} \\
T_f &= \frac{m_h c_h T_{i,h} + m_c c_c T_{i,c}}{m_h c_h + m_c c_c} \\
&= \frac{(10\text{ g})(1\text{ cal/(gC}^\circ\text{)})(30\text{ }^\circ\text{C}) + (40\text{ g})(1\text{ cal/(gC}^\circ\text{)})(90\text{ }^\circ\text{C})}{(30\text{ g})(1\text{ cal/(gC}^\circ\text{)}) + (40\text{ g})(1\text{ cal/(gC}^\circ\text{)})} \\
&= \frac{300\text{ cal} + 3600\text{ cal}}{30\text{ cal/}^\circ\text{C} + 40\text{ cal/}^\circ\text{C}} \\
&= \frac{3900\text{ cal}}{70\text{ cal/}^\circ\text{C}} \\
&= 55.7\text{ }^\circ\text{C}.
\end{aligned}
$$

10. The final temperature is equal to

$$
\begin{aligned}
T_f &= \frac{m_h c_h T_{i,h} + m_c c_c T_{i,c}}{m_h c_h + m_c c_c} \\
&= \frac{(200\text{ g})(1\text{ cal/(gC}^\circ\text{)})(20\text{ }^\circ\text{C}) + (300\text{ g})(0.584\text{ cal/(gC}^\circ\text{)})(30\text{ }^\circ\text{C})}{(200\text{ g})(1\text{ cal/(gC}^\circ\text{)}) + (300\text{ g})(0.584\text{ cal/(gC}^\circ\text{)})} \\
&= \frac{4000\text{ cal} + 5256\text{ cal}}{200\text{ cal/}^\circ\text{C} + 175\text{ cal/}^\circ\text{C}} \\
&= \frac{9256\text{ cal}}{375\text{ cal/}^\circ\text{C}} \\
&= 24.7\text{ }^\circ\text{C}.
\end{aligned}
$$

11. The final temperature is equal to

$$
\begin{aligned}
T_f &= \frac{m_h c_h T_{i,h} + m_c c_c T_{i,c}}{m_h c_h + m_c c_c} \\
&= \frac{(600\text{ g})(1\text{ cal/(gC}^\circ\text{)})(25\text{ }^\circ\text{C}) + (400\text{ g})(0.11\text{ cal/(gC}^\circ\text{)})(80\text{ }^\circ\text{C})}{(600\text{ g})(1\text{ cal/(gC}^\circ\text{)}) + (400\text{ g})(0.11\text{ cal/(gC}^\circ\text{)})} \\
&= \frac{15000\text{ cal} + 3520\text{ cal}}{600\text{ cal/}^\circ\text{C} + 44\text{ cal/}^\circ\text{C}} \\
&= \frac{18520\text{ cal}}{644\text{ cal/}^\circ\text{C}} \\
&= 28.8\text{ }^\circ\text{C}.
\end{aligned}
$$

Chapter 16

Numerical analysis

1. The area of the solar panels is equal to

$$
\begin{aligned}
A &= \frac{P}{1400\text{ W/m}^2} \\
&= \frac{8500\text{ W}}{1400\text{ W/m}^2} \\
&= 6.07\text{ m}^2.
\end{aligned}
$$

2. The area of the solar panels is equal to

$$\begin{aligned} A &= \frac{P}{e(1400\ \text{W/m}^2)} \\ &= \frac{8500\ \text{W}}{(0.15)(1400\ \text{W/m}^2)} \\ &= \frac{8500\ \text{W}}{210\ \text{W/m}^2} \\ &= 40.5\ \text{m}^2. \end{aligned}$$

3. The efficiency of the solar panels is equal to

$$\begin{aligned} A &= \frac{P}{e(1400\ \text{W/m}^2)} \\ e &= \frac{P}{A(1400\ \text{W/m}^2)} \\ &= \frac{150\ \text{W}}{(1\ \text{m}^2)(1400\ \text{W/m}^2)} \\ &= \frac{150\ \text{W}}{1400\ \text{W}} \\ &= 0.107 \\ &= 10.7\%. \end{aligned}$$

4. The initial difference between the temperature of the cookies and the surroundings is equal to

$$\begin{aligned} T_{cookies} - T_{surroundings} &= 200\ ^\circ\text{C} - 20\ ^\circ\text{C} \\ &= 180\ ^\circ\text{C}. \end{aligned}$$

1. The new difference between the temperature of the cookies and the surroundings is equal to

$$\begin{aligned} T_{cookies} - T_{surroundings} &= 110\ ^\circ\text{C} - 20\ ^\circ\text{C} \\ &= 90\ ^\circ\text{C}. \end{aligned}$$

Because this is one half the initial temperature difference, the new rate of cooling will also be one half the initial rate of cooling, or 2 ^{O}C/min.

2. The new difference between the temperature of the cookies and the surroundings is equal to

$$\begin{aligned} T_{cookies} - T_{surroundings} &= 65\ ^\circ\text{C} - 20\ ^\circ\text{C} \\ &= 45\ ^\circ\text{C}. \end{aligned}$$

Because this is one quarter the initial temperature difference, the new rate of cooling will also be one quarter the initial rate of cooling, or 1 ^{O}C/min.

3. The new difference between the temperature of the cookies and the surroundings is equal to

$$\begin{aligned} T_{cookies} - T_{surroundings} &= 42.5\ ^\circ\text{C} - 20\ ^\circ\text{C} \\ &= 22.5\ ^\circ\text{C}. \end{aligned}$$

Because this is one eighth the initial temperature difference, the new rate of cooling will also be one eighth the initial rate of cooling, or 0.5 ^{O}C/min.

Chapter 17

Numerical analysis

1. The amount of heat required to heat the water and melt it is equal to

$$\begin{aligned} Q &= mc\Delta T + mLH_f \\ &= mc(T_f - T_i) + mLH_f \\ &= (100\text{ g})(0.5\text{ cal/(gC}^\circ\text{)})(0\text{ }^\circ\text{C} - (-10\text{ }^\circ\text{C})) + (100\text{ g})(80\text{ cal/g}) \\ &= (100\text{ g})(0.5\text{ cal/(gC}^\circ\text{)})(10\text{ C}^\circ) + 8000\text{ cal} \\ &= 500\text{ cal} + 8000\text{ cal} \\ &= 8500\text{ cal}. \end{aligned}$$

2. The amount of heat required to boil the water is equal to

$$\begin{aligned} Q &= mLH_v \\ &= (500\text{ g})(540\text{ cal/g}) \\ &= 270000\text{ cal}. \end{aligned}$$

3. The amount of heat required to melt the lead is equal to

$$\begin{aligned} Q &= mLH_f \\ &= (50\text{ g})(5.5\text{ cal/g}) \\ &= 275\text{ cal}. \end{aligned}$$

4. —
    1. The freezing/melting point of the substance is -100 °C.
    2. The condensation/vaporization point of the substance is 50 °C.
    3. The latent heat of fusion of the substance is 50 cal/g.
    4. The latent heat of vaporization of the substance is 200 cal/g.
    5. The specific heat capacity of the solid form of the substance is equal to

$$\begin{aligned} Q &= mc\Delta T \\ c &= \frac{Q}{m\Delta T} \\ &= \frac{Q}{m(T_f - T_i)} \\ &= \frac{50\text{ cal}}{(1\text{ g})(-100\text{ }^\circ\text{C} - (-200\text{ }^\circ\text{C}))} \\ &= \frac{50\text{ cal}}{(1\text{ g})(100\text{ C}^\circ)} \\ &= \frac{50\text{ cal}}{1\text{ gC}^\circ} \\ &= 0.5\text{ cal/(gC}^\circ\text{)}. \end{aligned}$$

    6. The specific heat capacity of the liquid form of the substance is equal to

$$\begin{aligned} Q &= mc\Delta T \\ c &= \frac{Q}{m\Delta T} \\ &= \frac{Q}{m(T_f - T_i)} \\ &= \frac{100\text{ cal}}{(1\text{ g})(50\text{ }^\circ\text{C} - (-100\text{ }^\circ\text{C}))} \\ &= \frac{100\text{ cal}}{(1\text{ g})(150\text{ C}^\circ)} \\ &= \frac{100\text{ cal}}{150\text{ gC}^\circ} \\ &= 0.667\text{ cal/(gC}^\circ\text{)}. \end{aligned}$$

    7. The specific heat capacity of the gaseous form of the substance is equal to

$$
\begin{aligned}
Q &= mc\Delta T \\
c &= \frac{Q}{m\Delta T} \\
&= \frac{Q}{m(T_f - T_i)} \\
&= \frac{100 \text{ cal}}{(1 \text{ g})(300 \text{ °C} - 50 \text{ °C})} \\
&= \frac{100 \text{ cal}}{(1 \text{ g})(250 \text{ C°})} \\
&= \frac{100 \text{ cal}}{250 \text{ gC°}} \\
&= 0.4 \text{ cal/(gC°)}.
\end{aligned}
$$

Chapter 18

Numerical analysis

1. The change in internal energy of the gas is equal to

$$
\begin{aligned}
\Delta E &= Q - W \\
&= 200 \text{ J} - 150 \text{ J} \\
&= 50 \text{ J}.
\end{aligned}
$$

2. The change in internal energy of the system is equal to

$$
\begin{aligned}
\Delta E &= Q - W \\
&= -5000 \text{ J} - 3000 \text{ J} \\
&= -8000 \text{ J}.
\end{aligned}
$$

3. The heat added to the gas is equal to

$$
\begin{aligned}
\Delta E &= Q - W \\
Q &= \Delta E + W \\
&= E_{final} - E_{initial} + W \\
&= 6000 \text{ J} - 4000 \text{ J} + 2000 \text{ J} \\
&= 4000 \text{ J}.
\end{aligned}
$$

4. The Carnot efficiency of the engine is equal to

$$
\begin{aligned}
e_{carnot} &= \frac{T_h - T_c}{T_h} \\
&= \frac{773 \text{ K} - 373 \text{ K}}{773 \text{ K}} \\
&= \frac{500 \text{ K}}{773 \text{ K}} \\
&= 0.647 \\
&= 64.7\%.
\end{aligned}
$$

5. The Carnot efficiency of the engine is equal to

$$
\begin{aligned}
e_{carnot} &= \frac{T_h - T_c}{T_h} \\
&= \frac{573 \text{ K} - 303 \text{ K}}{573 \text{ K}} \\
&= \frac{270 \text{ K}}{573 \text{ K}} \\
&= 0.471 \\
&= 47.1\%.
\end{aligned}
$$

6. The Carnot efficiency of the heat pump is equal to

$$\begin{aligned} e_{carnot} &= \frac{T_h - T_c}{T_h} \\ &= \frac{313\text{ K} - 263\text{ K}}{313\text{ K}} \\ &= \frac{50\text{ K}}{313\text{ K}} \\ &= 0.160 \\ &= 16.0\%. \end{aligned}$$

7. The Carnot efficiency of the engine is equal to

$$\begin{aligned} e_{carnot} &= \frac{T_h - T_c}{T_h} \\ &= \frac{800\text{ K} - 300\text{ K}}{800\text{ K}} \\ &= \frac{500\text{ K}}{800\text{ K}} \\ &= 0.625 \\ &= 62.5\%. \end{aligned}$$

8. —
    1. The total number of states is equal to 16.
    2. Two of the 16 states are highly ordered, which is equal to 12.5%.
9. —
    8. The total number of states is equal to 36.
    9. Six of the states are highly ordered, which is equal to 16.667%.
10. —
    8. The total number of states is 216.
    9. Six of the states are highly ordered, which is equal to 2.778%.

Chapter 19

Numerical analysis

1. —
    1. The amplitude of the wave is equal to 4.
    2. The wavelength of the wave is equal to 2 m.
    3. The wave speed of the wave is equal to

$$\begin{aligned} v &= \lambda f \\ &= (2\text{ m})(4\text{ Hz}) \\ &= 8\text{ m/s}. \end{aligned}$$

2. The frequency of the wave is equal to

$$\begin{aligned} f &= \frac{oscillations}{t} \\ &= \frac{10}{2\text{ s}} \\ &= 5\text{ Hz}. \end{aligned}$$

3. The frequency of the wave is equal to

$$\begin{aligned} f &= \frac{1}{T} \\ &= \frac{1}{0.5\text{ s}} \\ &= 2\text{ Hz}. \end{aligned}$$

4. The frequency of the wave is equal to

$$
\begin{aligned}
v &= \lambda f \\
f &= \frac{v}{\lambda} \\
&= \frac{20\ \text{m/s}}{2\ \text{m}} \\
&= 10\ \text{Hz}.
\end{aligned}
$$

5. The wave speed of the wave is equal to

$$
\begin{aligned}
v &= \lambda f \\
&= (0.1\ \text{m})(100\ \text{Hz}) \\
&= 10\ \text{m/s}.
\end{aligned}
$$

6. —
   1. The frequency of the second harmonic is equal to

$$
\begin{aligned}
f_n &= n f_1 \\
&= (2)(440\ \text{Hz}) \\
&= 880\ \text{Hz}.
\end{aligned}
$$

   2. The frequency of the third harmonic is equal to

$$
\begin{aligned}
f_n &= n f_1 \\
&= (3)(440\ \text{Hz}) \\
&= 1320\ \text{Hz}.
\end{aligned}
$$

7. The frequency of the first harmonic is equal to

$$
\begin{aligned}
f_n &= n f_1 \\
f_1 &= \frac{f_n}{n} \\
&= \frac{900\ \text{Hz}}{3} \\
&= 300\ \text{Hz}.
\end{aligned}
$$

8. The scenario is depicted in Figure C.14.

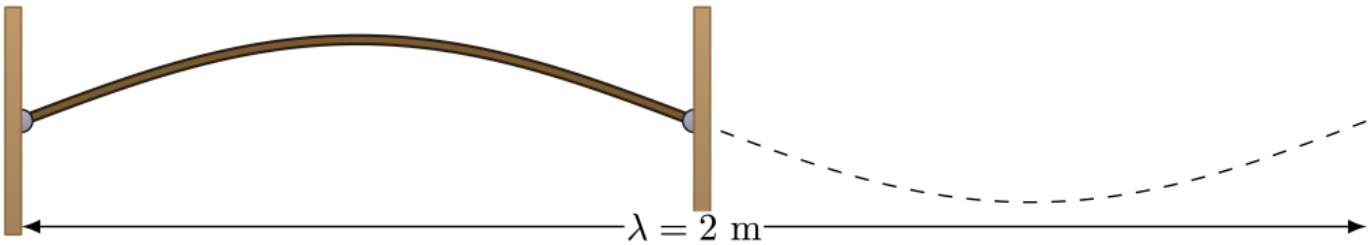


Figure C.14 – The first harmonic of a wave having a wavelength of 2 m. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

The length of the rope is equal to 1 m.

