Big Ideas in Cosmology (Coble et al.)

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

The LibreTexts mission is to unite students, faculty and scholars in a cooperative effort to develop an easy-to-use online platform for the construction, customization, and dissemination of OER content to reduce the burdens of unreasonable textbook costs to our students and society. The LibreTexts project is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning by developing an Open Access Resource environment. The project currently consists of 14 independently operating and interconnected libraries that are constantly being optimized by students, faculty, and outside experts to supplant conventional paper-based books. These free textbook alternatives are organized within a central environment that is both vertically (from advance to basic level) and horizontally (across different fields) integrated.

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

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. More information on our activities can be found via Facebook (https://facebook.com/Libretexts), Twitter (https://twitter.com/libretexts), or our blog (http://Blog.Libretexts.org).

This text was compiled on 02/24/2025

TABLE OF CONTENTS

About the Book

The traditional textbook, bound to the physical limits of the printed page, cannot illustrate the full complexity, movement, and beauty of the cosmos. In 2012, NASA granted astronomer Lynn Cominsky a multi-year grant with a mandate to develop an interactive experience that could better capture the dynamic nature of the universe. Dr. Cominsky was joined by like-minded astronomers and practicing educators Kim Coble, Kevin McLin, Anne Metevier, Carolyn Peruta, and Janelle Bailey. This collaboration built the fully digital publication the Big Ideas in Cosmology. The development process began with extensive research and vigorous peer review, to ensure the quality of the content and the level of instruction. Field testing was conducted at Sonoma State University to evaluate the effectiveness of the publication's methods and pedagogy, which revealed high levels of engagement and comprehension for science and non-science majors alike.

Licensing

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

About the Authors

Kim Coble (San Francisco State University)

Kim Coble is a Professor of Physics and Astronomy at SFSU. Her current research centers on understanding students' ideas about cosmology, recognizing the strengths that diverse learners bring to the classroom and to STEM professions, and creating innovative, active learning environments that engage students in realistic scientific practices. Her previous research focused on observations of the cosmic microwave background. She is currently the chair of the Education Committee of the American Astronomical Society (AAS). She was a member of the AAS Task Force on Diversity and Inclusion in Graduate Astronomy Education and the Committee for the Status of Minorities in Astronomy, served on the Committee on Diversity of the American Association of Physics Teachers (AAPT), and was an organizer of the Inclusive Astronomy 2015 conference. At SFSU she is the director for the Learning Assistant program, a member of the Faculty Agents of Change, as well as a faculty collaborator for the Center for Science and Math Education. She was formerly a Professor at Chicago State University, an NSF Astronomy and Astrophysics Fellow, and obtained her PhD from the University of Chicago.

Kevin McLin (Chico State University)

Kevin McLin is an observational astronomer who has studied galaxies and their environments in the local universe. His primary interests have moved from research to teaching over the past two decades, and he has been involved in various programs in science education and outreach with students, teachers and the public during that period. He especially enjoys sharing his love of the sky, and uses both naked eye and telescopic observations to engage people with the universe. He currently teaches in both the Department of Science Education and the Department of Physics at California State University, Chico.

Lynn Cominsky (Sonoma State University)

CHAPTER OVERVIEW

0: Review of Mathematics

This content reviews basic concepts from mathematics and science such as scientific notation, units and measurement in the metric system, and problem solving skills. It also introduces special units used in astronomy.

This page titled 0: Review of Mathematics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.0: Scientific Notation

Learning Objective

- Students will be able to express numbers in scientific notation.

In astronomy, we are often working with extreme numbers—large or small. The best way to express such numbers is to use scientific notation (sometimes called exponential notation or powers of 10). Scientific notation makes computation much easier, and scientists use it because it allows them to be precise.

For example, we know that the distance between Earth and the Sun is 93 million miles. If we write 93 million out as a number, it is:

$$93,000,000$$

In scientific notation, this rather large number is written as:

$$9.3 \times 10^7$$

Here is how it works: the first number, called the coefficient, is always a number greater than or equal to 1 but less than 10 (in our example, it is 9.3). The second number is always a multiple of 10. It can be written in exponential form (in our example, it is 10^7). Sometimes, it is written as “E” or “e"—as on a scientific calculator—but it still means the same thing. So, the example from above would appear as:

$$9.3\mathrm{E}7$$

and would be read as “nine point three times ten to the seventh.” In this book we will favor "E" over "e," but both are in common use.

The exponent is how many powers of 10. In this case, the exponent would be 7. This means there are seven powers of 10, i.e.,

$$\begin{aligned} 10^7 &= 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \\ &= 1\mathrm{E}7 \\ &= 1.0 \times 10^7. \end{aligned}$$

To determine the exponent for a large number, count the number of digits to the left of decimal place, except for the last digit on the left.

For small numbers, we use negative exponents. For example, the size of an atom is about 0.0000000001 meters. In scientific notation, this rather small number is written as:

$$1.0 \times 10^{-10} \text{ or } 1.0\mathrm{E}{-10}$$

To determine the exponent for a small number, count the number of digits to the right of the decimal place (in this case, 10).

Since we will be doing calculations using scientific notation, this site has a built-in scientific calculator. For information on how to use the site’s scientific calculator, see the calculator instructions. You may also choose to use your own handheld or computerized calculator. Whichever calculator you use, make sure you know how to use it properly. For example, you should always use the exponent key (E or EE on many calculators) and not the 10^ key, which can give you an error if used improperly. Finally, some calculators or computers will express positive exponents with a “+” sign instead or no sign, e.g., 1E5 is the same as 1E+5.

SCIENTIFIC NOTATION

In this activity, you will practice converting between regular decimal notation, scientific notation, and scientific notation with the “E” convention (as on your scientific calculator). Fill in the blanks in the chart.

Worked Examples

Some examples are shown in the chart below; the numbers in the three columns on a single row are equal. In other words, 1,000,000 is equal to 10^6 is equal to 1E6.

DECIMAL NOTATION	SCIENTIFIC NOTATION	EXPONENTIAL NOTATION
1000000	10^6	1E6
0.000001	10^{-6}	1E−6
378	3.78×10^2	3.78E2
0.0378	3.78×10^{-2}	3.78E−2

Questions

This page titled 0.0: Scientific Notation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.1: Manipulating Numbers in Scientific Notation

Learning Objectives

- Students will be able to perform basic math with numbers in scientific notation.
- Students will be able to use a scientific calculator effectively.

The ability to work with and understand scientific notation is one of the most important skills for science, whether it's astronomy, chemistry, or biology.

When thinking about astronomy and cosmology, it is easier to deal with large quantities, such as distances and masses, by using scientific notation. But you should also be thinking about what these numbers signify. Knowing that each power of ten increases the amount by ten times, and becoming familiar with the prefixes of the "series of 3's" can help you grasp how large of a number you're dealing with. The series of 3's is rather simple: 10^0 is one, 10^3 is a thousand, 10^6 is a million, 10^9 is billion, and 10^{12} is a trillion. Every third power changes the base name of the quantity. Adding one to each power gives you the "tens," so 10^1 is ten, 10^4 is ten-thousand, 10^7 is ten-million, etc. Add two each power, and you have the "hundreds," so 10^2 is a hundred, 10^5 is a hundred-thousand, 10^8 is a hundred-million, and so on.

Sometimes, we need to do more complex manipulations and calculations with numbers in scientific notation. For example, if you want to multiply two numbers using scientific notation, you multiply the coefficients and add the exponents, as in the following examples:

$$10^2 \times 10^5 = 10^{2+5} = 10^7$$

$$(2 \times 10^2) \times (3 \times 10^5) = (2 \times 3) \times 10^{2+5} = 6 \times 10^7$$

$$(4 \times 10^2) \times (5 \times 10^{-5}) = (4 \times 5) \times 10^{2-5} = 20 \times 10^{-3} = 2 \times 10^{-2}$$

If you want to divide two numbers using scientific notation, you divide the coefficients and subtract the exponents, as in the following examples:

$$\frac{10^2}{10^6} = 10^{2-6} = 10^{-4}$$

$$\frac{4 \times 10^2}{2 \times 10^6} = \left(\frac{4}{2}\right) \times 10^{2-6} = 2 \times 10^{-4}$$

Taking the inverse of a number with an exponent changes the sign of the exponent. For example:

$$\frac{1}{10^3} = 10^{-3}$$

When raising a number with an exponent to a higher power, you multiply the exponents. For example:

$$(10^2)^3 = 10^6$$

$$(4 \times 10^2)^3 = 4^3 \times 10^6 = 64 \times 10^6 = 6.4 \times 10^7$$

MANIPULATING NUMBERS IN SCIENTIFIC NOTATION

This page titled 0.1: Manipulating Numbers in Scientific Notation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.2: The Metric System

Learning Objectives

- Students will know the units of distance, time, mass, and temperature in the metric (SI) system.
- Students will understand the prefixes that may modify these units.

In order to have a standard system for measurements, scientists usually use the metric system (the International System of Units, or SI). Unless you live in the United States, you probably already use meters to measure length and kilograms to measure mass.

If the quantities to be measured are too big or too small, then a more suitable unit can be devised simply by adding a Greek prefix. The prefix “centi” means “one-hundredth,” so a centimeter is one-hundredth of a meter. For smaller objects, the millimeter can be used, with “milli” indicating “one-thousandth.” Going in the other direction, kilometers can be used to measure larger things. The prefix “kilo” means “one thousand,” so a kilometer is 1,000 meters. Other Greek prefixes are used to indicate other multipliers of the basic unit. Table 0.1 is included to help you understand some of the prefixes we will be using.

Table 0.1 Prefixes and Their Meanings

PREFIX	MEANING (IN USA)	EXPONENT	SYMBOL
Tera	trillion	10^{12}	T
Giga	billion	10^{9}	G
Mega	million	10^{6}	M
kilo	thousand	10^{3}	k
centi	one-hundredth	10^{-2}	c
milli	one-thousandth	10^{-3}	m
micro	one-millionth	10^{-6}	μ
nano	one-billionth	10^{-9}	n
pico	one-trillionth	10^{-12}	p

The distances we encounter while studying the Universe are much bigger (or smaller) than those we typically encounter on Earth. For example, we might measure distances in our daily lives using meters or kilometers. Meters, which are about the same as a yard, are suitable to measure the size of a person or a house, each of which has dimensions of a few meters. If we want to describe the size of a town or the distance between towns, then the meter is too small. A town might be 10,000 m across, so using kilometers would be more natural: It is much easier to describe the size of a town as 5 km across than to say it is 5,000 m across! Going smaller, centimeters can be used to measure things with sizes comparable to the width of a finger, and nanometers can be used to describe microscopic things.

All metric units work this way, so a kilogram (kg) is 1,000 grams. A gram is a small mass; a sugar cube has a mass of about one gram. A textbook has a mass of about a kilogram, and a human has a mass of about 80 kilograms. It is important to distinguish the mass and the size of an object scientifically. In everyday life, we might describe an object as “massive” and “large” and mean the same thing, but scientifically, “massive” means something has a large mass and “large” means it has a large size.

Mass and size are related to each other through an object’s density. Density is an object’s mass per unit volume and is a measure of how tightly packed the atoms of a substance are. If the substance is less dense, like cotton candy, it is fluffier, and if a substance is more dense, like lead, there is more mass squeezed into a smaller volume. The SI units for density are kg/m^3. Density can also be measured in g/cm^3.

While the Fahrenheit or Celsius temperature scales are most often used in everyday life, the Kelvin temperature scale is used by astronomers and physicists. This is because the Kelvin scale is more natural in terms of what temperature is really measuring, the random internal motions of the particles making up a material. Zero Kelvin, or absolute zero , is defined to be the point at which random motions in a substance would cease. As the motions become more energetic, the temperature would go up accordingly. We say “motions would cease,” not “motions cease,” because in real materials, the motions never actually cease; according to quantum mechanics, it is impossible for atoms to reach zero energy. Even if we could cool them to arbitrarily low temperatures, there would always be some small but nonzero residual energy left in the system.

Table 0.2 below provides the temperatures of some familiar phenomena in Kelvin, Celsius, and Fahrenheit to give you a sense of how these temperature scales relate. For example, zero Kelvin corresponds to −273.15 °C and −459.67 °F. Kelvin units are defined to have the same size as Celsius degrees; they just have different zero-points. Since water freezes at 0 °C and boils at 100 °C, the corresponding temperatures in Kelvin (K) are 273.15 K and 373.15 K. We often do not concern ourselves with the “0.15” and just say that water freezes at 273 K and boils at 373 K. The SI unit of temperature is the Kelvin (K) not the degree Kelvin (°K); since it is an absolute scale, there is no such thing as a “degree Kelvin.”

You can probably already guess why Kelvin temperatures are not used in most daily weather reports. We tend to live out our lives within the temperatures between the freezing and boiling points of water (for the most part) and so Celsius is a much more convenient scale. But when we want to easily understand how the temperature of an object is related to its internal energy state, then we have to use Kelvin to measure temperature. Many of the temperatures we will encounter in astronomy are quite high or low by earthly standards. For instance, even a “cool” stars still has a surface temperature of several thousand Kelvin and an interior temperature of several million Kelvin. On the other hand, when we talk about the average temperature of the Universe as a whole, it is only about 3K.

Table 0.2 Temperature Scales

	KELVIN (K)	CELSIUS (°C)	FAHRENHEIT (°F)
Absolute Zero	0	−273	−459
Water Freezes	273	0	32
Room Temperature	300	23	81
Water Boils	373	100	212

This page titled 0.2: The Metric System is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.3: Special Units Used in Astronomy

Learning Objectives

- Students will know the astronomy-specific units for distance and when to use them: AU , light-years, etc.

What about distances in space? In addition to the SI units, astronomers use several specialized units. Our nearest neighbor, the Moon, is about 380,000 km away from Earth. The Sun is about 150 million km away. Neither of those distances is very suitable for the kilometer unit. If we go to other objects, the numbers get even bigger. For example, the planet Jupiter is about five times as distant as the Sun. One way to avoid using inconveniently large numbers for the Solar System is to invent a larger unit of measure, which is exactly what astronomers have done. The astronomical unit (AU) is defined to be the average distance from Earth to the Sun. So, we would say that the planet Jupiter is about 5 AU from the Sun, while Saturn is about 10 AU from the Sun. That system works well for distances within the Solar System. It is several hundred AU across, so the numbers never become too big and cumbersome. If we want to convert back to kilometers for some reason, then we just have to remember that each AU is about 150 million, or 1.5×10^8, km.

What about for larger distances, like those to the stars? For that we will need an even bigger unit. Even the most nearby stars are about 100,000 AU away. Before we talk about what that bigger unit is, we will have a look again at distances in the Solar System, not in terms of AU, but in terms of light travel.

It turns out that light takes about 8 minutes to travel from the Sun to Earth. So instead of saying that the Sun is 150 million km away (not very convenient), or that it is 1 AU away (more convenient, but not particularly illuminating), we might say that the Sun is 8 light-minutes away from Earth. A light-minute is the distance that light travels in one minute, equal to 1.8×10^7 km. There is no ambiguity with this definition because light travels at a constant speed of about 3×10^5 km/s.

Notice that we are talking about a distance here, *not* about a time or a speed. Specifically, a light-minute is the distance that light will travel in one minute. So, just to be clear, a light-minute is a distance, not a time. It is much easier to deal in light-minutes than in kilometers for sizes encountered in the Solar System. If we use light-minutes as our unit of measure, then we would say that Jupiter is about 40 light-minutes from the Sun, Saturn is about 80 light-minutes, and the Solar System is several thousand light-minutes across (or several light-hours, or about one light-day, perhaps).

If we only had to worry about distances in the Solar System, we could content ourselves with the light-minute, or even the AU. But the Universe is much bigger than that. For instance, if we want to describe the distance to the Sun's nearest stellar neighbor (Alpha Centauri, also called Rigel Kentaurus), then the most convenient unit might be the light-year. After our discussion above, you will not be surprised to learn that a light-year is the distance that light will travel in one year, i.e., about 9.5 trillion kilometers (6 trillion miles). That is a pretty hard distance to visualize, even for astronomers. Most do not even try; they just realize that Alpha Centauri is about 4 light-years from the Sun. That means it takes light 4 years to travel between the Sun and Alpha Centauri; 4 is a much easier number to deal with than 9 trillion!

Astronomers also have special units for quantities like speed, time, and brightness. We will encounter some of them in later chapters, but for right now, we will concentrate our attention on the sizes, distances, and masses of objects that we will confront.

This page titled 0.3: Special Units Used in Astronomy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.4: Comparing Sizes and Converting Units

Learning Objectives

- Students will be able to compare two quantities with the same units by using a ratio
- Students will be able to convert units, e.g. of distance measure

Often in astronomy, to give us a sense of scale, we compare the relative sizes of, or distances to, objects. For example, we could compare how much farther the Sun is from Earth compared to how far the Moon is from Earth. We do this by taking a ratio of the two quantities. When astronomers say things like “compare,” or “how much bigger?”, “how much smaller?”, etc., we mean find the ratio. In this case:

$$\frac{\text{Distance to Sun}}{\text{Distance to Moon}}=\frac{1.5\times 10^8\ \cancel{km}}{3.8\times 10^5\ \cancel{km}}=395\text{ times farther.}$$

Play Animation

This means the Sun is 395 times farther away from Earth than the Moon is. Notice how the units cancel out: km on the top of the fraction cancel km on the bottom of the fraction. For this to happen, it is important that both quantities have the same units, for example, km and km or $miles$ and $miles$; you could not mix $miles$ and km in a comparison.

Therefore, sometimes it is necessary to convert units. By converting units, we change the way a measurement is described, even though it remains the same measurement. For instance, you can say that a ruler is 12 inches long or 1 foot long. You are using different units, but still describing the same physical length of the ruler.

There is a mathematical way to convert units, and all it entails is multiplying your original measurement by 1, expressed as a fraction. This fraction must have your original unit in the denominator (bottom part of the fraction) and the new units in the numerator (top part of the fraction).

As an example, say you have measured a bridge to be 1.4 km long, but you want to know how many meters that is. You can look up in a chart that 1,000 m = 1 km. So, you take 1.4 km and multiply it by the number 1, expressed as 1,000 m / 1 km.

$$1.4\ \cancel{km}\times\left(\frac{1,000\,m}{1\ \cancel{km}}\right)=1,400\,m$$

Play Animation

You can see that the kilometers cancel each other out, leaving only the new units. This works for any kind of conversion you can think of, as long as you are converting between equivalent physical measurements (length to length, temperature to temperature, etc.).

As another example, perhaps you needed to know the length of the bridge in miles. You can look up the fact that 0.62 miles = 1 kilometer. You can then calculate:

$$1.4\ \cancel{km}\times\left(\frac{0.62\,mi}{1\ \cancel{km}}\right)=0.87\,miles$$

Play Animation

Again, the original units are in the denominator of the conversion factor and the new desired units are in the numerator.

Table 0.3 Common Distance Unit Conversations

1 inch = 2.54 cm
1,000 m = 1 km

1 km = 0.62 mi
1 AU = 1.5×10^8 km
1 ly = 9.5×10^{12} km = 6.3×10^4 AU

CONVERTING UNITS AND COMPARING DISTANCES

Given the conversion factors in Table 0.3 answer the questions below. Numerical answers should be given in scientific notation. The expected units will follow the answer box. For example, if you get an answer of 6.5×10^4 km, you should enter 6.5e4 in a box with km next to it. When asked to show your work, you should explain your answers thoroughly.

This page titled 0.4: Comparing Sizes and Converting Units is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.5: Problem Solving and Thinking About Your Answers

> **Learning Objectives**
>
> - Students will learn a method for numerical problem solving that they will use throughout the modules.

In learning science, you will be practicing your critical thinking and problem-solving skills. To aid you as you develop your skills, we have created a set of guidelines. One thing to keep in mind if you get frustrated is that some problems are complex and are supposed to take a long time to answer. Be patient with yourself, and go through problems methodically, step by step. Write everything out in the space provided, and keep careful notes of all measurements, calculations, and ideas. It is okay to brainstorm if you do not know where to start. The goal is for you to come to your own understanding of the material.

A step-by-step process that we will use for numerical problems will help guide you through such problems in a systematic way. We suggest always writing down the following steps in the space provided (at a minimum) and we will provide examples throughout the modules for you. The steps include:

- Given: Write down the information you are given, including the symbols for variables, their values, and what they mean.
- Find: Write down the information you are trying to find, including the symbols for variables, their values, and what they mean. What is the goal?
- Concept(s): Write down any relevant equations, conversion factors, or anything else that you might need.
- Solution: Now that you have set up the problem clearly, you can start plugging in numbers.
- Think about it: Does your answer make sense?

For some problems, you might also want to draw diagrams to help you figure things out. Once you get to the solution stage, you are doing mathematical manipulations. At this stage, two key principles will get you a long way: (1) whatever you do to one side of an equation, you must do to the other, and (2) whatever you do to the top of a fraction, you must do to the bottom (and vice versa). You should also keep careful track of your units. Finally, make sure you know the ins and outs of your calculator, and that you are performing calculations correctly with it.

For all problems, conceptual and numerical, the last step is critical: think about your answer. Does it make sense? For conceptual questions, ask yourself "would my answer make sense if I read it out loud to a friend?" For numerical questions, you can check yourself by making sure the final units make sense and by checking if the number you get is in the ballpark that you would expect. For example, if you calculate the lifetime of the Sun and get 300 seconds, does that make sense to you? Do you need to revisit your calculations? Or, if you keep track of your units and the units come out wrong, you also know you need to revisit the problem. For example, if instead of getting 3.15×10^7 s for the number of seconds in a year, you get 3.15×10^7 1/sec, , you should return to the problem.

This page titled 0.5: Problem Solving and Thinking About Your Answers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.6: Measurement Accuracy and Uncertainty

Learning Objectives

- Students will be able to explain the meaning of measurement precision and sources of error

Another scientific skill you will be practicing in these modules is measurement; in many of the activities you will be using actual astronomical data. In any scientific experiment or observation, there are unavoidable sources of measurement imprecision. Each time you make a measurement and quote a value, you must also quote how accurately you measured the number.

The factors that influence the accuracy of the measurement, known technically as sources of error or uncertainty, can be due to equipment, measuring devices, the conditions under which the experiment was performed, etc. “The tick marks on the ruler are 1 mm apart. Therefore, we were able to measure the length of the glass rod to an accuracy of 0.5 mm, because I can read the ruler to an accuracy which is half as large as the distance between the tick marks.” “The graph was a straight line overall, with some slight variations of a few percent.” Mistakes and misconceptions do not count. Some examples of laboratory mistakes include things like: “We dropped the glass rod, and it broke into 100 pieces. Therefore, our measurement of its length is uncertain.” “I don’t know how to use my calculator; therefore, my calculations are in error.” Or, our personal favorite: “Human error.” If you make a mistake, you should start the activity over; astronomical data are very forgiving that way, as long as there is a copy of the original on the computer.

Finally, sometimes you will be asked to calculate other numbers from numbers that you measured. When doing this type of calculation, it is important to keep in mind that just because your calculator can give you 10 digits does not mean that you measured something that precisely! The key is to determine how accurate your measurements are and carry that through your calculations. For example, say you need to split up a room into three equal-length sections, and when you measure the total length of the room, you get 10.138 meters. But when you divide that number by three on your calculator, you get 3.3793333333 …. With a meter stick or measuring tape, there is no way you could be accurate down to the nanometer scale. Instead, you could accurately say that each section would be 3.379 ± 0.001 m. The $\pm$ symbol means “plus or minus” and indicates that you are certain each section of the room should be between 3.378 and 3.380 m long.

This page titled 0.6: Measurement Accuracy and Uncertainty is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.7: Wrapping It Up

Activity: POWERS OF TEN AND UNITS

In this activity, you will use the *Scale the Universe* tool to compare the sizes and distances of various objects in the Universe, including people and other things from everyday life.

For each of the questions in the sections below, first make predictions and record them, one section at a time. Then check them using the *Scale the Universe* tool, one section at a time. You can pause the animation at any time and use the slider bar to navigate to objects of different sizes. The measurement scales included are in meters. Resolve any discrepancies, recording your observations below.

To access the tool, simply click "play activity."

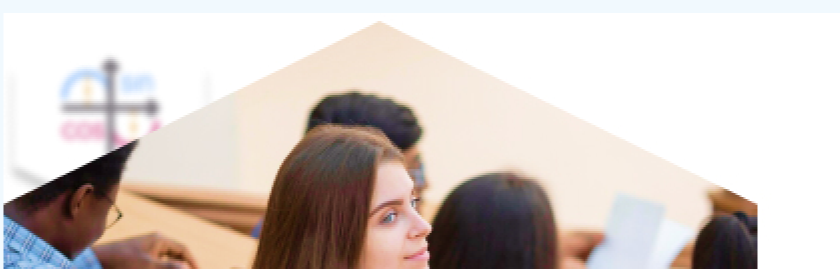
Play Activity

Part I. Everyday Scales

https://phys.libretexts.org/@go/page/31205


f(x)
s
a
b


f(x)
s
a
b




f(x)
s
a
b

Part II. Larger Than Life

Part III. Smaller Scales

This page titled 0.7: Wrapping It Up is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

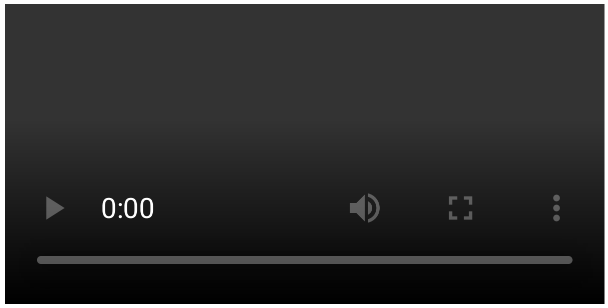
1: Size and Scope

Chapter 1 gives an overview of objects in our Solar System, Galaxy, and Universe, such as: elements in the periodic table, planets, stars, galaxies, galaxy clusters, and more. It engages you in size and distance ranking between astronomical objects and in understanding scales and models of astronomical systems.

This page titled 1: Size and Scope is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.0: Size and Scope Introduction

When you think of the Universe, what comes to mind? Do you have an idea of how big the Universe is? Do you know what sorts of objects it contains? This chapter will explore the Universe in the most general terms. Even if you think you can make a reasonably descriptive list of what's in the Universe (and you will be asked to do so), you might be surprised to find that some of the items do not fall where you originally thought they would.



Video Transcript

> ***The Universe: Transcript***
>
> *The Universe. When you say that word, what comes to mind? Do you imagine the objects that it contains, or possibly its size? Look around you. All of this is the Universe: the chair that you sit in, the monitor before you, the cars and people passing by outside. The Earth under your feet, and you and your feet as well, are all a part of the Universe.*
>
> *But all of this - the objects and materials that you interact with every day, the Earth itself - are less than a fraction of a fraction of a percent of what the Universe is. The scale of the Universe is so vast, containing everything we know and more, that new units of measurement and different ways of thinking and communicating are needed in order to understand it.*
>
> *In this chapter we will begin to explore the size and scale of the Universe, and to acquire the scientific tools we will need for our exploration.*

This page titled 1.0: Size and Scope Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.1: Your Model of the Universe

Learning Objectives

- You will articulate your own model of the universe
- You will be able to compare your model with the one in the powers of ten video

WHAT DO YOU THINK: MODELING

The Stargazers Club meets every summer at Crystal Lake for a weekend of camping, hiking, and, of course, stargazing. Alicia and Brianna drive over to pick up Chris. When they get there, they find him in his garage working on a large diorama, complete with miniature trees, grass, and even a polished laminate for the water.

Brianna: "What are you doing?"
Chris: "I'm making a model of Crystal Lake,"
Brianna: "Why?"
Chris: “Because I want to have a perfect model of the lake, so I know the best places to hike and fish and where the best places to set up the telescopes are.”
Alicia: “Well, it can’t be perfect.”
Chris: “Why not?”
Alicia: “Because if you want to know where the best places to fish are … you’ll need to not only know where the fish actually are, but you’ll have to put miniature fish into your lake there. And for the telescopes, you may find where good open spaces are, but your model doesn’t take into account how cloudy it is on any given day.”

A model is a representation of an object or a set of ideas that helps us understand the object. Sometimes models can be actual physical replicas, like the model planes and ships that many of us built as children. Sometimes, the models can be a picture (like a map). Other times, the models are represented by equations or sets of equations. Sometimes, the equations can be solved by hand, and sometimes, they must be solved by using computer simulations.

If we ask you to describe the planet Earth in terms of its geography, its size, and its contents, you could probably do a reasonably good job. For instance, you probably know that Earth's surface is divided into continents, with oceans occupying the area between them. You might indicate on your model that these continents have different political regions called countries, and each of these contains provinces or states, cities, towns, and roads. In addition, the continents have different geographical features, like mountains, plains, deserts, forests, lakes, etc. You could probably give a rough estimate for the sizes of the continents, cities, mountains, and other familiar features. Granted, there is a range in the sizes of any of these, but you probably have some idea what that range is; if we asked you to create a “descriptive list” of Earth and its contents, you would probably create a fairly detailed list.

Could you make a similar list for the contents of the Universe as a whole? Would you be able to rank list items from large to small and give a characteristic size for each? Could you make a list such that “things below are contained in things above?” If not, don't worry. Throughout this chapter, we will be discussing the size and scope of the Universe. Before we do that though, it is important that you sketch out what you already think about the Universe. We want to make you more conscious of your own current beliefs about science and the Universe, so that the information you learn as you work through the modules will be more meaningful. Over time, you will want to reflect on your ideas and how you have come to them, and see how they change as you work through the modules.

Your Model of the Universe

We would like you to create your own model of the Universe in the form of a written description. Your model should convey what you think the Universe is, what objects it contains, how they are arranged, and the size and distance scales involved.

As a start, make a list of the features and objects your model will contain. Then describe as many size and distance scales as possible. Finally, describe the relationships between the objects you have listed.

Remember, the idea here is not to see who is “right” or “wrong” in their beliefs. The most important thing for this activity is that you be as honest and clear as possible about what you really think. This will give you a starting point for your learning.

You probably found that you were not able to create your model Universe to scale. Depending on what you included, you might have found that some items were much, much smaller than other items. For instance, if you had decided that one kind of object contained in the Universe was an atom, and that another kind of object was a star, it would already be impossible to represent both of them to scale on a single drawing, at least if you used a linear scale. This scale problem is one that is frequently encountered in science. The Powers of Ten video in the next activity illustrates the range of scales we face while studying the Universe.

Powers of Ten

In addition to illustrating the relevant size scales in the Universe, the video also lists many different objects in the Universe. After watching the video, answer the questions below.

How do the lists you made for your model of the Universe compare with the contents of the Universe that the film mentioned? Were there objects in the film that you did not think of, or objects you considered that the film left out? How did your size and distance scales compare to those in the video?

This page titled 1.1: Your Model of the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.2: Our Solar System

Learning Objectives

- You will know the objects in the Solar System: the Sun; Planets; Moons; Comets, Asteroids, and Small Debris
- You will be able to compare and contrast the properties of these objects

What Do You Think: Contents of the Solar System

Three students are discussing which objects are in our Solar System.

> **Annie:** "A solar system has different things in it like galaxies and planets and stars and stuff like that. Our solar system has the planets Mercury, Venus, Earth, and so on. The planets have moons so I think moons, too."
> **Brenda:** "I disagree. I think a galaxy has stars inside it. Each one of the stars has planets orbiting around it, and that's what a solar system is. So, a galaxy has solar systems in it, but a solar system doesn't have galaxies in it."
> **Charles:** "I think that the terms 'solar system' and 'galaxy' mean the same thing."