9. —
   1. A wave with three nodes represents the second harmonic.
   2. The wavelength of the wave is equal to 30 cm.
   3. The frequency of the wave is equal to

$$\begin{aligned} v &= \lambda f \\ f &= \frac{v}{\lambda} \\ &= \frac{340 \text{ m/s}}{(30 \text{ cm})\left(\frac{1 \text{ m}}{100 \text{ cm}}\right)} \\ &= \frac{340 \text{ m/s}}{0.3 \text{ m}} \\ &= 1130 \text{ Hz}. \end{aligned}$$

10. —
    1. A wave with five antinodes represents the fifth harmonic.
    2. The scenario is depicted in Figure C.X.

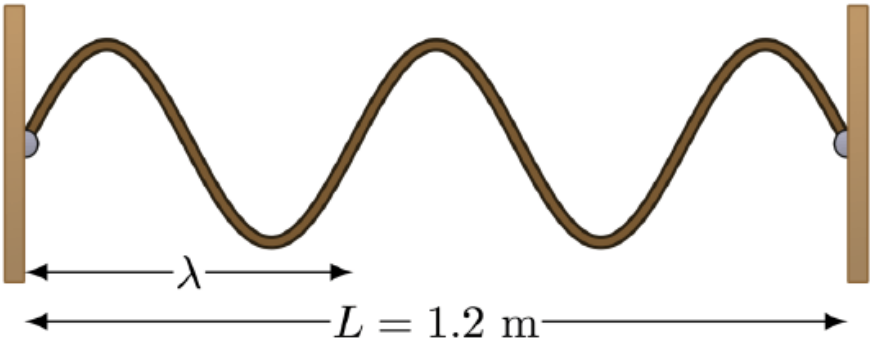


Figure C.15 – The fifth harmonic of a wave having a length of 1.2 m. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

There are 2.5 wavelengths between the two fixed ends. Therefore, the wavelength of this harmonic is equal to

$$\begin{aligned} L &= 2.5\lambda \\ \lambda &= \frac{L}{2.5} \\ &= \frac{1.2 \text{ m}}{2.5} \\ &= 0.48 \text{ m}. \end{aligned}$$

    3. The frequency of the wave is equal to

$$\begin{aligned} v &= \lambda f \\ f &= \frac{v}{\lambda} \\ &= \frac{8 \text{ m/s}}{0.48 \text{ m}} \\ &= 16.7 \text{ Hz}. \end{aligned}$$

Chapter 20

Numerical analysis

1. —
    1. The speed of sound at this temperature is equal to

$$\begin{aligned} v &= 331 \text{ m/s} + (0.6 \text{ m/(s}^\circ\text{C)})T \\ &= 331 \text{ m/s} + (0.6 \text{ m/(s}^\circ\text{C)})(0 \text{ }^\circ\text{C}) \\ &= 331 \text{ m/s}. \end{aligned}$$

    2. The speed of sound at this temperature is equal to

$$\begin{aligned} v &= 331\ \text{m/s} + (0.6\ \text{m/(s}^\circ\text{C)})T \\ &= 331\ \text{m/s} + (0.6\ \text{m/(s}^\circ\text{C)})(-30\ ^\circ\text{C}) \\ &= 331\ \text{m/s} - 18\ \text{m/s} \\ &= 313\ \text{m/s}. \end{aligned}$$

3. The speed of sound at this temperature is equal to

$$\begin{aligned} v &= 331\ \text{m/s} + (0.6\ \text{m/(s}^\circ\text{C)})T \\ &= 331\ \text{m/s} + (0.6\ \text{m/(s}^\circ\text{C)})(40\ ^\circ\text{C}) \\ &= 331\ \text{m/s} + 24\ \text{m/s} \\ &= 355\ \text{m/s}. \end{aligned}$$

2. The temperature of the lab is equal to

$$\begin{aligned} v &= 331\ \text{m/s} + (0.6\ \text{m/(s}^\circ\text{C)})T \\ T &= \frac{v - 331\ \text{m/s}}{0.6\ \text{m/(s}^\circ\text{C)}} \\ &= \frac{341\ \text{m/s} - 331\ \text{m/s}}{0.6\ \text{m/(s}^\circ\text{C)}} \\ &= \frac{10\ \text{m/s}}{0.6\ \text{m/(s}^\circ\text{C)}} \\ &= 16.7\ ^\circ\text{C}. \end{aligned}$$

3. The temperature of the oven is equal to

$$\begin{aligned} v &= 331\ \text{m/s} + (0.6\ \text{m/(s}^\circ\text{C)})T \\ T &= \frac{v - 331\ \text{m/s}}{0.6\ \text{m/(s}^\circ\text{C)}} \\ &= \frac{522\ \text{m/s} - 331\ \text{m/s}}{0.6\ \text{m/(s}^\circ\text{C)}} \\ &= \frac{191\ \text{m/s}}{0.6\ \text{m/(s}^\circ\text{C)}} \\ &= 318\ ^\circ\text{C}. \end{aligned}$$

4. The beat frequency is equal to

$$\begin{aligned} f_{beat} &= |f_1 - f_2| \\ &= |450\ \text{Hz} - 452\ \text{Hz}| \\ &= |-2\ \text{Hz}| \\ &= 2\ \text{Hz}. \end{aligned}$$

5. The beat frequency is equal to

$$\begin{aligned} f_{beat} &= |f_1 - f_2| \\ &= |680\ \text{Hz} - 685\ \text{Hz}| \\ &= |-5\ \text{Hz}| \\ &= 5\ \text{Hz}. \end{aligned}$$

6. The two possible frequencies are equal to

$$\begin{aligned} f_{beat} &= |f_1 - f_2| \\ &= \pm(f_1 - f_2) \\ &= f_{1,a} - f_2 \\ f_{1,a} &= f_{beat} + f_2 \\ &= 6\ \text{Hz} + 250\ \text{Hz} \\ &= 256\ \text{Hz}, \\ f_{beat} &= -f_{1,b} + f_2 \\ f_{1,b} &= f_2 - f_{beat} \\ &= 250\ \text{Hz} - 6\ \text{Hz} \\ &= 244\ \text{Hz}. \end{aligned}$$

Chapter 21

Conceptual comprehension

1. The sound wave that produced the given waveform is most likely a musical sound. This is due to the fact that the Fourier analysis exhibits mostly discrete frequencies.
2. The sound wave that produced the given waveform is most likely a noisy sound. This is due to the fact that the Fourier analysis exhibits mostly random frequencies.

Numerical analysis

1. The frequency of the peak value is equal to

$$\begin{aligned} f &= \frac{1}{T} \\ &= \frac{1}{3.8\text{ ms}\left(\frac{1\text{ s}}{1000\text{ ms}}\right)} \\ &= \frac{1}{0.0038\text{ s}} \\ &= 263.158\text{ Hz}. \end{aligned}$$

2. The frequency of the peak value is equal to

$$\begin{aligned} f &= \frac{1}{T} \\ &= \frac{1}{7.35\times10^{-4}\text{ s}} \\ &= 1360.544\text{ Hz}. \end{aligned}$$

3. The sound intensity in dB is equal to

$$\begin{aligned} I_{dB} &= 10\log_{10}\left(\frac{I_{W/m2}}{10^{-12}\text{ W/m}^2}\right) \\ &= 10\log_{10}\left(\frac{10^{-5}\text{ W/m}^2}{10^{-12}\text{ W/m}^2}\right) \\ &= 10\log_{10}\left(10^7\right) \\ &= (10)(7) \\ &= 70\text{ dB} \end{aligned}$$

4. —
    1. A 50 dB sound is 10x more intense than a 40 dB sound.
    2. A 50 dB sound is 100x more intense than a 30 dB sound.
    3. A 50 dB sound is 1,000x more intense than a 20 dB sound.
    4. A 50 dB sound is 10,000x more intense than a 10 dB sound.
5. The new intensity is 30 dB more intense than the original sound, giving it a total intensity of 110 dB.

Chapter 22

Numerical analysis

1. —
    1. Charge A would move to the right.
    2. Charge B would move to the left.
2. The magnitude of the force between the two charges is equal to

$$\begin{aligned}|F| &= k\frac{|q_1||q_2|}{d^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{(8\times10^{-6}\ \text{C})(12\times10^{-6}\ \text{C})}{\left((5\ \text{cm})\left(\frac{1\ \text{m}}{100\ \text{cm}}\right)\right)^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{9.6\times10^{-11}\ \text{C}^2}{(0.05\ \text{m})^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{9.6\times10^{-11}\ \text{C}^2}{0.0025\ \text{m}^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)(3.8\times10^{-8}\ \text{C}^2/\text{m}^2) \\ &= 346\ \text{N}.\end{aligned}$$

1. Charge A would move to the left.
2. Charge B would move to the right.

3. The magnitude of the electrostatic force is equal to

$$\begin{aligned}|F| &= k\frac{|q_1||q_2|}{d^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{(1.6\times10^{-19}\ \text{C})(1.6\times10^{-19}\ \text{C})}{(1\times10^{-10}\ \text{m})^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{2.56\times10^{-38}\ \text{C}^2}{1\times10^{-20}\ \text{m}^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)(2.56\times10^{-18}\ \text{C}^2/\text{m}^2) \\ &= 2.3\times10^{-8}\ \text{N}.\end{aligned}$$

4. The magnitudes of each charge due to the other two charges are equal to

$$\begin{aligned}|F_{6uC,-2uC}| &= k\frac{|q_{6uC}||q_{-2uC}|}{d^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{(6\times10^{-6}\ \text{C})(2\times10^{-6}\ \text{C})}{(0.1\ \text{m})^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{1.2\times10^{-11}\ \text{C}^2}{0.01\ \text{m}^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)(1.2\times10^{-9}\ \text{C}^2/\text{m}^2) \\ &= 10.8\ \text{N}, \\ |F_{6uC,8uC}| &= k\frac{|q_{6uC}||q_{8uC}|}{d^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{(6\times10^{-6}\ \text{C})(8\times10^{-6}\ \text{C})}{(0.2\ \text{m})^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{4.8\times10^{-11}\ \text{C}^2}{0.04\ \text{m}^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)(1.2\times10^{-9}\ \text{C}^2/\text{m}^2) \\ &= 10.8\ \text{N}, \\ |F_{-2uC,8uC}| &= k\frac{|q_{-2uC}||q_{8uC}|}{d^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{(2\times10^{-6}\ \text{C})(8\times10^{-6}\ \text{C})}{(0.1\ \text{m})^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)\frac{1.6\times10^{-11}\ \text{C}^2}{0.01\ \text{m}^2} \\ &= (9\times10^9\ \text{Nm}^2/\text{C}^2)(1.6\times10^{-9}\ \text{C}^2/\text{m}^2) \\ &= 14.4\ \text{N}.\end{aligned}$$

1. The net force acting on the +6 μC charge due to the other two charges is equal to

$$\begin{aligned}F_{net,6uC} &= +|F_{6uC,-2uC}| - |F_{6uC,8uC}| \\ &= 43.2\ \text{N} - 43.2\ \text{N} \\ &= 0\ \text{N}.\end{aligned}$$

2. The net force acting on the -2 μC charge due to the other two charges is equal to

$$\begin{aligned} F_{net,-2uC} &= -|F_{6uC,-2uC}| + |F_{-2uC,8uC}| \\ &= -43.2 \text{ N} + 57.6 \text{ N} \\ &= 14.4 \text{ N}. \end{aligned}$$

3. The net force acting on the +8 μC charge due to the other two charges is equal to

$$\begin{aligned} F_{net,8uC} &= +|F_{6uC,8uC}| - |F_{-2uC,8uC}| \\ &= 43.2 \text{ N} - 57.6 \text{ N} \\ &= -14.4 \text{ N}. \end{aligned}$$

Chapter 23

Conceptual comprehension

1. —
    1. Light bulbs A and B will both have the same amount of brightness. This is because they both have the equal amounts of voltage (because both light bulbs are identical, they will each share one half of the total voltage of the source) and equal amounts of current (because the circuit is a series circuit, the amount of current is the same at all points in the circuit), therefore they will have equal power and equal brightness.
    2. Light bulbs C and D will both have the same amount of brightness. This is because they both have the equal amounts of voltage (because the circuit is a parallel circuit, both light bulbs share the same voltage level as the voltage source) and equal amounts of current (because both light bulbs are identical, they have the same value of $I = V/R$), therefore they will have equal power and equal brightness.
    3. The series light bulbs will be dimmer than the parallel light bulbs. This is because each series light bulb has half the voltage as each parallel light bulb, and half the current as each parallel light bulb. Therefore, each series light bulb has one quarter the power as each parallel light bulb.
2. In a series circuit with three light bulbs, if one light bulb burns out, the other two light bulbs will turn off. This is because current can no longer flow through the circuit. In a circuit with three light bulbs wired in parallel, if one light bulb burns out, the other two light bulbs will not be affected. This is because only the path of the burnt out light bulb is effectively removed from the circuit, leaving the other two paths unaffected.
3. When you add more light bulbs to a series circuit, the brightness of all of the light bulbs will become dimmer (although each individual light bulb will still have the same amount of brightness as the others, if they are all identical). When you add more light bulbs to a parallel circuit, the brightness of all of the light bulbs will be unaffected. This is because each branch will have the same voltage as the source, and the same value of $I = V/R$, leaving the power in each branch unaffected by the number of branches.

Numerical analysis

1. The current through the resistor is equal to

$$\begin{aligned} V &= IR \\ I &= \frac{V}{R} \\ &= \frac{3 \text{ V}}{100 \ \Omega} \\ &= 0.03 \text{ A}. \end{aligned}$$

2. The resistance of the circuit elements is equal to

$$\begin{aligned} V &= IR \\ R &= \frac{V}{I} \\ &= \frac{12 \text{ V}}{0.5 \text{ A}} \\ &= 24 \ \Omega. \end{aligned}$$

3. The voltage is equal to

$$\begin{aligned} V &= IR \\ &= (0.12\ \text{A})(60\ \Omega) \\ &= 7.2\ \text{V}. \end{aligned}$$

4. The power consumed by the device is equal to

$$\begin{aligned} P &= IV \\ &= (5\ \text{A})(120\ \text{V}) \\ &= 600\ \text{W}. \end{aligned}$$

5. The current through the light bulb is equal to

$$\begin{aligned} P &= IV \\ I &= \frac{P}{V} \\ &= \frac{60\ \text{W}}{240\ \text{V}} \\ &= 0.25\ \text{A}. \end{aligned}$$

6. The power consumed by the resistor is equal to

$$\begin{aligned} P &= \frac{V^2}{R} \\ &= \frac{(6\ \text{V})^2}{40\ \Omega} \\ &= \frac{36\ \text{V}^2}{40\ \Omega} \\ &= 0.9\ \text{W}. \end{aligned}$$

7. —
    1. The total resistance of the circuit is equal to

$$\begin{aligned} R_{total} &= R_1 + R_2 \\ &= 800\ \Omega + 1200\ \Omega \\ &= 2000\ \Omega. \end{aligned}$$

While not explicitly asked for, it is also useful to calculate the current through both resistors, which is equal to

$$\begin{aligned} V &= IR \\ I &= \frac{V}{R} \\ &= \frac{10\ \text{V}}{2000\ \Omega} \\ &= 0.005\ \text{A}. \end{aligned}$$

    2. The power consumed by the 800 ohm resistor is equal to

$$\begin{aligned} P &= I^2R \\ &= (0.005\ \text{A})^2(800\ \Omega) \\ &= (2.5 \times 10^{-5}\ \text{A}^2)(800\ \Omega) \\ &= 0.02\ \text{W}. \end{aligned}$$

    3. The power consumed by the 1200 ohm resistor is equal to

$$\begin{aligned} P &= I^2R \\ &= (0.005\ \text{A})^2(1200\ \Omega) \\ &= (2.5 \times 10^{-5}\ \text{A}^2)(800\ \Omega) \\ &= 0.03\ \text{W}. \end{aligned}$$

8. —
    1. The total resistance of the circuit is equal to

$$\begin{aligned} R_{total} &= R_1 + R_2 + R_3 \\ &= 10\ \Omega + 15\ \Omega + 20\ \Omega \\ &= 45\ \Omega. \end{aligned}$$

While not explicitly asked for, it is also useful to calculate the current through the three resistors, which is equal to

$$\begin{aligned} V &= IR \\ I &= \frac{V}{R} \\ &= \frac{5\ \text{V}}{45\ \Omega} \\ &= 0.111\ \text{A}. \end{aligned}$$

    2. The power consumed by the 10 ohm resistor is equal to

$$\begin{aligned} P &= I^2 R \\ &= (0.111\ \text{A})^2 (10\ \Omega) \\ &= (0.0123\ \text{A}^2)(10\ \Omega) \\ &= 0.123\ \text{W}. \end{aligned}$$

    3. The power consumed by the 15 ohm resistor is equal to

$$\begin{aligned} P &= I^2 R \\ &= (0.111\ \text{A})^2 (10\ \Omega) \\ &= (0.0123\ \text{A}^2)(15\ \Omega) \\ &= 0.185\ \text{W}. \end{aligned}$$

    4. The power consumed by the 20 ohm resistor is equal to

$$\begin{aligned} P &= I^2 R \\ &= (0.111\ \text{A})^2 (20\ \Omega) \\ &= (0.0123\ \text{A}^2)(20\ \Omega) \\ &= 0.246\ \text{W}. \end{aligned}$$

9. —
    1. The total current will be distributed equally through each of the two branches. Therefore, each resistor will have a current of 0.05 A.
    2. The power consumed by each resistor is equal to

$$\begin{aligned} P &= IV \\ &= (0.05\ \text{A})(5\ \text{V}) \\ &= 0.25\ \text{W}. \end{aligned}$$

Chapter 24

Conceptual comprehension

3. Because the painted end of a compass needle is a north magnetic pole, and the painted end of a compass needle repels from the south pole of the Earth, the south pole of the Earth is a north magnetic pole. Like poles repel.

Chapter 25

Conceptual comprehension

1. This is a step-up transformer, due to the fact that the secondary contains more windings than the primary.
2. This is a step-down transformer, due to the fact that the secondary contains less windings than the primary.

Numerical analysis

1. The number of turns in the secondary coil is equal to

$$\begin{aligned}\frac{V_P}{N_P} &= \frac{V_S}{N_S} \\ N_S &= V_S\frac{N_P}{V_P} \\ &= (550\text{ V})\frac{400}{110\text{ V}} \\ &= (550\text{ V})(3.636\text{ V}^{-1}) \\ &= 2000.\end{aligned}$$

2. The number of turns in the secondary coil is equal to

$$\begin{aligned}\frac{V_P}{N_P} &= \frac{V_S}{N_S} \\ N_S &= V_S\frac{N_P}{V_P} \\ &= (4800\text{ V})\frac{200}{240\text{ V}} \\ &= (4800\text{ V})(0.833\text{ V}^{-1}) \\ &= 4000.\end{aligned}$$

3. The number of turns in the primary coil is equal to

$$\begin{aligned}\frac{V_P}{N_P} &= \frac{V_S}{N_S} \\ N_P &= V_P\frac{N_S}{V_S} \\ &= (120\text{ V})\frac{150}{600\text{ V}} \\ &= (120\text{ V})(0.25\text{ V}^{-1}) \\ &= 30.\end{aligned}$$

4. The number of turns in the secondary coil is equal to

$$\begin{aligned}\frac{V_P}{N_P} &= \frac{V_S}{N_S} \\ N_S &= V_S\frac{N_P}{V_P} \\ &= (120\text{ V})\frac{400}{480\text{ V}} \\ &= (120\text{ V})(0.833\text{ V}^{-1}) \\ &= 100.\end{aligned}$$

5. The number of turns in the primary coil is equal to

$$\begin{aligned}\frac{V_P}{N_P} &= \frac{V_S}{N_S} \\ N_P &= V_P \frac{N_S}{V_S} \\ &= (150\ \text{V})\frac{50}{30\ \text{V}} \\ &= (150\ \text{V})(1.667\ \text{V}^{-1}) \\ &= 250.\end{aligned}$$

6. The power consumed by each coil is equal to

$$\begin{aligned}P_P = P_S &= I_S V_S \\ &= (4\ \text{A})(1000\ \text{V}) \\ &= 4000\ \text{W}.\end{aligned}$$

7. The current through the secondary circuit is equal to

$$\begin{aligned}P_P &= P_S \\ I_P V_P &= I_S V_S \\ I_S &= \frac{I_P V_P}{V_S} \\ &= \frac{(5\ \text{A})(120\ \text{V})}{600\ \text{V}} \\ &= \frac{600\ \text{W}}{600\ \text{V}} \\ &= 1\ \text{A}.\end{aligned}$$

8. The current through the primary circuit is equal to

$$\begin{aligned}P_P &= P_S \\ I_P V_P &= I_S V_S \\ I_P &= \frac{I_S V_S}{V_P} \\ &= \frac{(2\ \text{A})(120\ \text{V})}{240\ \text{V}} \\ &= \frac{240\ \text{W}}{240\ \text{V}} \\ &= 1\ \text{A}.\end{aligned}$$

Chapter 26

Conceptual comprehension

1. Light is capable of traveling from the Sun, through the vacuum of space, to the Earth. If light required a medium in which to travel, we would not be able to experience the Sun's light on Earth.

Numerical analysis

1. —
    1. The amount of time it takes for an electromagnetic wave to travel from the Earth to the Moon is equal to

$$\begin{aligned}v &= \frac{d}{t} \\ t &= \frac{d}{v} \\ &= \frac{3.8 \times 10^8\ \text{m}}{3 \times 10^8\ \text{m/s}} \\ &= 1.267\ \text{s}.\end{aligned}$$

2. The amount of time it takes for an electromagnetic wave to travel from the Earth to Mars is equal to

$$\begin{aligned} v &= \frac{d}{t} \\ t &= \frac{d}{v} \\ &= \frac{76 \times 10^{9}\ \text{m}}{3 \times 10^{8}\ \text{m/s}} \\ &= 253.333\ \text{s}\left(\frac{1\ \text{min}}{60\ \text{s}}\right) \\ &= 4.222\ \text{min}. \end{aligned}$$

3. The amount of time it takes for an electromagnetic wave to travel from the Earth to Voyager 1 is equal to

$$\begin{aligned} v &= \frac{d}{t} \\ t &= \frac{d}{v} \\ &= \frac{24 \times 10^{12}\ \text{m}}{3 \times 10^{8}\ \text{m/s}} \\ &= 80000\ \text{s}\left(\frac{1\ \text{min}}{60\ \text{s}}\right) \\ &= 1333.333\ \text{min}\left(\frac{1\ \text{hr}}{60\ \text{min}}\right) \\ &= 22.222\ \text{hr}. \end{aligned}$$

2. The wavelength of the electromagnetic wave is equal to

$$\begin{aligned} \lambda &= \frac{c}{f} \\ &= \frac{3 \times 10^{8}\ \text{m/s}}{500\ \text{THz}\left(\frac{1 \times 10^{12}\ \text{Hz}}{1\ \text{THz}}\right)} \\ &= \frac{3 \times 10^{8}\ \text{m/s}}{500 \times 10^{12}\ \text{Hz}} \\ &= 6 \times 10^{-7}\ \text{m}\left(\frac{1\ \text{nm}}{1 \times 10^{-9}\ \text{m}}\right) \\ &= 600\ \text{nm}. \end{aligned}$$

3. The frequency of the electromagnetic wave is equal to

$$\begin{aligned} f &= \frac{c}{\lambda} \\ &= \frac{3 \times 10^{8}\ \text{m/s}}{2\ \text{mm}\left(\frac{1 \times 10^{-3}\ \text{m}}{1\ \text{mm}}\right)} \\ &= \frac{3 \times 10^{8}\ \text{m/s}}{2 \times 10^{-3}\ \text{m}} \\ &= 1.5 \times 10^{11}\ \text{Hz}\left(\frac{1\ \text{GHz}}{1 \times 10^{9}\ \text{Hz}}\right) \\ &= 150\ \text{GHz}. \end{aligned}$$

4. The wavelength of the WDCB radio waves is equal to

$$\begin{aligned} \lambda &= \frac{c}{f} \\ &= \frac{3 \times 10^{8}\ \text{m/s}}{90.9\ \text{MHz}\left(\frac{1 \times 10^{6}\ \text{Hz}}{1\ \text{MHz}}\right)} \\ &= \frac{3 \times 10^{8}\ \text{m/s}}{90.9 \times 10^{6}\ \text{Hz}} \\ &= 3.300\ \text{m}. \end{aligned}$$

5. The frequency of that particular wavelength of red light is equal to

$$\begin{aligned} f &= \frac{c}{\lambda} \\ &= \frac{3 \times 10^8 \text{ m/s}}{633 \text{ nm} \left(\frac{1 \times 10^{-9} \text{ m}}{1 \text{ nm}}\right)} \\ &= \frac{3 \times 10^8 \text{ m/s}}{633 \times 10^{-9} \text{ m}} \\ &= 4.739 \times 10^{14} \text{ Hz} \left(\frac{1 \text{ THz}}{1 \times 10^{12} \text{ m}}\right) \\ &= 473.9 \text{ THz}. \end{aligned}$$

Chapter 27

Conceptual comprehension

1. The color yellow will appear when shining equal intensities of red and green light onto a white surface. This is because yellow is the secondary color generated when red and green are mixed with color addition.
2. To generate colors such as orange or purple, use color addition to mix unequal intensities of red, green, and blue light until the desired effect is achieved. You can try this yourself using the PhET Color Vision simulation.
3. If you look through both a red and blue filter simultaneously, no light should reach your eyes and what you see should appear black. This is because the red filter will absorb (block) green and blue light, allowing only red light to transmit through. Then, the blue filter will absorb (block) that red light. This is depicted in Figure C.16.

> A diagram of two filters is shown. On the left is a red filter. Depicted to the left of the filter are three separate arrows depicting red, green, and blue light traveling from left to right. To the right of the red filter, only one arrow depicting red light traveling from left to right is shown. To the right of this is a blue filter. No light arrows are shown to the right of the blue filter.

Figure C.16 – A red filter blocks green and blue light. A subsequent blue filter will block that red light, allowing no light to transmit through the combination of red and blue filters. This figure, created by Izaak Neutelings and modified by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-SA 4.0.

4. When white light is shone through a green filter, only green light will transmit through. The green filter will absorb red and blue light. If other colors of light are shone through a green filter, only the green components of that light will transmit. If there are no green components (such as with red or blue light), then no light will be transmitted through the green filter. You can try this yourself using the PhET Color Vision simulation. You can also try this using the Physics Classroom Filtering Away simulation.
5. When white light is shone onto a cyan object, the object will appear cyan. This is because the object will absorb red light and reflect green and blue (cyan) light back to the viewer, explained by the concept of color subtraction. If red light is shone onto the cyan object, all of that red light will be absorbed, causing the object to appear black to the viewer. If green light is shone onto the cyan object, that green light will be reflected, causing the object to appear green. You can try this yourself using the Physics Classroom Stage Lighting simulation (if you want to see the effect on a cyan object, click on the stage repeatedly until the glass of water appears).
6. A blue object will appear to be black when illuminated with red light. This is because the blue object will absorb blue and green light and only reflect blue light, none of which is present in a red light source. You can try this yourself using the Physics Classroom Stage Lighting simulation (if you want to see the effect on a blue object, click on the stage repeatedly until the blueberry appears).