While we think of the Solar System as an enormous structure in human terms, in astronomical terms, a solar system is very small. Although solar systems are not important to the overall structure of the Universe, the one in which we live is very important from a human perspective, and it will help us to set the scale by which we can understand larger cosmic constituents. We will therefore briefly discuss the objects that are contained in our Solar System:

- The Sun
- The terrestrial planets: Mercury, Venus, Earth, Mars
- The gas giant planets: Jupiter, Saturn, Uranus, Neptune
- The dwarf planets: Pluto, Eris, Makemake, Haumea, and Ceres
- The moons of the planets

- Comets, asteroids, other small bits of rock, dust, and gases

We have sent spacecraft and probes to many of these places. However, the only place that humans have traveled to beyond near-Earth orbit is the Moon. Distances and sizes of objects in our Solar System are shown in Figure 1.2.1 and Figure 1.2.2.

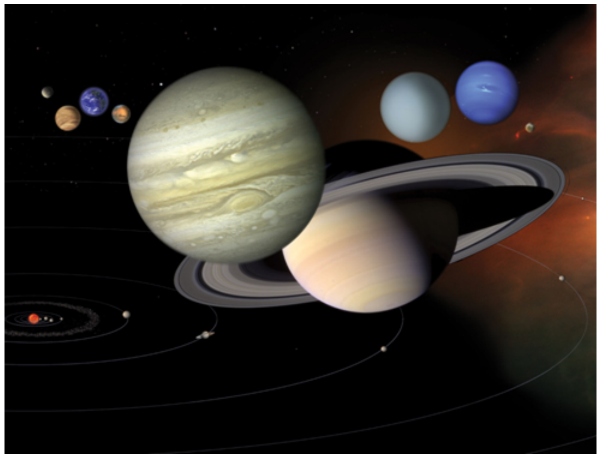


Figure 1.2.1: Planets in our Solar System. The relative planet sizes are shown to scale at the top part of the figure. The relative orbital distances are shown to scale in the bottom part of the figure. The planet sizes are not to scale relative to the orbital distances (the planet sizes depicted are too large). The planets, in increasing distance from the Sun, are: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune. Also shown are the asteroid belt and the dwarf planet Pluto. Credit: NASA/JPL.

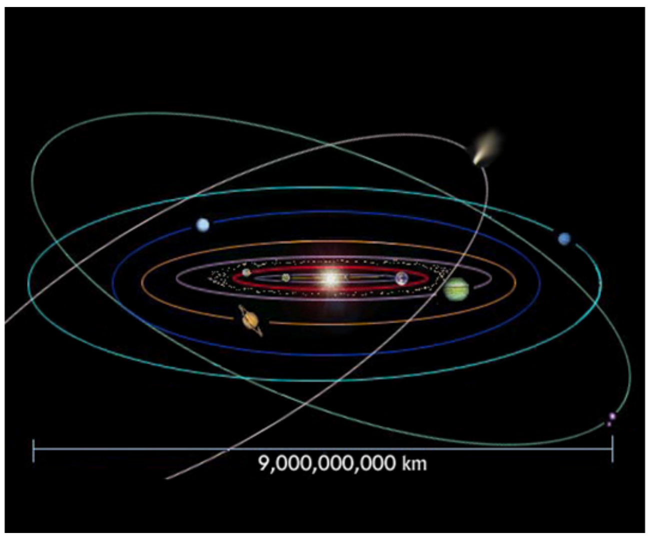


Figure 1.2.2: Orbits in our Solar System (not to scale). The planets, in increasing distance from the Sun, are: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune. Also shown are the asteroid belt, a comet, and the dwarf planet Pluto. Credit: NASA/HEASARC/Maggie Masetti.

1.2.1: The Sun

Table 1.1 Solar Data

Diameter of the Sun	1.4×10^6 km
Distance from the Earth to the Sun	1 AU = 1.5 × 108 km
Mass of the Sun	1.99 × 1030 kg

At the center of our Solar System is a single star, the Sun. Stars are giant balls of gas powered by nuclear fusion at their cores during the main part of their lifetimes. They come in a variety of sizes, colors, and brightness. Our Sun is a fairly average star; it is actively converting hydrogen to helium in its core, where the temperature is about 15 million K. The temperature decreases to 5800 K on the Sun’s surface. Compared to other stars, the sun is of medium size with a diameter of 1.4×10^6 km. It is one astronomical unit (1AU), 1.5×10^8 km, or 8 light-minutes from Earth. The Sun is the largest (in diameter) and most massive object in our Solar System. With a mass of 1.99×10^{30} kg (which is about 330,000 times more massive than Earth), the Sun contains 99.8% of the total mass of the Solar System. There is a strong gravitational force between the Sun and the other objects in the Solar System, and all other objects in the Solar System revolve around the Sun.

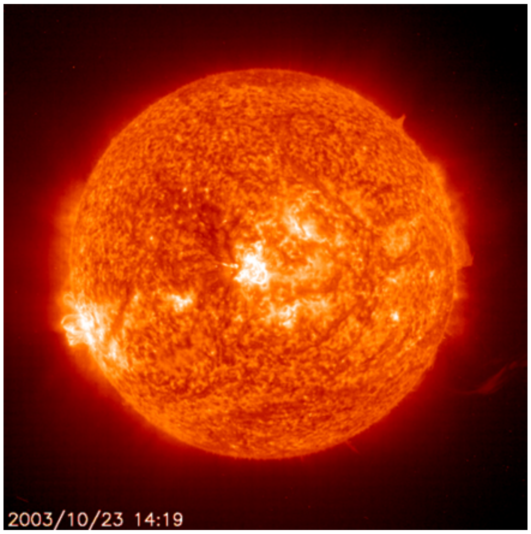


Figure 1.2.3: The Sun, as viewed by NASA's SOHO satellite. Credit: NASA/ESA/SOHO.

1.2.2: The Planets

The International Astronomical Union has defined a planet as a celestial body that (1) orbits the Sun, (2) has enough mass to retain a spherical shape, and (3) has cleared the neighborhood around its orbit. Criteria 3 means that the only objects occupying the space around a planet are its natural satellites or moons.

In our Solar System, there are two classes of major planets: terrestrial and gas giants. These meet all three criteria. The Solar System also contains dwarf planets. Dwarf planets (Pluto is the best known example) meet criteria 1 and 2, but do not meet criterion 3.

The word terrestrial is a derivative of *terra*, the Latin word for ground or soil (earth). The terrestrial planets in our Solar System are Mercury, Venus, Earth, and Mars. These planets reside in the inner part of our Solar System, closest to the Sun. The terrestrial planets are smaller in diameter and less massive than the gas giants. Terrestrial planets have solid, rocky surfaces with a molten, metallic core containing elements such as iron and nickel. They have a thin atmosphere, if any, containing carbon dioxide and other gases. Other characteristics the inner planets share include the presence of few to no moons (natural satellites), and they do not have rings. The orbital period of the terrestrial planets is shorter than that of the outer planets—it takes less time for them to go around the Sun—and their periods are measured in days.

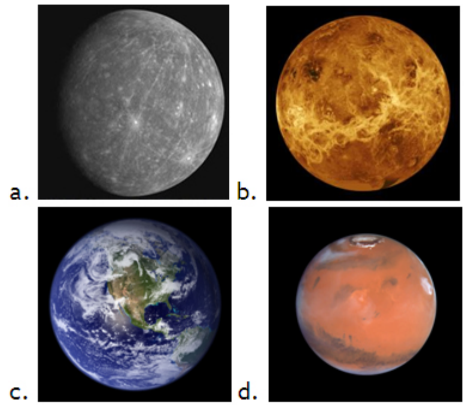


Figure 1.4: (a) NASA's satellite Messenger took this image of Mercury when it flew by in 2006. (b) The surface of Venus, as imaged by radar by NASA's Magellan spacecraft and others. (c) Western hemisphere of Earth as seen from the Terra satellite. (d) A Hubble Space Telescope view of Mars and the volcanic region named Elysium. Credits: (a) and (b) NASA/JPL, (c) NASA/GSFC, (d) NASA/STScI

Table 1.2: Terrestrial Planet Table of Comparison*

PLANET	DIAMETER (EARTH) (KM)	MASS (EARTH) (KG)	DENSITY (G/CM^3)	AVG. TEMP (K)	KNOWN MOONS	ORBITAL PERIOD (DAYS)	DISTANCE FROM SUN (AU)
Mercury	0.38 4.880×10^3	0.06 3.30×10^{23}	5.43	700	0	88	0.39
Venus	0.95 1.210×10^4	0.81 4.896×10^{24}	5.24	740	0	224	0.72
Earth	1 1.276×10^4	1 5.972×10^{24}	5.52	290	1	365	1
Mars	0.53 6.794×10^3	0.11 6.421×10^{24}	3.94	220	2	687	1.52

*To give you a sense of scale, the sizes and masses of the planets are given in both SI units and relative to Earth. The comparisons relative to Earth are made by taking a ratio.

The gas giant planets are Jupiter, Saturn, Uranus, and Neptune. They are sometimes called the jovian planets, which means Jupiter-like. They reside in the outer part of our Solar System. These planets are composed of a combination of gases, primarily hydrogen and helium, and they do not contain as many metallic elements as the terrestrial planets do. These planets are very large and very massive (more than 1,000 Earths could fit inside Jupiter). Although the gas giants are thought to have rocky cores, they do not have solid surfaces and their atmospheres are thick. For example, if you journeyed deep inside Jupiter, the pressure of the atmosphere would crush you—similar to the pressure of the ocean on a deep sea diver, only to a much greater degree. The gas giant planets have fast, strong winds that create storms. This phenomenon is evidenced by the striped appearance of the planets, the Great Red Spot of Jupiter, and the Great Dark Spot on Neptune. Neptune has the fastest known winds in our Solar System, reaching speeds of up to 2,000 km/hr. The gas giants have many natural satellites, and some of their moons are similar in size to the terrestrial planets. The orbital periods of the gas giants are much longer than those of the terrestrial planets and are measured in years. For example, Mercury circles the Sun every 88 days, but it takes Jupiter over 11 years to complete its orbit.

One of the most intriguing features of the gas giants is their rings. All of the gas giants in our Solar System have rings, including Jupiter. The rings of Jupiter and Uranus are composed mainly of small, rocky material with very little ice (Uranus) to no ice (Jupiter). The composition of Neptune's rings is still unknown. The rings of these three planets are relatively narrow and dark—they do not reflect much light. The rings of Saturn are wider and contain larger pieces of ice and debris than the rings of Jupiter and Uranus. Due to their composition and size, Saturn's rings reflect more light, making them appear bright. This brightness, combined with the angle at which the rings orbit the planet, allows them to be seen from Earth. The ring material is held in place by gravitational attraction with the planet.

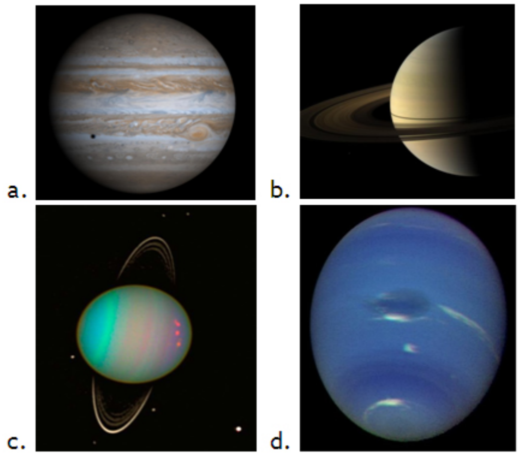


Figure 1.5: (a) Jupiter as viewed by NASA's Cassini spacecraft. (b) Saturn as viewed by NASA's Cassini spacecraft. Saturn is the planet most famous for its rings, though all of the gas giant planets have rings. (c) Uranus is the only planet that rotates on its side. (d) Neptune as seen from the Voyager 2 spacecraft. The streaks are caused by extremely fast winds.Credits: (a) and (b) NASA/JPL, (c) NASA/STScI, (d) NASA/JPL

Table 1.3 Gas Giant Planet Table of Comparison*

PLANET	DIAMETER (EARTH) (KM)	MASS (EARTH) (KG)	DENSITY (G/CM^3)	AVG. TEMP (K)	KNOWN MOONS	ORBITAL PERIOD (YEARS)	DISTANCE FROM SUN (AU)
Jupiter	11.2 1.43×10^5	318 1.9×10^{27}	1.33	125	63	11.9	5.20
Saturn	9.5 1.21×10^5	95 5.68×10^{26}	0.70	95	62	29	9.54
Uranus	4.0 5.1×10^4	14.5 8.7×10^{25}	1.30	60	27	84	19.19
Neptune	3.8 4.95×10^4	17.1 1.27×10^{25}	1.76	60	13	165	30.07

*To give you a sense of scale, the sizes and masses of the planets are given in both SI units and relative to Earth. The comparisons relative to Earth are made by taking a ratio.

The composition of the planets is determined by where they formed. Being closer to the Sun, the terrestrial planets formed at higher temperatures than the gas giants. At higher temperatures, the hydrogen and helium have more energy and are moving too fast to clump together. In fact, these lighter atoms tended to be blown out of the inner part of the solar nebula. The gas giants, being farther from the Sun, formed at temperatures colder than the freezing point of water. At such low temperatures, the gases have less energy and can clump together. In addition, there was more volume, and therefore more mass, in the outer parts of the solar nebula - think of how the area of a disk increases as one moves farther from its center. As the gas giants built up, their increasing mass gave them a stronger gravitational pull, and they were able to attract and retain even more gases and grow very large very fast. As a result, the giant planets are much bigger than the terrestrial planets and have much more hydrogen and helium gases than the terrestrial worlds.

The category of dwarf planet is a relatively new classification of planets. It was created in 2006 after the discovery of Eris, an object that is larger than Pluto and located farther from the Sun. Since Eris is larger than Pluto, some scientists felt it should be classified as a planet. Others felt Pluto should be removed from the planet category. After much debate, Pluto has been reclassified as a dwarf planet. In addition to Eris and Pluto, there are, at present, three other objects in this category: Makemake, Haumea, and Ceres.

Dwarf planets meet two of the three IAU requirements for planets: they orbit the Sun (they are not satellites of another planet) and have sufficient mass to have a nearly spherical or round shape. However, the dwarf planets have not cleared their orbits of other objects, and there can be a great many other objects in addition to their moons near their orbital path. Three of the known dwarf planets have moons—Pluto has three, Haumea has two, and Eris has one. These planets are much smaller and less massive than terrestrial and gas giant planets. With the exception of Ceres, the dwarf planets are far away from the Sun—beyond Neptune—making them dark and cold. Some dwarf planets may have surfaces containing frozen gases - they are gases on Earth - such as methane. The orbital periods of these distant planets are very long and therefore are measured in years.

In 2006, NASA launched the *New Horizons* spacecraft on a mission to study Pluto and its neighbors. The mission goals include studying the atmospheres, surfaces, and interiors, as well as the general environment of the target objects. The spacecraft *Dawn*, launched in 2007, is en route to study Ceres. Both missions are expected to reach their destinations in 2015.

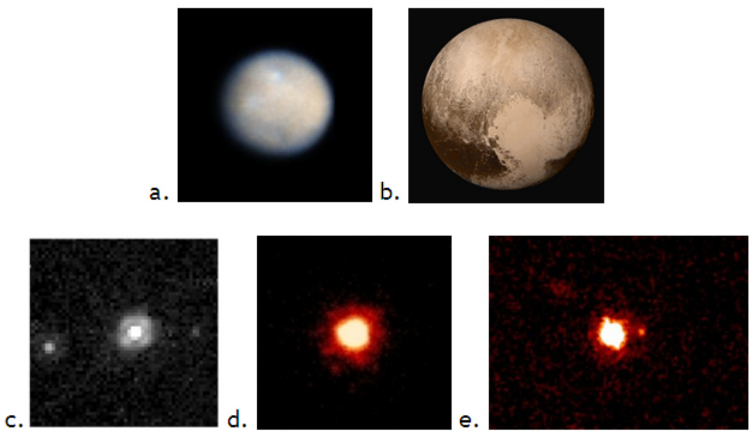


Figure 1.6: (a) Hubble Space Telescope image of Ceres. (b) Pluto in its true color—brown—which is believed to be frozen methane gas. (c) Haumea and its two moons. (d) Image of Makemake. (e) Eris is the largest of the dwarf planets. Credits: (a) NASA/ESA/J. Parker (Southwest Research Institute), P. Thomas (Cornell University), L. McFadden (University of Maryland, College Park), and M. Mutchler and Z. Levay (STScI); (b) NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute; (c) Fraser and Brown (2009), Astrophysical Journal Letters, 695, L1; (d) Alex Willman, Princeton; (e) copyright W. M. Keck Observatory

Table 1.4: Dwarf Planet Table of Comparison*

DWARF PLANET	DIAMETER (EARTH) (KM)	MASS (EARTH) (KG)	DENSITY (G/CM^3)	AVG. TEMP (K)	KNOWN MOONS	ORBITAL PERIOD (YEARS)	DISTANCE FROM SUN (AU)
Ceres	975	1.6×10^{-4} 9.6×10^{20}	2.08	167	0	~4.5	~2.7
Pluto	0.18 2.4×10^3	2.0×10^{-3} 1.2×10^{22}	1.75	40	3	248	39.46
Haumea	0.11 $\sim 1.96 \times 10^3$	7.0×10^{-4} 4.2×10^{21}	2.6–3.3	32	2	285	~43.5
Makemake	0.13 $\sim 1.5 \times 10^3$	6.7×10^{-4} 4.0×10^{21}	~2	30	0	~310	~45
Eris	0.21 $\sim 2.5 \times 10^3$	2.8×10^{-3} 1.7×10^{22}	2.3	?	1	557	~67.5
*To give you a sense of scale, the sizes and masses of the planets are given in both SI units and relative to Earth. The comparisons relative to Earth are made by taking a ratio.							

1.2.3: Moons

A moon, also called a natural satellite, is a celestial object that orbits a planet. Earth's moon is the most familiar to us, but there are hundreds of known moons in our Solar System. We will highlight just a few of them here. Several are shown in Figure 1.7 and listed in Table 1.5. Satellites vary in composition and size from one to the next, and some moons in the Solar System are similar to the planet Earth.

Galileo saw the four largest moons of Jupiter—Io, Europa, Ganymede, and Callisto—when he turned his telescope to the sky in 1609. That is why they are often called the Galilean moons. Io has many active volcanoes and contains a variety of colorful sulfuric compounds, both on its surface and in its atmosphere. Europa is covered in ice; many scientists think there might be a liquid ocean beneath the ice. Ganymede is the largest of the Galilean satellites.

Saturn has dozens of satellites, the largest of which is Titan. Titan also has a very dense atmosphere filled with organic (carbon-based) particles that are not found on other moons. In 2006, the Cassini spacecraft flew by Titan and the Huygens probe landed there, finding evidence of methane lakes on the surface. Scientists think that Titan may provide clues to the early stages of the way life formed on Earth. Like Earth, Titan has an atmosphere composed mainly of nitrogen; Titan has the densest atmosphere of any moon in the Solar System. The moons Rhea and Iapetus are believed to be composed of three-fourths ice and one-fourth rock. The moon Enceladus has trace amounts of water erupting from its surface.

Moons can form when an object impacts a planet's surface and material breaks off but cannot escape the gravitational field of the parent planet; this is how Earth's moon formed. Additionally, planets can capture objects, which then orbit the planet; this is probably what happened in the case of Neptune's moon Triton, as well as the two moons of Mars, Phobos and Deimos. Finally, moons can also form in place alongside their parent planet.

Most of the moons in our Solar System orbit around their parent planet in the same direction (counter-clockwise) as the planets orbit around the Sun. However, Neptune's moon Triton orbits clockwise. Astronomers believe this is due to a collision, possibly with another moon, which also allowed Triton to be captured by Neptune as it passed the planet. In 1977, NASA launched the *Voyager 2* spacecraft to explore the regions around Uranus and Neptune, including their moons.

A moon is locked in orbit through the mutual gravitational force between the moon and the planet; the planet tugs on the moon, and the moon tugs on the planet. One example of the effect of the gravitational interaction between our Moon and Earth is the rise and fall of the ocean tides.

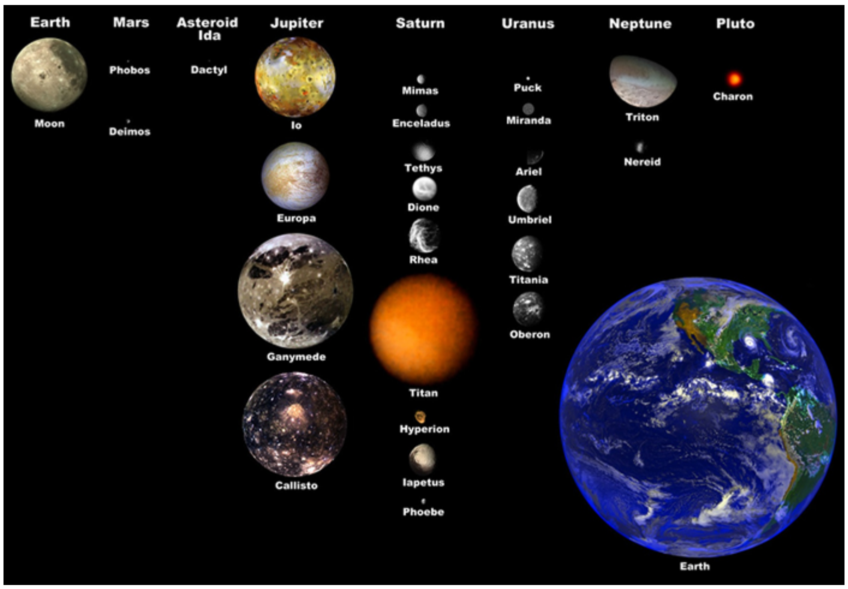


Figure 1.7: Some of the moons in our Solar System, to scale, compared to the sized of Earth. Credit: Wikimedia Commons.

Table 1.5 Some of the Moons in Our Solar System*

MOON	PARENT PLANET	DIAMETER (EARTH) (KM)	MASS (EARTH) (KG)	DENSITY (G/CM^3)
The Moon	Earth	0.27 3476	1.23×10^{-2} 7.35×10^{22}	3.35

MOON	PARENT PLANET	DIAMETER (EARTH) (KM)	MASS (EARTH) (KG)	DENSITY (G/CM^3)
Io	Jupiter	0.285 3630	1.50×10^{-2} 8.93×10^{22}	3.55
Europa	Jupiter	0.246 3138	8.03×10^{-3} 4.8×10^{22}	3.0
Ganymede	Jupiter	0.413 5262	2.48×10^{-2} 1.48×10^{23}	1.94
Callisto	Jupiter	0.376 4800	1.81×10^{-2} 1.08×10^{23}	1.8
Titan	Saturn	0.404 5150	2.26×10^{-2} 1.35×10^{23}	1.88
Rhea	Saturn	0.12 1530	4.17×10^{-4} 2.49×10^{21}	1.23
Iapetus	Saturn	0.114 1460	3.15×10^{-4} 1.88×10^{21}	1.08
Enceladus	Saturn	0.039 498	2.0×10^{-5} 1.2×10^{20}	1.6
Triton	Neptune	0.212 2700	3.58×10^{-3} 2.14×10^{22}	2.07

*To give you a sense of scale, the sizes and masses of the moons are given in both SI units and relative to Earth. The comparisons relative to Earth are made by taking a ratio.

1.2.4: Comets, Asteroids, and Small Debris

Comets are icy objects of the Solar System with highly elliptical - and often highly inclined - orbits. A comet has a solid core that is surrounded by a coma—a cloud-like ball of gases - as the comet nears the Sun. A tail of gas and dust also is emitted as the object's orbit brings it near the Sun. Sunlight is reflected off of the gaseous tail, allowing the comet to be seen from Earth. The sizes of these objects vary in diameter from 1 km to as large as 170 km. The famous Halley's Comet, to take one example, is approximately 15 km in diameter.

Comets originate from two main locations—the Kuiper belt and the Oort cloud. The Kuiper belt is the region of the outer Solar System that begins just past Neptune. It occupies the area of space that is from 30 to 50 AU from the Sun. Objects of the Kuiper belt are also called Trans-Neptunian Objects. These are generally believed to be short-period comets (<200 yrs). With the exception of Ceres, the dwarf planets are all members of the Kuiper Belt.

The Oort cloud is located about 50,000 AU from the Sun. It is where long-period (>200 yrs) comets come from. The object Sedna is thought to be a comet from the Oort cloud.

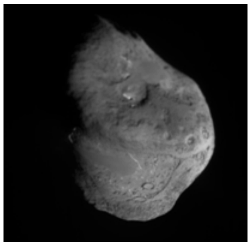


Figure 1.8: Spacecraft Deep Impact image of the nucleus of Comet Temple 1. Credit: NASA/JPL/Caltech/University of Maryland.

Asteroids are rocky celestial bodies that orbit the Sun but are too small to be called planets; they typically do not have strong enough gravity to pull themselves into spherical shapes, and are instead irregularly shaped. They do not have any atmosphere. Typical sizes range from a few centimeters to hundreds kilometers, with smaller asteroids being more numerous. The largest asteroid so far discovered is Ceres (now also classified as a dwarf planet) and the smallest bodies are the size of very small rocks—called small debris. Most of the asteroids are found in the area between Mars and Jupiter known as the asteroid belt. This area is located between 2 and 4 AU from the Sun. Most asteroids are in stable orbits around the Sun, with typical orbital periods of about 6 years long.

Figure 1.9: The asteroid Ida and its moon Dactyl as seen by the Galileo spacecraft in 1993. Credit: NASA/JPL.

Meteors are small pieces of rocky or metallic debris from asteroids or comets that enter Earth's atmosphere. Some meteors are debris thrown up by collisions between asteroids and the Moon or other planets. We see meteors as bright streaks of light traveling across the sky. They are sometimes colloquially called shooting or falling stars—but they are not stars at all. They appear as bright streaks in the sky because friction heats up the atmospheric gases leaving a glowing trail behind them as they pass through. Many small meteors enter and disintegrate in Earth's atmosphere all the time. Although most meteors that enter the atmosphere burn up completely while still very high in the sky, some do find their way to the ground. When this happens they are called meteorites. Meteors range in size from as small as a grain of sand ($\leq$ 1 mm) to the size of boulders (a few meters).

Figure 1.10: The Geminid meteor shower. Credit: NASA/JPL

This page titled 1.2: Our Solar System is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.3: Our Galaxy - The Milky Way

Learning Objectives

- You will know the objects in our Galaxy: Stars; Star Clusters; Nebulae
- You will know the shape of our Galaxy and that its major components are the bulge, disk, and halo
- You will know where we fit within our Galaxy

WHAT DO YOU THINK: LOOKING AT GALAXIES

Members of the Stargazers Club are trying to find interesting objects in their telescopes during a star party.

> **Danielle**: I really want to find our Galaxy, but I'm not sure I am pointing the right direction.
> **Emma**: Well, you can't get our whole Galaxy in the scope at the same time, it is too big.
> **Danielle**: Then what am I seeing here?
> **Faith**: I think that's the Andromeda Galaxy. It is a different galaxy than ours.

If you look up into the sky on a summer night, you may be able to see a bright strip of stars and gas across the sky (Figure 1.3.1). This is part of the Milky Way, our Galaxy (denoted with a capital “G” to distinguish it from other galaxies). Our Galaxy, like other galaxies, is a massive collection of hundreds of billions of stars, gas, dust, and mysterious dark matter. Here we will describe the various constituents and components of the Galaxy.

Figure 1.3.1: Photograph of the Milky Way rising over the MacDonald Observatory in Texas. Credit: Shutterstock.

1.3.1: STARS AND STAR SYSTEMS

Of the hundreds of billions of stars in the Galaxy, you can only see about 2,000 of them at any given time, and only if you are in a place with very clear and very dark skies. But even if you have viewed the sky from a less pristine location, you might have noticed that not all stars look the same. Viewing them at night, you can see some stars appear slightly red or blue or yellow, though their colors are subtle. Some stars are brighter than others, and some are fainter. The colors of stars correspond to their surface temperatures, which range from about 3,500 K to more than 30,000 K. All are at least tens of millions of kelvin at their cores, with some being billions of kelvin. Stars can be anywhere from 10,000 times dimmer to in excess of 100,000 times brighter than the Sun.

During the main part of their lives, stars are in a state of what is called hydrostatic equilibrium. The inward pressure from gravity due to a star's own mass is balanced by the outward thermal pressure sustained by nuclear reactions in the star's core. As long as the star has enough nuclear fuel in its core - hydrogen during the first phase of its life - it will be able to generate the pressure needed to resist collapsing under its own weight.

Stars can be found alone or in groups. Many stars exist in pairs, called binary star systems. It was surprising for astronomers to find out that the Sun is uncommon among stars of its type, having no orbiting companion star.

Open clusters are groups of several hundred stars that all lie within about 30 light-years of each other. The stars in an open cluster are loosely gravitationally bound. Within the Milky Way, we see that open clusters often consist of young, recently formed stars. The Galaxy has thousands of open clusters. The Pleiades (in Japanese, Subaru) is one of the most recognizable in the night sky (Figure 1.3.1).

Globular clusters are larger (~50 – 500 ly radius) and spherically shaped. They contain far greater numbers of stars and are tightly gravitationally bound. The stars in globlular clusters are much older (Figure 1.3.1) than typical stars in the disk of the Milky Way. While open clusters contain anywhere between a hundred to a thousand stars, globular clusters have tens of thousands or hundreds of thousands, and there are a few known that contain over a million stars. Globular clusters are so densely packed that their central regions have an average of about 1,000 stars per cubic light-year! Compare that to the region around the Sun, where there are only 12 stars within 10 light-years.

Figure 1.3.2: The Pleiades star cluster, also known as the Seven Sisters, or in Japanese as Subaru, is 440 light-years from Earth. It is an example of an open cluster. Credit: NASA/ESA/AURA/Caltech. (right) The globular cluster 47 Tucanae is about 15,000 light-years from Earth, and 120 light-years across. Credit: South African Large Telescope.

Extremely exciting to many astronomers, as well as members of the public, has been the discovery of extra-solar planets. These are planets that orbit stars other than the Sun. The first extra-solar planets were discovered in 1992 orbiting a star called PSR B1257+12, which is 980 light-years away from Earth. By the year 2014, more than 1500 extra-solar planets have been confirmed, most within about 300 light-years of Earth. Some of the star systems have multiple planets orbiting the central star. Most of the planets detected thus far have been more massive than Jupiter; this is likely because current detection techniques make more massive planets easier to detect. No Earth-like planets have been found … yet (meaning both Earth-sized and the right temperature for liquid water; each of these criteria has been met separately as of 2012). Extra-solar planets are typically found indirectly, either by observing the effects of their gravitational interaction with their central star or by observing a tiny dip in brightness of the parent star as the planet passes in front of it. A few rare extra solar planets have been imaged directly. Figure 1.3.3 (left) shows the location of most extra-solar planets found so far in the Galaxy, and Figure 1.3.3 (right) is a diagram of one extra-solar planetary system.

For the latest information and an updated planet count, see JPL Planet Quest.

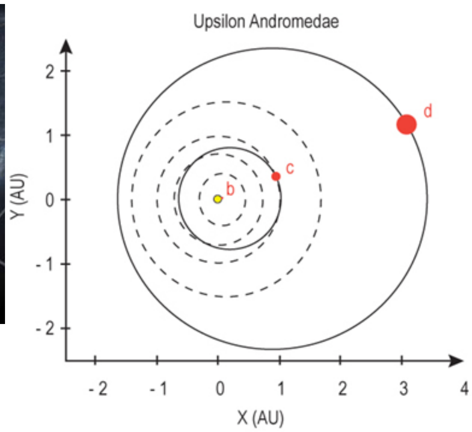


Figure 1.3.3: (left) Most of the extra-solar planets found so far are within a few hundred light-years of earth. Credit: NASA/JPL. (right) Scale drawing of the Upsilon Andromedae system, showing the orbits of the three planets at points b, c, and d. The masses of the planets are 0.72 $M^{Jupiter}$, 1.98 $M^{Jupiter}$, and 4.11 $M^{Jupiter}$, respectively. The yellow dot at (0,0) represents the parent star. The dashed lines represent the orbits of the four inner planets of our Solar System for comparison. Credit: NASA/SSU/Aurore Simonnet.