Chapter 30

Conceptual comprehension

1. The color of a star relates to its temperature. A hot star will appear more blue than a red star. (DAVE: Is this correct? Is this incandescence or is it due to emission spectra? Yes I'm an idiot and I don't know.)

Numerical analysis

1. The wavelength of this light wave is equal to

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ \lambda &= \frac{hc}{E} \\ &= \frac{(6.626 \times 10^{-34}\ \text{Js})(3 \times 10^{8}\ \text{m/s})}{2.18 \times 10^{-18}\ \text{J}} \\ &= \frac{1.9878 \times 10^{-25}\ \text{Jm}}{2.18 \times 10^{-18}\ \text{J}} \\ &= 9.1 \times 10^{-8}\ \text{m} \left(\frac{1\ \text{nm}}{1 \times 10^{-9}\ \text{m}} \right) \\ &= 91\ \text{nm}. \end{aligned}$$

This corresponds to a frequency of

$$\begin{aligned} f &= \frac{c}{\lambda} \\ &= \frac{3 \times 10^{8}\ \text{m/s}}{9.1 \times 10^{-8}\ \text{m}} \\ &= 3.297 \times 10^{15}\ \text{Hz} \left(\frac{1\ \text{PHz}}{1 \times 10^{15}\ \text{Hz}} \right) \\ &= 3.297\ \text{PHz}. \end{aligned}$$

2. —
    1. There are 15 total electron transitions. This is depicted in Figure C.17.

An energy-level diagram consisting of 6 levels. The ground state (E0, 0 J), E1 which has an energy level of 3.48E-19 J, E2 which has an energy level of 5.44E-19 J, E3 which has an energy level of 9.66E-19 J, E4 which has an energy level of 2.18E-18 J, and E5 which has an energy level of 8.70E-18 J. Each of the individual transitions between levels is depicted as an arrow. For example: there is an arrow pointing down from E5 to E0, from E5 to E1, from E5 to E2, and so on. There are 15 transition arrows indicating all of the possible electron transitions

Figure C.17 – There are 15 total electron transitions in this energy-level diagram. This figure, created by Alyssa J. Pasquale, Ph.D., is licensed under CC BY-NC-SA 4.0.

    2. The amount of energy required to jump from the ground state (E0) to E2 is equal to

$$\begin{aligned} E &= E_2 - E_0 \\ &= 5.44 \times 10^{-19}\ \text{J} - 0\ \text{J} \\ &= 5.44 \times 10^{-19}\ \text{J}. \end{aligned}$$

    3. The wavelength corresponding to the transition between the ground state (E0) to E2 is equal to

$$
\begin{aligned}
E &= \frac{hc}{\lambda} \\
\lambda &= \frac{hc}{E} \\
&= \frac{(6.626 \times 10^{-34}\ \text{Js})(3 \times 10^{8}\ \text{m/s})}{5.44 \times 10^{-19}\ \text{J}} \\
&= \frac{1.9878 \times 10^{-25}\ \text{Jm}}{5.44 \times 10^{-19}\ \text{J}} \\
&= 3.65 \times 10^{-7}\ \text{m} \left(\frac{1\ \text{nm}}{1 \times 10^{-9}\ \text{m}} \right) \\
&= 365\ \text{nm}.
\end{aligned}
$$

4. The largest energy transition is the jump between levels E5 and E0, which corresponds to an energy of

$$
\begin{aligned}
E &= E_5 - E_0 \\
&= 8.7 \times 10^{-18}\ \text{J} - 0\ \text{J} \\
&= 8.7 \times 10^{-18}\ \text{J}.
\end{aligned}
$$

5. The wavelength corresponding to the largest energy transition is equal to

$$
\begin{aligned}
E &= \frac{hc}{\lambda} \\
\lambda &= \frac{hc}{E} \\
&= \frac{(6.626 \times 10^{-34}\ \text{Js})(3 \times 10^{8}\ \text{m/s})}{8.7 \times 10^{-18}\ \text{J}} \\
&= \frac{1.9878 \times 10^{-25}\ \text{Jm}}{8.7 \times 10^{-18}\ \text{J}} \\
&= 2.3 \times 10^{-8}\ \text{m} \left(\frac{1\ \text{nm}}{1 \times 10^{-9}\ \text{m}} \right) \\
&= 23\ \text{nm}.
\end{aligned}
$$

6. The smallest energy transition is the jump between levels E2 and E1, which corresponds to an energy of

$$
\begin{aligned}
E &= E_2 - E_1 \\
&= 5.44 \times 10^{-19}\ \text{J} - 3.48 \times 10^{-19}\ \text{J} \\
&= 1.96 \times 10^{-19}\ \text{J}.
\end{aligned}
$$

7. The wavelength corresponding to the largest energy transition is equal to

$$
\begin{aligned}
E &= \frac{hc}{\lambda} \\
\lambda &= \frac{hc}{E} \\
&= \frac{(6.626 \times 10^{-34}\ \text{Js})(3 \times 10^{8}\ \text{m/s})}{1.96 \times 10^{-19}\ \text{J}} \\
&= \frac{1.9878 \times 10^{-25}\ \text{Jm}}{1.96 \times 10^{-19}\ \text{J}} \\
&= 1.014 \times 10^{-6}\ \text{m} \left(\frac{1\ \mu\text{m}}{1 \times 10^{-6}\ \text{m}} \right) \\
&= 1.014\ \mu\text{m}.
\end{aligned}
$$

Chapter 31

Numerical analysis

1. The velocity of the Olympic sprinter is equal to

$$\begin{aligned} v &= \frac{d}{t} \\ &= \frac{100\text{ m}}{9.58\text{ s}} \\ &= 10.438\text{ m/s}. \end{aligned}$$

The wavelength of the Olympic sprinter is therefore equal to

$$\begin{aligned} \lambda &= \frac{h}{p} \\ &= \frac{h}{mv} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{(94\text{ kg})(10.438\text{ m/s})} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{981.211\text{ kgm/s}} \\ &= 6.753 \times 10^{-37}\text{ m}. \end{aligned}$$

2. The wavelength of the cyclist and her bicycle is equal to

$$\begin{aligned} \lambda &= \frac{h}{p} \\ &= \frac{h}{mv} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{(100\text{ kg})(11.2\text{ m/s})} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{1120\text{ kgm/s}} \\ &= 5.91 \times 10^{-37}\text{ m}. \end{aligned}$$

3. The wavelength of the ice hockey puck is equal to

$$\begin{aligned} \lambda &= \frac{h}{p} \\ &= \frac{h}{mv} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{(160\text{ g})\left(\frac{1\text{ kg}}{1000\text{ g}}\right)(51\text{ m/s})} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{(0.160\text{ kg})(51\text{ m/s})} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{8.16\text{ kgm/s}} \\ &= 8.12 \times 10^{-35}\text{ m}. \end{aligned}$$

4. The wavelength of the electron is equal to

$$\begin{aligned} \lambda &= \frac{h}{p} \\ &= \frac{h}{mv} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{(9.1 \times 10^{-31}\text{ kg})(1.45 \times 10^{8}\text{ m/s})} \\ &= \frac{6.626 \times 10^{-34}\text{ Js}}{1.3195 \times 10^{-22}\text{ kgm/s}} \\ &= 5.022 \times 10^{-12}\text{ m}\left(\frac{1\text{ pm}}{1 \times 10^{-12}\text{ m}}\right) \\ &= 5.022\text{ pm}. \end{aligned}$$

5. The wavelength of the neutron is equal to

$$\begin{aligned}\lambda &= \frac{h}{p} \\ &= \frac{h}{mv} \\ &= \frac{6.626 \times 10^{-34}\ \text{Js}}{(1.67 \times 10^{-27}\ \text{kg})(2290\ \text{m/s})} \\ &= \frac{6.626 \times 10^{-34}\ \text{Js}}{3.824 \times 10^{-24}\ \text{kgm/s}} \\ &= 1.73 \times 10^{-10}\ \text{m}\left(\frac{1\ \text{pm}}{1 \times 10^{-12}\ \text{m}}\right) \\ &= 173\ \text{pm}.\end{aligned}$$

6. The wavelength of the helium atom is equal to

$$\begin{aligned}\lambda &= \frac{h}{p} \\ &= \frac{h}{mv} \\ &= \frac{6.626 \times 10^{-34}\ \text{Js}}{(6.64 \times 10^{-27}\ \text{kg})(1780\ \text{m/s})} \\ &= \frac{6.626 \times 10^{-34}\ \text{Js}}{1.182 \times 10^{-23}\ \text{kgm/s}} \\ &= 5.606 \times 10^{-11}\ \text{m}\left(\frac{1\ \text{pm}}{1 \times 10^{-12}\ \text{m}}\right) \\ &= 56.06\ \text{pm}.\end{aligned}$$

7. The wavelength of the sodium atom is equal to

$$\begin{aligned}\lambda &= \frac{h}{p} \\ &= \frac{h}{mv} \\ &= \frac{6.626 \times 10^{-34}\ \text{Js}}{(3.82 \times 10^{-26}\ \text{kg})(870\ \text{m/s})} \\ &= \frac{6.626 \times 10^{-34}\ \text{Js}}{3.3234 \times 10^{-23}\ \text{kgm/s}} \\ &= 1.994 \times 10^{-11}\ \text{m}\left(\frac{1\ \text{pm}}{1 \times 10^{-12}\ \text{m}}\right) \\ &= 19.94\ \text{pm}.\end{aligned}$$

Chapter 32

Numerical analysis

Chapter 33

Conceptual comprehension

1. Americium-241 must undergo beta-minus decay to become plutonium. This is the only decay process by which a nucleus can gain a proton. This particular decay process can be summarized by the equation

$$^{241}_{94}\text{Pu} \rightarrow ^{241}_{95}\text{Am} + ^{0}_{-1}\text{e}.$$

2. After any given interval of time, more undecayed nickel-63 will be present compared to undecayed nickel-56. This is because nickel-63 has a much longer half-life than nickel-56, so it will take much longer for the sample to decay.

Numerical analysis

1. The number of nucleons in the new nucleus is equal to

$$\begin{aligned} A_{new} &= A - 4 \\ &= 241 - 4 \\ &= 237. \end{aligned}$$

The number of protons in the new nucleus is equal to

$$\begin{aligned} Z_{new} &= Z - 2 \\ &= 95 - 2 \\ &= 93. \end{aligned}$$

This new nucleus corresponds to neptunium-237.

2. The number of nucleons in the new nucleus is equal to

$$\begin{aligned} A_{new} &= A - 4 \\ &= 223 - 4 \\ &= 219. \end{aligned}$$

The number of protons in the new nucleus is equal to

$$\begin{aligned} Z_{new} &= Z - 2 \\ &= 88 - 2 \\ &= 86. \end{aligned}$$

This new nucleus corresponds to radon-219.

3. The number of nucleons in the new nucleus is equal to 14, as beta decay will not change the number of nucleons in a nucleus. The number of protons in the new nucleus is equal to

$$\begin{aligned} Z_{new} &= Z + 1 \\ &= 6 + 1 \\ &= 7. \end{aligned}$$

This new nucleus corresponds to nitrogen-14.

4. —
    1. Polonium-218 undergoes beta-minus decay to become astatine-218, which can be summarized by the decay equation

$$^{218}_{84}\mathrm{Po} \rightarrow ^{218}_{85}\mathrm{At} + ^{0}_{-1}\mathrm{e}.$$

    2. Bismuth-210 undergoes alpha decay to become thallium-214, which can be summarized by the decay equation

$$^{210}_{83}\mathrm{Bi} \rightarrow ^{214}_{81}\mathrm{Tl} + ^{4}_{2}\mathrm{He}.$$

    3. Aluminum-26 undergoes beta-plus decay to become magnesium-26, which can be summarized by the decay equation

$$^{26}_{13}\mathrm{Al} \rightarrow ^{26}_{12}\mathrm{Mg} + ^{0}_{1}\mathrm{e}.$$

    4. Nitrogen-11 undergoes proton emission to become carbon-10, which can be summarized by the decay equation

$$^{11}_{7}\mathrm{N} \rightarrow ^{10}_{6}\mathrm{C} + ^{1}_{1}\mathrm{p}.$$

5. DAVE: This is tricky as stated. Lead-206, 207, and 208 are the only stable isotopes of lead but that's not clear from the question. I think we should make that clear. "Which of the following stable isotopes of lead could be the result: lead-206, lead-207, or lead-208?" There are multiple decay chains for uranium-234 so if a student is just looking this up they may find a graphic like this which shows three different lead outcomes: https://en.wikipedia.org/wiki/Decay_...um_series).svg
6. See above. (Correct answer: four beta-minus decays, 7 alpha decays.)

7. The number of half-lives that have elapsed after 22.8 days is equal to

$$\begin{aligned} n &= \frac{\text{time}}{\text{half-life}} \\ &= \frac{22.8 \text{ days}}{11.4 \text{ days}} \\ &= 2. \end{aligned}$$

Therefore, the amount of undecayed radium-223 left after this period of time is equal to

$$\begin{aligned} m_{new} &= (0.5)^n m_{original} \\ &= (0.5)^2 (100 \text{ g}) \\ &= 0.25 (100 \text{ g}) \\ &= 25 \text{ g}. \end{aligned}$$

The number of half-lives that have elapsed after 45.6 days is equal to

$$\begin{aligned} n &= \frac{\text{time}}{\text{half-life}} \\ &= \frac{45.6 \text{ days}}{11.4 \text{ days}} \\ &= 4. \end{aligned}$$

Therefore, the amount of undecayed radium-223 left after this period of time is equal to

$$\begin{aligned} m_{new} &= (0.5)^n m_{original} \\ &= (0.5)^4 (100 \text{ g}) \\ &= 0.0625 (100 \text{ g}) \\ &= 6.25 \text{ g}. \end{aligned}$$

8. Two half-lives have elapsed to result in only 25% undecayed element remaining after 8 days. This means that a single half-life is equal to 4 days.
9. In one year, the flight crew can fly a number of hours equal to

$$\begin{aligned} t &= \frac{\text{exposure limit}}{\text{exposure rate}} \\ &= \frac{5 \text{ rem}}{0.06 \text{ rem/hr}} \\ &= 83.333 \text{ hr}. \end{aligned}$$

Chapter 34

Conceptual comprehension

1. In a commercial nuclear fission reactor that generates electricity, the most desirable neutron multiplication number is equal to one. This is to prevent the reaction from dying out (which will happen if k is less than one) or from growing out of bounds and causing an explosion (which will happen if k is greater than one).
2. Two neutrons will be generated in this reaction. This is the only outcome consistent with conservation of nucleons. The reaction equation can be expressed as

$$^{235}_{92}\text{U} + ^{1}_{0}\text{n} \rightarrow ^{236}_{92}\text{U} \rightarrow ^{137}_{55}\text{Cs} + ^{97}_{37}\text{Rb} + ^{2}_{0}\text{n}.$$

3. The atomic mass number of the zirconium isotope involved in this reaction is equal to 97. This is the only outcome consistent with conservation of nucleons. The reaction equation can be expressed as

$$^{235}_{92}\text{U} + ^{1}_{0}\text{n} \rightarrow ^{236}_{92}\text{U} \rightarrow ^{137}_{52}\text{Te} + ^{97}_{40}\text{Zr} + ^{2}_{0}\text{n}.$$

Numerical analysis

1. The total number of fission events occurring during the 6th cycle is equal to

$$\begin{aligned} \text{events} &= 2^{\text{cycle}} \\ &= 2^6 \\ &= 64. \end{aligned}$$

The total number of fissions that have taken place since the start of the reaction is equal to

$$\begin{aligned} n &= 64 + 32 + 16 + 8 + 4 + 2 + 1 \\ &= 127. \end{aligned}$$

2. The amount of time it will take for 1023 fission events to occur is equal to

$$\begin{aligned} t &= (\text{events})(\text{event rate}) \\ &= (1023)(10 \text{ ns}) \\ &= 10230 \text{ ns}. \end{aligned}$$

3. It will take 9 cycles to create 1023 fission events when k=2. This result can be demonstrated in the table below.

fission events occurring	1	2	4	8	16	32	64	128	256	512
total fission events	1	3	7	15	31	63	127	255	511	1023

Therefore, the amount of time it takes for 1023 fission reactions to occur is equal to

$$\begin{aligned} t &= (\text{events})(\text{event rate}) \\ &= (9)(10 \text{ ns}) \\ &= 90 \text{ ns}. \end{aligned}$$

This is

$$\frac{10230 \text{ ns}}{90 \text{ ns}} \\ 113.667$$

times faster than the critical reaction.

4. The rate at which mass is converted to energy is equal to mass divided by time. This rate is equal to

$$\begin{aligned} E &= mc^2 \\ \frac{E}{t} &= \frac{m}{t}c^2 \\ \frac{m}{t} &= \frac{P}{c^2} \\ &= \frac{3.86 \times 10^{26} \text{ kgm}^2/\text{s}^3}{(3 \times 10^8 \text{ m/s})^2} \\ &= \frac{3.86 \times 10^{26} \text{ kgm}^2/\text{s}^3}{9 \times 1016 \text{ m}^2/\text{s}^2} \\ &= 4.289 \times 10^9 \text{ kg/s}. \end{aligned}$$

5. The final mass of all byproducts is equal to

$$\begin{aligned} m_{final} &= m_{\text{Np}} + m_{\text{He}} \\ &= 240.056162 \text{ u} + 4.001506 \text{ u} \\ &= 244.057668 \text{ u}. \end{aligned}$$

Therefore, the energy released in this reaction is equal to

$$\begin{aligned} E &= \Delta mc^2 \\ &= |m_{final} - m_{initial}|c^2 \\ &= |244.057668 \text{ u} - 244.064204 \text{ u}|(931.5 \text{ MeV/u}) \\ &= |-0.006536 \text{ u}|(931.5 \text{ MeV/u}) \\ &= (0.006536 \text{ u})(931.5 \text{ MeV/u}) \\ &= 6.088 \text{ MeV}. \end{aligned}$$

6. The initial mass is equal to

$$\begin{aligned} m_{initial} &= m_{\text{n}} + m_{\text{U}} \\ &= 1.008665 \text{ u} + 235.043930 \text{ u} \\ &= 236.052595 \text{ u}. \end{aligned}$$

The final mass is equal to

$$\begin{aligned} m_{final} &= m_{\text{Xe}} + m_{\text{Sr}} + 2m_{\text{n}} \\ &= 143.93851 \text{ u} + 89.907738 \text{ u} + 2(1.008665 \text{ u}) \\ &= 143.93851 \text{ u} + 89.907738 \text{ u} + 2.01733 \text{ u} \\ &= 235.863578 \text{ u}. \end{aligned}$$

Therefore, the energy released in this reaction is equal to

$$\begin{aligned} E &= \Delta mc^2 \\ &= |m_{final} - m_{initial}|c^2 \\ &= |235.863578 \text{ u} - 236.052595 \text{ u}|(931.5 \text{ MeV/u}) \\ &= |-0.189017 \text{ u}|(931.5 \text{ MeV/u}) \\ &= (0.189017 \text{ u})(931.5 \text{ MeV/u}) \\ &= 176.069 \text{ MeV}. \end{aligned}$$

7. —
    1. The initial mass is equal to

$$\begin{aligned} m_{initial} &= 2(m_{\text{H-2}}) \\ &= 2(2.013553 \text{ u}) \\ &= 4.027106 \text{ u}. \end{aligned}$$

The final mass is equal to

$$\begin{aligned} m_{final} &= m_{\text{H-3}} + m_{\text{H-1}} \\ &= 3.015501 \text{ u} + 1.007276 \text{ u} \\ &= 4.022777 \text{ u}. \end{aligned}$$

Therefore, the energy released in this reaction is equal to

$$\begin{aligned} E &= \Delta mc^2 \\ &= |m_{final} - m_{initial}|c^2 \\ &= |4.022777 \text{ u} - 4.027106 \text{ u}|(931.5 \text{ MeV/u}) \\ &= |-0.004329 \text{ u}|(931.5 \text{ MeV/u}) \\ &= (0.004329 \text{ u})(931.5 \text{ MeV/u}) \\ &= 4.032 \text{ MeV}. \end{aligned}$$

    2. The initial mass is equal to

$$\begin{aligned} m_{initial} &= 2(m_{\text{He-3}}) \\ &= 2(3.016029 \text{ u}) \\ &= 6.032058 \text{ u}. \end{aligned}$$

The final mass is equal to

$$\begin{aligned} m_{final} &= m_{\text{He-4}} + 2(m_{\text{H-1}}) \\ &= 4.001506 \text{ u} + 2(1.007276) \text{ u} \\ &= 4.001506 \text{ u} + 2.014552 \text{ u} \\ &= 6.016058 \text{ u}. \end{aligned}$$

Therefore, the energy released in this reaction is equal to

$$\begin{aligned} E &= \Delta mc^2 \\ &= |m_{final} - m_{initial}|c^2 \\ &= |6.016058 \text{ u} - 6.032058 \text{ u}|(931.5 \text{ MeV/u}) \\ &= |-0.016 \text{ u}|(931.5 \text{ MeV/u}) \\ &= (0.016 \text{ u})(931.5 \text{ MeV/u}) \\ &= 14.904 \text{ MeV}. \end{aligned}$$

8. The average nuclear binding energy per nucleon for nickel-62 is 8.79 MeV. Therefore, the amount of energy it takes to disassemble a nickel-62 nucleus into its individual protons and neutrons is equal to

$$\begin{aligned} E &= (\text{average energy})(\text{nucleons}) \\ &= (8.79 \text{ MeV/nucleon})(62 \text{ nucleons}) \\ &= 544.98 \text{ MeV}. \end{aligned}$$

9. The average mass per nucleon of a carbon-12 nucleus is equal to 1 u. (This is the definition of 1 u.)
10. 12 μg of carbon-12 can be converted to a mass using the unit of unified atomic mass, which is equal to

$$\begin{aligned} m &= (12 \ \mu\text{g}) \left(\frac{1 \text{ g}}{1 \times 10^{-6} \ \mu\text{g}} \right) \left(\frac{1 \text{ kg}}{1000 \text{ g}} \right) \left(\frac{1 \text{ u}}{1.66054 \times 10^{-27} \text{ kg}} \right) \\ &= 7.226 \times 10^{18} \text{ u}. \end{aligned}$$

The amount of energy required to disassociate this mass is equal to

$$\begin{aligned} E &= mc^2 \\ &= (7.226 \times 10^{18} \text{ u})(931.5 \text{ MeV/u}) \\ &= 6.732 \times 10^{21} \text{ MeV} \left(\frac{1.60218 \times 10^{-19} \text{ J}}{1 \text{ eV}} \right) \left(\frac{1 \times 10^{6} \text{ eV}}{1 \text{ MeV}} \right) \left(\frac{1 \text{ MJ}}{1 \times 10^{6} \text{ J}} \right) \\ &= 1078.515 \text{ MJ}. \end{aligned}$$

The amount of gasoline required to supply this energy is equal to

$$\begin{aligned} V &= (1078.515 \text{ MJ}) \left(\frac{1 \text{ gal}}{140 \text{ MJ}} \right) \\ &= 7.704 \text{ gal}. \end{aligned}$$

This page titled Appendix C: Answers to selected questions is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

INDEX

A

acceleration
- Chapter 10: Projectile and satellite motion
- Chapter 3: Linear motion
- Chapter 4: Newton's second law
- Chapter 5: Newton's third law
- Chapter 9: Gravity

acoustics
- Chapter 21: Musical sounds
- V: Sound and waves

action reaction pairs
- Chapter 5: Newton's third law

alpha decay
- Chapter 33: Nuclear physics and radioactivity

amorphous_solids
- Chapter 12: Solids

amplitude
- Chapter 20: Sound
- Chapter 21: Musical sounds

angles
- Chapter 28: Reflection and refraction

angular momentum
- Chapter 8: Rotational Motion

applications
- Chapter 29: Light waves

applied force
- Appendix C: Answers to selected questions
- Chapter 4: Newton's second law

Archimedes principle
- Chapter 13: Liquids

atmosphere
- Chapter 14: Gases

atmospheric phenomena
- Chapter 14: Gases

atmospheric scattering
- Chapter 27: Color

atomic model
- Chapter 32: Quantum physics

atomic nature of matter
- III: Matter

atomic nucleus
- Chapter 34: Fission and fusion

atoms
- Chapter 11: The atomic nature of matter

average speed
- Chapter 3: Linear motion

average velocity
- Chapter 3: Linear motion

B

base units
- Appendix B: Units and unit conversions

batteries
- Chapter 23: Electric circuits

beta decay
- Chapter 33: Nuclear physics and radioactivity

Bohr model
- Chapter 32: Quantum physics

boiling point
- Chapter 17: Phase change

Boyle's law
- Chapter 14: Gases

buoyancy
- Chapter 14: Gases

buoyant force
- Chapter 13: Liquids

C

categories of matter
- Chapter 11: The atomic nature of matter

chain reaction
- Chapter 34: Fission and fusion

change in momentum
- Chapter 6: Momentum

chapter 19
- V: Sound and waves

chapter 20
- V: Sound and waves

chapter 21
- V: Sound and waves

charge
- Chapter 22: Electrostatics

chemical composition
- Chapter 11: The atomic nature of matter

circuit elements
- Chapter 23: Electric circuits

circular motion
- Chapter 8: Rotational Motion

classical physics
- Chapter 31: Light Quanta

climate change
- Chapter 15: Temperature and Heat

collisions
- Chapter 6: Momentum
- Chapter 7: Work and energy

color
- Chapter 27: Color
- VI: Electricity and magnetism

color perception
- Chapter 27: Color

color theory
- Chapter 27: Color

communication technology
- Chapter 26: Light

compass
- Chapter 24: Magnetism

components
- Appendix A: Vectors and vector arithmetic

condensation
- Chapter 17: Phase change

conduction
- Chapter 16: Heat transfer

conductors
- Chapter 16: Heat transfer

conservation
- Chapter 7: Work and energy

conservation of momentum
- Chapter 6: Momentum

convection
- Chapter 16: Heat transfer

conversion
- Appendix B: Units and unit conversions

Coulomb's law
- Chapter 22: Electrostatics

crystalline_solids
- Chapter 12: Solids

current
- Chapter 23: Electric circuits
- Chapter 25: Electromagnetic induction

D

deexcitation
- Chapter 30: Light emission

density
- Chapter 12: Solids
- Chapter 13: Liquids
- Chapter 14: Gases

derived units
- Appendix B: Units and unit conversions

direction
- Appendix A: Vectors and vector arithmetic

dispersion
- Chapter 29: Light waves

displacement
- Appendix A: Vectors and vector arithmetic
- Appendix C: Answers to selected questions
- Chapter 3: Linear motion

distance
- Chapter 9: Gravity

dynamic equilibrium
- Chapter 2: Newton's first law

dynamics
- Chapter 6: Momentum

E

Earth's magnetic field
- Chapter 24: Magnetism

elasticity
- Chapter 12: Solids

electric circuits
- Chapter 22: Electrostatics
- Chapter 23: Electric circuits
- VI: Electricity and magnetism

electric field
- Chapter 25: Electromagnetic induction

electric fields
- Chapter 22: Electrostatics

electrical energy
- Chapter 25: Electromagnetic induction

electrical engineering
- Chapter 23: Electric circuits

electromagnet
- Chapter 24: Magnetism

electromagnetic energy
- Chapter 31: Light Quanta

electromagnetic induction
- Chapter 25: Electromagnetic induction
- VI: Electricity and magnetism

electromagnetic spectrum
Chapter 26: Light
Chapter 27: Color
electromagnetic waves
Chapter 26: Light
electromagnetism
Chapter 26: Light
electronics
Chapter 23: Electric circuits
electrostatic attraction
Chapter 22: Electrostatics
electrostatic repulsion
Chapter 22: Electrostatics
electrostatics
Chapter 22: Electrostatics
VI: Electricity and magnetism
emission spectrum
Chapter 30: Light emission
energy
Appendix C: Answers to selected questions
Chapter 17: Phase change
Chapter 18: Thermodynamics
Chapter 34: Fission and fusion
Chapter 4: Newton's second law
Chapter 7: Work and energy
II: Kinematics
energy levels
Chapter 30: Light emission
energy transfer
Chapter 15: Temperature and Heat
engineering
Chapter 1: About science
entropy
Chapter 18: Thermodynamics
equilibrium
Appendix A: Vectors and vector arithmetic
Chapter 13: Liquids
Chapter 2: Newton's first law
evidence
Appendix C: Answers to selected questions
excitation
Chapter 30: Light emission
excited state
Chapter 30: Light emission
experimentation
Chapter 1: About science
external forces
Chapter 5: Newton's third law