1.3.2: NEBULAE: GAS AND DUST BETWEEN THE STARS

In addition to the pinpoints of starlight, we also see many objects in the sky that are fuzzy—their light is spread out. Some of these extended objects are clouds of gas and dust, known as nebulae (singular: nebula). In the early days of Western astronomy, nearly anything that looked fuzzy (including star clusters and galaxies) was called a nebula, which is Latin for “cloud.” As telescopes became more powerful, more detailed observations of these objects were possible, allowing them to be sorted and given their modern names and designations. Today, a nebula refers only to a cloud of gas and dust. By gas, we mean atoms and small molecules, primarily hydrogen. By dust we mean a mixture of tiny particles or grains composed of silicates, graphite, iron, and other compounds.

There are dark interstellar clouds located primarily in the disk-shaped plane of the Galaxy. These provide the raw materials from which stars are made. On average, interstellar space is quite empty; there is about 1 particle per cm^3. It is also quite cold: below 100 K on average. Contrast that with Earth’s atmosphere, which has 10^{19} particles per cm^3 and temperatures around 300 K. The nebulae where stars form are relatively dense compared to rest of interstellar space; they have about $10^4 – 10^9$ particles per cm^3.

If a cloud becomes dense and cold enough it will begin to collapse in upon itself due to the self-gravitational attraction of one part of the cloud to another, triggering stars to form in that region. Star formation can also be triggered by anything that compresses an interstellar cloud. These triggers can be a collision with a different cloud, or an internal collision in a cloud with disorganized internal motions. Certain new stars that form can light up the clouds in which they were born, causing the clouds to heat up to temperatures around 10,000 K and glow brightly. These ionized gas clouds, known as emission nebulae, tend to look red in color. This is the result of strong emission by hydrogen. Figure 1.3.4 shows emission nebulae and dark nebulae in a star-forming region. The colors seen in images like this are not generally very realistic and are used to set off different regions of emission (by different types of atoms) from one another.

Figure 1.3.4: Pillars of gas and dust in a star-forming region as seen by the Hubble Space Telescope. Credit: NASA/STSCI/KPNO/T. Rector (University of Alaska)

Nebulae can also result when stars die. So-called planetary nebulae occur when a low-mass star expels its outer layers of gas after it runs out of fuel for nuclear fusion (Figure 1.3.4). This expelled gas is primarily composed of hydrogen and helium, but often contains heavier elements that were formed in the parent star during its lifetime. In contrast, a massive star produces a nebula called a supernova remnant. It is the outer regions of a massive star that catastrophically exploded when it ran out of nuclear fuel (Figure 1.3.4). A supernova is a much more violent event than formation of a planetary nebula: If a planetary nebula is like a dandelion losing its seeds in a light breeze, then a supernova remnant is like setting off a bomb in a sunflower. Some of the expanding material can reach speeds up to 10% the speed of light (~30,000 km/s), and it creates a shockwave that will plow through all of the dust and gas in its path, heating it up to million kelvin temperatures. Over millions of years, the expanding supernova remnant will have run into enough surrounding dust and gas to slow and cool down. The supernova shockwave itself can help spur the gases surrounding the exploding star into collapse, thus creating new stars. The new generation of stars will be enriched with heavier elements (than hydrogen and helium) that were created by the star that exploded. The new stars will also incorporate even heavier elements that were created in the supernova explosion itself.

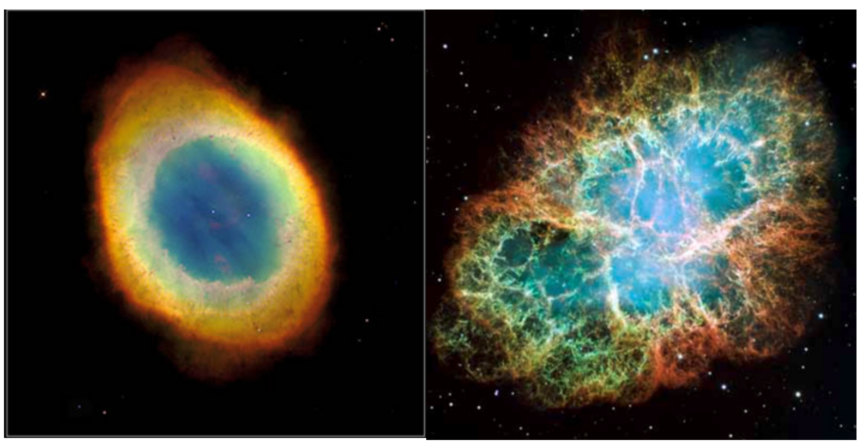


Figure 1.3.4: (left) The Ring Nebula. A planetary nebula is formed by the expulsion of the outer layers of a star similar to our Sun at the end of its life. Credit: NASA/STScI. (right) The Crab Nebula, the remnants of a massive star that exploded as a supernova in 1054 AD. Credit: NASA/STScI.

Sometimes, many of the stages of the lives of stars, including molecular clouds and the nebulae where stars form, clusters of stars, and dying stars and their nebulae, can be seen together in the same complex, as shown in Figure 1.3.5.

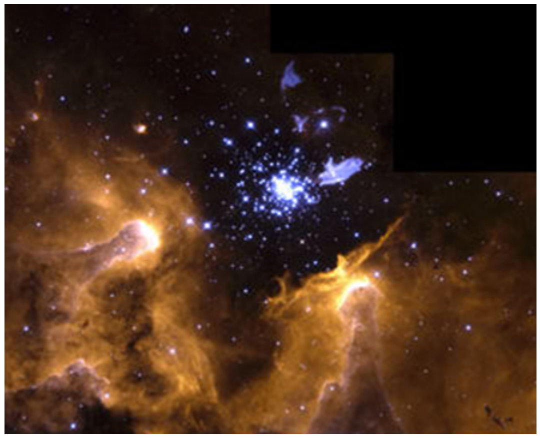


Figure 1.3.5: "Many stages of star birth, life, and death can be seen in the nebula NGC 3603. For more information, see the press release." Credit: NASA/STScI. .

1.3.3: THE SHAPE OF THE GALAXY

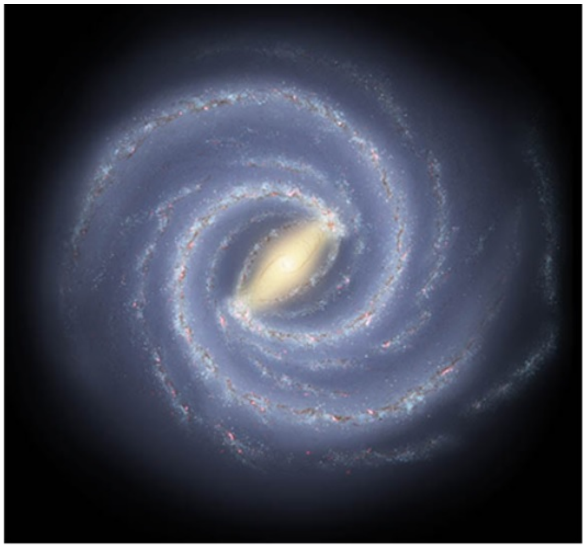
At first glance, the Milky Way (and similar galaxies) is shaped like a round, flat disk. Because we are inside the Galaxy, and because of its shape, we see it as a bright swath across the whole sky rather than a separate fuzzy region. We have an “edge-on” view. Furthermore, all of the stars that we are able to see as we look into the night sky are within our own Galaxy.

Because we cannot travel outside of our Galaxy and look down at it, we do not know exactly what it looks like. However, by making careful maps of the stars and gas within it, using observations at many different wavelengths of light, we can get a basic sense of its structure. Also, by observing other galaxies outside our own, we can get a sense for how normal or unique our Galaxy is in the Universe. From our observations of our own and other galaxies, we know that the Milky Way is comprised of several parts: the bulge, the disk, and the halo.

The bright, puffy region that we observe at the center of the Milky Way is a central bulge composed of many stars. At the very center of the Galaxy astronomers have determined that there is a supermassive black hole, about 4 million times the mass of the Sun. Yet this black hole is squeezed into an area smaller than the size of Mercury’s orbit around the Sun. One way astronomers have deduced this is by observing the orbits of stars at the center of the Galaxy, and by calculating the mass of the very central region of the nucleus based on the movements of those stars.

Outside of the bulge, astronomers have mapped several spiral arms. These lie within the flat plane comprising the disk of the Galaxy. Most of the gas and dust within the Galaxy lies within the disk, and most of the open clusters of young stars lie within the arms themselves. These stars, gas, and dust revolve around the flat central plane and bulge of the Galaxy. When the gas clouds enter a spiral arm, they can bump into each other and trigger the formation of new stars. The entire disk of the Galaxy is roughly 100,000 light-years across. The Sun is located within the disk, about 30,000 light-years from the galactic center.

While most of the stars and gas within the Milky Way lie in the Galaxy’s disk, there is actually a lot of matter that is distributed in a larger spherical region that is called the halo. The halo is very large, about 500,000 light-years in diameter. Most of the globular clusters in our Galaxy lie within the halo, as does a significant amount of dark matter. Dark matter is matter that does not emit light, so we can not see it with our telescopes, but we know it is there because of its gravitational effects. In fact, as we shall see in later chapters, the dark matter dominates the total mass of the galaxy, comprising by far the bulk of the entire mass of the galactic system.



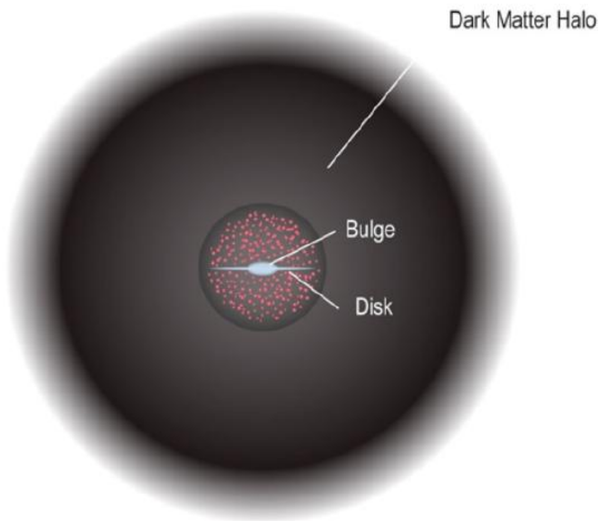


Figure 1.3.6: Drawing of the bulge and disk of our Milky Way Galaxy (face-on view). Credit: NASA/JPL-Caltech/R. Hurt. (right) Drawing of our Milky Way Galaxy including the dark matter halo (side view). Credit: NASA/SSU/Aurore Simonnet.

? Exercise: SIZE AND SCALE OF OUR GALAXY

Adapted from Prather, Edward E.; Slater, Timothy F.; Adams, Jeffrey P.; Brissenden, Gina; and the CAPER Team, *Lecture Tutorials for Introductory Astronomy*, 2nd Edition, © 2008, pp. 123-126. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.

Figure A.1.1 shows a picture of the spiral galaxy NGC 3184. This is much like what astronomers think the Milky Way Galaxy would look like if we could view it from outside, which, of course, we cannot do from our vantage point within the Galaxy. Nonetheless, we will assume that this is a picture of our own Galaxy and then use this model to try to understand the size scale of the Milky Way. Answer the following set of questions by referring to the picture, noting the size scales indicated by the arrows.

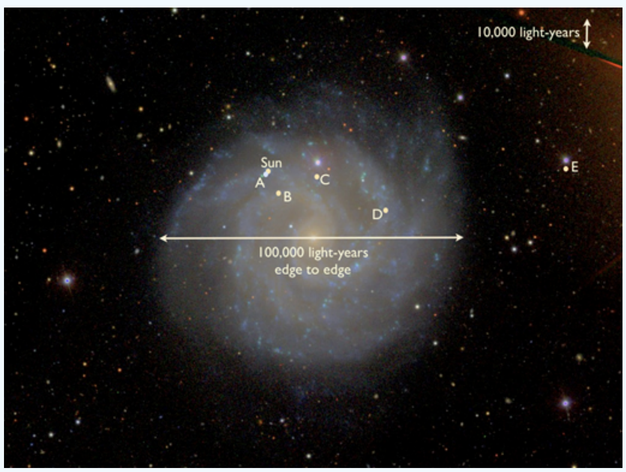


Figure A.1.1: An image of NGC 3184 – to be used as a model for our Milky Way Galaxy. Credit: Sloan Digital Sky Survey.

This page titled 1.3: Our Galaxy - The Milky Way is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.4: Other Galaxies and Large-scale Structures in the Universe

Learning Objectives

- You will know the types of galaxies: spiral, elliptical, irregular
- You will know that galaxies can be grouped into galaxy groups and clusters and form web-like structures on the largest scales

WHAT DO YOU THINK 1.4.1: DIFFERENT TYPES OF GALAXIES

Three students are talking after class about the topic for next week: galaxies.

> **Germaine**: Our professor said that next week we'll talk about different types of galaxies. Aren't they all the same, like a cinnamon swirl?
> **Holly**: I've seen a bunch of pictures of galaxies and some look different, more like a big ball.
> **Ignacio**: I think there are different types. Those spiral ones—like Germaine's cinnamon swirl—are by themselves, but there are others that string together like spider webs.

Being able to measure distances was key to figuring out what galaxies are. Well into the early 20th century, there was still much debate about the nature of these objects. Some astronomers believed that all of the fuzzy patches in the sky were relatively nearby clouds of gas and dust, or in other words, nebulae. As technology improved and distance measurement techniques became more accurate, astronomers realized that some of the fuzzy-looking extended objects we see in the sky are indeed other galaxies: They are distant collections of hundreds of billions of stars, with attendant gas and dust. The reason they look fuzzy is that they are so far away that we cannot resolve the stars within them with our eyes or with binoculars. In fact, the astronomer Edwin Hubble (1889–1953, for whom the Hubble Space Telescope was named) was the first to prove that one of the fuzzy patches in the sky, Andromeda (M31), was actually a neighboring galaxy. Hubble announced his discovery in 1925; until this time, astronomers debated whether or not all of the objects in the sky were part of our own Galaxy.

Under dark skies, the Andromeda Galaxy can be seen even with the naked eye in the Northern Hemisphere. In the Southern Hemisphere, two other galaxies can be seen with the naked eye: the Large and Small Magellanic Clouds. In cosmic terms, these galaxies are relatively close to the Milky Way —Andromeda, for example, is "only" 2.5 million light-years away. Most galaxies are much farther away. Their distance makes them faint, and so we need telescopes to observe them.

Figure $1.4.1$: The Hubble Ultra Deep Field illustrates some of the variety of sizes and shapes of galaxies. A few foreground stars within our own Galaxy can also be seen in the image: look for bright objects with cross-shaped spikes (an artifact introduced by the structure of the telescope). Credit: NASA/STScI.

Figure $1.4.1$, the Hubble Ultra Deep Field, demonstrates the vast number of galaxies we can observe when we train a powerful telescope on the sky for a long period of time. To collect this image, astronomers pointed the Hubble Space Telescope at an apparently blank area of the sky and collected many snapshots totaling an exposure time of one million seconds. The long exposure time of this image allows it to reveal very distant, faint galaxies.

In the Hubble Ultra Deep Field, we see an example of the assortment of galaxies found in the Universe. They are classified into three main types: spiral, elliptical, and irregular (Figure $1.4.2$). Like our Galaxy, other spiral galaxies usually have on-going star formation. As a result, they still have younger, bluer stars. They are typically found in regions of the Universe that are less densely packed with galaxies. The Milky Way and nearby Andromeda Galaxy are relatively large, massive spirals galaxies.

Elliptical galaxies look like balls. Sometimes they are squashed or elongated, like footballs or even cigars. All of these shapes go under the name of spheroidal. They are usually found in denser regions of the Universe, and they have less gas, dust and star formation than spiral galaxies. Elliptical galaxies usually contain mostly older, redder stars than spirals do.

The majority of galaxies are irregular in shape, meaning they do not exhibit spiral or ellipsoidal character. In fact, they don't really show any discernable general trend in their shape, hence the name irregular. Both the Large and Small Magellanic Clouds are irregular galaxies.

Finally, a few rare galaxies are called active galaxies. They emit a tremendous amount of energy from their cores, and therefore appear very bright. This energy output is thought to be due to an in-falling stream of material onto the central black hole in the galaxy. Active galaxies can by any type of galaxy, though they are usually very large spiral or elliptical galaxies. We will study them in more detail in later chapters.

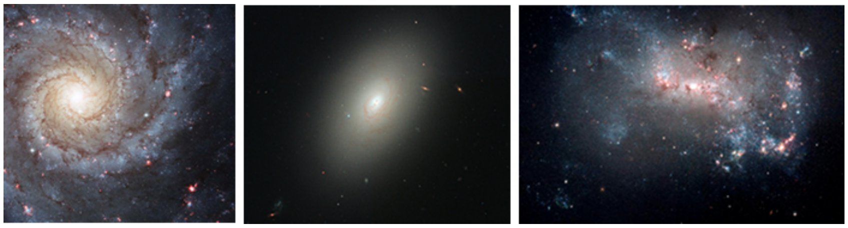


Figure $1.4.2$: (a) Spiral galaxy M74. (b) Elliptical galaxy NGC 4150. (c) Irregular galaxy NGC 4449. Credit: NASA/STScI.

1.4.1: GALAXY GROUPS AND CLUSTERS

Like stars, but on a much larger scale, the majority of galaxies are found in small "groups." Galaxies within a group are gravitationally bound to each other; in other words, they orbit each other due to their mutual gravitational pull. The Milky Way and

Andromeda galaxies are members of a group that we call the Local Group (Figure 1.4.3). The Local Group actually contains 40–50 galaxies, most of which are called “dwarfs” because they are very small compared to large galaxies like the Milky Way.

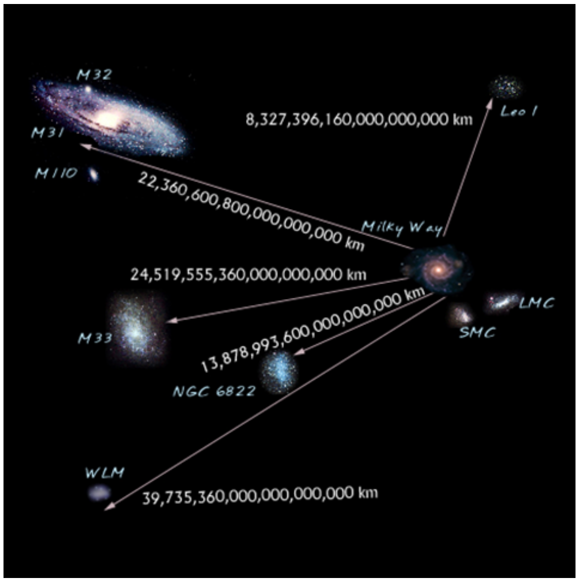


Figure 1.4.3: Diagram of the galaxies that make up the Local Group, with sizes approximately to scale (but distances between them are not to scale). Credit: NASA/HEASARC/Maggie Masetti.

Some galaxies are grouped into clusters (Figure 1.4.4). Clusters of galaxies contain hundreds to thousands of galaxies, and some clusters are themselves grouped together into larger “superclusters.” Galaxy clusters also contain clouds of hot gas (30–100 million kelvin), and this gas comprises more mass than all of the galaxies in these cluster. The hot gas can be observed with X-ray telescopes. Like galaxy groups, clusters of galaxies are held together by gravity. Although the galaxies and gas in a cluster are quite massive, they do not have enough mass for gravity to keep the cluster bound together, so we know there must be a large amount of additional mass - dark matter - in a given galaxy cluster to keep the cluster from flying apart.

Figure 1.4.4: Hubble ACS image of galaxy cluster Cl0024+17. The majority of yellowish-looking galaxies in this image, and some of the blue galaxies, are members of the cluster. Other galaxies in the image are foreground and background galaxies. You can also see a few foreground stars from our own Galaxy in the image. Credit: NASA/HST/ACS.

1.4.2: LARGE-SCALE STRUCTURE

On the largest observable scales, we have seen a web of galaxies and galaxy clusters filling the Universe. The galaxies and clusters are found in thin, filamentary structures and sheets, something like a three-dimensional spider web, or maybe a better analogy is the pattern of bubbles in a bubble bath. In the space between the filaments and sheets (the regions inside the bubbles in the analogy) are regions devoid of galaxies (Figure 1.4.5).

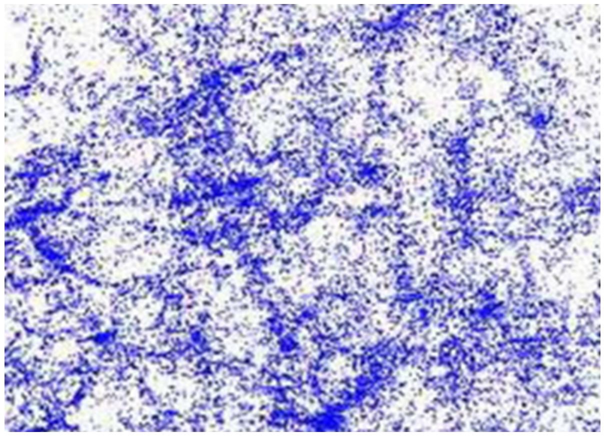


Figure 1.4.5: A small sample of the 2dF galaxy survey, approximately half a billion light-years across. Each blue dot represents a galaxy. The large-scale structure of space is web-like, with large filaments and superclusters of matter, and great empty spaces known as voids. Credit: Anglo-Australian Observatory

Table 1.6 Galactic Scales and Larger

Mass of a galaxy	10^8–10^{12} solar masses
Disk of the Milky Way Galaxy	100,000 ly
Halo of the Milky Way Galaxy	500,000 ly
Disk of the Andromeda Galaxy	220,000 ly
Distance to Andromeda Galaxy	2.5 million ly
Size of a galaxy cluster	3–30 million ly
Mass of a galaxy cluster	10^{14}–10^{15} solar masses
Typical size of voids in the cosmic web	30–500 million ly

This page titled 1.4: Other Galaxies and Large-scale Structures in the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.5: The Smallest Stuff- Particles, Atoms, and Molecules

Learning Objectives

- You will know that matter is composed of atoms, which are composed of particles

What Do You Think 1.5.1: Particles and Atoms

Three students are studying for their upcoming exam.

Jue: I am confused by all these different particles. Atoms make up electrons and protons, is that right?
Keiko: No, electrons and protons make up atoms—atoms are bigger than either electrons or protons.
Lourdes: That's right. Neutrons are part of atoms too.

Our bodies, the air we breathe, and the planets and stars are all composed of atoms. An atom is composed of a dense central nucleus made up of protons and neutrons, which is surrounded by an electron cloud. Since the electron cloud is much larger than the nucleus, most of an atom is empty space. An atom is about 10^{-10} m in size, while the nucleus, at 10^{-15} m, is 100,000 times smaller. To put this into perspective, if the nucleus were half the length of an American football field, the size of an atom would be about the distance across the continental United States.

The type of atom, or chemical element, depends on the number of protons it contains—this is called its atomic number. Changing this number would mean changing the element and would require a nuclear reaction to do so. Figure 1.27 shows the Periodic Table of the Elements. It organizes all of the known elements, and this version gives the symbol for each element and its atomic number (number of protons) and atomic mass (number of protons plus neutrons, averaged over different forms of each atom called isotopes

- see below). The mass of an atom depends almost entirely on the number of protons and neutrons; the electrons are so much lighter that their contribution to the mass is negligible. All known elements are shown in this table, including many that are familiar to you such as carbon (C), oxygen (O), and iron (Fe), and some that may be less familiar because they occur less frequently in nature, for example, lithium (Li) and beryllium (Be).

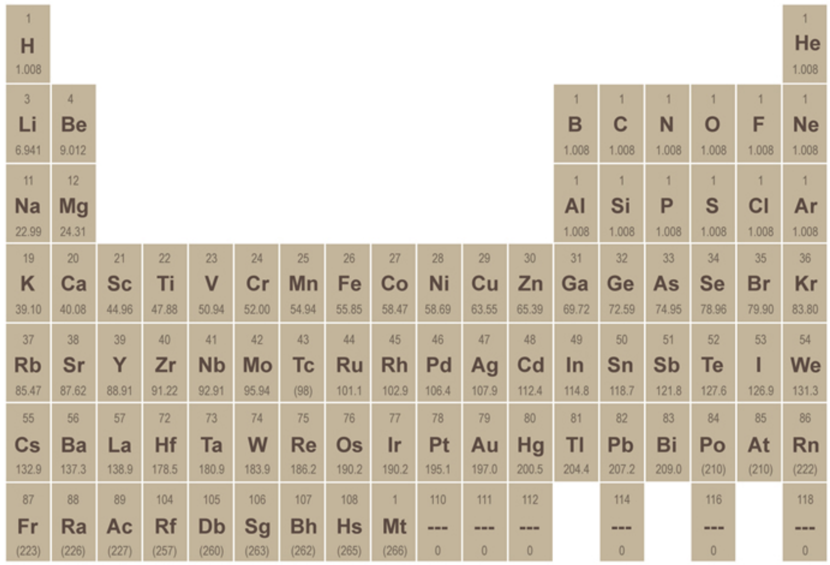


Figure 1.27: Periodic table of the elements. Letters denote atomic symbol, number above denotes atomic number, and number below denotes atomic mass. Credit: NASA/SSU/Aurore Simonnet. In the interest of space, we have not included atomic numbers 58-71 and 90-103.

The atomic number of an atom determines its chemical properties, because each proton is paired with an electron orbiting the nucleus, and the outer electrons of an atom determine the atom’s chemistry. Since the positive electric charge of the proton has the same strength as the negative electric charge of the electron, the total charge of an atom is zero. In some cases, the atom can lose some of its electrons, or more rarely, gain extra electrons. In either case, the atom will have a net electrical charge and is then called an ion.

Two atoms with the same number of protons but a different number of neutrons are known as isotopes of the same element. For example, the most common form of hydrogen has just a single proton for a nucleus (atomic number 1, mass number 1). An isotope of hydrogen that has one proton and one neutron in the nucleus is called deuterium (atomic number 1, mass number 2). Both of these isotopes of hydrogen have a single electron orbiting the nucleus (whose negative electric charge balances the positive charge of the proton). As another example, helium usually has two protons and two neutrons (atomic number 2, mass number 4), but there is an isotope of helium that has two protons and one neutron (atomic number 2, mass number 3). Both of these isotopes of helium have two electrons. Isotopes of an element are often abbreviated with the element’s name or symbol and the mass number. For example, the isotope of helium with just one neutron is abbreviated helium-3 or ^{3}He; the more common form of helium, used in helium balloons, is abbreviated helium-4 or ^{4}He. In Figure 1.27, the mass number reported is a weighted average over all the different isotopes for that species. For this reason, the mass numbers shown are not integers.

While we often think of atoms as the building blocks of matter, they are themselves composed of the subatomic particles protons, neutrons, and electrons. As it turns out, there is a series of more fundamental, or elementary, particles, from which all matter in the Universe is made. Protons and neutrons are composed of smaller particles yet, called quarks. There are six types of quarks. Two of the most common types are the up quark (u) and the down quark (d). Protons are composed of two up quarks and one down quark (uud), and neutrons are composed of one up quark and two down quarks (ddu), as shown in Figure 1.28. There are hundreds of particles that do not make up chemical elements that can be built from other combinations of two or three quarks.

Each fundamental particle also has a corresponding antimatter particle or antiparticle. Antiparticles are similar to particles except that their charges are reversed. So, for example, an anti-proton has the same mass as a proton, but it has a negative electric charge instead of a positive one. Further, the antiproton would be composed of corresponding groups of anti-quarks, so two anti-up quarks and one anti-down quark. A corresponding scheme would make up the anti-neutron.

 https://phys.libretexts.org/@go/page/31210

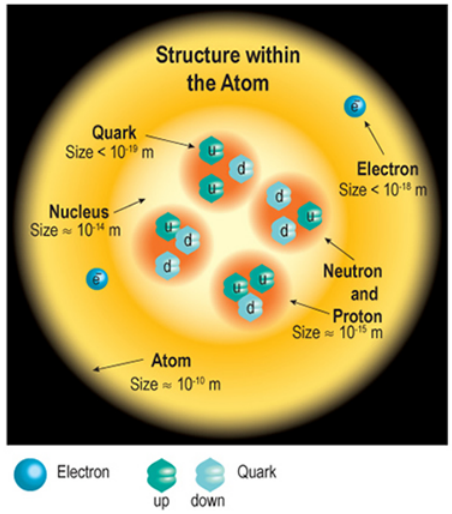


Figure 1.28: The nucleus of an atom is composed of protons and neutrons, which are in turn composed of quarks. Surrounding the nucleus is a large cloud of orbiting electrons. Credit: NASA/SSU/Aurore Simonnet.

Unlike protons and neutrons, which are made up of smaller particles, the electron is itself an elementary particle. Electrons are part of a class of particles known as leptons. As with quarks, there are six types of leptons (and six corresponding anti-leptons). Neutrinos are another example of leptons. Neutrinos have such a small mass and interact so weakly with matter that even though billions of them pass through your body every day - as many coming up from the ground after passing entirely through the Earth as coming down from the sky before then passing on through you and the Earth - that you will never notice them.

The third and final class of fundamental particles consists of the field bosons, which transmit the four fundamental forces of nature. These four forces are: the electromagnetic force, which keeps electrons bound to the nucleus and thereby determines the structure of atoms; the strong nuclear force, which holds the nucleus and the particles in it together; the weak nuclear force, which is important in radioactive decay and other interactions; and gravity, which dominates on astronomical scales. For example, the particle associated with the electromagnetic force is the photon, or particle of light. For the strong nuclear force, an exchange of gluons holds quarks together to form larger particles. If all this sounds very complicated, it sort of is. But don't worry. We will return to these concepts again in later chapters when we discuss the conditions that were present at the beginning of the Universe. Hopefully with further study this will all become less confusing.

The three classes of elementary particles: quarks, leptons, and field bosons, are summarized in Figure 1.29.

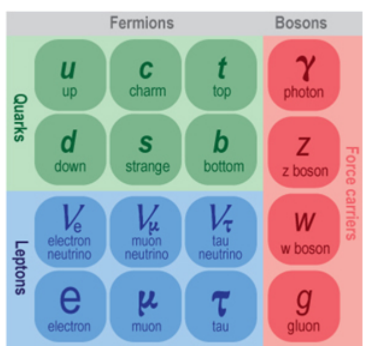


Figure 1.29: Three classes of elementary subatomic particles make up the building blocks of the Universe: quarks, leptons, and field bosons. Credit: NASA/SSU/Aurore Simonnet.

Going larger, molecules are structures that contain two or more atoms bound together by sharing electrons between them. This sharing is called a chemical bond. The atoms of any particular molecule have a definite pattern or structural arrangement that is determined by these bonds. Molecules can have all the same atom or a combination of different atoms. For example, oxygen (O_2) is composed of two oxygen atoms, and water (H_2O) is made of two hydrogen atoms and one oxygen atom. The size of a molecule is dependent upon the number of bound atoms and the distance between these atoms. Most molecules are so small they are measured in units such as nanometers or micrometers, but sizes can range anywhere from 10^{-10} m for molecular hydrogen, the smallest molecule, up to 5 cm in the case of a human DNA strand (if it were uncurled). DNA is one example of a rather large molecule that is present in cells, the building blocks of life. There are about 10^{14} cells in the human body: a typical human cell size is 10^{-5} m.