F

Faraday's law
Chapter 25: Electromagnetic induction
first law
II: Kinematics
first law of thermodynamics
Chapter 18: Thermodynamics
fissile nuclei
Chapter 34: Fission and fusion
fission
VII: Quantum and nuclear physics
fissionable nuclei
Chapter 34: Fission and fusion
fluid dynamics
Chapter 13: Liquids
fluid mechanics
Chapter 13: Liquids
force
Chapter 2: Newton's first law
Chapter 4: Newton's second law
Chapter 6: Momentum
forces
Appendix A: Vectors and vector arithmetic
Chapter 10: Projectile and satellite motion
Chapter 5: Newton's third law
free fall
Chapter 10: Projectile and satellite motion
Chapter 9: Gravity
freezing point
Chapter 17: Phase change
frequency
Chapter 19: Vibrations and Waves
Chapter 20: Sound
Chapter 21: Musical sounds
friction
Chapter 4: Newton's second law
Chapter 7: Work and energy
fusion
VII: Quantum and nuclear physics

G

gamma radiation
Chapter 33: Nuclear physics and radioactivity
gas
Chapter 14: Gases
gases
III: Matter
gravitational force
Appendix C: Answers to selected questions
Chapter 13: Liquids
Chapter 9: Gravity
gravity
Chapter 10: Projectile and satellite motion
Chapter 2: Newton's first law
Chapter 4: Newton's second law
Chapter 5: Newton's third law
Chapter 9: Gravity
II: Kinematics
ground state
Chapter 30: Light emission
grounding
Chapter 22: Electrostatics

H

heat
Chapter 15: Temperature and Heat
IV: Thermodynamics
heat exchange
Chapter 16: Heat transfer
heat flow
Chapter 18: Thermodynamics
heat transfer
Chapter 16: Heat transfer
Chapter 17: Phase change
Chapter 18: Thermodynamics
IV: Thermodynamics
heterogeneous mixtures
Chapter 11: The atomic nature of matter
homogeneous mixtures
Chapter 11: The atomic nature of matter
horizontal motion
Chapter 10: Projectile and satellite motion
human hearing
Chapter 21: Musical sounds

I

image
Chapter 28: Reflection and refraction
image formation
Chapter 28: Reflection and refraction
impulse
Chapter 6: Momentum
incandescence
Chapter 30: Light emission
induction
Chapter 25: Electromagnetic induction
inertia
Chapter 2: Newton's first law
innovations
Chapter 1: About science
instantaneous velocity
Chapter 3: Linear motion
insulators
Chapter 16: Heat transfer
intensity
Chapter 21: Musical sounds
interaction
Chapter 5: Newton's third law
interference
Chapter 29: Light waves
internal combustion engines
Chapter 18: Thermodynamics
ionizing radiation
Chapter 33: Nuclear physics and radioactivity
isotopes
Chapter 33: Nuclear physics and radioactivity
Chapter 34: Fission and fusion

K

kinematics
Chapter 10: Projectile and satellite motion
Chapter 3: Linear motion
kinetic energy
Chapter 7: Work and energy

L

lasers
Chapter 30: Light emission
latent heat
Chapter 17: Phase change
law of reflection
Chapter 28: Reflection and refraction
light
Chapter 26: Light
Chapter 28: Reflection and refraction
Chapter 29: Light waves
Chapter 31: Light Quanta
VI: Electricity and magnetism
light emission
Chapter 30: Light emission
VI: Electricity and magnetism
light quanta
VII: Quantum and nuclear physics
light receptor cells
Chapter 27: Color
light waves
Chapter 27: Color
VI: Electricity and magnetism

linear motion
Chapter 3: Linear motion
II: Kinematics
liquids
III: Matter
longitudinal waves
Chapter 19: Vibrations and Waves
loudness
Chapter 20: Sound
Chapter 21: Musical sounds

M

magnetic attraction
Chapter 24: Magnetism
magnetic confinement
VII: Quantum and nuclear physics
magnetic field
Chapter 25: Electromagnetic induction
magnetic fields
Chapter 24: Magnetism
magnetic interactions
Chapter 24: Magnetism
magnetic objects
Chapter 24: Magnetism
magnetic repulsion
Chapter 24: Magnetism
magnetism
Chapter 24: Magnetism
VI: Electricity and magnetism
magnitude
Appendix A: Vectors and vector arithmetic
mass
Chapter 4: Newton's second law
Chapter 9: Gravity
materials science
Chapter 15: Temperature and Heat
materials_science
Chapter 12: Solids
mathematical concepts
Chapter 1: About science
matter
Chapter 11: The atomic nature of matter
Chapter 13: Liquids
Maxwell equations
Chapter 26: Light
measurement
Appendix B: Units and unit conversions
mechanical energy
Chapter 7: Work and energy
mechanical equilibrium
Appendix C: Answers to selected questions
mechanics
Chapter 2: Newton's first law
Chapter 8: Rotational Motion
Chapter 9: Gravity
metric system
Appendix B: Units and unit conversions
mirror
Chapter 28: Reflection and refraction
mixtures
Chapter 11: The atomic nature of matter
molecular_arrangement
Chapter 12: Solids
molecules
Chapter 11: The atomic nature of matter
moment of inertia
Chapter 8: Rotational Motion
momentum
Appendix C: Answers to selected questions
Chapter 4: Newton's second law
Chapter 6: Momentum
Chapter 7: Work and energy
II: Kinematics
motion
Chapter 2: Newton's first law
Chapter 4: Newton's second law
Chapter 5: Newton's third law
motion analysis
Chapter 3: Linear motion
music
Chapter 21: Musical sounds
musical instruments
Chapter 20: Sound
Chapter 21: Musical sounds
musical sounds
V: Sound and waves

N

negative charge
Chapter 22: Electrostatics
net force
Appendix C: Answers to selected questions
neutron bombardment
Chapter 34: Fission and fusion
Newton
Chapter 2: Newton's first law
II: Kinematics
Newton's law
Chapter 9: Gravity
Newton's laws
Chapter 10: Projectile and satellite motion
Chapter 6: Momentum
newtons third law
Chapter 5: Newton's third law
nitrogen
Chapter 14: Gases
nuclear fission
Chapter 34: Fission and fusion
nuclear fusion
Chapter 34: Fission and fusion
nuclear physics
VII: Quantum and nuclear physics
nuclear reactions
Chapter 33: Nuclear physics and radioactivity

O

Ohm's law
Chapter 23: Electric circuits
oxygen
Chapter 14: Gases

P

permanent magnets
Chapter 24: Magnetism
phase change
Chapter 17: Phase change
IV: Thermodynamics
photons
Chapter 31: Light Quanta
physics
Appendix B: Units and unit conversions
Chapter 1: About science
Chapter 3: Linear motion
Chapter 5: Newton's third law
Chapter 6: Momentum
Chapter 8: Rotational Motion
V: Sound and waves
physics textbook
Chapter 26: Light
pitch
Chapter 20: Sound
planetary motion
Chapter 9: Gravity
polycrystalline_materials
Chapter 12: Solids
positive charge
Chapter 22: Electrostatics
potential energy
Chapter 7: Work and energy
power
Chapter 23: Electric circuits
Chapter 7: Work and energy
power transmission
Chapter 25: Electromagnetic induction
prefixes
Appendix B: Units and unit conversions
pressure
Chapter 13: Liquids
Chapter 14: Gases
projectile motion
Chapter 10: Projectile and satellite motion
properties
Chapter 29: Light waves
properties of substances
Chapter 11: The atomic nature of matter
propulsion
Chapter 5: Newton's third law
pure substances
Chapter 11: The atomic nature of matter

Q

quantity
Appendix A: Vectors and vector arithmetic
quantum entanglement
Chapter 32: Quantum physics
quantum mechanics
Chapter 32: Quantum physics
quantum physics
Chapter 31: Light Quanta
VII: Quantum and nuclear physics
quantum theory
Chapter 31: Light Quanta

R

radiation exposure
Chapter 33: Nuclear physics and radioactivity
radiation shielding
Chapter 33: Nuclear physics and radioactivity
radioactive decay
Chapter 33: Nuclear physics and radioactivity
radioactivity
VII: Quantum and nuclear physics
ray optics
Chapter 29: Light waves

real image
Chapter 28: Reflection and refraction
reflection
Chapter 28: Reflection and refraction
Chapter 29: Light waves
VI: Electricity and magnetism
refraction
Chapter 28: Reflection and refraction
Chapter 29: Light waves
VI: Electricity and magnetism
resistance
Chapter 23: Electric circuits
resonance
Chapter 20: Sound
rotational dynamics
Chapter 8: Rotational Motion
rotational motion
Chapter 8: Rotational Motion
II: Kinematics
rotational speed
Chapter 8: Rotational Motion

S

scalar
Appendix A: Vectors and vector arithmetic
scattering
Chapter 27: Color
Schrödinger equation
Chapter 32: Quantum physics
science
Chapter 1: About science
scientific community
Chapter 1: About science
scientific law
Appendix C: Answers to selected questions
scientific method
Chapter 1: About science
scientific theory
Appendix C: Answers to selected questions
scientific units
Appendix B: Units and unit conversions
second law
II: Kinematics
second law of thermodynamics
Chapter 18: Thermodynamics
SI
Appendix B: Units and unit conversions
skepticism
Chapter 1: About science
solids
Chapter 12: Solids
III: Matter
sound
Chapter 21: Musical sounds
V: Sound and waves
sound propagation
Chapter 20: Sound
sound waves
Chapter 19: Vibrations and Waves
Chapter 20: Sound
specific heat capacity
Chapter 15: Temperature and Heat
spectroscopy
Chapter 30: Light emission
speed of light
Chapter 26: Light
speed of sound
Chapter 20: Sound
stability
Chapter 8: Rotational Motion
states_of_matter
Chapter 12: Solids
static equilibrium
Chapter 2: Newton's first law
structural_properties
Chapter 12: Solids
superposition
Chapter 32: Quantum physics

T

technology
Chapter 1: About science
temperature
Chapter 15: Temperature and Heat
Chapter 16: Heat transfer
Chapter 17: Phase change
Chapter 18: Thermodynamics
IV: Thermodynamics
temperature measurement
Chapter 15: Temperature and Heat
tension
Chapter 4: Newton's second law
textbook
V: Sound and waves
textbook chapter
III: Matter
thermal conductivity
Chapter 16: Heat transfer
thermal energy
Chapter 16: Heat transfer
thermal expansion
Chapter 15: Temperature and Heat
thermodynamics
Chapter 15: Temperature and Heat
Chapter 16: Heat transfer
Chapter 17: Phase change
Chapter 18: Thermodynamics
IV: Thermodynamics
thermometers
Chapter 15: Temperature and Heat
thermonuclear fusion
VII: Quantum and nuclear physics
third law
II: Kinematics
timbre
Chapter 21: Musical sounds
torque
Chapter 8: Rotational Motion
transformer
Chapter 25: Electromagnetic induction
transverse waves
Chapter 19: Vibrations and Waves

U

uncertainty principle
Chapter 31: Light Quanta
Chapter 32: Quantum physics
units
Appendix B: Units and unit conversions
universal gravitation
Chapter 9: Gravity

V

vaporization
Chapter 17: Phase change
vector
Appendix A: Vectors and vector arithmetic
Chapter 2: Newton's first law
vector addition
Appendix A: Vectors and vector arithmetic
vector quantities
Chapter 3: Linear motion
velocity
Chapter 10: Projectile and satellite motion
Chapter 6: Momentum
vertical motion
Chapter 10: Projectile and satellite motion
vibrations
Chapter 20: Sound
V: Sound and waves
virtual image
Chapter 28: Reflection and refraction
visible light
Chapter 27: Color
voltage
Chapter 23: Electric circuits
Chapter 25: Electromagnetic induction
volume
Chapter 13: Liquids

W

wave
Chapter 29: Light waves
wave amplitude
Chapter 19: Vibrations and Waves
wave frequency
Chapter 26: Light
wave function
Chapter 32: Quantum physics
wave optics
Chapter 29: Light waves
wave period
Chapter 19: Vibrations and Waves
wave properties
Chapter 19: Vibrations and Waves
wave speed
Chapter 19: Vibrations and Waves
wave wavelength
Chapter 26: Light
wavelength
Chapter 19: Vibrations and Waves
wavelengths
Chapter 27: Color
waves
Chapter 19: Vibrations and Waves
V: Sound and waves
weather forecasting
Chapter 14: Gases
work
Chapter 18: Thermodynamics
Chapter 7: Work and energy
II: Kinematics

This page titled Index is shared under a license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

DETAILED LICENSING

Overview

Title: Conceptual Physics (Pasquale, Fazzini, and Bennett)

Webpages: 58

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC 4.0: 82.8% (48 pages)
- Undeclared: 17.2% (10 pages)

By Page

- Conceptual Physics (Pasquale, Fazzini, and Bennett) - *CC BY-NC 4.0*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - About This Book - *CC BY-NC 4.0*
    - Disclaimer - *CC BY-NC 4.0*
    - Licensing - *Undeclared*
  - I: Introduction - *CC BY-NC 4.0*
    - Chapter 1: About science - *CC BY-NC 4.0*
  - II: Kinematics - *CC BY-NC 4.0*
    - Chapter 2: Newton's first law - *CC BY-NC 4.0*
    - Chapter 3: Linear motion - *CC BY-NC 4.0*
    - Chapter 4: Newton's second law - *CC BY-NC 4.0*
    - Chapter 5: Newton's third law - *CC BY-NC 4.0*
    - Chapter 6: Momentum - *CC BY-NC 4.0*
    - Chapter 7: Work and energy - *CC BY-NC 4.0*
    - Chapter 8: Rotational Motion - *CC BY-NC 4.0*
    - Chapter 9: Gravity - *CC BY-NC 4.0*
    - Chapter 10: Projectile and satellite motion - *CC BY-NC 4.0*
  - III: Matter - *CC BY-NC 4.0*
    - Chapter 11: The atomic nature of matter - *CC BY-NC 4.0*
    - Chapter 12: Solids - *CC BY-NC 4.0*
    - Chapter 13: Liquids - *CC BY-NC 4.0*
    - Chapter 14: Gases - *CC BY-NC 4.0*
  - IV: Thermodynamics - *CC BY-NC 4.0*
    - Chapter 15: Temperature and Heat - *CC BY-NC 4.0*
    - Chapter 16: Heat transfer - *CC BY-NC 4.0*
    - Chapter 17: Phase change - *CC BY-NC 4.0*
    - Chapter 18: Thermodynamics - *CC BY-NC 4.0*
  - V: Sound and waves - *CC BY-NC 4.0*
    - Chapter 19: Vibrations and Waves - *CC BY-NC 4.0*
    - Chapter 20: Sound - *CC BY-NC 4.0*
    - Chapter 21: Musical sounds - *CC BY-NC 4.0*
  - VI: Electricity and magnetism - *CC BY-NC 4.0*
    - Chapter 22: Electrostatics - *CC BY-NC 4.0*
    - Chapter 23: Electric circuits - *CC BY-NC 4.0*
    - Chapter 24: Magnetism - *CC BY-NC 4.0*
    - Chapter 25: Electromagnetic induction - *CC BY-NC 4.0*
    - Chapter 26: Light - *CC BY-NC 4.0*
    - Chapter 27: Color - *CC BY-NC 4.0*
    - Chapter 28: Reflection and refraction - *CC BY-NC 4.0*
    - Chapter 29: Light waves - *CC BY-NC 4.0*
    - Chapter 30: Light emission - *CC BY-NC 4.0*
  - VII: Quantum and nuclear physics - *CC BY-NC 4.0*
    - Chapter 31: Light Quanta - *CC BY-NC 4.0*
    - Chapter 32: Quantum physics - *CC BY-NC 4.0*
    - Chapter 33: Nuclear physics and radioactivity - *CC BY-NC 4.0*
    - Chapter 34: Fission and fusion - *CC BY-NC 4.0*
  - Appendices - *Undeclared*
    - Appendix A: Vectors and vector arithmetic - *CC BY-NC 4.0*
    - Appendix B: Units and unit conversions - *CC BY-NC 4.0*
    - Appendix C: Answers to selected questions - *CC BY-NC 4.0*
  - Back Matter - *Undeclared*
    - Index - *Undeclared*
    - Glossary - *Undeclared*
    - Detailed Licensing - *Undeclared*
    - Glossary (1) - *CC BY-NC 4.0*

This page titled Conceptual Physics (Pasquale, Fazzini, and Bennett) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.

GLOSSARY

4

Absolute zero

Absolute zero is the thermodynamic limit corresponding to the lowest possible temperature. Absolute zero is defined to be 0 K.

Absorption

Absorption of light occurs when light waves completely transfer their energy to a medium as they attempt to pass through it. The light is therefore blocked from transmitting through the medium.

Absorption spectrum

An absorption spectrum shows the wavelengths of light that are absorbed by a material. It appears as a continuous spectrum with dark lines caused by light passing through a substance.

Acceleration

Acceleration quantifies the rate at which an object changes its velocity. (symbol: **a**, unit: m/s^2)

Adiabatic process

An adiabatic process is a specific type of thermodynamic process in which the heat added to or removed from a system is zero.

Alpha particle

An alpha particle is a helium nucleus (consisting of two protons and two neutrons).

Alternating current

Alternating current (AC) is a property of a voltage or current source in which the value of the voltage or current changes periodically with time.

Amorphous

Amorphous solids have atomic or molecular arrangements that are spatially disordered, or that otherwise lack a regular periodic arrangement.

Amplitude (wave)

Wave amplitude describes the maximum displacement from either side of the equilibrium position.

Apparent weight

Apparent weight is equal to the value of the support force acting on a person. This causes our perception of having weight.

Archimedes' principle

Archimedes' principle states that the buoyant force that acts on a substance is equal to the weight of the fluid that is displaced when that substance is placed into a fluid.

Atomic mass number

Atomic mass number is equal to the number of nucleons (protons and neutrons) in an atom. (symbol: A)

Atomic number

The atomic number uniquely defines an element and is equal to the number of protons in atoms of that element. (symbol: Z)

Ballistics

Ballistics is the study and analysis of how projectiles travel.

Beat

A beat is an interference phenomenon resulting from the time interference of two sound waves at very similar frequencies. It is perceived as a periodic change in amplitude of the composite wave.

Bernoulli's principle

Bernoulli's principle relates pressure to the speed and elevation of a fluid. If the elevation is constant, then as the speed of a fluid increases, the pressure decreases and vice versa.

Binding energy (nuclear)

Binding energy describes the energy required to dissociate a nucleus into its constituent nucleons.

Birefringence

Birefringence is a property of some materials that causes the index of refraction to be different based on the polarization of light.

Boyle's law

Boyle's law describes the relationship between pressure and volume of a gas when the temperature and number of molecules in a closed system is held constant. It states that the pressure times the volume is constant.

Brewster's angle

When unpolarized light reflects off of a surface at a particular angle, called Brewster's angle, the reflected light is polarized 100% in a direction parallel to the reflecting surface.

Capacitor

A capacitor is a circuit element that stores energy in the form of an electric field. It is commonly used in circuits to smooth out changes in voltage to create a DC signal.

Carnot efficiency

Carnot efficiency describes the maximum efficiency that can result from any heat engine, even if it is perfectly ideal in every other way: no friction or other losses. (symbol: e_carnot, unit: none)

Center of mass

The center of mass is the point in an object (or system of objects) around which it rotates freely. It is considered to be the average position of all the mass in a system.

Centripetal force

Centripetal force is a force, pointing toward the center of the rotational motion of an object, that causes an object to move in a circular path. (symbol: **F_c**, unit: N)

Chain reaction

A chain reaction is a self-sustaining fission process whereby one fission reaction provokes one or more subsequent chain reactions.

Charge

Electric charge is a fundamental property of matter that causes particles that carry a charge to experience a force when in the presence of an electric field.

Circuit

A circuit is an interconnection of elements such as batteries, light bulbs, resistors, capacitors, and diodes.

Classical physics

Classical physics describes the natural phenomena that can be explained by Newton's laws of motion, Maxwell's equations, and the laws of thermodynamics.

Coherent light

Coherent light means that all of the waves have the same frequency and phase.

Color addition

Color addition refers to the ability to use light to create secondary colors from primary colors (red, green, and blue).

Color subtraction

Color subtraction is a color model based on light reflection and absorption. Colors that are not reflected back into our eyes are subtracted out and modify the color that we perceive.

Compression

Compression is the force that exists when something is pushed or squeezed together.

Condensation

Condensation is the phase change from a gas to a liquid.

Conduction (electricity)

Conduction is the motion of electrical charge within a material or between different materials.

Conduction (heat)

Conduction is heat transfer that occurs due to collisions among atoms, electrons, or molecules.

Convection

Convection is heat transfer that comes about due to the motion of molecules themselves.

Critical mass

Critical mass describes the quantity of a fissionable material above which energy is released so quickly that an explosion can result.

Crystalline

A crystalline solid has constituent atoms or molecules that are arranged in an orderly, repeating manner.

Current

Current describes the flow of electric charge through a circuit. (symbol: I, unit: A)

Density

Density describes the compactness of matter in a substance, defined as the mass per unit volume. (symbol: D, unit: kg/m^3)

Deposition

Deposition is the phase change process when a gas turns directly into a solid without passing through the liquid form.

Diffraction

Diffraction is a bending of a wave as it travels around a corner or through a narrow opening.

Diode

A diode is a circuit element that acts like a one-way valve in that it only allows current to flow in one direction but not the other.

Direct current

Direct current (DC) is a property of a voltage or current source in which the value of the voltage or current remains constant over time.

Displacement

Displacement describes the straight-line distance from one point to another. Displacement is a vector quantity. (symbols: $\Delta\mathbf{d}$ [for general displacement], $\Delta\mathbf{x}$ [for horizontal displacement], $\Delta\mathbf{y}$ [for vertical displacement]. unit: m).

Doppler effect

The Doppler effect is the perceived change in frequency of a sound wave as heard by an observer when there is motion between the observer and the source.

Dynamic equilibrium

Dynamic equilibrium is a condition of mechanical equilibrium where either the velocity or rotational velocity of an object is constant and not equal to zero.

Efficiency

Efficiency quantifies how well work input is converted to work output. (symbol: e, unit: none)

Elastic collision

An elastic collision occurs when there is neither lasting deformation of objects in the collision, nor generation of heat during the collision. Both momentum and kinetic energy are conserved in elastic collisions.

Elasticity

Elasticity describes how much an object deforms when experiencing a stress or force.

Electric potential

Electric potential describes the amount of work it takes to move a charge through an electric field divided by the amount of charge. (symbol: EP, unit: V)

Electromagnetic radiation (heat)

Electromagnetic radiation is heat that is transferred from one place to another through light waves.

Electromagnetic spectrum

The electromagnetic spectrum describes the entire range of electromagnetic waves. Each subset of waves can be defined in different categories based on its frequency or wavelength.

Electromagnetic wave

An electromagnetic wave is a transverse wave that contains electric and magnetic fields oscillating at right angles to each other.

Electromagnetism

Electromagnetism describes the interactions that occur between charged objects.

Electron

An electron is a fundamental building block of matter that has a negative charge and is found surrounding the nucleus of an atom.

Electron deexcitation

Electron deexcitation is the process by which an electron loses energy and jumps down one or more energy levels. In the process, the electron either transfers energy to another particle, emits a photon, or generates heat.

Electron excitation

Electron excitation is the process by which an electron gains energy and moves up one or more energy levels.

Element

An element is a substance that consists of atoms whose nuclei have the same atomic number. An element cannot be broken down further by chemical means.

Emission spectrum

The collection of frequencies of light emitted by a substance via the electron deexcitation process is known as its emission spectrum. It appears as bright lines on a dark background.

Energy

Energy is defined as the capability of an object (or collection of objects) to do useful work. (symbol: E, unit: J)

Energy level

Energy levels describe the quantity of energy that each electron orbiting an atom can have. Each atom is uniquely defined by its discrete energy level values. The lowest energy level is known as the ground state.

Entropy

Entropy is a thermodynamic property that describes the amount of disorder in a system. In particular, it is related to the amount of energy that becomes unavailable to do work.

Fissile

A fissile nucleus is an isotope that can undergo fission after being bombarded with a slow neutron (known as a “thermal” neutron).

Fission

Fission is the splitting of heavy atomic nuclei into lighter pieces.

Fissionable

A fissionable nucleus describes an unstable nucleus that will undergo fission after being bombarded with a suitable neutron.

Fluorescence

Fluorescence is a process of light emission where an object absorbs light at a high frequency, causing electron excitation, and during the deexcitation process emits light at a lower frequency.

Force

A force is a push or a pull that causes an object to change its motion. More fundamentally, force is an interaction between two objects. (symbol: **F**, unit: N)

Forced vibration

A forced vibration occurs when one object causes another object to vibrate at a certain frequency.

Fourier analysis

Fourier analysis decomposes a wave into individual frequency components.

Free fall

Free fall is defined as a situation where gravity is the *only* force acting on an object. Forces such as friction and air drag are ignored and assumed to be equal to zero. The acceleration of an object in free fall is defined by the acceleration due to gravity (g).

Freezing

Freezing is the phase change from a liquid to a solid.

Frequency

Frequency describes how many oscillations occur per second in a wave. (symbol: f, unit: Hz)

Fusion

Fusion is the combining of light atomic nuclei into a single heavier nucleus.

Galvanometer

A galvanometer is a mechanical instrument capable of measuring electrical current by detecting a needle proportional to the amount of current flowing through the circuit.

Gas

Gas is a state of matter where molecules are very free to move about and generally do not interact with each other except during collisions. This means that the shape and volume of a gas is free to change. Gas is one of the four most common phases of matter.

Generator

A generator converts mechanical work to electrical energy.

Gravity

Gravity is the attractive force experienced by objects of mass. It is one of the four fundamental forces.

Grounding

Grounding is a special type of conduction where charges redistribute themselves to an area that can accept lots of charges without causing a large buildup of electrons or positive ions.

Half-life

Half-life is a description of the amount of time it takes, on average, for half of a particular unstable isotope to decay.

Heat

Heat is energy that is transferred from one object to another in response to a difference in temperature. (symbol: Q, unit: cal)

Huygens' principle

Huygens' principle states that every point on a wave can be treated as a source of a spherical wave that itself propagates through time and space. Each of these individual spherical waves interact with each other through interference, with the sum representing the total wave.

Hypothesis

A hypothesis is a proposed explanation of how the universe works that is based on observations and can be tested experimentally.

Impulse

Impulse is the measure of the change in momentum of an object or of a system of objects. (symbol: $\Delta\mathbf{p}$, unit: kgm/s or Ns)

Incandescence

Incandescence is a form of light emission and heat transfer process by which an object emits electromagnetic radiation with a peak frequency that corresponds to the temperature of that object.

Induction (electromagnetic)

Electromagnetic induction is the process by which changing magnetic fields produce electric fields, and vice versa.

Induction (electrostatic)

Induction is a redistribution of electric charge due to the presence of a nearby electric field.

Inelastic collision

In an inelastic collision, there is either lasting deformation of the objects in the collision, or heat/sound generated during the collision (or both). In an inelastic collision, momentum is conserved, but kinetic energy is not (as it is converted to other forms of energy during the collision).

Inertia

Inertia is a property of matter whereby objects resist changes in their motion. Inertia is related to the mass of an object.