Table 1.7 Sizes of Particles and Other Building Blocks

OBJECT	SIZE
Quarks	$< 10^{-18}$ m
Leptons (including electrons)	$< 10^{-18}$ m
Protons	10^{-15} m
Neutrons	10^{-15} m
Atomic nucleus	10^{-15} to 10^{-14} m
Atom	10^{-10} m
Molecule	Typical ~10^{-6} m, range 10^{-10} m – 0.05 m
Cell	Typical 10^{-5} m

This page titled 1.5: The Smallest Stuff- Particles, Atoms, and Molecules is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.6: Hierarchy- How the Objects Are Arranged

Learning Objectives

- You will be able to rank objects by size and distance
- You will be able to sort objects into the hierarchy of solar system – galaxy - universe

What Do You Think: Solar System or Galaxy

The Stargazers Club is taking their first observing trip of the year. They are looking at a band of light that goes across the sky.

- **Maggie**: I love how you can see the solar system out here! You could never see that back home.
- **Nelson**: I don’t think that’s the solar system—I think it is our own galaxy .
- **O’Shea**: I don't think we can see either one. How can we see our own solar system or our galaxy if we’re inside them?

In the last few sections, we have discussed the many different scales encountered in the Universe. In addition, we have mentioned many different kinds of objects. Scientists have discovered a hierarchy of structures in the Universe, with small objects being collected into successively larger objects. On the smallest scales studied (below 10^{-15} m), there are subatomic particles. These particles and their interactions produce structures like atomic nuclei and atoms. In turn, the atoms form larger structures, the molecules and crystals, which form objects like rocks, oceans, and other parts of the macroscopic world. In living things, molecules are combined to form the parts of cells, which can subsequently be built up into plants and animals. Collections of these sorts of objects come together to make up Earth. This hierarchical structure goes all the way up to the cosmic web. Between the scales of the planets, about 10^7 to 10^8 m, and the scales of galaxy filaments and voids, hundreds of millions of light-years (about 10^{24} m), there are many intermediate-sized objects. The following activities will help you clarify their order.

Ranking Sizes and Distances

A. SIZE RANKING

When you play the activity, you will see seven tiles, each with the name of an astronomical object. Rank these objects from the smallest to largest.

When you are ready, click and drag each tile into the box you think best represents the size ordering of the objects. The smallest objects should be on the left, the largest on the right. When you have finished, click the “check” button to see how many got correct.

If you place a tile in a box and later decide that a different object would be better suited in that location, just drag and drop the new tile into the box and it will replace the old tile.

Play Activity

B. DISTANCE RANKING

Now rank the same seven objects from closest to Earth to farthest from Earth.

When you are ready, click and drag each tile into the box you think best represents the size ordering of the objects. The closest objects should be on the left, the farthest on the right. When you have finished, click the “check” button to see how many got correct.

If you place a tile in a box and later decide that a different object would be better suited in that location, just drag and drop the new tile into the box and it will replace the old tile.

Play Activity

ACTIVITY: Matching Sizes and Distances

Drag the tile to match the correct size or distance.

When you are done, click "check answers," and modify your choices if needed.

Play Activity

ACTIVITY: Sorting the Solar System, Galaxy, Universe

Decide whether each object is a member of our Solar System, our Galaxy, or the Universe.

Once you feel that you have put each object into the correct bin, click "check," and modify your choices if needed.

Play Activity

ACTIVITY: Hierarchy of the Universe

A hierarchy is an ordered list that is used to classify related objects. For example, cities are contained within states, which are contained within countries, which are all on Earth.

Arrange the astronomical object tiles from the list in hierarchical order, according to which objects are contained within the other objects.

You will note that some boxes have arrows leading away from them while others do not. To place an object within its encompassing object, drag and drop the tile to any of the boxes below it connected by an arrow. Astronomical objects which do not contain any of the items in the list will go in the boxes without arrows leading away.

When you have finished, click "check answers," and modify your choices if needed.

Play Activity

This page titled 1.6: Hierarchy- How the Objects Are Arranged is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.7: Scale the Universe

Learning Objectives

- You will create a scale model of the Solar System and relate it to everyday and galactic scales

What Do You Think: Scale Model Solar System

One day while eating lunch, Adam, Carlos, and Jemisha are discussing the scale of the Solar System.

> **Adam**: “If this cantaloupe was the Sun, then I think Earth would be the about as big as a cherry.”
> **Carlos**: “No. I think Earth would be smaller than that. It would be about as big as this raisin.”
> **Jemisha**: “I think raisins and cherries are still too big. I think Earth would be the size of this little black poppy seed on my bagel.”

You might not be aware that all the planets of the Solar System revolve around Stockholm, Sweden. Or more precisely, that they revolve around the Ericsson Globe, an indoor hockey arena in the city. In a leap of imagination, the Swedes have used the Globe—which is a semi-spherical building resembling an over-sized golf ball with a 110-m diameter—as the basis for a scale-model solar system. The Globe plays the role of the Sun, and all of the other planets, along with their distances from the Sun, are scaled accordingly. This gives a model solar system 20 million times smaller than the actual one, as shown in Animated Figure 1.30.

To find Earth, you must travel about 7.5 km from the arena and look for a sculpted sphere 64 cm in diameter. Jupiter is nearly 40 km away and is more than 7 m across. To find the last known planet in the solar system, Neptune, you must travel more than 200 km and find a ball 2.5 m across. A few trans-Neptunian objects, like Pluto, Eris, Sedna, and others, are also included in the map. If you travel to the far north of the country, almost to the Finnish border, you will find a representation of the terminal shock of the solar wind. That is where charged particles from the Sun slow down as they run into gases from the interstellar medium. Clearly, it helps to have a car or a train schedule and a good set of directions to visit all the objects in this model solar system.

If you would rather view the Solar System on a more manageable scale, you could travel to the National Air and Space Museum in Washington, D.C. The museum has built a model, called Voyage Through the Solar System, pictured in Figure 1.31, using a much smaller scale: each of the objects is 10 billion times smaller than its real counterpart. On this scale, the Sun is about 16 cm across, Earth is 15 m away, and even the farthest planet, Neptune, is only 450 meters distant. The smaller scale makes the Voyage scale model much easier to take in at a glance. The price paid, of course, is that some of the objects are very small. Jupiter is a small sphere, about the size of a marble. Earth is tiny, barely more than one millimeter in diameter. And, of course, objects like the Moon, Pluto, and other subplanetary bodies are barely visible at all.

Figure 1.31: A view of the inner part of the Voyage Through the Solar System scale model in front of the National Air and Space Museum in Washington, D.C. Credit: Smithsonian Institution, Eric Long.

So, what is the best scale to use for a model solar system? The next activity will let you decide for yourself. You will be asked to choose a scale and then to build a solar system model using commonly available objects for the Sun and planets. If you choose too small a scale, you will have a difficult time finding objects to represent Mercury, Earth, Mars, etc. If you choose too large a scale, you will have a difficult time placing all the objects at their proper distance from the Sun. There is no getting around it. The distances between the planets are so vast compared to their sizes that finding a good scale to represent everything is a difficult task. And what about the stars? We will ask you to consider that when you have finished your model.

Scale Model

To get started, choose a scale. You should probably choose something smaller than 1:20 million unless you want your model to stretch all the way across Sweden! Use the tables of data for the Solar System, and scale all the values down by your chosen scale. For instance, if you use a 1:10 billion scale, then you would divide the size of the Sun by 10 billion, and your model Sun would be about 0.14 meters across. All other sizes and distances would be similarly diminished. You can use a calculator or spreadsheet to do this task. Or if you prefer, there is a nice online tool to help you at: Build a Solar System.

If you are extremely squeezed for space, then it will be impossible for you to build a model that allows you to both see the planets and their separations from each other at the same time. Some teachers, when faced with this problem, have come up with a convenient solution that at least allows their students to get an idea of the distances separating the planets: they use a scale model where one sheet of toilet paper is 1 AU. This scale is easy to fit into a classroom, but it has the drawback that the planets are all too small to see. Even the Sun is a mere dot on this scale.

When you have chosen a scale for your model, think of some objects around your house that you can use to represent the planets. For instance, on the Voyage Through the Solar System model in Washington, D.C. (it uses a scale of 1:10 billion), Jupiter is about the size of a small marble, whereas Earth is about the size of a poppy seed. The entire model is about half a kilometer in radius, so you can see that you will have to use very small objects to represent the planets if you want your model to be a reasonable size.

https://phys.libretexts.org/@go/page/31212

As we have mentioned, a scale model such as the one described here has been built on the south side of the Capitol Mall in Washington, D.C. It is located in front of the National Air and Space Museum. Other models to the same scale can be found in several cities around the United States. In addition to these Voyage models, there are many Solar System scale models around the country (and the world) that use different scales. We have compiled a listing of these models for you. If one of them is nearby, it might make a good destination for an outing.

SCALE MODEL SOLAR SYSTEMS IN THE UNITED STATES

- McCarthy Observatory, New Milford, CT (links to home page; click on Solar Sys link at top of page to view info)
- University of Colorado Boulder

SCALE MODEL SOLAR SYSTEMS AROUND THE WORLD

- The Australian Solar System (page has driving instructions with distance and times given)
- The Swedish Solar System (page has information and map)

This page titled 1.7: Scale the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

1.8: Wrapping It Up 1 - Size, Scope, and Units

Learning Objectives

- You will be able to put everything together to demonstrate your understanding of the size and scope of the universe.
- You will be able to compare human, universe, and small scales.

In this activity, you will be pulling together everything you have learned in this chapter to compare the sizes, distances, and other properties of objects in the Universe.

For each of the questions in the sections below, first make predictions, one section at a time. Then check them, one section at a time, by using the *Scale the Universe* tool or by reading the text. Resolve any discrepancies, recording your observations below. To access the *Scale the Universe* tool, click the “play activity” bubble. You can pause the animation at any time and use the slider bar to navigate to objects of different sizes.

Play Activity

Part I. Small Scales


f(x)
s
a
b


f(x)
s
a
b

Part II. Earth, Moon, and Sun

Part III. Our Solar System

Part IV. Galaxies and the Universe

Part V. The Universe and You

This page titled 1.8: Wrapping It Up 1 - Size, Scope, and Units is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.


f(x)
s
a
b


f(x)
s
a
b

Questions to be graded for accuracy


f(x)
s
a
b


f(x)
s
a
b

This page titled 1.9: Mission Report 1 - Size, Scope, and Units is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

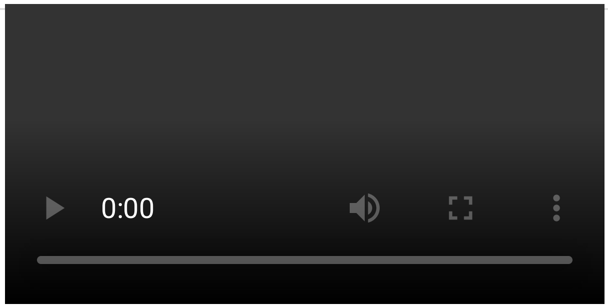
CHAPTER OVERVIEW

2: Light

Chapter 2 centers on the nature of light, which is humanity's primary source of information about the Universe. It discusses the dual nature of light as both a particle and a wave; the relationship between wavelength, frequency, and speed; the wavebands of the electromagnetic spectrum; and continuum and line spectra.

This page titled 2: Light is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.0: Light Introduction



Video Transcript

An Introduction to the Electromagnetic Spectrum: Transcript

Something surrounds you, bombards you, some of which you can't see, touch, or even feel. Every day, everywhere you go, it is odorless and tasteless yet you use it and depend on it every hour of every day. Without it the world you know could not exist. What is it? Electromagnetic radiation.

These waves spread across a spectrum from very short gamma rays to X-rays, Ultraviolet rays, visible light waves, even longer infrared waves, microwaves, to radio waves which can measure longer than a mountain range. This spectrum is the foundation of the information age and of our modern world. Your radio, remote control, text message, television, microwave oven, even a doctors X-ray all depend on waves within the electromagnetic spectrum.

Electromagnetic waves or EM waves are similar to ocean waves in that both are energy waves, they transmit energy. EM waves produced by the vibration of charged particles and have electrical and magnetic properties, but unlike ocean waves that require water, EM waves travel through the vacuum of space at the constant speed of light. EM waves have crests and troughs like ocean waves. the distance between crests is the wavelength. While some EM wavelengths are very long and are measured in meters, many are tiny and are measured in billionths of a meter (nanometers). The number of these crests that pass a given point within one second is described as the frequency of the wave. One wave or cycle per second is called a Hertz (Hz). Long EM waves such as radio waves have the lowest frequency and carry less energy. Adding energy increases the frequency of the wave and makes the wavelength shorter. Gamma rays are the shortest, highest energy waves in the spectrum.

So, as you sit watching TV not only are there visible light waves from the TV striking your eyes but also radio waves transmitting from a nearby station and microwaves carrying cellphone calls and text messages, and waves from your neighbors wifi, and GPS units in the cars driving by. There is a chaos of waves from all across the spectrum passing through your room, right now.

Almost everything that we know about the Universe ultimately comes from the light we observe. Looking at the night sky from a dark location is a breathtaking experience. But the Universe contains much more than is visible to the naked eye!

To go beyond the limitations of our eyes, we build telescopes and detectors that help us expand our physical perceptions. Some of these are the familiar visible-light telescopes seen at mountain-top observatories; they allow us to see fainter objects with more detail than our eyes alone could see. We also use sophisticated radio antennas and receivers—radio telescopes. In fact, there are different telescopes for all of the types of light in the electromagnetic spectrum: radio waves, microwaves, infrared light, visible light, ultraviolet light, x-rays, and gamma-rays. Each kind of light has a different amount of energy and interacts differently with matter. By looking at what happens to light when it is emitted or absorbed by various types of objects, or when the light emitter is moving through space, we can determine important physical properties of astronomical objects, such as temperature, density, and chemical composition.

In this chapter we will examine the nature of light, also called electromagnetic radiation because it is comprised of oscillating electromagnetic fields. Light travels as a wave, but it is often detected as a particle. These particles of light are called photons, and each carries a definite amount of energy related to wavelength of the photon.

Radiation is the term that astronomers use for any method of energy transport that can carry information across empty space (as opposed to the bulk transport of energy in matter, such as the bubbles in boiling water). This is what astronomers mean when they talk about radiation, not just radiation that is harmful to humans, though it is certainly the case that some types of light are quite harmful. Examples of harmful light are ultraviolet, x-rays and gamma rays.

In addition to simply looking with our eyes, we can also measure the properties of light in a variety of more quantitative ways. If you have ever seen a rainbow or looked through a prism, you know that light comes in a spectrum of many different colors. Using spectroscopy (spreading light into different colors), we can learn a wealth of information about astronomical objects. The information that can be encoded within spectra includes temperature, chemical composition and density. We can take pictures with specialized detectors similar to digital consumer cameras that record extremely faint astronomical sources and provide information about their positions in space and their energy output. The latter process is called photometry, and it is how astronomers acquire an understanding of energy flows into and out of celestial objects.

This page titled 2.0: Light Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.1: The Wave Nature of Light

> **Learning Objectives**
>
> - You will know that light can act like either a wave or a particle.
> - You will know that all types of light travel at the same speed.
> - You will be able to distinguish between wavelength, frequency, and speed. You will be able to perform calculations and understand conceptually the relationship between wavelength, frequency, and speed.

> **WHAT DO YOU THINK: COSMIC SPEED LIMIT**
>
> A group of students are working on their astronomy homework.
>
> - **Piper:** Ok, one of the things we have to remember is that all types of light travel at the same speed , the speed of light.
> - **Quentin:** Not radio, that travels at the speed of sound.
> - **Ricardo:** Yeah, and x-rays have a higher frequency and more energy, so they must travel faster.
> - **Sasha:** Radio waves are still light, so I think Piper is right. What I don't understand is how frequency and energy comes into play.

Much of your experience with the world comes from your senses, including your vision. Light enters your eyes, is converted to electrical signals, and is interpreted by your brain, allowing you to explore the world around you.

In the next interactive activity, you will predict what happens when different colors of light compete in a race.

PHOTON RACE, ROUND ONE

Ladies and gentlemen! Place your bets! Which colors of light do you think travel fastest through the vacuum of empty space?

Now that you have made your predictions, press the “start” button to enjoy the race.

Play Activity

After doing this activity, you should see that different colors of light all travel with the same speed in the vacuum of empty space. When you look at the objects around you, it might seem as if you are seeing the light from them infinitely fast because they are so close, but it turns out that light has a speed: 3×10^8 m/s (in SI units). This is a fundamental constant that scientists call “*c*”—and this speed is the same for light of any color. This is true not just for visible light, but for all forms of light in the electromagnetic spectrum.

You might have heard that light is a wave. In fact, one model of electromagnetic radiation is that of oscillating waves of electric and magnetic fields, i.e., electromagnetic waves. This is a classical model, developed at the end of the 19th century by a British physicist named James Clerk Maxwell (1831 – 1879).

Like water waves, or waves on a string, electromagnetic waves can be characterized by their amplitude, frequency, wavelength, and speed. The wavelength is the distance between adjacent peaks (or adjacent troughs, etc). Frequency is the number of complete waves, or wavelengths, that pass a given point each second. The amplitude of a water wave, or wave on a string, is its height, or how far it rises and drops from its midpoint. The amplitude of an electromagnetic wave is related to the strength of the electric and magnetic fields causing the wave. All light travels at the same speed, but each color has a different wavelength and frequency. Amplitude and wavelength are shown in Figure 2.1.

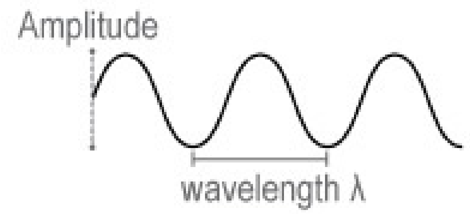


Figure 2.1: Waves are described by several properties, two of which are shown here. The amplitude of the wave is its height, or how far it rises and drops from its midpoint. The wavelength is the distance between adjacent peaks (or adjacent troughs, etc). Waves are also described by the speed at which they travel and by the frequency at which they oscillate. Credit: NASA/SSU/Aurore Simonnet.

Mathematically, the frequency, wavelength, and speed of any wave (not just light) are related by:

$$\text{wavelength} = \frac{\text{speed}}{\text{frequency}}$$

The units of frequency are named hertz in recognition of the contributions to the understanding of waves by Heinrich Hertz (1857 – 1894). Scientists also say “cycles per second” or “inverse seconds” when using the frequency unit hertz, which is written as s^{-1}, 1/s, or Hz. Wavelength and frequency are inversely proportional to each other, which means that if the frequency is *bigger* (higher), the wavelength will be *smaller* (shorter) and vice versa.

Given what we have learned about light so far, we can rewrite the equation above so that it applies to light waves:

$$\lambda = \frac{c}{f}$$

where c is the speed of light in m/s, λ (Greek letter “lambda”) is the wavelength in meters, and f is the frequency in Hz.

There is another way to think of this equation. If you realize that the frequency is how many waves pass per second, then the reciprocal of the frequency is the time needed for one wave to pass. For example, if ten waves pass each second (f = 10 Hz), then the time needed for one wave to pass is 1/10 Hz = 0.1 seconds. We call the time for one wave to pass the period, P, of the wave. So we know that $P = 1/f$. So for any wave, if it has a high frequency, then it has a short period, and vice versa. In other words, if a lot of waves are passing in each second (high frequency), then it doesn't take very much time for a single wave to pass, and the wave has a short period. On the other hand, if it takes a long time for one wave to pass by, then not very many of them will pass each second. We may rewrite the previous equation in terms of the period to get the following very similar equation.

$$\lambda = cP$$

This might not seem like a big improvement. But if you realize that the period is the time for one wave to pass, and that c is the speed it travels, then we know that the product cP, the speed times the time, is how far the wave travels over its period. But this is necessarily its wavelength, so this form of the equation makes the relationship between speed, period (or frequency) and wavelength a bit more natural. Unfortunately, scientists generally prefer to think about waves in terms of their frequencies, not their periods. This is unfortunate because, while the two quantities are equivalent in terms of their information about the wave, many beginning students have a hard time understanding what frequency is. But most can understand the notion of period - the time for a single oscillation of the wave - with much less effort. Either way of thinking about this equation is fine as long as you remember that the period is the reciprocal of the frequency.

While the wave model of light precedes the idea of electromagnetic waves by several centuries, Maxwell was the first to realize that the separate equations of classical electricity and classical magnetism could be combined into a single electromagnetic theory that predicted the existence of oscillating electromagnetic fields. With his electromagnetic theory, Maxwell was able to unify the (at that time) separate ideas of radio waves and visible light into a single theoretical framework. The equations also predicted the existence of higher-frequency waves, such as x-rays and gamma-rays. All of these were thereafter understood to be oscillating electromagnetic fields, with radio waves having long wavelengths and low frequencies, and gamma-rays having the opposite extreme of short wavelengths and high frequencies.

Knowing about the wave nature of light helps us interpret the different types of light that astronomical objects emit. In turn, this allows us to employ those waves to learn more about the processes that affect the formation and evolution of stars, galaxies, and the Universe itself.

The next activity, which uses the interactive Wave Generator tool, will give you some practice working with the ideas of wavelength, frequency, and speed. These are all important properties of waves.

WAVE GENERATOR

A wave generator is a laboratory device that creates wave signals whose properties can be changed by the user. In this activity, you will use a virtual wave generator, where the x-axis shows the wave's wavelength in meters and the y-axis shows the wave's amplitude. The amplitude of a wave essentially describes its intensity, or strength. Wavelength is the distance between two adjacent peaks of a wave. Frequency describes how many wave patterns or cycles pass by a certain point in 1 second.

You can make changes to the basic properties of the wave that is being generated in the following ways:

- Stop the wave generator to measure the wave at any time. Hover your cursor over the graph and you will see an x-y coordinate display of cursor position on the graph. Notice that the origin of the graph is on the far left, in the middle of the y-axis.
- Below the “stop” button, there are the wavelength and amplitude sliders. Use these sliders to increase or decrease the wave's properties in real time.
- There is also a stopwatch that counts time in seconds. You can start and stop the stopwatch, and also reset it to zero when you need to make a new measurement.

Use the *Wave Generator* to conduct the following measurements, and answer the following questions. All the waves generated have the same speed. Only the wavelength and amplitude can be changed with the sliders, though this can also change the frequency of the wave in accordance with the wavelength-frequency-speed relation above.

Play Activity

A. Measuring the wavelength of a wave

B. Measuring frequency

To determine the frequency, you will need to count the waves as they pass by a marker in a certain amount of time.

Worked Example (for a different wavelength):

Suppose 3 waves pass by the vertical slider bar (orange vertical line) on the wave generator in 3 seconds. What's the frequency?

- Given: Time = 3 seconds, number of waves = 3
- Find: Frequency in cycles per second
- Concept(s): f = number of waves/number of seconds

- Solution: f = 3 waves/3 seconds = 1 cycle per second or 1 Hz

Questions:

1. Now use the wavelength of 150 m from the previous section and the stopwatch to determine the frequency. You will need to count the waves as they pass by the orange vertical bar. You can move the bar along the x-axis to a position that is easy for you to watch.

- Start the stopwatch when the peak (or trough) of a wave passes the arrow.
- Let it run until at least 3 waves have passed by the arrow.
- Stop the stopwatch and read out the results for time passed.
- You may want to repeat this activity a few times to make sure you are reading the graph and the stopwatch accurately.

Show your work here:

C. Measuring wavelength, frequency, and speed.

We can use the relationship between speed, wavelength and frequency to calculate the speed of a wave.

Worked Example:

What is the speed of a wave with wavelength 100 m and frequency 1 Hz?

- Given: Wavelength is 100 m, frequency is 1 Hz
- Find: Speed of the wave in m/s
- Concept(s): speed = (wavelength)x(frequency)
- Solution: speed = 100 m × 1 Hz = 100 m s^{-1} = 100 m/s

Questions:

After completing the previous activity, you should have a good idea of what is meant by the frequency and wavelength of a wave. The relation between frequency and wavelength allows us to describe electromagnetic waves (in other words, light) using either frequency or wavelength without introducing any confusion. Scientists are comfortable speaking about light in terms of either frequency or wavelength. They tend to use whichever is most convenient at the time.

This page titled 2.1: The Wave Nature of Light is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.2: The Particle Nature of Light

Learning Objectives

You will be able to perform calculations and understand conceptually the relationship between energy and frequency

What Do You Think: Wavelength and Frequency

The same group of students continues on the next part of their homework; they are soon joined by another group, who are arguing over the correct answers.

- **Ting-li:** Hey all. We're having a debate about the homework, can you help us decide who's right?
- **Sasha:** We can try—what's the problem?
- **Ting-li:** Well, Uhura and I think that the longest wavelength light—radio waves —has the most energy .
- **Uhura:** It just makes sense, if one thing is bigger, so is the other. Vanessa thinks we're wrong. What do you think?
- **Sasha:** Oh, we just talked about that. Remember the equation for energy uses frequency , not wavelength. So if frequency is bigger, so is the energy.
- **Piper:** And frequency and wavelength are opposite each other. That means that big wavelengths have small frequencies, which means small energies.
- **Vanessa:** Right, like I said, the most energetic light is whatever has the highest frequency and shortest wavelength. Like gamma rays .

Do you agree with any of these students and if so, whom?

Piper

Sasha

Ting-li

Uhura

Vanessa

None

Explain

There is another way to describe light in addition to electromagnetic waves. In some cases, light behaves more like particles. Toward the end of the 19th century, Heinrich Hertz first observed that when a metal is exposed to light, it will absorb energy from the light and emit electrons. In subsequent experiments, physicists found that the amount of energy absorbed by the metal depends on the frequency of the light that is shining on it. This is known as the photoelectric effect. More specifically, the German physicist Max Planck (1858 – 1947) showed that the energy E of light with frequency f is given by the expression below.

$$E = hf$$

where the SI units for energy are joules (J) and the units for frequency are Hz. The constant of proportionality, h, is called **Planck's constant**. It is an extremely small number. In SI units, Planck's constant is 6.626×10^{-34} J s. The equation means that the *higher* the frequency of a photon, the *higher* its energy.

Until this time, models of electromagnetic radiation predicted that the energy of light should depend on its amplitude, or intensity; they could not explain the dependence on frequency. It was Albert Einstein (1879-1955) who clarified the matter when he explained the photoelectric effect. In deducing how the photoelectric effect works, Einstein showed that light can be modeled as discrete little packets, like particles, each having energy hf. More precisely, Einstein showed that light behaves as if it were made of discrete particles when it is absorbed.

Since the energy of a particle of light depends on its frequency, an incoming particle with a high enough frequency will have enough energy to liberate an electron from a metal. The higher the intensity of light shining on a metal, the more packets, or particles, the metal absorbs and the more electrons are emitted. However, if the frequency of the light is too low, then no electrons can be liberated, no matter how many of them are present. This is because no single light particle has the required minimum energy, and these particles are absorbed only one at a time.

In everyday life, the photoelectric effect explains the working of solar panels, for example. Because of the particle nature of light, we can capture the energy of sunlight. In astronomy, we make use of the particle nature of light by using the photoelectric effect to record the number of photons that come from a particular region of the sky. In this way, we can take an image of the sky and determine the brightness of an astronomical object.

Einstein's idea of light as discrete packets of energy was published in 1905. It was not until 21 years later that the chemist Gilbert Lewis published his own theory of particle light (now long abandoned) in which he called these particles of light photons. Though Lewis's theory was inconsistent with many behaviors of light known through experiments, his name for the particles, photons, was adopted immediately and has stuck to this day.

So, now when scientists talk about light, they might refer to electromagnetic waves or photons of a given energy, frequency, or wavelength. All are equivalent. While it can be confusing at first, this is how the nomenclature developed historically. No one has yet come up with a better system. After all, how do you describe something that seems to be both particle and wave, and which can have a wide range of frequencies or wavelengths?

Energy, Frequency, and Wavelength

Using the equations introduced in this and the previous section, calculate the wavelength, frequency, and energy of the light.

Example 2.2.1

A visible light photon has a wavelength of 500 nm (nm is short for nanometer, 10^{-9} m).

1. What is the frequency of this photon?
   - Given: λ = 500E-9 m
   - Find: frequency f
   - Concept(s): λ = c / f where c = 3E8 m/s
   - Solution: (500E-9 m) = (3E8 m/s)/ f
   - f = (3E8 m/s) / (500E-9 m) = 6E14 s^{-1} = 6E14 Hz
2. What is its energy?
   - Given: f = 6E14 Hz
   - Find: energy E
   - Concept(s): $E = hf$ where h = 6.626E-34 J s
   - Solution: E = (6.626E-34 J s) × (6e14 Hz) = 3.98E-19 J

Questions:

1. An ultraviolet photon has a frequency equal to 2.41 x 10^{15} Hz.

a. What is its wavelength?

m

Show your work:

b. What is its energy?

J

2. An x-ray photon has a wavelength of 10^{-11} m.

a. What is it's frequency?

Hz

Show your work:

b. What is its energy?

J

Show your work:

When students first learn about light, they often ask questions like, “So which is light, a particle or a wave?” These questions are understandable, but they miss the point. We cannot really see light (yes, our eyes detect it, but they are not able to measure its properties in detail). We can only measure light’s properties in careful laboratory experiments. In some experiments, light shows wave-like behaviors. That is, it can bend around sharp corners or small obstructions (think of water waves bending around the end of a jetty or around the pilings of a pier). At other times, as in the photoelectric effect, light behaves as if it were a particle. So, which is it really? To answer that question, let us think about what we are doing when we claim it is one or the other.

When we claim that light is a wave, what we are really saying is that it has similar characteristics to certain macroscopic objects (water or waves on a string are examples) that we call waves. Furthermore, we are saying that the same (fairly detailed) mathematical tools we use to describe water waves and waves on strings can be used to describe light. But does that really mean that light is a wave?

A better way to think about it might be that waves are one *model* that we can use to describe light. The model is very good at describing certain aspects of light. For other aspects, like when light is emitted or absorbed by a material, the wave model can fail completely. In that case, we have found that a different model, a particle model, works very well. But the particle model cannot explain the bending of light around edges.

So, light behaves in these two apparently incompatible ways. Does this mean that light is both a particle and a wave? Well, that does not seem to make a lot of sense. Perhaps it is better just to say that light is light. Sometimes it behaves like a wave, and sometimes like a particle; it depends on the experiment you do. What seems clear is that our two models of light, though they work quite well in their separate domains of applicability, fail to give us a fully satisfying picture of the nature of light. This will be a general aspect of many of the models we develop to explain new phenomena: We must generally try to explain new phenomena in terms of ideas already familiar, and those old ideas might be inadequate to fully explain the new cases we come across, at least in ways that make us comfortable.

This page titled 2.2: The Particle Nature of Light is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.3: The Electromagnetic Spectrum

Learning Objectives

- You will be able to rank the different bandpasses in the EM Spectrum by energy, frequency, and wavelength

What Do You Think: Radio Waves

On their way to a concert, three students decide to listen to the radio.

- **Angela:** “I wonder if radio telescopes work the same way as the radio we’re listening to. You know, by picking up sound waves that we can hear.”
- **Brianna:** “I don’t think so, there’s no air in space for the sound waves to travel through, so you wouldn’t be able to get any signals.”
- **Callie:** “Right, radio waves are light waves. The radio is really detecting light waves in the radio portion of the spectrum and converting them into sound waves that we can hear.”

Do you agree with any of these students, and if so, whom?

Angela

Briana

Callie

None

Explain.

While much can be learned from observing objects in visible light—the kind of light we see with our eyes—there is a tremendous amount of additional information carried only by other kinds of light. Over the past half-century or so, astronomers have learned to observe the Universe not only in visible light, but in all bands of the electromagnetic (EM) spectrum, from low-energy radio waves all the way up to high-energy gamma-rays. Figure 2.2 shows the (EM) spectrum.

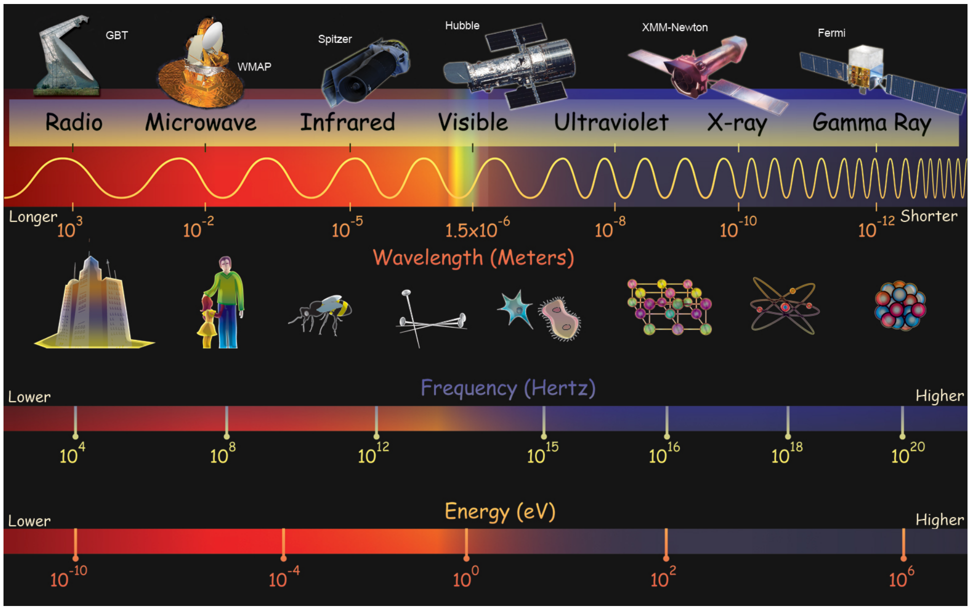


Figure 2.2: This figure shows the different bands of the electromagnetic spectrum as a function of wavelength, frequency, and energy. Also shown are common objects with sizes similar to the wavelength scale and telescopes that observe in each waveband. Credit: NASA/SSU/Aurore Simonnet.

The EM spectrum, of which visible light is a tiny part, is divided into somewhat arbitrary regions, or bands, that are based on how the light in that region is measured. These different methods of detecting or measuring the light result from the different wavelengths or frequencies of the light in each band. To give you an overview, and to put some useful labels on what we are discussing, we will briefly describe the different bands of the EM spectrum. It turns out that most portions of the EM spectrum are not visible to our eyes.

- ***Radio***—At the longest wavelength end of the EM spectrum are the radio waves. This is the kind of light used to carry radio signals to a stereo in your car and to your television (if you receive broadcast TV and not cable). Are you surprised to find that radio is a kind of light, and not a kind of sound? Many people are surprised by this fact when they first learn about it. Since astronomical radio waves have such long wavelengths, they are usually detected using telescopes that look like big dishes. There is no limit to the length of radio waves (on the long end), but the shortest radio waves have lengths of about a meter.
- ***Microwaves***—At the high-frequency end of the radio spectrum are microwaves. They are the types of waves used in microwave ovens, where they interact with the water molecules in your food to heat it up. They also carry the signals used by cell phones. Although the prefix “micro” usually means “one-millionth,” microwaves have typical wavelengths of centimeters (one-hundredth of a meter) to millimeters (one-thousandth of a meter).
- ***Infrared***—At slightly higher energies than microwave is infrared light. Infrared (IR) light is between microwaves and the red end of the visible light spectrum (infra is a prefix that means “below” — IR has a frequency below that of visible light). IR is sometimes called heat radiation, but this is a misnomer. As we will learn, all kinds of EM radiation are emitted by warm objects. But IR is emitted predominantly by objects we consider warm. Such objects include hot coals and warm-blooded animals (including people). Long-wavelength IR is also emitted by telescopes themselves. IR light has enough energy that it is not usually measured as a wave, but rather as a particle. IR photons have wavelengths in the range of hundreds to 1 or 2 microns (millionths of a meter).
- ***Visible Light***—Next is visible light, also known as optical light. We perceive many different colors of visible light, which correspond to different frequencies of the light. At the low-frequency end is red, followed by orange, yellow, green, blue, and violet. One common mnemonic used to remember these colors in the proper order is ROY G. BIV (where indigo has been inserted to make the phrase easier). Check it out the next time you see a rainbow. Optical astronomy has been done using telescopes made of mirrors and lenses since the early 1600s, when Galileo Galilei first used a telescope to study the night sky. Visible light spans the wavelength range 400–700 nanometers (billionths of a meter), or if you prefer, it has wavelengths clustered around half a micron.
- ***Ultraviolet***—Ultraviolet (UV) light is just beyond the blue end of the visible spectrum and it has somewhat higher frequencies than visible light. UV is what gives you a suntan (or sunburn with too much exposure). For the most part, exposure to light from radio through visible is generally without risk (but there are exceptions). At UV, this risk changes; too much exposure to light at

UV or higher frequencies can be harmful or even fatal because these particles have enough energy to dislodge electrons from the atoms in your body (in a manner similar to the photoelectric effect). Although some low-energy UV rays penetrate Earth's atmosphere, most higher-energy UV does not. Hence, most UV astronomy is done from detectors launched into space. UV photons have wavelengths typically measured in tens to hundreds of nanometers.

- ***X-rays***—X-rays have energies higher than UV. You probably know from visits to the doctor or dentist that x-rays can be used to create images of our insides. These images are not like normal photographs, however. They are really shadows. X-rays can penetrate many materials, but their penetrability depends on the density of the material. Thus, they pass easily through skin and muscles, but less easily through bones and teeth. As a result, x-ray images really just show contrast in the shadows caused by different parts of the body. They can also be used the same way to look into other materials, like machines and machine parts. X-ray images of astronomical objects are very different because the objects are emitting x-ray photons and the telescope is just catching them. Since Earth's atmosphere absorbs x-rays from space, x-ray astronomy is always done from instruments in space. X-rays have so much energy that we tend to describe them using the energy of individual photons, rather than specifying their wavelengths or frequencies. X-ray energies start at 1 keV (kilo or thousands of electron volts), which are a more convenient measurement than using the SI unit joules. (One eV is equal to 1.6×10^{-19} J.)
- ***Gamma-rays***—The highest energy light is called gamma-rays. At low energies, gamma-rays merge into x-rays. In principle, there is no upper limit to gamma-ray energies; their energies could go to higher and higher values. In practice, we will see there are practical limits to how high the energy of a gamma-ray can go. Gamma-rays are the least familiar kind of light because they are almost completely absent on the Earth's surface. This is a good thing. Gamma-rays interact so strongly with matter (including the matter comprising the proteins and DNA in our bodies) that a small exposure to them can be harmful, or even fatal. Gamma-ray energies start around 1 MeV (mega or million electron volts). The highest gamma-ray energies that have been directly measured from astronomical objects are in the TeV (tera or trillion electron volt) range, though indirectly detected gamma rays have been seen with energies more than a million times higher.

Now that you have learned about the different bands of the electromagnetic spectrum, you will do some activities exploring wavelength, frequency, energy, and speed.

RANKING THE EM SPECTRUM BANDS

A. Wavelength.

Rank the different bands of the EM spectrum in order from shorter wavelength (on the left) to longer wavelength (on the right).

Play Activity

B. Energy.

Rank the different bands of the EM spectrum in order from lower energy (on the left) to higher energy (on the right).

Play Activity

C. Frequency.

Rank the different bands of the EM spectrum in order from lower frequency (on the left) to higher frequency (on the right).

Play Activity

D. Relationships.

1. How did the three rankings you completed compare with one another?

PHOTON RACE REPRISE

Ladies and gentlemen! Place your bets! Which colors of light do you think travel fastest through the vacuum of empty space?

1. Rank the wavebands: gamma-ray, x-ray, optical, and radio, according to their speeds.

Now that you have made your predictions, press the "start" button to enjoy the race.

Play Activity

2. Describe how the outcome of the race compares with your initial predictions.

3. Explain how the equation that describes the relationship between a wave's wavelength and frequency ($\lambda = c / f$) relates to the outcome of the race.

UNITS OF ENERGY

In this activity, you will have a chance to practice converting the units of energy for photons using the conversion factor 1 eV = 1.6×10^{-19} J. Sometimes astronomers express energy in keV (10^3 eV) or MeV (10^6 eV).

Worked Example:

1. An x-ray photon has an energy of 1 keV. What is its energy in joules?

- Given: E = 1 keV = 1E3 eV
- Find: Energy in joules
- Concept(s): 1 eV = 1.6E-19 J
- Solution: E = 1 keV = 1E3eV × (1.6E-19 J / 1 eV) = 1.6E-16 J

Questions

1. A gamma-ray photon has an energy of 6.63×10^{-13} J. What is its energy in MeV?

MeV

Show your work:

SPECTRUM PUZZLE

Play Activity

Matching.

Above the table are several tiles, each with a specific energy, frequency, or wavelength. Drag and drop each tile into the appropriate row and column on the chart of the EM spectrum until all tiles have been placed. If the tile is placed correctly, a green check mark appears.

If you are having difficulty, you may want to refer to previous activities where you ranked the energies, frequencies, and wavelengths of the various bands of the EM spectrum.

Challenge round: Compute the values.

Once you complete part A, blank tiles will automatically appear in the rest of the chart.

Complete the chart by computing the energies, frequencies, and wavelengths for the remaining tiles and entering the values on the tiles:

- Click the "edit" button, which is the vertical bar on the right-hand side of the tile
- Enter the coefficient into the first small box
- Enter the exponent into the second small box
- Select the appropriate units
- Click the "edit" button again to check your answers

Try at least one example of each type to make sure you understand the calculations.

Show your work here:

This page titled 2.3: The Electromagnetic Spectrum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.4: What a Spectrum of Light Can Tell Us About Matter

Learning Objectives

- You will be able to describe how light can be measured as a function of energy (spectroscopy).

? What Do You Think: Rainbows

Some students are discussing a demo they recently saw in their class.

- **Demaryius:** "I loved seeing the rainbows in class today— that was so cool!"
- **Eugenia:** “Well, they weren’t really rainbows— they were spectra. But I agree that they are cool.”
- **Frank:** “What’s the difference?”
- **Eugenia:** “Spectra are when light is spread out over all of its different energies when it passes through some kind of scientific instrument. So I think spectra is the technical term for rainbow.”

Do you agree with any of these students and if so, whom?

Demaryius

Eugenia

Frank

None

Explain

Have you ever looked through a prism on a sunny day? Seen a rainbow reflected off a piece of jewelry or glass? Did you know that gorgeous rainbow of colors was already present in the sunlight before you held up your prism? All the prism did was spread the light out into a spectrum of different wavelengths, as shown in Figure $2.4.1a$.

By looking at the spectrum of light from an object, we can learn a wealth of information about the conditions of the object that produced the light, for example, its temperature, density, and composition. For this reason, spectroscopy, the study of spectra, is used in many sciences, including biology, geology, chemistry, and forensics, not just astronomy and physics. The term “spectrum” refers to how many photons are being emitted at each frequency (or wavelength or energy); it is just a fancy term to describe the distribution in energy. A familiar example of a distribution might be that of the grades in a class. The grade distribution describes how often each score occurred (i.e., how many students in the class received each score). In a similar manner, the spectrum of a light source describes how many photons of a given frequency (or wavelength or energy) the source is emitting. A spectrograph is the scientific instrument used to obtain a spectrum; a schematic of one is shown in Figure $2.4.1b$.

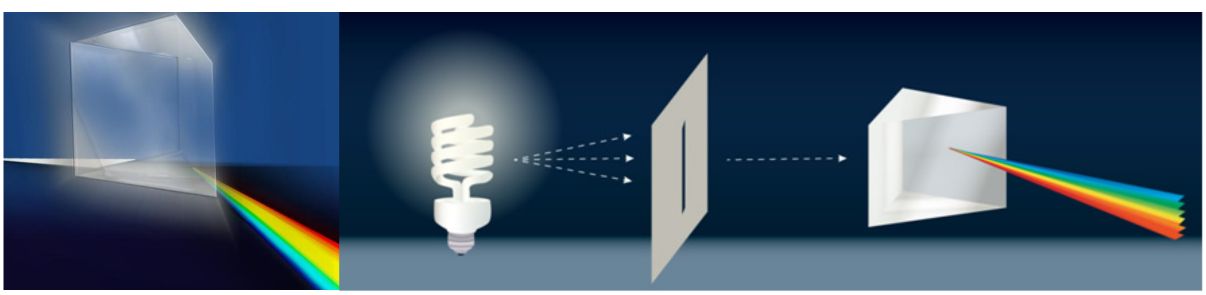


Figure $2.4.1$: Paste Caption Here

Figure 2.3 Prism and simple spectrograph (spectroscope). (a) A prism causes white light to be spread into its various colors because the different wavelengths of light have different speeds while they travel through the prism (different wavelengths travel at the same speed in a vacuum, but not otherwise). (b) Light enters a spectroscope through a small opening, such as a slit, and is then split into its constituent wavelengths by a prism-like device (which is usually a grating) so that each wavelength can be studied independently. The study of spectra has allowed astronomers to learn a tremendous amount about the Universe. Credit: NASA/SSU/Aurore Simonnet.

Some of the different distributions you might see include continuous spectra, emission-line spectra, and absorption-line spectra. Continuous (or continuum) spectra show continuous emission (no gaps) at all frequencies, or at least over a broad range of frequencies. Absorption-line spectra feature a bright continuum plus dark lines; something has absorbed light at very specific frequencies, taking those out of the mix. Emission-line spectra consist of bright lines on a dark background; the source is emitting light at very specific frequencies. See Figure 2.4.2 for illustrations of continuous, absorption, and emission spectra.

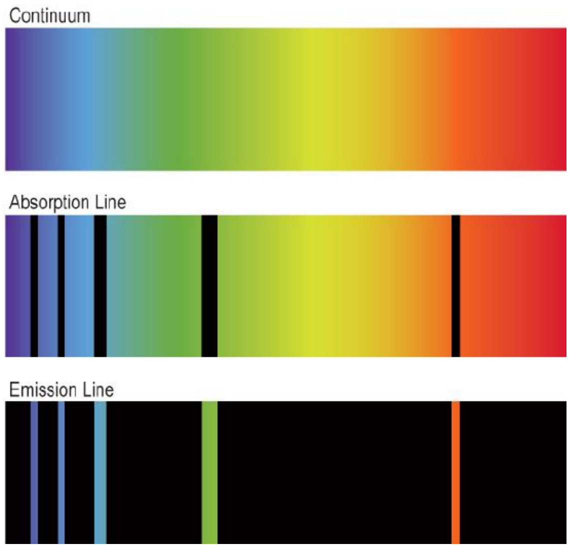


Figure 2.4.2: A continuous (or continuum) spectra has light at all frequencies. An absorption spectrum has dark gaps at wavelengths where atoms have absorbed some of the light. An emission spectrum has bright lines at wavelengths where atoms are emitting light. Credit: NASA/SSU/Aurore Simonnet.

This page titled 2.4: What a Spectrum of Light Can Tell Us About Matter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.5: Continuous Spectra - a Planck Spectrum Tells us the Temperature of Objects

Learning Objectives

- You will know that most light seen in the universe is thermal and can be represented by a Planck spectrum
- You will be able to perform calculations and understand conceptually the relationship between temperature and peak wavelength
- You will be able to perform calculations and understand conceptually the relationship between flux and temperature.

What Do You Think: Hot and Cold

Two students are sitting around after a barbecue, roasting marshmallows and discussing how the color and temperature of objects might be related.

- **Geraldine:** "I like my marshmallows charred, so I want to put mine in the hottest part of the flames. I heard that blue is the hottest, so that's where I'm going to roast my marshmallows."
- **Helen** "No; the convention is that red stands for hot and blue stands for cold. Like on a bathroom sink. So, if you want your marshmallows to get charred, I think you should put them near the redder flames."

Which student do you think has a better chance of burning her marshmallows?

Geraldine

Helen

Neither

Explain your reasoning:

For the most common type of continuous spectrum, which is called a Planck spectrum (named after Max Planck) or blackbody spectrum, the relative brightness of each color follows a distribution that depends on the temperature of the object. Any dense object, such as a person, a desk, or the filament in an incandescent light bulb, will radiate a Planck spectrum.

To better understand what a spectrum is and how it can be used, we will do a short interactive activity on the Planck spectrum.

Planck Spectrum

Play Activity

For quantitative work it is often convenient to graph a spectrum. This representation plots the brightness of light on the vertical axis and the corresponding frequency or wavelength on the horizontal axis. On the graph shown, a higher value for the relative brightness at a particular wavelength means the spectrum of light is brighter at that wavelength.

In this activity, you will see how the temperature of an object is related to its color, or more specifically, you will note the wavelength at which its Planck spectrum has its peak.

The colors are shown at the bottom of the graph in the activity, but you can also refer back to the discussion of wavebands in Section 2.3 in order to relate wavelength and color.

1. Set the temperature to 5834 K, about the same as the surface of the Sun.

a. What is the peak wavelength for this curve?

nm

b. What color does this correspond to?

red

yellow

blue/green

purple

2. Predict: If you set the temperature higher, the peak wavelength will?

be longer

be shorter

remain the same

3. Test your predictions by setting the temperature to 7,094 K. Describe your results and reconcile any differences. Be sure to note the peak wavelength and corresponding peak color.

4. Predict: If you set the temperature lower, the peak wavelength will?

be longer

be shorter

remain the same

5. Test your predictions by setting the temperature to 4,260 K. Describe your results and reconcile any differences. Be sure to note the corresponding peak color.

6. What waveband will the spectrum peak in if the temperature is set lower, for example 3,000 K?

7. At what waveband will the spectrum peak if the temperature is set much higher, for example 15,000 K?

8. Predict: If you set the temperature higher, will overall brightness be higher, lower, or remain the same?

9. Test your predictions by setting the temperature to 4,260 K, 5,834 K, and 7,094 K. Describe your results and reconcile any differences.

As you have just seen, the peak of the distribution of wavelengths of the emitted light shifts to shorter wavelengths as the temperature increases. So, for example, an object that glows with a blue color (shorter wavelength) is hotter than one glowing with a red color (longer wavelength). While an object with a Planck spectrum is emitting light at all wavelengths, the peak wavelength is the most prominent color. However, the peak of the spectrum need not be a color in the visible range. As a result, very hot objects, those with peaks well into the UV range, all appear bluish-white to our eyes; the part of their spectra in the visible is quite similar regardless of their respective temperatures.

Quantitatively, the temperature (T, measured on the kelvin scale) is inversely proportional to the peak wavelength (λ_{peak}) in meters:

$$T = 2.9 \times 10^{-3} \lambda_{peak}$$

This relationship, called **Wien's Law**, means if λ is *smaller*, T is *bigger*, and vice versa.

You also saw that a hotter object is also brighter overall (when comparing objects of the same size). Brightness is also known as intensity, or flux. It is the amount of light energy passing through a unit of surface area in a certain amount of time. The units for flux are watts/m^2, where 1 watt (W) = 1 J/s. The units of watts should be familiar to you from household light bulbs.

Quantitatively, we find the total flux of light (F) is proportional to the temperature (T) to the fourth power.

$$F = \sigma T^4$$

The constant σ = 5.67 × 10^{-8} W m^{-2} K^{-4} (Greek letter "sigma") is called the **Boltzmann constant**. This equation is called the **Stefan-Boltzmann Law**. It means, for example, that if the temperature is 2 times hotter, the flux will be 16 times greater.

So, in summary, a hotter object is bluer than a cooler object, and if two objects have the same size, then the hotter object will be brighter than the cooler object. Knowing this helps us interpret the color and brightness of different astronomical objects.

This page titled 2.5: Continuous Spectra - a Planck Spectrum Tells us the Temperature of Objects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.6: Lines Spectra- Emission and Absorption Lines

Learning Objectives

- You will be able to distinguish between emission and absorption lines in a spectrum
- You will know how spectral lines are produced
- You will be able to calculate the energy/frequency/wavelength of a photon absorbed or emitted by a hydrogen atom

What Do You Think: Spectral Lines

Some students are completing an activity on absorption and emission spectra and discussing their answers.

- **Isóbel:** "I'm having a hard time remembering which type of lines are created by which process. Emission lines come from gas clouds, right?"
- **José:** "I don't think so, don't they come from hot sources? Gas clouds are cold."
- **Kensi:** "Emission lines are created in hot sources, like a fluorescent lamp. Absorption lines are made when light passes through a colder source like gas clouds, and the gas absorbs particular wavelengths of light."

Do you agree with any of these students, and if so, whom?

Isóbel

José

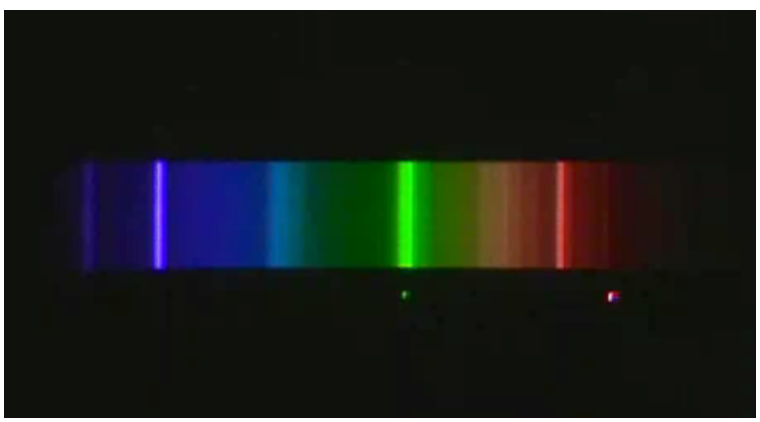
Kensi

None

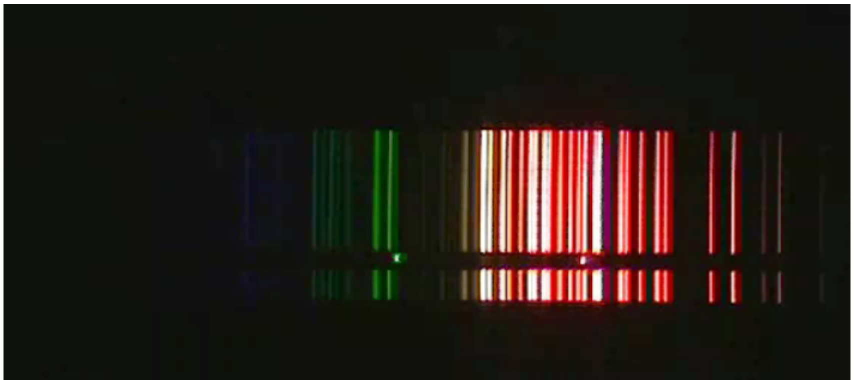
Explain

We now turn to non-continuous, or discrete, spectra. In discrete spectra, only a few frequencies are observed. We will learn about two kinds of discrete spectra: emission and absorption spectra. Figure 2.5 shows spectra from some everyday sources of light.

a.



b.



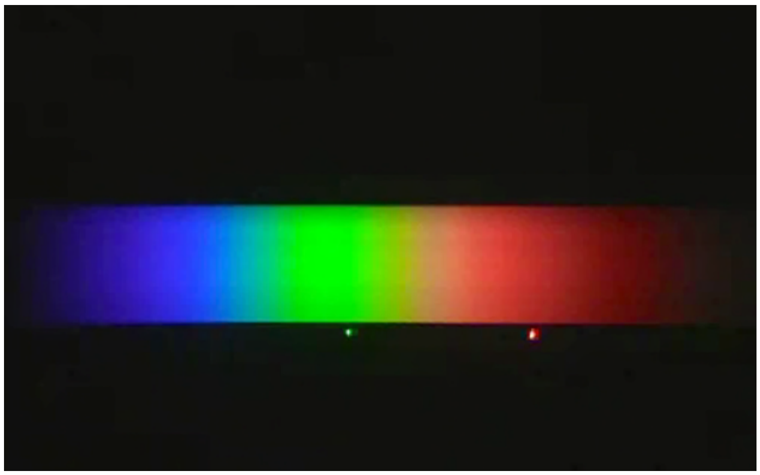


Figure 2.5: Spectra from: (a) fluorescent light, (b) neon light, (c) incandescent light. Credit: NASA/SSU/Steve Anderson.

If you view a fluorescent lamp with a diffraction grating (which spreads light out into its various wavelengths similar to the way a prism does), you will see that it is emitting a rainbow, just like an incandescent lamp. However, if you look carefully, you will notice that certain colors are especially bright. The rainbow is produced by a coating on the inside of the bulb. The coating converts some of the light from the gas into a continuous spectrum. But the coating does not convert all of the light. The extra bright parts of the spectrum are the places where the discrete spectrum from the gas really stands out. Sometimes, you can see the pure discrete or emission-line spectrum coming from a gas tube that has no coating. A "neon" sign can be a good way to achieve this result. Incidentally, not all neon signs contain the element neon. We will discuss that later. In contrast to a fluorescent bulb or a gas tube, an incandescent lamp emits a purely continuous Planck spectrum.

As shown in Figure 2.6, an absorption spectrum is created when a continuous source is viewed (head on) through a low-density cloud, and an emission spectrum is created when a source of energy is exciting a low-density cloud. This energy source could be, for example, a strong electric current (in the case of neon lights) or a nearby star (in the case of emission nebulae).

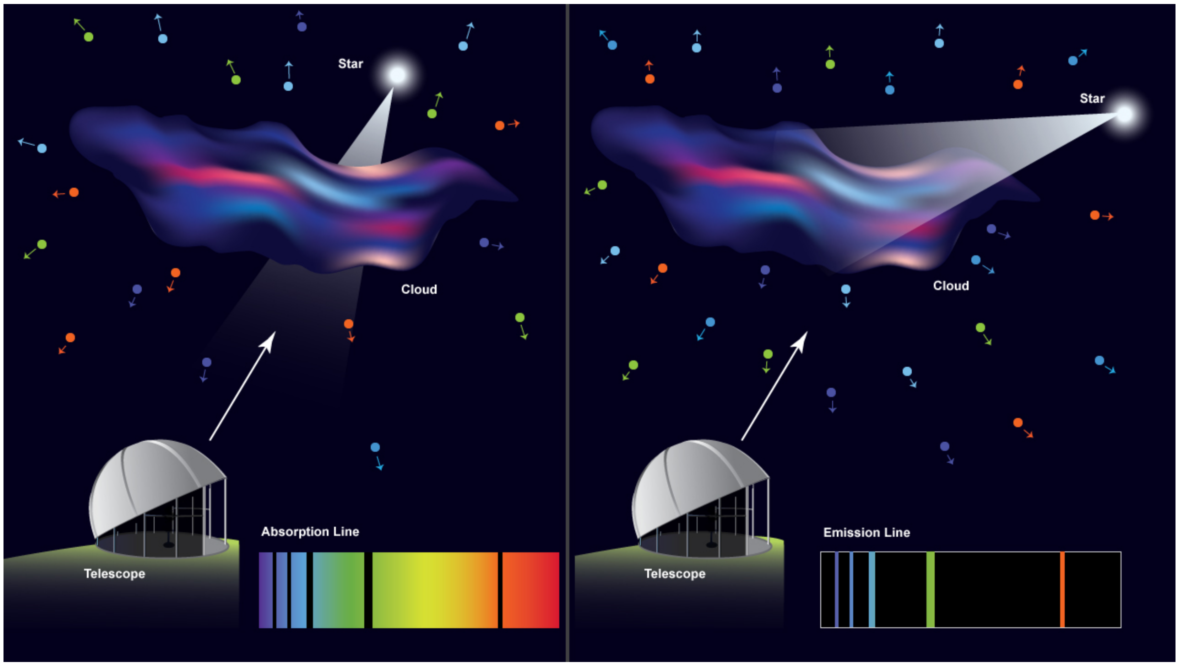


Figure 2.6: (a) When a bright object like a star is viewed through a cool cloud of gas, we observe gaps in its continuum , as depicted on the left. This is because some of the light is absorbed by the atoms, then re-emitted in random directions as indicated by the multi-colored "photons" (with arrows showing their direction of travel). Because some light has been removed, the spectrum is darker there, though it might not be completely black. (b) On the other hand, if we observe a cloud that is being illuminated by a bright star that is outside our line of sight, we see emission lines. In this case, the only photons to reach us are the scattered ones corresponding to the absorption in the previous image. Credit: NASA/SSU/Aurore Simonnet.

We now explore what happens when a cloud of thin gas is viewed at different angles in relationship to a source of continuous radiation. In the next activity, you will use the Spectrum Explorer simulation tool to explore how the spectrum you measure depends on the orientation of the light source. This activity will give you some experience with different emission and absorption spectra. After completing this activity you will know how astronomers are able to determine the composition of distant objects.

CREATION OF SPECTRA FROM A GAS CLOUD: DEPENDENCE ON VIEWING ORIENTATION

When we view a cloud of gas, the spectrum we see will depends on how we view it with respect to nearby light sources that illuminate it. In this activity, you will:

- Note your predictions.
- Make observations and record what you observe.
- Explain your predictions and your observations, noting any discrepancies.

Do not worry if you are not able to explain everything you see here yet; you will be revisiting the *Spectrum Explorer* several times.

Imagine there is a container, perhaps a glass tube, that can be empty or filled with hydrogen gas. There are also two light sources that can be turned on and off in such a way that they illuminate the tube. Additionally, we have a spectroscope with which to collect light and separate it to produce a spectrum. The resulting spectrum is displayed at the bottom of the screen. In an astronomical setting, a similar situation would occur if we viewed a star behind a gas cloud or if we viewed a gas cloud that had a star nearby but off to one side.

Play Activity

A. Nothing between observer and light sources.

1. Predictions: At first, both light sources are off and the glass tube has no gas in it. The clear spectroscope is thus dark. If you then turn on light source B, what do you expect to see in the scope? Choose one of the following:

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

2. Observations: Keeping the gas tube clear turn on light source B. What spectrum do you observe?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

3. Is this result consistent with what you predicted in Question 1? Resolve any discrepancies.

B. Tube filled with hydrogen. Head on.

1. Predictions: What do you expect to see with light source B on and hydrogen filling the tube?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

2. Observations: Select hydrogen to fill the tube by dragging down the menu on the glass tube icon and selecting hydrogen. What do you see in the spectroscope display?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

3. Is this the result you predicted you would see? This series of lines is called an absorption spectrum. Resolve any discrepancies between your predictions and observations.

4. Describe the differences in the spectrum when the hydrogen is present and when it is absent. What do you think is occurring in the hydrogen to cause the different appearance? (Hint: why do you think we call this type of spectrum an *absorption* spectrum?)

5. If we had used a gas other than hydrogen, how might things have been the same and how might they have been different? Recall that other atoms have more electrons than hydrogen. We will have a chance to work with other types of gasses in a later activity, so if you are not sure about the answer here, don't worry. We will explore these ideas in more detail later in the chapter.

C. No gas in the tube, light source side on.

Now turn off light source B and empty the tube of gas. You should see only darkness in the spectroscope.

1. Predictions: If you turn on light source A, what do you think the spectroscope will show?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We willl still see darkness.

2. Observations: Now turn on light source A. What does the spectroscope show?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of bright lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

3. Is this the result you expected to see? Resolve any discrepancies.

D. Gas in the tube, light source side on.

1. Predictions: What do you think you will see if we fill the tube with hydrogen gas?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

2. Observations: Select hydrogen from the drop-down menu to fill the tube with gas. Now turn on light source A. What does the spectroscope show?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

3. Is the result you expected it would be? This type of spectrum is called an emission spectrum. Record what you think is happening in the hydrogen to produce the spectrum you observe. Again, don't worry too much about getting the right answer here. We will have additional opportunities to learn about how atoms absorb and emit light.