Infrasound

Infrasound is the set of sound waves that have frequencies below 20 Hz. They are not audible to humans.

Intensity

Intensity describes the amount of energy per unit of area that reaches a certain location every second (unit: W/m^2)

Interference

Interference is the phenomenon that occurs when two waves overlap. The total amount of wave energy present at a point is equal to the sum of each individual wave's energy at that point.

Ion

An ion is a charged atom; there is an unequal number of protons and electrons in the atom.

Ionization

Ionization is the process by which a gas is converted into a plasma.

Ionizing radiation

Ionizing radiation refers to electromagnetic waves (UV, x-rays, and gamma rays) that contain sufficient energy to ionize atoms. These waves can cause damage to human DNA and other tissue.

Isotope

Elements that have the same number of protons but different numbers of neutrons are known as isotopes.

Kinetic energy

Kinetic energy is the energy that an object has due to its motion. (symbol: KE, unit: J)

Kinetic friction

Kinetic friction is the force that occurs between surfaces as they slide past each other. The force acts parallel to the contact surfaces.

Latent heat of fusion

Latent heat of fusion describes the amount of heat required to convert one gram of a substance from a solid form to a liquid form. Alternatively, it defines the amount of heat that must be removed from one gram of a substance to convert it from a liquid to a solid.

Latent heat of vaporization

Latent heat of vaporization describes the amount of heat required to convert one gram of a substance from a liquid form to a gaseous form. Alternatively, it defines the amount of heat that must be removed from one gram of a substance to convert it from a gas to a liquid.

Law

A scientific law is a statement that describes relationships among the quantities that we observe or measure.

Law of reflection

The law of reflection states that the angle of reflection of a wave is equal to the angle of incidence of that wave.

Liquid

A liquid is a state of matter in which the constituent molecules will change their shape or arrangement but cannot be easily compressed to change their volume. Liquid is one of the four most common phases of matter.

Longitudinal wave

In a longitudinal wave, the oscillations of the wave travel in the same direction as the wave itself.

Loudness

Loudness has to do with the subjective quality of the amplitude of a sound wave.

Magnetic domain

A magnetic domain is an area inside of a material where the magnetic effect of each individual atom couples to those of neighboring atoms, causing the internal magnets to be aligned.

Magnetism

Magnetism is a phenomenon that occurs when moving electric charges generate attractive and repulsive forces.

Mass

Mass is a property of physical objects that relates to resistance to changes in motion: inertia. (symbol: m, unit: kg)

Mechanical advantage

Mechanical advantage is a force multiplication that can be provided by a simple machine. (symbol: MA, unit: none)

Mechanical energy

Mechanical energy is the sum of kinetic energy and potential energy of a system. When ignoring dissipative forces such as air drag and friction, mechanical energy will be conserved.

Mechanical equilibrium

Mechanical equilibrium is a condition where the net force and net torque acting on an object are equal to zero. Two cases of mechanical equilibrium are static and dynamic. An object in mechanical equilibrium has a constant velocity (constant speed in one direction).

Melting

Melting is the phase change that occurs when a substance converts from a solid to a liquid.

Mixture

A mixture is a substance where the properties of that substance (such as density, melting point, and specific heat capacity) vary based on the concentration of each component in the mixture.

Momentum

Momentum is a vector property that quantifies the motion of an object. It is sometimes called "inertia in motion" as it is the product of mass and velocity. (symbol: **p**, unit: kgm/s or Ns)

Motor

A motor converts electrical energy to mechanical work.

Natural frequency

The natural frequency of an object refers to the frequency that the object will tend to produce.

Net force

The net force is the vector sum of all forces acting on an object.

Neutron

A neutron is a subatomic particle that has no charge and resides in the nucleus of an atom.

Newton's first law

Newton's first law is sometimes called the law of inertia. Newton's first law states that "an object in motion tends to stay in motion, and an object at rest tends to stay at rest, unless acted upon by an outside force."

Newton's second law

Newton's second law states that the acceleration of an object is proportional to the net force acting on the object, and inversely proportional to the mass of the object. In equation form, Newton's second law can be stated as $\mathbf{F_{net}} = m\mathbf{a}$.

Newton's third law

Newton's third law states that for every action, there is an equal and opposite reaction. All forces exist in action-reaction pairs that are equal in magnitude and opposite in direction.

Nuclide

A nuclide is a nucleus of a particular isotope.

Observer effect

The observer effect describes the process that occurs when measuring a quantum process. When making an observation, the process is influenced in some way, and therefore affects the outcome.

Ohm's law

Ohm's law describes the relationship between voltage, resistance, and current in an electric circuit. Ohm's law states that $V = IR$.

Opaque

An object is opaque if it does not allow any light to transmit through it.

Optical aberration

An optical aberration occurs when light (incident on a mirror or traveling through a lens) is not focused at a single point. Instead, light rays are spread out around the focal point.

Orbit (electron)

The orbit of an electron describes the properties of the electron wave as it exists around the nucleus of an atom.

Orbit (satellite)

The orbit of a satellite describes its path around a planet or star.

Parallel circuit

A parallel circuit is one where circuit elements are all contained in their own path in the circuit.

Path

Path describes the total distance that an object travels as it moves from one point to another, measured along its trajectory. It is a scalar quantity. (symbols: d, x [for horizontal path], y [for vertical path], unit: m)

Perfectly inelastic collision

In a perfectly inelastic collision, the maximum amount of kinetic energy is lost in a collision. In a perfectly inelastic collision between two objects, the two objects will travel together after the collision.

Period (wave)

The period of a wave describes how long it takes for one complete oscillation to occur. (symbol: T, unit: s)

Phosphorescence

Phosphorescence is a light emission phenomenon similar to fluorescence. However, the deexcitation process takes place on a timescale that extends after the excitation source is removed.

Photoelectric effect

The photoelectric effect is the emission of electrons when light of a particular frequency (especially UV light) strikes a metal target.

Physics

Physics is a branch of science that focuses on the fundamentals of the workings of our universe.

Plasma

Plasma is ionized gas. Plasma is the most common phase of matter in the universe.

Polarization (light)

Polarization is the property of transverse waves which allows them to have their electric field components oscillating in a single direction.

Potential energy

Potential energy is energy that exists based on the arrangement or configuration of a substance.

Power

Power quantifies how quickly work is done. (symbol: P, unit: W)

Pressure

Pressure defines the amount of force applied over an area of a substance. (symbol: P, unit: Pa)

Principle of flotation

The principle of flotation states that any object that floats displaces a weight of fluid equal to its own weight.

Principle of least time

The principle of least time describes how light travels. Light does not necessarily take the shortest distance between two points. Instead, it will travel whatever path takes the least amount of time.

Proton

A proton is a subatomic particle that has a positive charge and resides in the nucleus of an atom.

Pseudoscience

Pseudoscience is the misuse of scientific phrases and terminology, usually to mislead others into purchasing a product that sounds “high tech.”

Pure substance

In a pure substance, the properties of a substance (density, melting point, specific heat capacity, etc.) are inherent to the material.

Quantum physics

Quantum physics describes the physical laws used to describe the properties of matter and energy that occur on a very small scale.

Quark

A quark is a fundamental building block of matter, described by the standard model of physics. It is the building block of protons and neutrons.

Radioactivity

Radioactivity is the process by which an unstable nucleus releases energy and emits a particle.

Recombination

Recombination is the process by which a plasma changes phase to a gas.

Reflection

Reflection occurs when a wave bounces off of a surface and subsequently changes its direction.

Refraction

Refraction occurs when a wave changes speed (and therefore bends) at an interface between two different media.

Resistance

Resistance is the property of a material that impedes the flow of electrons throughout that material. (symbol: R, unit: Ω)

Resonance

Resonance occurs when an object is forced to vibrate at its natural frequency, which could lead to vibrations of increasing amplitude.

Rotational inertia

Rotational inertia is a measure of how much an object resists changes to its rotational speed.

Rotational speed

Rotational speed quantifies how quickly (or slowly) something rotates about an axis. (symbol: ω, unit: RPM, degrees/s, rad/s, etc.)

Scalar

A scalar quantity is a variable that can be conveyed with a numerical quantity indicating magnitude or strength.

Scattering (wave)

Scattering is a process where the direction that a wave travels in is changed without affecting the wavelength.

Science

Science is the systematic pursuit of knowledge that comes about through asking questions, making careful observations, and carrying out experiments about how the universe works.

Scientific method

The scientific method is the process by which scientists come to answer questions about the universe.

Series circuit

A series circuit is one where all of the circuit elements are contained in a single path.

Short circuit

A short circuit is a circuit that contains a very low-resistance path along which electrons will travel. It can result in excess current, which may cause the circuit to overheat.

Simple machine

A simple machine is used to change either the direction or magnitude (or both) of an applied force. There are six types of simple machines: levers, wheels and axles, pulleys, inclined planes, wedges, and screws.

Solar constant

The solar constant is the rate at which we obtain energy from the sun. The solar constant is approximately 1,400 W/m^2.

Solid

A solid is a substance where the molecules or atoms are very tightly bound together. This gives a solid a very rigid volume and shape. Solid is one of the four most common phases of matter.

Sound wave

A sound wave is a longitudinal wave that is caused by vibrations in air pressure between two points.

Specific heat capacity

Specific heat capacity defines how difficult it is to change the temperature of a substance. It describes the amount of heat required to change the temperature of a certain mass of a substance by a certain temperature. (symbol: c, unit: cal/(g°C))

Speed

Speed is the scalar quantity that describes the rate at which an object changes its position. Speed is equal to position divided by time. (symbols: s, |v|, unit: m/s)

Spring constant

The spring constant defines the stiffness of an elastic object such as a spring. The spring constant is equal to the force required to deflect a spring a certain distance. (symbol: k, unit: N/m)

Standard atomic weight

The standard atomic weight is the weighted average of the mass of naturally occurring isotopes of the element.

Standing wave

A standing wave is a wave that oscillates in time but that does not move in space. Standing waves occur due to interference between two waves that have equal amplitude and frequency that travel in opposite directions.

Static equilibrium

Static equilibrium is a special case of mechanical equilibrium where both the velocity and rotational speed of an object are constant and equal to zero.

Static friction

Static friction is the force that must be overcome in order to get an object at rest to start moving.

Sublimation

Sublimation is the process by which a substance goes directly from a solid to a gas without going through the liquid phase.

System

A system is any collection of objects that we define. The concept of a system is used to determine what is internal or external when discussing concepts such as momentum conservation.

Tangential velocity

Tangential velocity quantifies the linear distance that something travels as it rotates per unit of time. The direction of the tangential velocity is tangent to the motion of the rotation. (symbol: **v_T**, unit: m/s)

Technology

Technology is the outcome of using scientific principles to solve problems. This can be a product, innovation, or other object whose creation is based on science such as physics, chemistry, or biology.

Temperature

Temperature defines the average kinetic energy of an object. It quantifies the “hotness” or “coldness” of something. (symbol: T, unit: °C or K)

Tension

Tension is the force that exists when something is pulled or stretched apart.

Theory

A theory is an explanation of how the universe works that is supported by multiple, repeated experiments.

Thermodynamics

Thermodynamics is the study of heat flow and its relationship to temperature, work, energy, and entropy.

Timbre

The timbre of a sound wave describes the qualities of that wave that make it distinct from other sound waves that have the same frequency and amplitude.

Torque

Torque is a twist or a turn that causes an object to change its rotational motion. Torque creates rotational acceleration and causes the rotation of an object to speed up, slow down, or change direction. (symbol: τ, unit: Nm)

Total internal reflection

Total internal reflection describes the phenomenon that occurs when light travels from an object with a high index of refraction to an object with lower index of refraction at an angle greater than the so-called critical angle. Light that is incident at angles greater than the critical angle will not refract at all, but will instead completely reflect off of the interface.

Transformer

A transformer is an electronic device that is capable of transferring energy from one circuit (known as a primary circuit) to another (known as a secondary circuit) with no physical connection between the two circuits. It can increase the voltage, decrease the voltage, or keep the voltage constant between primary and secondary devices.

Translucent

An object is translucent if it allows light to pass through, but scatters that light in the process.

Transmission (wave)

Transmission occurs when light waves pass through an object.

Transmutation

The term transmutation refers to the change of one nuclide into a completely different nuclide.

Transparent

An object is transparent if it allows visible light to pass through without getting scattered or absorbed.

Transverse wave

In a transverse wave, the vibrations of the wave travel perpendicular to the direction of motion of the wave.

Ultrasound

Ultrasound is the set of sound waves that have frequencies above 20,000 Hz. They are not audible to humans.

Uncertainty principle

The uncertainty principle theorizes the extent to how well we can know information about related properties.

Vaporization

Vaporization is the phase change that occurs when a substance converts from a liquid to a gas.

Vector

A vector quantity is a variable that must be conveyed with both a numerical quantity (indicating magnitude or strength) and a direction.

Velocity

Velocity is the vector quantity that describes the rate at which an object changes its position. Velocity is equal to displacement divided by time. (symbol: **v**, unit: m/s)

Venturi effect

The venturi effect states that as the area of fluid flow decreases, the speed of fluid flow must increase.

Vibration

A vibration is a periodic back-and-forth motion that remains fixed in one location.

Voltage

Voltage is a difference in electric potential between two points in a circuit. It can be considered as a "push" that causes current to flow. (symbol: V, unit: V)

Wave

A wave is a periodic oscillation that transfers energy from one place to another.

Wavelength

The wavelength of a wave describes the shortest distance between two identical repeating points on a wave. (symbol: λ, unit: m)

Weight

The weight of an object is proportional to its mass as well as the force of gravity acting on that object, related by Newton's second law. More accurately, weight arises due to a supporting force acting on the object.

Work

Work is energy that is transferred to an object by exerting a force on that object over a distance. (symbol: W, unit: J)

Work-energy theorem

The work-energy theorem states that work is equal to the change in kinetic energy of an object, $W = \Delta KE$.

This page titled Glossary (1) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Alyssa J. Pasquale, David R. Fazzini, and Carley Bennett via source content that was edited to the style and standards of the LibreTexts platform.