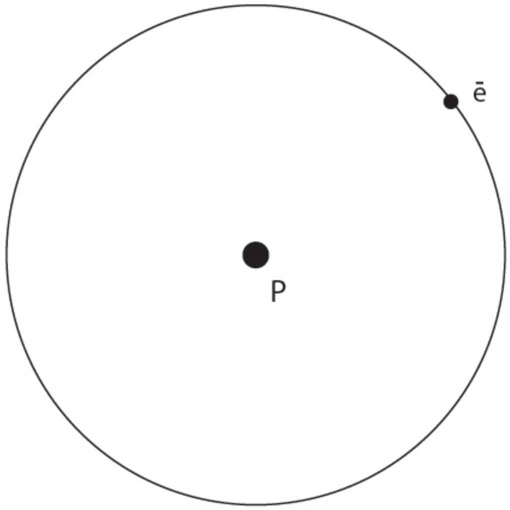
4. Leaving the hydrogen in the tube, alternately switch on light source A and then B. You can only have one light source on at a time, so switching A on will turn B off, and vice versa. Record what you see happening. If you have an idea of what is occurring in the gas to produce the spectra you see, write that explanation down here.

In this activity, we have imagined a laboratory with gas contained in a glass tube and light sources that we could control with a switch. In an astrophysical context, we cannot control the light sources or the gas in the "tube." For astronomical sources, the glass tube might be replaced by a cloud of gas in space. The light source would be nearby stars, either behind or adjacent to the gas clouds. Alternatively, the glass tube could be the cooler top layers of a star's atmosphere. In that case the light source would be the hotter, inner parts of a star's atmosphere. These sorts of observations of spectra from astrophysical sources led to a revolutionary change in our understanding of matter and energy at the beginning of the 20th century.

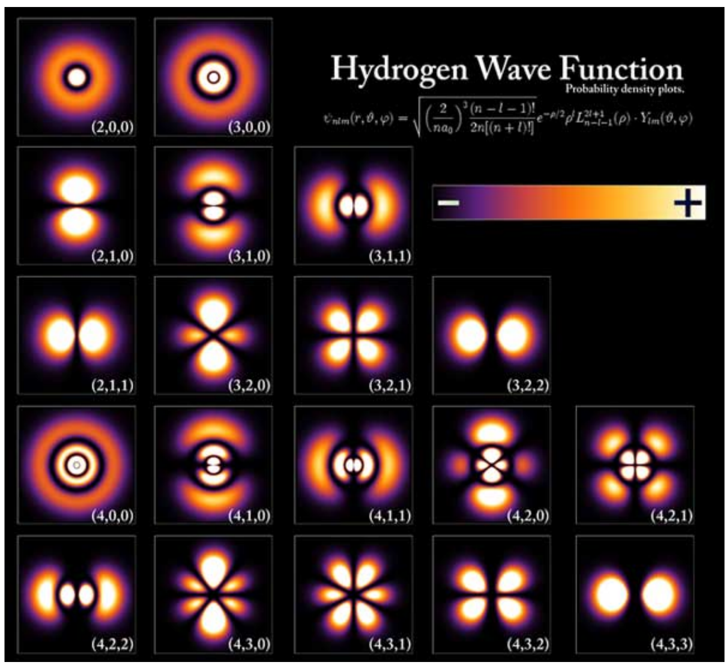
Now that you have observed how various types of spectra are produced in a general sense, let us explain your observations in detail.

Line spectra were first observed and studied in the latter half of the 19th century, but it was only in the 1920s that a full explanation was finally developed. To explain the energy emitted by atoms, a new branch of physics was developed, called quantum mechanics. In quantum mechanics, not only does light come in discrete little packets called photons, but the energies and other properties of electrons in atoms are also discrete, or quantized. So, for example, in an atom of hydrogen, the electron orbiting the nucleus can only have certain discrete energies. Other energies are simply not allowed. This may sound strange, but that is that is observed. We can use the results from quantum mechanics to consider what this theory implies for the spectrum emitted by an atom.

a.


ē
P

b.


Hydrogen Wave Function
Probability density plots.
(2,0,0)
(3,0,0)
(2,1,0)
(3,1,0)
(3,1,1)
(2,1,1)
(3,2,0)
(3,2,1)
(3,2,2)
(4,0,0)
(4,1,0)
(4,1,1)
(4,2,0)
(4,2,1)
(4,2,2)
(4,3,0)
(4,3,1)
(4,3,2)
(4,3,3)

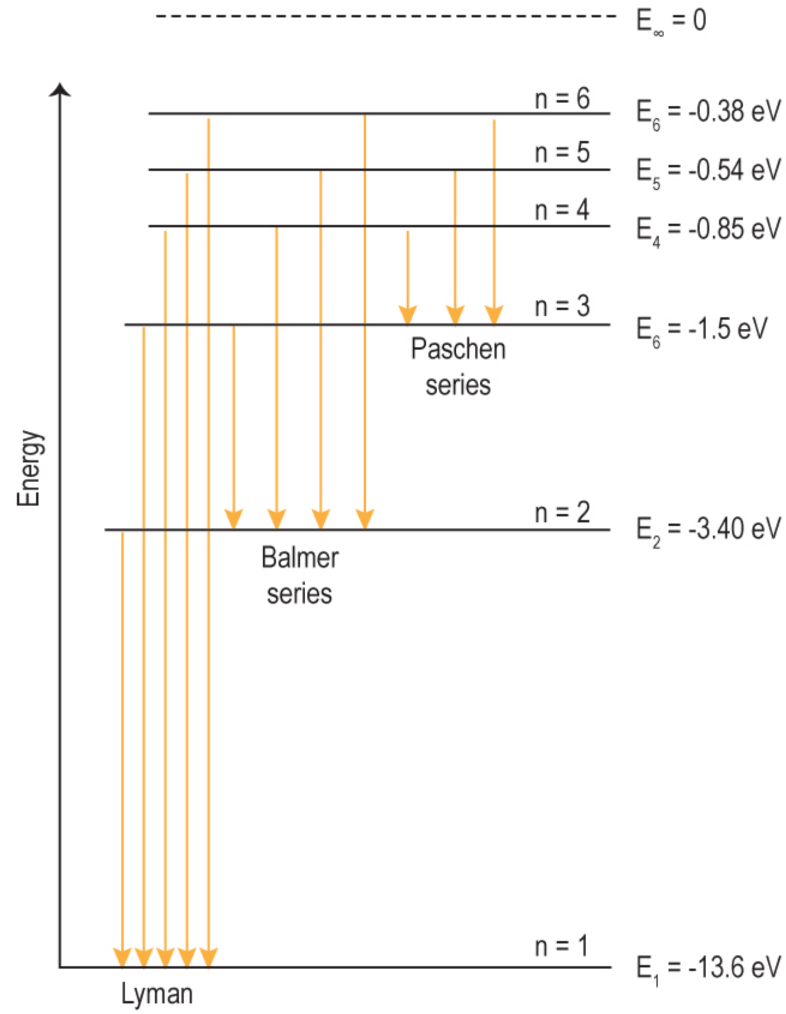


Figure 2.7: (a) A schematic representation of a hydrogen atom is shown. The proton and electron do not actually form a tiny "solar system" as shown. Rather, the electron occupies orbitals (not orbits) as depicted in (b). The first three energy levels are shown. Energy levels above the ground state (n=1) split into sub-states. The orbitals are three-dimensional, shown here in relief. (c) Some examples of atomic transitions in hydrogen are shown. Higher energy levels are plotted upward. Other atoms are similar to hydrogen, but they have more electrons and therefore generally have more complicated spectra. Credits: (a) NASA/SSU/Aurore Simonnet, (b) Creative Commons License, and (c) NASA/SSU/Aurore Simonnet.

We will use hydrogen as a first example. Because it has only a single electron, its spectrum is much simpler to analyze and understand.

Generally, the electron in a hydrogen atom sits at its lowest-possible allowed energy state, called the ground state. We can illustrate why this is so using an example from the macroscopic world around us. Even there, objects tend to be in their lowest possible energy. So, for instance, a cup of coffee might rest on a table, but if the table is removed, the coffee will fall to the floor. If the table happens to be on the second floor of a building, perhaps hanging over the edge of a balcony looking down to the first floor, then the coffee could fall all the way down to the first floor (Figure 2.8). Each of these motions would correspond to an energy transition. The coffee would move from a high-energy state to a low-energy state. Coffee (or anything else) will always move to a lower-energy state unless something prevents it from doing so (like the imposition of a table or the grasp of your hand).

Figure 2.8: When an object falls in a gravitational field, it loses gravitational potential energy (energy it had by virtue of its height). Since energy is conserved, this energy must go someplace. In a falling object, it goes into the energy of the object's motion. The energy can be further transformed, perhaps by shattering the chemical bonds in the coffee cup depicted here when it impacts the ground. The coffee cup is moving from a state of high potential energy to a state with low potential energy. Credit: NASA/SSU/Aurore Simonnet.

Atoms are similar to coffee cups in that electrons will always fall to their lowest-allowed energy state within the atom. However, unlike a cup of coffee, which could in principle fall all the way to the center of Earth (where the gravitational field is zero) if it were unobstructed, there is a lowest-possible energy state below which the electron cannot fall. This state is called the ground state.

When an electron is in the ground state of hydrogen, its average position is slightly separated from the proton, though its precise position at any given moment is not well-defined. Even though the electron does not have a well-defined position in space, the ground state, as with any other energy state, has a very well-defined energy: it is 13.6 eV below the free electron states for a hydrogen atom. That is, an energy of 13.6 eV is required to completely free the electron from the proton, thus ionizing the hydrogen.

There are additional allowed energy states within a hydrogen atom that are above the ground state but still below the energy required to free the electron. For instance, the next possible energy above the ground state sits 3.4 eV below the free states. All of the additional allowed states have higher energies still. Collectively, these allowed states above the ground state are called excited states, and several of them are depicted in Figure 2.7-c. This terminology derives from the fact that the electron must gain energy, or be "excited," in order to move into one of them from the ground state. Typically, none of the excited states are occupied; any electron in one of the excited states will spontaneously fall to an unoccupied lower energy state after a very brief time. Hydrogen has only one electron, so if that electron is in a highly excited state, it will quickly fall down to the unoccupied ground state.

The energy levels of hydrogen can be calculated by the following expression.

$$E_n = -13.6 \frac{1}{n^2} \text{eV}$$

The index parameter n is used to label the level of interest: n = 1 is the ground state, n = 2 is the first excited state, n = 3 the second excited state, and so on. The energies are given in electron volts (eV). They are negative because they sit below the lowest free-state energy, which is defined to be zero.

You might be wondering where the difference in energy goes when an electron falls from some excited state to the ground state. Energy must always be conserved, which means it can be transferred or transformed, but never created or destroyed. So if the electron has lost energy, it must have put that energy somewhere; that somewhere is into a photon. When an electron drops from a higher-energy state to a lower one, it emits a photon with an energy (hf) that is exactly the energy difference (ΔE) between the higher- and lower-energy states:

$$E_{photon} = hf = \Delta E_{electron}$$

So, for example, if an electron in the first excited state drops to the ground state, it will emit light of energy given by the difference in energy of the two states. A similar thing happens if you drop a coffee cup off the kitchen table: The cup accelerates toward the floor, moving faster and faster because as it falls, it converts gravitational energy into the energy of motion. The electron in an atom does not convert the difference in energy into accelerated motion when it undergoes an energy transition. Instead, the energy difference between each state is converted into a photon with energy equal to that difference. This process explains how line spectra are produced.

Of course, for light to be emitted, an atom must contain an excited electron in the first place. This can be done, for instance, if the atoms undergo collisions. When the collisions are violent enough, some of that energy can be converted into excitation energy in each of them. The excited atoms will then spontaneously de-excite, emitting light in the process. This process is exactly how fluorescent and neon lights work. They contain a mixture of gases (typically not hydrogen), and when we apply an electric field to the gas, collisions occur within it. The result of the collisions is excitation of electrons. These excited electrons spontaneously de-excite, producing light as they do so.

Collisions are not the only way to excite an atom. If light with the exact energy difference between two energy levels shines on an atom, then the atom can absorb some of that light. An electron gets excited from the lower-energy level to the higher level in the process. For this process to occur there must be an electron in the lower-energy level to start with. There must also be space in the higher level for the electron to move into.

The gaps (dark lines) in an absorption spectrum are formed because light of the specific energy (or wavelength, if you prefer) in the dark lines is absorbed by the atoms in a gas. This process is called photo-excitation. The photo-excited atoms will quickly de-excite (the electron will fall back to its original energy level) and re-emit the absorbed photon. However, the emitted photon will be emitted in a random direction. Only occasionally will the light be emitted in the direction of the original photon. As a result, you will see a dimmer area in the narrow wavelength region where the photon energy corresponds to the atomic energy transition. On either side of that dark line, the spectrum will be unaffected, showing its normal brightness. Stars typically show absorption spectra because the cooler gas layers near their surface absorb some of the light emitted by the hotter layers below.

EMISSION AND ABSORPTION IN A HYDROGEN ATOM

The figure in the exercise panel depicts a cartoon hydrogen atom. In the center is an electron in one of its possible orbital states. Next to it on the right is a diagram of the energy level that the electron occupies.

Your job is to identify whether the atom makes a transition that involves the emission or absorption of a photon.

Press the “next transitions” button to cause the hydrogen atom to undergo a transition to a higher - or maybe lower - energy state. Use the “absorption” and “emission” buttons to indicate what kind of a transition you just witnessed.

Repeat this process until you feel confident in your answers.

Play Activity

1. Describe what happens to the electron’s energy levels for:

a. an absorption

b. an emission

2. How do the changes in the drawing of the atom (in the center) correspond to the changes in the energy levels shown (on right)?

TRANSITIONS IN A HYDROGEN ATOM

In the following activity, you will calculate the energies of photons that are emitted in various transitions for electrons in a hydrogen atom.

Worked Example:

Calculate the (a) energy, (b) frequency, and (c) wavelength of the photon emitted when a hydrogen atom’s electron in the first excited state drops to the ground state. (d) What waveband of the electromagnetic spectrum is this?

A. Energy

- Find: E_{photon} (energy of the photon)
- Given:
  $n = 2$ = first excited state
  $n = 1$ = ground state
  E_1 = energy of the ground state = -13.6 eV
- Concept(s):
  $E_n = (-13.6\text{ eV})(1/n^2)$
- $E_{photon} = \Delta E = E_2 - E_1$ (energy of the photon equals change in energy of electron)
- Solution:
  E_2 = energy of first excited state
  $E_2 = (-13.6\text{ eV})(1/2^2) = (-13.6\text{ eV})(1/4) = -3.4\text{ eV}$
  $E_{photon} = (-3.4\text{ eV}) - (-13.6\text{ eV}) = 10.2\text{ eV}$

B. Frequency:

- Find: f
- Given: E = 10.2 eV, h = 4.136E-15 eV s
- Concept(s): $E = hf$
- Solution: $f = E/h$ = (10.2 eV)/(4.136E-15 eV s) = 2.466e15 Hz

C. Wavelength:

- Find: λ
- Given: f = 2.466e15 Hz
- Concept(s): $\lambda = c/f$
- Solution: λ = (3E8 m/s)/(2.466E15 Hz) = 1.22E-7 m = 122 nm

D. Recall that visible light has wavelengths between about 400 and 700 nm. This wavelength is shorter, in the UV range.

Questions:

Calculate the (a) energies, (b) frequencies, (c) wavelengths, and (d) wavebands for other transitions in hydrogen.

1. What are your results for $n = 3$ to $n = 1$?

a. Energy:

eV

Show your work:

b. Frequency:

Hz

Show your work:

c. Wavelength:

m

Show your work:

d. Waveband:

This page titled 2.6: Lines Spectra- Emission and Absorption Lines is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.7: Determining the Composition of an Unknown Gas

Learning Objectives

- You will know that chemical elements leave distinct "fingerprints" on the light from astronomical sources.

WHAT DO YOU THINK: FINGERPRINTS

Three students are talking about today's lecture on spectroscopy.

- **LaTasia:** "What did our professor mean when she said that stars have spectral fingerprints? Is every star unique, like people's fingerprints are?"
- **Malia:** "I don't think the stars have spectra that are absolutely unique, but you can use the spectra to identify certain stars—at least the ones that are close enough to get a spectrum ."
- **Noelle:** "Each kind of element and molecule does have a unique spectrum, and that's what is the "fingerprint". You can identify what elements or molecules are in the object by looking at what lines are present in the spectrum."

Do you agree with any of these students, and if so, whom?

LaTasia

Malia

Noelle

None

Explain

So far, we have looked at how light can be emitted and absorbed by atoms. We have looked specifically at hydrogen because it is simple. However, other atoms can also absorb and emit light when their electrons change energy levels. These atoms have more complicated energy configurations though. Unlike hydrogen, where the energies are determined solely by the interaction between the electron and the nucleus, other atoms always have more than one electron. Since the electrons are charged particles, they interact with each other, as well as with the nucleus. As a result, most atoms have a much richer set of possible energies than hydrogen. Even helium, with two electrons instead of one, has a much more complicated spectrum than hydrogen. In an atom with many electrons, the energy levels are usually extremely complicated. Consider neon, which has ten electrons. It has a spectrum that is much more complex than either hydrogen or helium. In the next interactive activity you will explore a few examples for several common elements.

CREATION OF ABSORPTION AND EMISSION SPECTRA FROM A MIXTURE OF GASES

We are now ready to return to our simulated *Spectrum Explorer* tool. This time we will not study only hydrogen. We will look at more complicated atoms as well. One caveat to keep in mind is that not all the lines in an atom are the same strength. Some lines appear strong, while others are quite faint. We have chosen to display only the strongest lines for this activity. This keeps things simpler. Some atoms have many hundreds or thousands of lines (or more!). We don't need to look at every single line to understand how spectroscopy is used in astronomy.

Answer the following questions using the *Spectrum Explorer* where appropriate. Again, keep track of your answers to questions. Note your predictions and observations, and be sure to resolve any discrepancies between the two.

Play Activity

A. Light source head on.

As before, the glass tube starts out devoid of gas, and both light sources are off.

1. What will you see when source B is turned on without any gas in the tube?

A series of bright lines

A series of bright lines on top of a continuous rainbow

A series of dark lines in an otherwise continuous rainbow

A continuous rainbow with no lines

Nothing will change. We will still see darkness.

2. Now select hydrogen from the drop-down menu for the tube. With light source B on, verify that you see the same spectrum that you saw before with hydrogen. Is this an absorption spectrum or an emission spectrum?

Absorption

Emission

B. Other gases, light source head on.

Now we have several other gases available to choose from to fill the tube instead of hydrogen.

1. Now select helium to fill the tube instead of hydrogen. Recall that helium has 2 electrons rather than a single electron like hydrogen. How does the helium spectrum compare to that of hydrogen? Do you have an explanation for the differences?

2. If we replace helium with neon, which has 10 electrons rather than 2, what do you think will happen? Record your predictions.

3. Now replace the helium with neon. Record your observations. Was your prediction confirmed or not?

4. Now try some of the other gases and note the spectra produced. Do you see any relationships between the number of lines and the number of electrons? (Note that we are showing you only the strongest lines, while ignoring faint ones. Also, we are only showing lines in the visible part of the spectrum. Atoms also have energy levels in other wavelength regions, so your conclusion here might not be generally true).

5. Are any of the lines common to one or more different gases? If so, note these lines below. (The wavelength is given along the bottom of the spectroscope display.)

C. Other gases, light source to the side.

In the previous examples, the light passed through the gas tube on its way to the spectroscope. We thus observed an absorption spectrum for each type of gas. We will now explore what will happen if we turn off light source B and turn on source A.

1. Select the tube to be filled with helium, and turn on light source A. What type of spectrum does the spectroscope show?

2. Switch between light source A and light source B, leaving the helium inside the tube. What do you notice about the spectra produced? Are they at all related to each other?

3. Repeat the procedure for the other gases by switching from light source A to light source B and back again. Do they behave the same as hydrogen or helium, or do you note differences for different gases?

4. You have had an opportunity to study both absorption and emission spectra from various gases. Write down any general patterns you see for the gases emission and absorption spectra. Give a general rule for when you will see an absorption spectrum for an object and when you will see an emission spectrum. Give a brief explanation of what is happening to the atoms in the gas to produce these spectra.

5. Explain how using spectral techniques might be useful in determining the composition of distant stars, galaxies, and other celestial objects. (Hint: What would the spectrum look like if we could put a mixture of gases into the tube, rather than just one gas at a time?)

Fortunately, no matter how complicated the spectrum of an atom is, it will always be different from the spectrum of any other atom, just as with a set of fingerprints. Each element exhibits a unique pattern of spectral lines. Whenever we see that particular pattern in the light from a source, we know that element is present. The patterns of spectral lines allow us to measure the composition of stars and other objects at very great distances, even all the way across the visible Universe. Not only can we tell what the objects are made of, but the relative strengths of the lines can tell us about physical conditions in the emitting or absorbing gas. Quantities like temperature, density, magnetic field strength, and the relative abundance of elements can all be determined by careful examination of electromagnetic spectra. Complete the next activity to get some experience matching an unknown spectrum to the known lines from the previous activities.

DETERMINING THE COMPOSITION OF AN UNKNOWN GAS

In the preceding *Spectrum Explorer* activities, you learned how absorption and emission spectra are created from gases and light sources. You also learned that gases made of different kinds of atoms have unique patterns of spectral lines. These patterns of lines can be used to determine the composition of a gas.

In this activity, you will use what you have learned previously to identify the composition of an unknown mixture of gases.

Play Activity

To begin, click on the "generate" button in the *Spectrum Explorer*. You will see a pattern of lines appear on the lower strip; this is your unknown gas spectrum. The unknown contains two types of atoms, so you will not be able to match all the lines with just a single type of gas.

To match this unknown, you will compare the line pattern to lines from known gases. Use the gas tube pull-down to check the patterns of different gases against your unknown. Check each gas in succession, keeping track of which gases match or do not match.

When you have matched all of the lines in the unknown, click in the "identify" button, and select the gases you think are in the unknown. You will be given feedback about whether you are correct.

Describe a method for using spectral techniques to determine the composition of distant stars, galaxies, and other celestial objects.

In 1842, the French philosopher Auguste Comte (1798–1857) declared that the "chemical and mineralogical structure" of the stars would never be known. He was referring to what we call planets, but he thought the same was true for stars. Given the immense distances to even the closest stars, which were known even to Comte, it is hardly surprising that he held this opinion. Yet, within just a couple of decades of his death, advances in atomic physics allowed scientists to begin studying the spectral lines from the Sun and other stars, much as you have done using the *Spectrum Explorer*. The composition and inner structure of the stars began to be revealed. And of course, we now have physical samples returned from the Moon, and robots regularly visit the surfaces and atmospheres of the planets to carry out experiments. Perhaps there is a cautionary tale here about holding too fast to our ideas of what might be possible in the distant, or even not too distant, future.

This page titled 2.7: Determining the Composition of an Unknown Gas is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.8: Wrapping It Up 2 - The Properties of Light

2.8.1: Part I. Wave Generator

We will once again employ our wave generator to create wave signals whose properties you can adjust.

Recall, the x-axis shows the wave's wavelength in meters and the y-axis shows the wave's amplitude. The amplitude of a wave essentially describes its intensity, or strength. Wavelength is the distance between two adjacent peaks of a wave. Frequency describes how many wave patterns or cycles pass by a certain point in 1 second.

You can make changes to the basic properties of the wave that is being generated in the following ways:

- Stop the wave generator to measure the wave at any time. Hover your cursor over the graph and you will see an x-y coordinate display of cursor position on the graph. Notice that the origin of the graph is on the far left, in the middle of the y-axis.
- Below the "stop" button, there are the wavelength and amplitude sliders. Use these sliders to increase or decrease the wave's properties in real time.
- There is also a stopwatch that counts time in seconds. You can start and stop the stopwatch, and also reset it to zero when you need to make a new measurement.

Use the *Wave Generator* to conduct the following measurements, and answer the following questions.

Play Activity

2.8.1.1: A. Measuring the wavelength of a wave.

1. Adjust the wavelength slider until you think the wavelength equals 100 m.

2. Stop the wave generator, and measure the distance between two peaks (or two troughs). Continue to adjust the slider until you are as close as possible to a wavelength of 100 m. After you make your final measurement, click on the "reveal wavelength" button. How close were you to 100 m?

3. Express the accuracy of your measurements as a percent. For example if your wave was 95 m instead of 100 m, your accuracy would be (100 – 95)/100 = 5%.

2.8.1.2: B. Measuring frequency and speed

To determine the frequency, you will need to count the waves as they pass by the orange vertical bar.

- Start the stopwatch when the peak (or trough) of a wave passes the arrow.
- Let it run until at least 3 waves have passed by the arrow.
- Stop the stopwatch, and read out the results for time passed.
- You may want to repeat this activity a few times to make sure you are reading the graph and the stopwatch accurately.

1. Show your work here:

2. What is the frequency that you find for a wave with a wavelength of 100 m?

Hz

3. Using the relationship between speed, frequency, and wavelength and your measured values of wavelength and frequency, what is the speed of the wave?

m/s

4. Predict: How do you think the frequency of a wave with wavelength of 50 m will compare to the frequency of a wave with wavelength 100 m?

5. Predict: How do you think the speed of a wave with wavelength of 50 m will compare to the speed of a wave with wavelength 100 m?

6. Test your predictions using the wave generator. Discuss your results.

2.8.1.3: C. Measuring wavelength and amplitude

1. Predict: How do you think adjusting the amplitude will affect the wavelength? For example, if you increase the amplitude, will the wavelength increase, decrease, or remain the same?

2. Check your predictions using the wave generator: with the wavelength as close to 100 m as you're able to get it, increase the amplitude. After letting the wave generator run long enough, measure the wavelength again. Is there any change?

3. Adjust the amplitude again, this time making it smaller. What happens to the wavelength?

The wavelength increases.

The wavelength decreases.

The wavelength remains the same.

4. How does decreasing the amplitude affect the frequency of a wave?

The frequency increases.

The frequency decreases.

The frequency remains the same.

5. How does decreasing the amplitude affect the speed of a wave?

The speed increases.

The speed decreases.

The speed remains the same.

6. Summarize: how does changing the amplitude (strength) of a wave affect its speed, frequency, and wavelength, if at all?

2.8.2: PART II. LIGHT WAVES

The wave generator cannot accurately portray the frequencies or speed of a light wave because light travels at a much higher speed than we can easily show in an interactive activity: $c = 3 \times 10^8$ m/s. In this next section, you will apply what you have learned about waves in general to specific examples using light waves.

2.8.2.1: A. Wavelength and frequency

1. Calculate: Using the equation $\lambda = c / f$, what will the frequency be for a wave with a wavelength of 100 m? (A light wave with this wavelength is known as a radio wave.)

Hz

2. Calculate: Using the equation $\lambda = c / f$, what will the frequency be if the wavelength is 150 m?

Hz

3. If you doubled the wavelength from 150 m to 300 m, by what factor would the frequency change? Will it be greater or smaller than the frequency you calculated in question 2?

4. In a vacuum, can a light wave have a frequency of 10^{14} Hz and a wavelength of 3 m? Why or why not? Explain.

2.8.2.1.1: B. Measuring light waves and color

1. Red visible light has a wavelength of 7×10^{-7} m. What is its frequency?

Hz

2. If a light wave had a frequency twice that of red visible light, what would be its wavelength in meters? (Light with a frequency this high is known as ultraviolet light.)

m

3. A radio wave is 10 m long; what is its frequency?

Hz

4. A gamma-ray has a frequency equal to 6×10^{20} Hz. What is its wavelength?

m

2.8.2.2: C. Energy, wavelength, frequency, and speed of light

Use the relations $\lambda = c / f$ and $E = hf$ to answer the following conceptual questions. We will compare a photon of red light ($\lambda = 7 \times 10^{-7}$ m) and a photon of ultraviolet light ($\lambda = 1 \times 10^{-8}$ m).

1. How does the wavelength of ultraviolet light compare to that of red light?

The wavelength of ultraviolet light is greater than the wavelength of red light.

The wavelength of ultraviolet light is less than the wavelength of red light.

The wavelength of ultraviolet light is the same as the wavelength of red light.

2. How does the frequency of ultraviolet light compare to that of red light?

The frequency of ultraviolet light is greater than the frequency of red light.

The frequency of ultraviolet light is less than the frequency of red light.

The frequency of ultraviolet light is the same as the frequency of red light.

3. How does the speed of ultraviolet light compare to that of red light?

The speed of ultraviolet light is greater than the speed of red light.

The speed of ultraviolet light is less than the speed of red light.

The speed of ultraviolet light is the same as the speed of red light.

4. How does the energy of ultraviolet light compare to that of red light?

The energy of ultraviolet light is greater than the energy of red light.

The energy of ultraviolet light is less than the energy of red light.

The energy of ultraviolet light is the same as the energy of red light.

5. If I increase the energy of light, how does that affect its speed?

The speed increases.

The speed decreases.

The speed remains the same.

This page titled 2.8: Wrapping It Up 2 - The Properties of Light is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

2.9: Mission Report 2 - The Properties of Light

A. Please rate on a scale of 1 to 5 how effective you think the Wrapping It Up activity was in helping you understand the material. (1: not effective at all → 5: very effective)

* 1 2 3 4 5

B. What were the main ideas that you learned in conducting the Wrapping It Up activity? Be specific and detailed in your response. Please address the following questions: What did you learn? How did you learn it? What is still unclear? (At least 150-200 words)

*

C. If the Wrapping It Up activity included measurements or data, please describe what factors influenced the accuracy of your results. (Do *not* include mistakes, only unavoidable measurement imprecision.) If you obtained any numerical values for the accuracy of your measurements during the activity, note those here. If there were no measurements or data, say so explicitly.

*

D. Questions to be graded for accuracy

1. Explain the difference between the frequency and speed of a light wave.

*

2. Explain the relationships between the frequency, wavelength, and energy of a light wave.

*

3. What are the wavebands of the electromagnetic spectrum? List them in order from lowest to highest energy.

*

4. An electron jumps from the n = 3 to the n = 2 level of hydrogen. What type of spectrum is produced and why? What is the energy of the photon in eV?

*

This page titled 2.9: Mission Report 2 - The Properties of Light is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

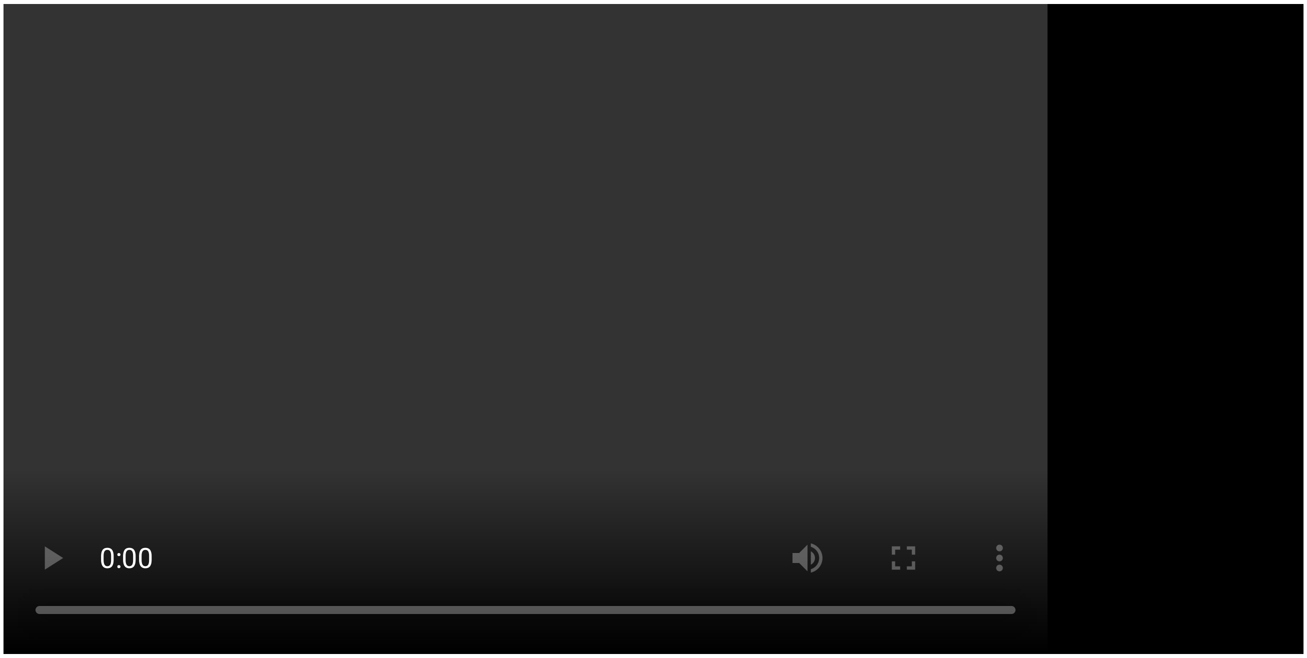
3: Telescopes

Chapter 3 focuses on telescopes, from their design to how their physical location affects their use. We will also explore the use of filters and how data are retrieved from images. Lastly, the chapter looks at how the brightness of astronomical objects is determined.

This page titled 3: Telescopes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

3.0: Telescope Introduction

Investigating light is no easy task, as telescopes on the ground all suffer from the same predicament: Earth's atmosphere degrades the images they produce. It makes objects appear blurrier and dimmer than they actually are. Furthermore, most kinds of light are not detectable from the surface of Earth at all; they are absorbed by our atmosphere before ever reaching the ground. This is why we put many types of specialized detectors on satellites that are then launched into space.



Video Transcript

> This video contains no audio. Used with permission from Stephane Guisard (sguisard.astrosurf.com) and Jose Francisco Salgado (josefrancisco.org).

By the end of this chapter, you will compare and contrast images of a celestial object obtained in three different wavebands of light by three different telescopes. There is really much more out there than meets the eye!

This page titled 3.0: Telescope Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

3.1: Designing Telescopes Across the Spectrum

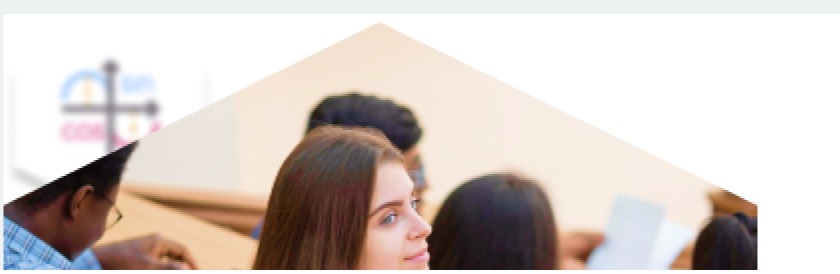
WHAT DO YOU THINK: X-RAY VISION

Three students are discussing a recent trip to the dentist's office.

- **Jason**: "The dentist must have x-ray vision to see what is going on in my mouth. I have a lot of cavities."
- **Keisha:** "No one really has x-ray vision. The dentist takes an x-ray of your mouth by putting film or a digital camera inside your mouth and then shooting a beam of x-rays through your teeth onto the film."
- **Mark:** "Well then, how does Superman see through things?"


f(x)
s
a
b


f(x)
s
a
b



As we have seen, there are many kinds of light across the electromagnetic spectrum. Only a tiny portion of this light is what we commonly think of as light; it is the kind that we can see with our eyes (visible light). Furthermore, there are many objects in space that we cannot see despite the fact that they emit visible light. They are just too faint. We deal with the limitations of our eyes by building telescopes and light detectors. These tools, like many of the tools we build, help us expand our physical capabilities beyond what is possible by our bodies alone. In this section, we look at telescopes, our "eyes on the sky." Our purpose is to illustrate how different telescopes are used, as well as some of the physical limitations on their use.

3.1.1: Magnification is Not Important

Most terrestrial telescopes (or binoculars, which are just paired telescopes mounted such that each eye looks through one of them) are used for magnification. In other words, they are used to make the object you are observing look larger. For instance, birders want to be able to see larger images of the birds they watch so that they can identify the birds and observe their behaviors without having to get so close that they might disturb them. For astronomical telescopes magnification can also be important, but it is not usually of primary concern to astronomers. What is far more important is the ability of telescopes to collect large amounts of light and focus it onto a small area.

3.1.2: Big Telescopes Collect a Lot of Light

If you go out at night in a dark place away from city lights, after taking 20 minutes or so to let your eyes become accustomed to the darkness, you will be able to see 2,000 or 3,000 stars at any given time. Some stars are fairly bright, but most are quite faint. The Milky Way—the galaxy in which we live—contains hundreds of billions of stars, not a few thousand. So why is our view of it so limited? Basically, it is because the sensitivity of our eyes is too feeble to allow us to detect the vast majority of stars—they are simply too faint for us to see. However, if we use even a small telescope or pair of binoculars, the number of stars we can see jumps dramatically, to many tens or even hundreds of thousands. This is the power of astronomical telescopes: They allow us to see fainter objects. And because there are many, many (Many!) more faint objects than bright ones, telescopes reveal multitudes of objects that we could never see with our unaided eyes. As a powerful illustration of this concept, go outside on a dark night and hold your hand out at arm's length with the fingers spread apart and the palm of your hand pointing away from you. Look at the nail on your smallest finger. If your eyes had the sensitivity of one of today's large telescopes, you could see that every patch of sky the size of your nail would contain millions of galaxies. In fact, every patch that size does contain them: they are there, but they are just too faint for you to see.

To understand how telescopes are able to reveal faint objects, we need to think a little bit about geometry. The pupil of your eye, when dilated, is perhaps 1 centimeter in diameter. All the light that can be focused on your retina must pass through that small area. On the other hand, even small telescopes, say with a 2-inch lens (about 5 cm diameter) have much more area. The area of a circle is proportional to its diameter squared. That means that a 5-cm telescope has 25 times the area of your eyeball and can therefore collect 25 times more light than your eyes can on their own. A telescope focuses all the light passing through its primary lens into its eyepiece, allowing you to see an amplified (brighter) version of whatever you are viewing.

Mathematically, the relationship between the collecting area (A) and the radius (r) of the instrument is roughly given by a well-known equation.

$$A = \pi r^2$$

Since the radius is half of the diameter (D), you can also use this alternative form.

$$A = \pi \left(\frac{D}{2}\right)^2$$

This means that the collecting is proportional to the diameter of the telescope squared.

Light Collecting Area

For this activity, you will determine how the light collecting area of an instrument depends on its size, from the human eye to several well-known telescopes.

You can use the slider to see how the collecting area changes as the diameter changes.

Play Activity

Example 3.1.1

What is the collecting area of a 2-m telescope?

Solution

- Given: diameter $D = 2\,m$
- Find: area A
- Concept: $A = \pi(D/2)^2$
- Solution: $A = (3.14)(1\,m)^2 = 3.14 m^2$
- Check: using the slider and setting the diameter to $2\,m$, we also find that the area is $3.14\,m^2$

Math Exploration 3.1: Light Collecting Area

1. What is the collecting area of a 1-m telescope?

- Given: diameter $D = 1\,m$
- Find: area A
- Concept: $A = \pi(D/2)^2$
- Solution: $A = (3.14)(0.5\,m)^2 = 0.79 m^2$

2. How much more light does a 10-m telescope, like the Keck Telescope in Hawaii, collect than a human eye (in darkness)?

- Given: D_{Keck} = 10 m (diameter of Keck), D_{eye} = 0.01 m (diameter of pupil)
- Find: ratio of areas: A_{Keck}/A_{eye}
- Concept(s): A proportional to D^2
- Solution: $A_{Keck}/A_{eye} = (10)^2/(0.01)^2 = 10^6$
- ⇒ Keck collects about 1 million times more light than a human eye collects.

Example 3.1.2

How many times bigger is the collecting area of the 2-m telescope than a 1-m telescope?

Solution

- Either you can take the ratio of the areas found from using the slider tool, or you can use the concept that the area goes as the diameter squared. Both methods give an answer of the larger telescope collecting 4 times more light.
- Method 1: area of 2-m telescope / area of 1-meter telescope = 3.14/0.79 = 4
- Method 2: ratio of areas is the square of the ratio of diameters: $(2/1)^2 = 4$

3.1.3: Resolution

Another advantage telescopes have over our eyes is their ability to distinguish small details. This property is called **resolution**.

Think about the smallest detail you can see. If you have typical eyes that do not require correction with glasses, then that detail has an angular size of about 1 arcminute. Notice we are talking about angular size, not actual size. This is because as we change our distance from an object, its physical size will not change, but its apparent size, that is, the angle it subtends, will become smaller if the object moves away and bigger if the object moves closer. Resolution is a measure of angular size.

Before we continue, we have to define what exactly we mean by an arcminute. You probably already know that a circle contains 360 degrees. Each degree contains 60 arcminutes, so if you imagine dividing the circle angle into 360 equal pie-wedge pieces, and then you further divide one of those wedges into 60 equal pieces, that is an arcminute. The arcminutes can be further subdivided into 60 equal parts called arcseconds—it works just like time. So 1 degree contains 3,600 arcseconds (just as 1 hour contains 3,600 seconds). Perhaps this is not an easy thing to visualize. Some examples might be in order.

The full moon subtends an angle of about half a degree, or 30 arcminutes. A person who looks at the Moon can discern details as small as about 1/30th the diameter of the Moon—an arcminute.

If we use a telescope, we can distinguish much smaller details than that. This is due to two effects. The most obvious effect is that the telescope magnifies the image of the Moon causing the objects that subtend only one arcminute to our bare eyes to grow to subtend perhaps 30 or more arcminutes in the image we see in the telescope. However, if we used a telescope with a diameter no larger than our eyes' diameter, all we would see would be a large blurry image. The telescope would magnify the scene, but it would still not show any details smaller than those visible to our eyes alone. This is one consequence of the wavelike properties of light.

As with light gathering power, a major advantage of a telescope is not its ability to magnify. Rather, it is the larger diameter of its primary lens. The larger the diameter allows the telescope to clearly discern smaller details - smaller in angular size, not necessarily absolute physical size. We say it is able to resolve these smaller details. This is the result of light diffraction, or the bending and interference of light waves in the telescope. The resolution of a telescope is proportional to the ratio of the wavelength of the light it collects and the diameter of the telescope's lens.

$$\theta = 1.22 \frac{\lambda}{D}$$

the angle, θ is the smallest angle resolvable, λ is the wavelength of the light, and D is the diameter of the telescope. From the expression, you can see that as long as the ratio of the wavelength to the diameter is small, the resolution of the telescope will be good (i.e., the telescope will be able to discern very small details). To use this equation, the angle θ must be in radians, not degrees. Radians are a unit of angular measure derived from working with circles: 1 radian = 57.3 degrees because there are 2π radians = 360 degrees in a full circle.

Resolution

In this activity, you will predict the number of astronomical objects that make up the blurred image that you see in the window.

You can use the sliders to adjust the diameter (D) of the primary mirror of the telescope being used. You can also adjust the wavelength of light (\lambda\) emitted from the object.

Without proper resolution, astronomers cannot see the finer details in an image. Keep in mind that the resolution is best when the angle θ, which depends on the ratio of λ/D, is small. (You will not need to perform any calculations. The activity is designed to promote understanding of the concept behind resolution, not test your skills in arithmetic.)

Play Activity

1. Keep track of your predictions as you go through the activity. Complete at least three different trials or more until you get at least one correct. Record the positions of the wavelength and diameter sliders when the best resolution is reached for each trial below. State the number of objects that became visible and whether or not each of your predictions was correct.

Example:

Resolution of a Telescope

In the following activity, you will calculate the resolution for several different telescopes.

Worked Example

1. What resolution is obtainable by a 1-m diameter telescope observing visible light of wavelength 500 nm (green light)? Express your answer as an angle in arcseconds.

- Given: λ = 500 nm = 500E-9 m; D = 1 m
- Find: θ
- Concept(s): $\theta = 1.22\left(\frac{\lambda}{D}\right)$
- Solution: θ= 1.22 (500E-9 m)/(1 m) = 6.1E-7 radians

The expression gives us the angle in radians, but we would like to have it in arcseconds. Knowing that there are π (roughly 3.14) radians in 180 degrees, and that there are 3,600 arcseconds in each degree, we can convert our angle to arcseconds as follows:

$$\begin{aligned}\theta &= (6.1\text{E-7}\ \cancel{radians})\left(\frac{180\ \cancel{degrees}}{\pi\ \cancel{radians}}\right)\left(\frac{3600\,arcsec}{1\ \cancel{degree}}\right)\\ &= 0.13\,arcsec\end{aligned}$$

This is a very small angle, much too small to been seen by the eye alone.

Questions:

https://phys.libretexts.org/@go/page/30467

3.1.4: Field of View

The **field of view** (FOV) is the total area that can be seen through the eyepiece of a telescope. For a camera, it is the angular size of the area that a camera can image at any given time. Our eyes also have a field of view, often called the field of vision. The field of view varies for different instruments, such as different size telescopes, binoculars, and the eyes of different animals (humans included). The area that is covered by a field of vision is described by an angle. For a telescope, the FOV is so small that it is typically measured in arcminutes or arcseconds. Examples of some typical fields of view are listed in the following chart.

Table 3.1 Fields of View

OPTICAL INSTRUMENT	FIELD OF VIEW
A Human Eye	~180°
A Horse’s Eye	~200°
Binoculars (10 x 50)	~5°
Small Amateur Telescope	~1.5°
The Hubble Telescope’s Wide Field Camera 3	164 × 164 arcseconds

3.1.5: Capturing the Light With Detectors

In addition to collecting and focusing much more light than an eyeball, telescopes have another advantage; most of them do not use our eyes as a detector. Instead, a sensitive electronic instrument is placed where the eyepiece would go. This has several advantages. The first is efficiency. While our eyes only capture a small percent of the photons that enter them, modern detectors of visible light are more than 90 percent efficient. A second advantage of telescopes and detectors is time of exposure. Our eyes only collect photons for a fraction of a second before dumping the data and starting over. This is good if we want to see our environment change in real time. It is not good if we want to do astronomy. Most astronomical objects do not change much over timescales of seconds or minutes (or sometimes even millennia!), so it is advantageous to simply stare at them and collect photons for many seconds, minutes, hours, or even days. This is what the electronic "eyes" —detectors—we put on telescopes can do. They are able to stare at an object for as long as we like, building up an image until even extremely faint objects begin to show themselves. A final huge advantage is that with electronic cameras we can store any images we collect and look at them later. Much, much later if we like: we still have photographic plates of the sky from the end of the 19th century that can be referred to if the need arises.

Modern detectors for visible light are not photographic film. They consist of semiconductor crystals (mostly silicon) that turn incoming photons into electrical charge. The charge can then be read off by electronics to determine how much light has hit the detector. These sorts of detectors are called Charge-Coupled Devices (or CCDs for short) or Complimentary Metal Oxide Semiconductors (CMOS). Consumer level digital cameras in common use by photographers, as well as those found in cell phones, are made with CMOS light detectors. For the most part, CCDs are used only for scientific applications, including astronomy. Both types of detectors can be used for detecting IR, UV, and low-energy x-ray light, not just visible light. An example of a CCD chip is shown in Figure 3.1. A CMOS detector would look essentially the same.

Figure 3.1: This image shows a charge-coupled device, or CCD chip. The dark rectangle at the center of the chip is the array of pixels that detect individual photons. The pins read out the electrical charge produced by the light to a digital camera, and electronic circuitry inside the camera converts the charge into an image that can be viewed on a computer screen. Credit: NASA.

Using CCDs and CMOS sensors to detect light has one major disadvantage: the detectors are not able to distinguish different wavelengths of light; that is, they are like black-and-white film; they record only intensity, not color. To get around this problem, astronomers have developed systems of color filters that can be placed in front of the camera, thereby providing some crude spectral information about objects within the images. These filters were originally created to be used with black-and-white film, but they can be used with modern CCDs as well. If one wishes, a color image can be created by collecting images through each of the filters and then combining them into a single image with software. Consumer cameras perform this task by mounting a pattern of similar color filters permanently in front of the sensor. This technique allows color information to be collected in a single exposure, though at the cost of poorer angular resolution.

When astronomers make an image of an object (using visible light), they almost always take the image through one of the standard filters. And as mentioned, any color images we see in the news are made by combining several images, usually at least three, taken with a different filter. If we use filters that all fall in the visible part of the spectrum, then the different colors can be combined so as to give a composite image that corresponds roughly to what our eye might see, if only it had the sensitivity of the telescope/CCD combination. However, for images taken with IR or UV filters, the colors are arbitrarily assigned, usually to highlight different parts of the object with different chemical composition or other properties. These non-visible forms of light have no colors of their own, naturally.

The detectors used for x-ray, gamma-ray, and radio telescopes are inherently sensitive to the energy (or wavelength of frequency) of the light they detect, not just its intensity. As a result, it is not necessary to use filters for those kinds of telescopes.

3.1.6: Different Telescopes and Detectors for Different Wavelengths of Light

Telescopes help us see faint objects in optical light, but what about the other kinds of light? Might there be something of interest going on in other parts of the spectrum? The answer to this question is yes! But that was not always obvious. In fact, only slowly did scientists come to understand that celestial objects can be a source of light across the electromagnetic spectrum, all the way from radio waves to gamma-rays.

The earliest evidence for non-visible radiation from the cosmos came in the early 1930s. At that time, an engineer named Karl Jansky (1905 – 1950), working at the AT&T Bell Labs, was attempting to track down sources of radio interference that might compromise the new systems of radio communication then being developed (Figure 3.2). Engineers expected that lightning storms might degrade their radio signals. Jansky was attempting to characterize such noise - scientists refer to any such interference that interferes with their signals as "noise." He did find the anticipated noise, but he also found something else: a source of static that tracked across the sky with the stars as Earth rotated. At first, he thought it must come from the Sun. Upon further study, he discovered that the noise came from the Milky Way. This discovery, reported in 1933, was largely ignored by astronomers; they were all focused on visible light.

The first person to follow up on Jansky's discovery was Grote Reber (1911 – 2002), a radio engineer from the suburbs of Chicago. Reber built a radio telescope in his backyard to systematically survey the sky in radio waves. Despite its modest beginnings, radio astronomy has grown to become a vital part of astrophysics. The sky is now known to be bright with radio emission at many different radio frequencies and coming from many sources, not just the Milky Way.

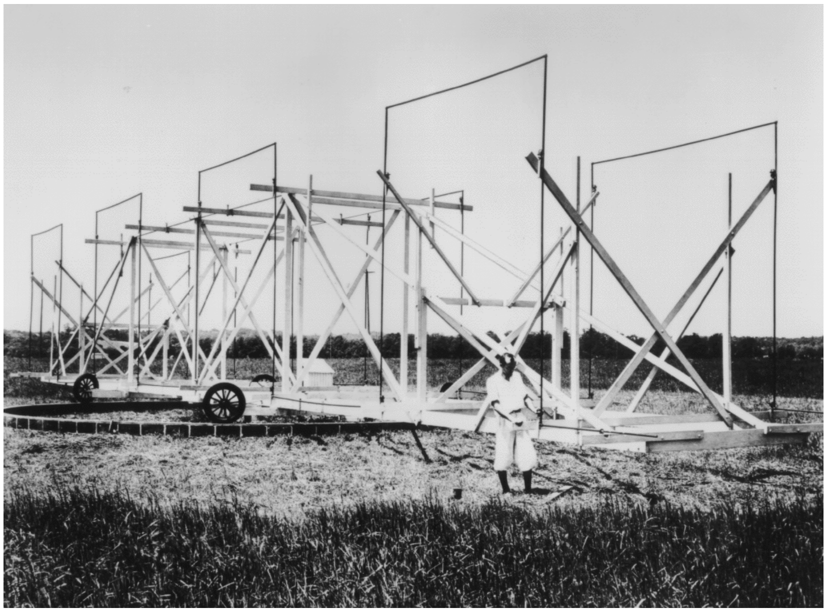


Figure 3.2: Karl Jansky and one of the radio antennas he built to study sources of radio signals. Jansky's antennas gave the first indication that there are sources of radio emission beyond the confines of Earth. Credit: NRAO.

The beginnings of x-ray astronomy are similar to those of radio astronomy, though they came several decades later. Most astronomers were still focusing their attention on optical wavelengths. They did not think that cosmic sources of x-rays other than the Sun would even be visible. Despite this, in 1962, a team of scientists led by Riccardo Giacconi launched an x-ray detector on a rocket above Earth's atmosphere, purportedly in an effort to detect fluorescence from solar x-rays on the Moon's surface. Actually, what Giacconi's team at American Science and Engineering really wanted to learn about was what sorts of then-unknown astrophysical sources might emit x-rays. Since that was a more difficult argument to make to the agencies that funded their efforts, they justified their efforts with lunar x-ray fluorescence. However, they found something much more interesting: cosmic x-ray sources well away from our solar system.

As a result of the success in discovering the first extra-solar x-ray source (called Sco X-1, located in the constellation Scorpius), many other rocket flights occurred into the 1960s, and many more x-ray sources were discovered. (Giacconi would receive the Nobel Prize in physics in 2002 for his "discovery of cosmic x-ray sources.") The team also built a NASA satellite known as Uhuru. Uhuru (which means "freedom" in Swahili) was launched from an offshore platform near the coast of Kenya in 1970. The launch occurred on Kenyan Independence Day (Figure 3.3). Uhuru did the first complete survey of the sky in x-rays and discovered many sources outside of the Galaxy. Some of the new x-ray sources corresponded to known radio or optical sources. Others did not.

Figure 3.3: The Uhuru satellite being tended to before launch. At left is Marjorie Townsend, one of the first engineers hired by NASA after its formation in the late 1950s. She went on to manage the program of Small Astronomy Satellites (SAS) that explored the Universe from space. Uhuru was also called SAS-1 because it was the first of the series. Also shown is scientist Bruno Rossi. Credit: NASA.

An interesting problem in the early days of x-ray and radio astronomy was in correlating the sources that were seen in radio, x-ray, and visible light. Almost none of the new sources discovered, either in radio or x-rays, were the stars that dominate the night sky in optical light. Sometimes, visible counterparts to the sources could be found, and often, these turned out to be galaxies. Sometimes, they appeared to be normal-looking stars, but astronomers suspected that something else might be there. Stars do not emit x-rays or radio waves in the amounts these new sources did. Some of the sources had no visible-light counterparts at all, and so they remained a complete mystery.

The field of astrophysics has undergone a revolution as new energy bands have been explored. It has become clear that if we want to have a complete view of the Universe, we must be able to observe it in all available wavelengths of radiation. Multi-wavelength astronomy, as it is now called, is incredibly powerful, because it allows us to probe a wide range of physical processes going on within astronomical objects. It has even led to the discovery of entirely new types of astronomical objects. Not only can we look at individual objects in multiple wavelengths, but we have also done surveys of the entire sky in all bands of the EM spectrum, as shown in Figures 3.4.

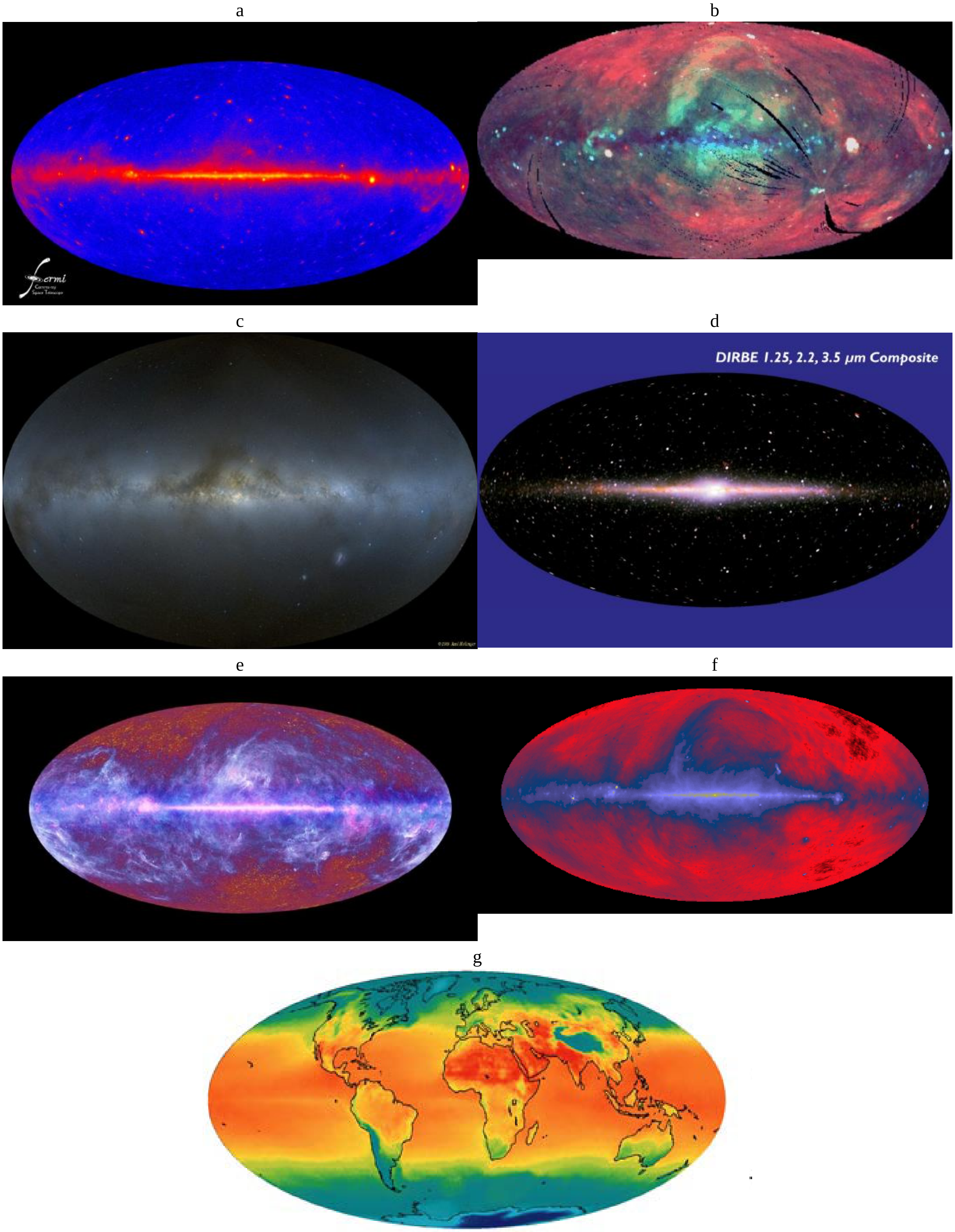


Figure 3.4: The images show all-sky maps at various wavelengths. As we will see in Section 2.5, images in non-optical wavelengths can be colorized to represent different wavelengths or intensities of light. (a) Gamma-ray sky as seen by the Fermi

Gamma-ray Space Telescope (Credit: NASA/DOE/Fermi LAT Collaboration), (b) X-ray sky as seen by ROSAT (Credit: NASA/GSFC/MPE/ROSAT/S. Digel and S. Snowden), (c) visible (Credit: Axel Mellinger), (d) near infrared sky as seen by DIRBE (Credit: NASA/COBE/DIRBE), (e) far infrared through microwave sky as seen by Planck (Credit: ESA, HFI and LFI consortia), (f) radio continuum, 408 MHz (Credit: Max Planck Institute for Radio Astronomy, generated by Glyn Haslam). (g) The style of map projection for all of these maps is known as an Aitoff projection. It allows us to see a 360-degree view of the sky. If a temperature map of the Earth is made in an Aitoff projection, this is what we would see (Credit: NASA/WMAP).

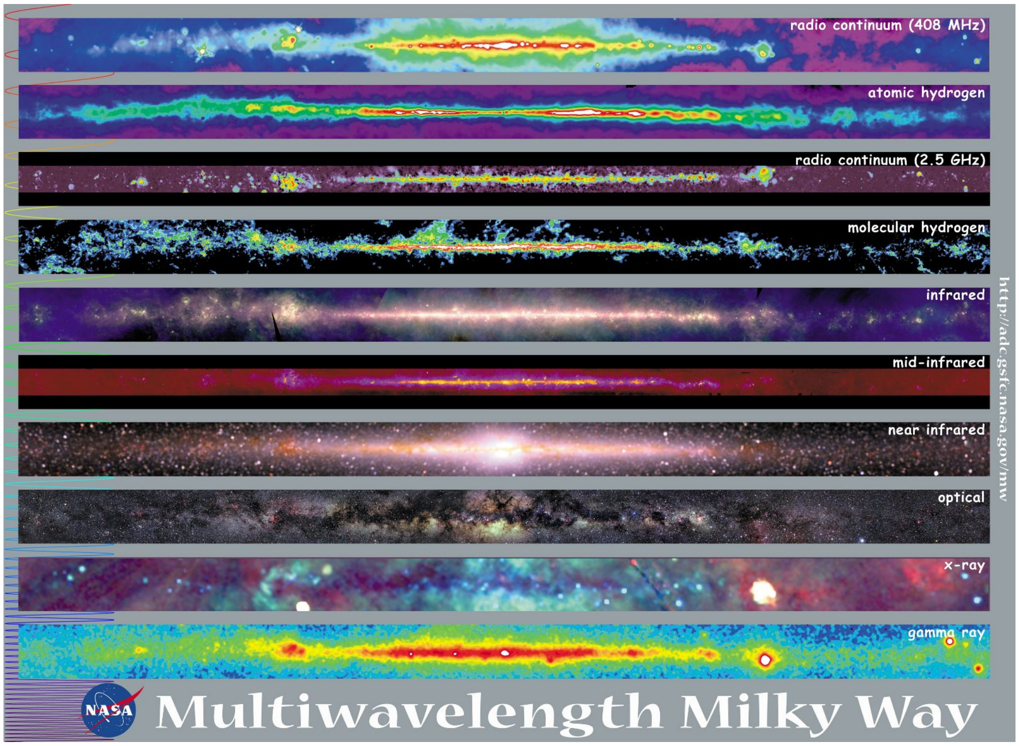


Figure 3.5: This image depicts the plane of the Milky Way from the radio region through gamma-rays. The top three strips are radio, followed by three infrared images, then visible, x-ray, and gamma-ray. Credit: NASA/GSFC.

As another example, consider the following images of a cluster of galaxies located about 2.6 billion light years away in the constellation of Camelopardus, shown in Figure 3.6. If we observe in visible light, we see light from the galaxies that is dominated by starlight. The surface temperature of stars corresponds to a peak wavelength of a Planck spectrum in or near the visible wavelengths. However, if we look at the cluster in x-rays, we see hot, diffuse gas between the galaxies, shown as the diffuse blue haze. And, if we look in the radio, we see jets of material shooting out of the cluster. These are the large red blobs in the image.

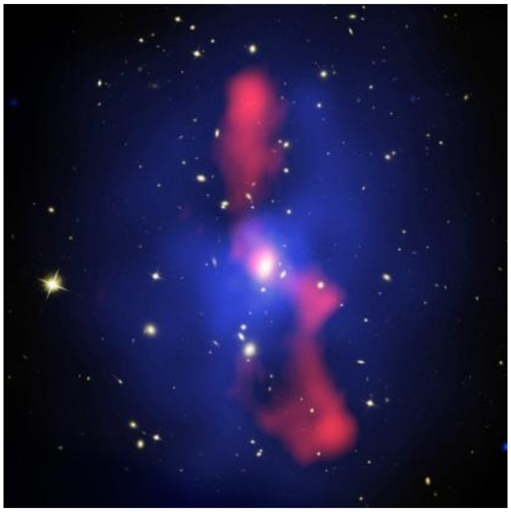


Figure 3.6: Multi-wavelength composite of galaxy cluster MS 0735.6+7421. Shown are visible (starlight from galaxies, white), x-ray (hot gas between galaxies, blue), radio (jets of material shooting out of central galaxy, red). (Credit: Credit: X-ray: NASA/CXC/Univ. Waterloo/B.McNamara; Optical: NASA/ESA/STScI/Univ. Waterloo/B.McNamara; Radio: NRAO/Ohio Univ./L.Birzan et al.)

The technology needed to detect electromagnetic waves of different frequencies and make astronomical measurements is diverse. CCDs and CMOS sensors have become the standard kind of detector for optical observations; they are also used in the infrared, ultraviolet, and low-energy x-ray bands. Radio telescopes typically use very large dishes to focus radio waves onto a radio receiver. Gamma-ray detectors are modeled after those used in particle physics. You will learn about these other types of detectors as you progress through the modules.

This page titled 3.1: Designing Telescopes Across the Spectrum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

3.2: Location, Location, Location

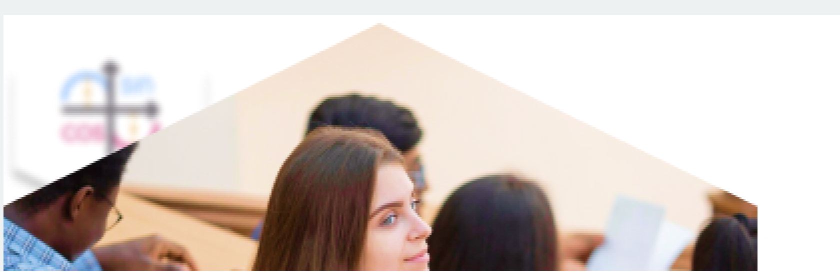
TELESCOPE LOCATION

Members of the Stargazers Club decide to look at some galaxies. It is a very dark night at Crystal Lake, and they are disappointed that their photos do not look as good as ones they have seen on the internet.

- **Aaron:** “When I look at a galaxy through the 8-inch telescope that we brought, I can’t see anything but a fuzzy blob. What am I doing wrong?”
- **Bonnie** hooks up her digital camera to her telescope and tries to take a photo. She complains: “Well, at least this looks like a galaxy. But it is certainly nowhere near as sharp as pictures from the Hubble telescope.”
- **Camille:** “That’s because Hubble has a bigger mirror than our telescope, so it can catch more light.”


f(x)
s
a
b


f(x)
s
a
b



Two more students join the conversation.

- **Mark:** “You know, if we had really big telescopes, like the Keck Telescope in Hawaii, I bet we could see those galaxies really well.”
- **Nina:** “Wait, so if they went to all the trouble to build a huge telescope like Keck, why didn’t they put it in space?”

3.2.1: EARTH'S ATMOSPHERE BLOCKS CERTAIN WAVELENGTHS

Not all wavelengths of light penetrate through Earth's atmosphere to the ground. Certain wavelengths of radio are absorbed, as are those in the far infrared. In fact, most of what penetrates our atmosphere and reaches the ground is the very narrow band of visible light and a large portion of the radio spectrum. Everything else is stopped high above. This is illustrated by Figure 3.7.

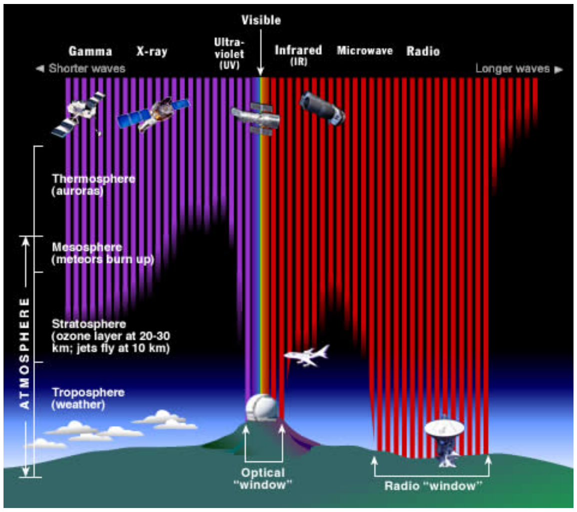


Figure 3.7: NASA/ESA image showing absorption in Earth's atmosphere across the EM spectrum. Most wavelengths of light do not reach Earth's surface. Only the narrow band of visible light and another band in radio can be viewed from the ground. Other wavelengths must be studied by high-altitude balloons and rockets, or better, from orbiting space observatories. Credit: NASA/ESA/CXC/STScI.

To observe high-energy radiation from celestial objects, specialized telescopes must be put on balloons or sub-orbital rockets that carry them to the highest reaches of the atmosphere. Better yet, we can launch them into orbit as free-flying satellites. From there, they can observe the cosmos completely free of atmospheric absorption. Almost all gamma-ray, x-ray, and ultraviolet observatories are on satellites launched into space, as are some of the infrared observatories. High-energy photons are so reactive that long before they reach the surface, they are absorbed by molecules in the upper atmosphere. This is a good thing. Exposure to all of these high-energy photons is harmful, or in some cases, fatal to humans. Only some of the lower-energy ultraviolet penetrates to the ground and contributes to sunburn - and skin cancer.

Not all of the blocked light must be studied from space, however. Sometimes, it is just necessary to get high enough in the atmosphere that you are above the absorbing layer. This is true for much of the near-infrared and microwave regions. The primary atmospheric constituent that absorbs these bands is water, but other molecules absorb them as well. Water in the atmosphere is mostly below about 3,000 m, so telescopes that are used to study the near-infrared and microwave parts of the spectrum can be placed either on high mountaintops or on airplanes that fly at high altitudes.

3.2.2: EARTH'S ATMOSPHERE MAKES IMAGES BLURRY

Even for wavelengths of light that are not blocked, Earth's atmosphere blurs the images. This blurring occurs because of turbulence and temperature differences, like the waviness you see in the air above a road on a hot day, or over a barbecue grill. For this reason, even optical telescopes are built on high mountaintops, a placement that puts them above a substantial fraction of the atmosphere and atmospheric pollution. This is also the reason the Hubble Space Telescope was placed in orbit above Earth's atmosphere. Even though HST has a modest-sized mirror (2.4 m) compared to most modern research telescopes, it has superior resolution because of its location (Figures 3.8 and 3.9).

Figure 3.8: The Hubble Space Telescope orbits above Earth, as seen by Space Shuttle astronauts. Credit: NASA.

Figure 3.9: The Whirlpool Galaxy as seen by (a) Hubble Space Telescope and (b) the 4-m Mayall telescope at Kitt Peak National Observatory in Arizona. The Hubble image shows much more fine detail, despite the smaller size of the telescope (2.4 vs. 4 meters). Advances in what are known as adaptive optics techniques are now allowing ground-based telescopes to rival Hubble in image sharpness. Credits: (a) NASA/ESA/STScI/AURA and (b) T.A.Rector and Monica Ramirez/NOAO/AURA/NS.

GOING FURTHER 3.1: HUBBLE SPACE TELESCOPE

The Hubble Space Telescope (HST) is an amazing **observatory** that acts as our eyes in space, allowing us to see the Universe in incredible detail—you have probably seen some HST images already. Named for astronomer Edwin Hubble, the HST is part of NASA's spectrum. The Hubble Space Telescope was launched during April 1990 from the Space Shuttle Discovery and celebrated its 20th birthday in April 2010. Unlike the other three Great Observatories, HST was designed to be serviced and repaired by astronauts while in space. This has been accomplished with great success on several occasions during Hubble's lifetime. Because of this ability, Hubble's equipment has been continually upgraded, allowing for clearer high-resolution images of the deepest reaches of the Universe. Now, you might be wondering, "How far into space can the Hubble see?" The HST has seen galaxies farther away than 12 billion light years! One such image is called the Hubble Ultra Deep Field (HUDF), shown in Figure B.2. The HUDF has provided astronomers around the world with a wealth of new and exciting information about the Universe.

Figure B.1 This image shows the Hubble Space Telescope in orbit above Earth, during servicing mission 3B in March 2002.

HST is located 353 miles (569 km) above Earth, circling the planet every 97 minutes. The telescope has a 2.4-m-diameter primary mirror and a 0.3-m (~12 inches)-diameter secondary mirror that direct the collected light onto the focal plane where it is analyzed by the science instruments. Compared to telescopes on the ground, these are not large mirrors; but because Hubble is above Earth's atmosphere, the mirrors can collect light without interference. HST is powered by two solar arrays that convert sunlight into electricity, much like solar pictures you often see in news reports. Two of these, the Wide-Field Planetary Camera 3 (WFPC 3) and the Advanced Camera for Surveys (ACS), observe in the optical. The other, called the Near Infrared Camera and Multi-Object Spectrometer (NICOMOS), observes in the near infrared; it can both image and take spectra. In addition to the imagers, there are two spectrometers. The first is the Space Telescope Imaging Spectrometer (STIS), and as its name implies, it can both image and take spectra, though it is primarily a spectrometer. The other spectrometer is the Cosmic Origins Spectrograph (COS). Both work in the optical and ultraviolet parts of the spectrum. Of these instruments, COS and WFPC 3 were delivered by the Space Shuttle on the final Hubble servicing mission in May 2009. STIS received a major overhaul during that same mission.

Figure B.2 Hubble Ultra Deep Field. Credit: NASA/ESA/S. Beckwith and the HUDF team.

To learn more about HST, visit the following:

- *http://hubblesite.org*
- *http://hubblesite.org/the_telescope/hubble_essentials/*
- *http://hubble.nasa.gov/technology/summary.php*

GOING FURTHER 3.2: CHANDRA X-RAY OBSERVATORY

On July 23, 1999, NASA launched the Advanced X-ray Astophysics Facility, or AXAF, as the second in its Great Obervatories program. The observatory was first proposed in 1977, and the intervening 22 years had seen its design go through several changes. Nonetheless, the observatory was, and continues to be, the most sesitive x-ray observatory ever built in terms of angular and spectral resolution. It studies x-rays with energies between about 0.12 and 12 keV (wavelengths from 0.1 to 10nm).

Figure B.3 The parts of the Chandra Observatory. To see an interactive version of this figure, click here. Mouse over the different parts to learn more details. Credit: NASA/CXC.

After its launch and 60-day shakedown phase, the observatory was christened the Chandra X-ray Observatory, after Subramanyan Chandrasekhar, the late Indian astrophysicist, former professor at the University of Chicago and Nobel Prize laureate. Chandra, as he was known, is considered to have been one of the foremost astrophysicists of the 20th century.

The Chandra X-ray Observatory hosts both imaging and spectroscopic capabilities. Its eight Wolter mirrors focus x-rays onto a focal plane where detectors from Chandra's four different instruments can be placed. The high-resolution camera (HRC) allows detailed images to be taken. Its angular resolution is about half an arcsecond, or about twice that of the average human eye. Unlike the HRC, the advanced CCD imaging spectrometer (ACIS) uses CCD arrays to detect x-rays. Because Chandra is in a highly eccentric orbit around Earth, taking it from 16,000 km to 133,000 km altitudes, the telescope is able to make much longer exposures than telescopes in low Earth orbit.

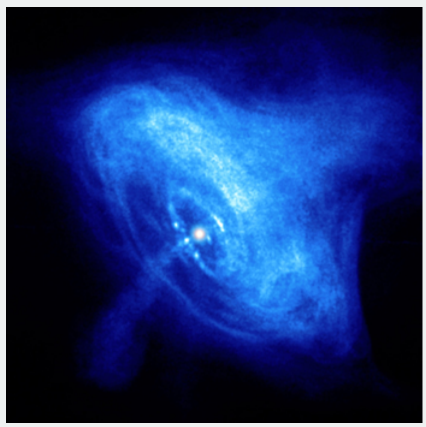


Figure B.4 This image and accompanying movie by the Chandra X-ray Observatory show the pulsar wind at the center of the Crab Nebula. The Nebula is the remnant of a star that exploded in the year 1054. At its center, it contains the collapsed core of the star, a rapidly spinning, extremely dense object called a neutron star. The neutron star spins about 30 times a second, and its strong magnetic field produces an outward flow of particles called a pulsar wind. Motions in this wind are an appreciable fraction of the speed of light. Note the jet of emission extending along the axis of the conical wind emission.

Click on http://chandra.harvard.edu/photo/2002/0052/ to learn more about this photo. Credit: NASA/CXC/ASU/J. Hester et al.

To learn more, visit the Chandra website:

- *http://chandra.harvard.edu/*

GOING FURTHER 3.3: SPITZER SPACE TELESCOPE

The Spitzer Space Telescope is the last in the series of NASA's four Great Observatories. It observes in the infrared (IR) part of the spectrum between 3 and 180 microns (1 micron = 10^{-6} m). Spitzer was launched in August 2003.

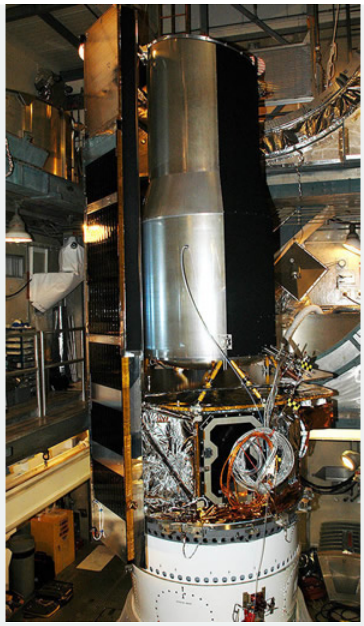


Figure B.5 This image shows the Spitzer Space Telescope sitting atop its rocket prior to launch. The telescope was launched from Cape Canaveral, Florida, in August 2003. Visible in this image is the solar array/shield to the left of the silver telescope body. The opposite side of the tube is black to aid it in cooling. Credit: NASA/JPL-Caltech.

Spitzer allows astronomers to study what they deem “cool” and “warm” objects. These are objects that have temperatures ranging from around zero degrees Celsius to a few thousand degrees Celsius. Such objects are primarily emitters of IR radiation. These objects include stars cooler than the Sun, substellar objects called brown dwarfs, planets, and warm clouds of dust in the Milky Way and other galaxies. But Spitzer is useful for more than just looking at warm objects. IR wavelengths of light can easily pass through clouds that are entirely opaque at visible wavelengths. Astronomers can use IR observations to peer behind these clouds to study stellar nurseries and the center of our Galaxy. In addition, the most distant galaxies in the Universe have their light stretched to IR wavelengths by the expansion of the Universe; thus, we must observe them in IR wavelengths if we wish to see the stellar activity within them.

Spitzer must be cooled to only a few degrees above absolute zero because it observes radiation that is emitted by warm objects. As a result, if the telescope is not cooled, its own radiation will swamp the emission from the astronomical objects it is used to study. This cooling requirement is what places the strongest constraint on the duration of the mission: The spacecraft is cooled by boiling off liquid helium to a temperature only a few degrees above absolute zero (–273 ^{o}C). Once the helium is exhausted, the cooled portion of the mission comes to an end. Spitzer reached this milestone in May 2009, after which it has been operated in “warm mode.” Some of the science remains the same in warm mode, namely, at the shorter IR wavelengths. However, the telescope is no longer able to study longer IR wavelengths beyond about 5 microns.

The cool stage of the mission was greatly extended by using an innovative orbit. Rather than orbit Earth, which is quite a warm object, Spitzer was put into a solar orbit that causes it to slowly drift away from Earth. This tactic had several advantages. First, it greatly expanded the fraction of the sky that was available to observe at any given time. Spitzer generally cannot point closer than about 90 degrees away from bright objects like the Sun, Earth, and Moon. However, because its orbit carried it so far from Earth and Moon, those bodies became unimportant for pointing the telescope. In addition, the telescope could be cooled all the way down to –242 ^{o}C by space itself, before any helium was needed. This allowed less helium to be used, lowering the weight of the spacecraft and, as a result, the cost of the mission.

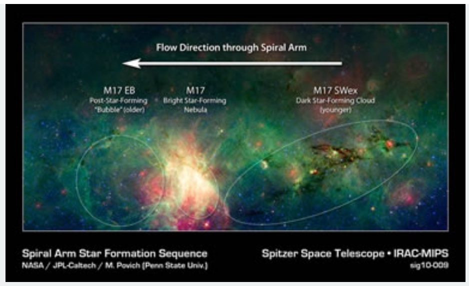


Figure B.6 Star Formation in the Milky Way. Here, we see the region of the Milky Way around the HII region M17. M17 itself, an area of current star formation, is shown red in this image. On the left of it is an area of earlier star formation that has since ceased forming stars. To the right is an area of dark clouds just beginning to collapse into stars. Because our eyes cannot see IR light, the colors here have been assigned arbitrarily: 3.6, 8.0, and 24.0 microns are shown as blue, green, and red, respectively. Credit: NASA/JPL/Caltech/M. Povich (Penn State University).

To learn more, visit the Spitzer website:

- *http://www.spitzer.caltech.edu/*

GOING FURTHER 3.4: GEMINI OBSERVATORY AND ADAPTIVE OPTICS

The Gemini Observatory is a pair of 8.1-meter telescopes, one located in the northern hemisphere and one located in the south. The telescopes are located in two of the best observing sites in the world, Mauna Kea in Hawaii and the Andes Mountains of northern Chile. They are optical/near infrared telescopes and were built and are operated by a consortium of seven partner countries: the United States, United Kingdom, Canada, Chile, Australia, Brazil, and Argentina. Both telescopes began science observations circa 2000.

Figure B.7 In this image, the Gemini North telescope on Mauna Kea, Hawaii, is seen against a backdrop of the Milky Way. The observatory sometimes uses a laser beam as part of its adaptive optics system, allowing it to correct for atmospheric distortions and obtain much sharper images than would otherwise be possible. The images can even rival those from the Hubble Space Telescope for sharpness. Gemini North has a twin 8-m telescope, Gemini South, located in the Chilean Andes. Credit: Gemini Observatory/AURA.

The Gemini telescopes were designed to carry out diverse observing programs in the near-infrared part of the spectrum, as well as the optical. Their locations in high and dry settings means that they are not much hampered by atmospheric water vapor, which strongly absorbs IR radiation. In addition, their mirrors are coated with silver, using a special process that enhances their IR performance. They are also able to use techniques of adaptive optics to correct for atmospheric turbulence, greatly increasing the sharpness of their images. Both Gemini locations have dark skies, making it possible for the telescopes to view faint objects.

The Gemini telescopes have a complement of state-of-the-art cameras and instruments that allow astronomers to take images and spectra of stars, galaxies, and other astronomical objects. One type of new technology used in both telescopes is adaptive optics, which can greatly improve the resolution of the images we take with ground-based telescopes. In an adaptive-optics system, astronomers use a wavefront sensor to model the blurring effects of Earth's atmosphere on light from a reference star (or at some telescopes, the blurring is measured from a spot made by a laser). This information is fed to a deformable mirror, which can rapidly change its shape and be used to counteract the atmospheric blurring, resulting in a sharper image.

Turbulence from the atmosphere changes at a rapid rate, so the adaptive-optics system must constantly update the correction applied by the deformable mirror. In typical adaptive-optics systems, these corrections are made hundreds to thousands of times per second; you can actually hear the deformable mirror whirring as it rapidly adjusts and readjusts. Adaptive optics has only been used regularly in telescopes since about 2000 and has mainly been developed for use with near-infrared observations. Astronomers are now pushing toward visible-light adaptive optics systems, and future space-based observatories may also make use of similar technologies. For instance, adaptive optics could correct for optical distortions due to the motion of the telescopes themselves as they orbit Earth or travel toward other planets.

Figure B.8 N44 Superbubble. This hot bubble of gas is made to glow by nearby hot stars. The ultraviolet light from the stars ionize the hydrogen gas and other species. The image is a composite of three images taken in narrow-band filters centered on the emission of hydrogen (H-alpha), doubly ionized oxygen (OIII), and singly ionized sulfur (SII). Credit: Gemini Observatory/AURA.

To learn more, visit the Gemini website:

- *http://www.gemini.edu/*

GOING FURTHER 3.5: VERY LARGE ARRAY RADIO TELESCOPE

The Very Large Array (VLA) radio observatory is one of the premiere astronomical facilities in the world. It is also one of the most productive scientific instruments of the past 30 years. The VLA consists of 27 separate parabolic antennas, each with a diameter of 25 m. The antennas can be moved along a system of railroad tracks, allowing the array to adjust its size from a 1-km diameter, in “D” configuration, up to a 36-km diameter in the “A” configuration. Two additional configurations have intermediate sizes. The array is sensitive to different angular sizes on the sky in the different configurations, having a maximum angular resolution of 0.04 arcseconds in the A array, comparable to that of the Hubble Space Telescope. The array is located at an elevation of 7,000 feet on the plains of San Agustin in New Mexico, far away from sources of radio noise like radio and TV stations and electrical appliances.

Figure B.9 Radio telescopes in the Very Large Array (VLA). These antennas are located in New Mexico. The signals from the antennas are combined, or synthesized, to effectively create a single telescope with much higher resolution than any single dish. Click on the image to learn more. Credit: courtesy of NRAO/AUI and NRA.

The VLA is a type of telescope called an interferometer. It combines the signals from its antennas electronically to create a single image. To achieve this feat, the distances between any two pairs of antennas must be known to a fraction of a wavelength. The timing of arrival of the radio signals at each antenna must also be known. The wavefront can then be reconstructed to provide an image of comparable quality to one made by a single dish with a diameter equal to the largest antenna separation in the array. Of course, since the synthetic dish created by all combined antennas is mostly empty, the VLA sensitivity is much lower than a filled dish of the same diameter. Only its angular resolution is comparable. Nonetheless, it is far less costly to build many small antennas and then to combine their signals than it is to attempt to build one very large antenna: Consider the difficulties encountered in trying to build a single, steerable dish 27 km in diameter, and it will be clear why astronomers chose an array of smaller dishes instead.

The VLA is sensitive to radio wavelengths from about 4 m down to 7 mm (frequencies from 70 MHz to 50 GHz, respectively), though the coverage is not complete in this region. In addition to making broadband observations, the telescope can observe individual spectral lines from atoms and molecules. This capability allows detailed studies to be made of the velocity structure of gas in galaxies, including our own, as well as its spatial distribution, for example.

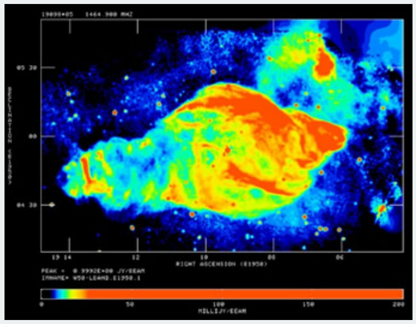


Figure B.10 Radio Galaxy Cygnus A. This VLA image shows the giant radio lobes of this galaxy, along with the relativistic jets of material that are being ejected from the black hole at its center. Radio galaxies are one type of active galaxy. Credit: Image Courtesy of NRAO/AUI.

To learn more, visit the National Radio Astronomy Observatory website:

- *http://www.nrao.edu/*

Including the section about the VLA:

- *http://www.nrao.edu/index.php/about/facilities/vlaevla*

3.2.3: COST

While it is possible, and sometimes even desirable, to go high into the atmosphere or into space, there are disadvantages in terms of cost and convenience. For example, servicing Hubble required expensive and risky Space Shuttle missions. Compare this requirement to simply driving to the top of a mountain, even if that mountain is in a remote desert location in northern Chile. Typically, the difficulty and expense related to working in space means that only telescopes that absolutely must be placed there to achieve the desired science goals are launched into orbit or beyond.

3.2.4: DARK SKIES

Most observatories are built in remote locations. This is true even for telescopes that detect optical and radio light, which easily reaches the ground. These locations are necessary because of city lights, weather, and radio interference from electrical circuits, cell phones, etc. The growing problem of “light pollution” (as illustrated in Figure 3.10) has led to great losses in sensitivity from historical telescopes that are now located too close to population centers. It has also led to laws in cities such as Tucson (near the National Optical Astronomy Observatory’s Kitt Peak) to try to limit the amount of light that can be directed into the sky at night. Tucson passed these ordinances in 1972, and they have been modified many times since then. As a result, Tucson is one of the few major cities where it is still possible to see the Milky Way galaxy with the naked eye.

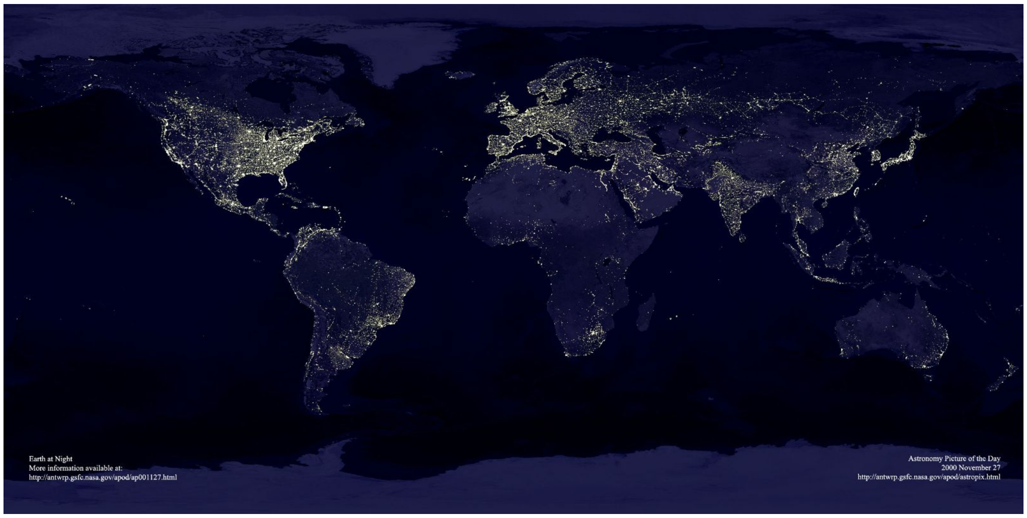


Figure 3.10: Earth at night. The night sky, available to nearly everyone only a century ago, is now a vanishing resource. City lights create a bright glow in the sky, preventing most people in cities and even small towns from seeing all but the brightest stars in the sky. Viewing the Milky Way and other faint objects is out of the question. Professional observatories must be located in remote areas with dark skies, but as development encroaches on these sites, optical astronomy becomes more difficult. Credit: C. Mayhew & R. Simmon (NASA/GSFC), NOAA/ NGDC, DMSP Digital Archive.

The Globe at Night project tracks sky brightness around the world. The International Dark Sky Association advocates and educates for dark skies through environmentally responsible outdoor lighting.

FUNDING NEW TELESCOPES

To get funding to build telescopes and use them to do observations, astronomers usually write proposals that are reviewed by committees composed of other astronomers. In this activity, your role is to be a member of a review committee, and your goal is to rate the scientific merit and feasibility of each proposal, to provide advice to the government official who controls the funding.

The table below summarizes the wavelength of the light, the target object the astronomers will study, and the size and location of the telescope that they propose to build.

Proposals for the Review Committee

PROPOSAL NUMBER	WAVELENGTH	TARGET OBJECT	DIAMETER OF TELESCOPE	LOCATION (ALTITUDE)
1	Radio	Jupiter	100 m	West Virginia (900 m)
2	Infrared	Surface of a neutron star	1 m	Chicago (175 m)
3	Microwave	Sun	3.5 m	Underground mine (1,000 m below Earth’s surface)
4	Visible	Distant galaxies	2.4 m	Earth orbit (300 km)
5	Ultraviolet	Quasars	1 m	Mountain in Hawaii (4,000 m)
6	X-ray	Supernova	25 cm	High-altitude balloon (25 km)
7	Gamma-ray	Exoplanet	10 cm	Suborbital rocket (550 km peak trajectory)

1. To help you decide which proposal(s) to fund, consider the following criteria:

a. Does the proposed object emit light in the indicated wavelength band?
b. Will the proposed object be able to be detected if you build a telescope of the indicated size and put it in the specified location? (You might want to revisit Figure 3.7 for help with this part.)

Fill in the table below with a yes or no answer to each of these two criteria, and then decide if you would recommend funding the proposal.

This page titled 3.2: Location, Location, Location is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

3.3: Imaging Astronomical Objects

WHAT DO YOU THINK: COLOR IMAGES

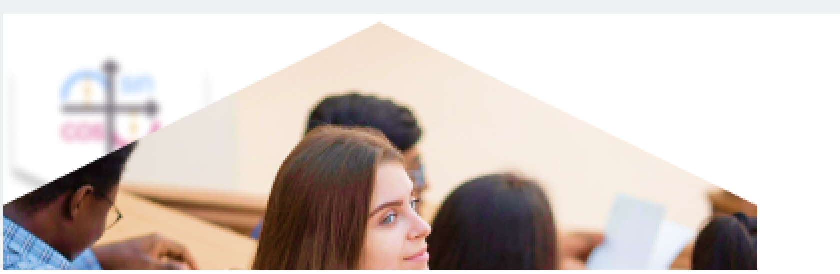
A group of students are browsing the internet before class.

One of the students asks, "Do you guys ever look at the Astronomy Picture of the Day website?"

- **Cyle:** "Of course. I love APoD, especially those colorful nebulae and supernova remnants. It's too bad there aren't any of those close by that you can see just with your eyes."
- **Donna:** "Those aren't the real colors. They have to do things to those images, like color in the clouds and gasses."
- **Eric:** "I thought they had to take different images of the same object and merge them together."
- **Fiona**: "Then how do they know what colors x-rays are?"


f(x)
s
a
b


f(x)
s
a
b



Taking color pictures with optical telescopes such as Hubble, or any ground-based telescope with CCD detectors, is very different and much more complex than using film in a traditional camera. Electronic detectors do not read out information in color—rather, the energies of the photons must be assigned colors in a computer process known as image processing. For visible light images, the color choices are sometimes assigned to try to faithfully reproduce what our eyes could see (if they could stare at the object for a long time without blinking or otherwise taking “snapshots”). Many full-color images are combinations of data taken in separate exposures of red, green, and blue visible light. When mixed together, these three colors of light can simulate almost any color of light that is visible to human eyes. That is how televisions, computer monitors, and video cameras recreate colors.

The standard colors that are mixed together on a television screen are called R, G, and B for red, green, and blue. This is done with a set of filters that pass light of wavelengths centered around 650, 520, and 450 nm, respectively. They block all other colors. Each filter is typically about 100 nm wide. While RGB filters are adequate for creating color images on a screen, these are not the filters that astronomers typically use.

Astronomical filters were not developed to create color images, though they can be used for that purpose. Rather, they were designed to study the physics of stars and other astrophysical objects. For instance, by comparing the brightness of a star in two filters, it is possible to determine its temperature. This is because the separate filters sample different points on the star’s Planck spectrum. Because Planck curves with different temperatures are unique, these two points are sufficient to uniquely determine the shape of the curve, and thus, its temperature. But stars are not perfect Planck emitters. They have absorption lines. (Sometimes, they even have emission lines.) Filters can be designed to be especially sensitive to these absorption lines, and thus, provide the ability to distinguish one type of star from another by means of the absorption features. Filters allow this determination to be made by simple imaging techniques rather than more complicated spectral techniques, usually saving a lot of time at the telescope.

Standard photometric filter sets have been developed over the past 50 years. The most common set is called the Johnson/Cousin system. It was developed in the 1960s and uses U, B, V, R, and I filters, for ultraviolet, blue, visible, red, and infrared. These filters typically have widths of about 100 nm, give or take, and they are centered at 365, 445, 551, 658, and 806 nm, respectively. Additional filters have been developed that push farther into the near and mid-infrared, going out into the region between 1,000 and 5,000 nm (1 and 5 microns). Other filter sets have been developed as well, usually with some specific use in mind. For instance, both Hubble and the Sloan Digital Sky Survey developed special filter sets based on their instrumentation and science goals.

In addition to these broadband filters, there are narrowband ones that only pass light close to a particular wavelength. These narrowband filters typically have widths less than 10 nanometers and are centered on an emission line from hydrogen, oxygen, sulfur, etc. Many of the beautiful pictures we see of nebulae (gas clouds) are taken using several of these narrowband filters to highlight emission from different atomic species.

All astronomical observations are done using these (or other) standard filter sets. This standardization allows one set of observations to be easily compared to another, a very important ability to have when doing science. Whenever a new filter standard is created, a lot of effort goes into understanding how it is related to others so that the new observations can be compared to older ones. Of course, it would be much easier to always use the same sets of filters for all observations, but sometimes, the science goals make that impractical or impossible.

MAKING MULTICOLOR IMAGES FROM RED, GREEN, AND BLUE LIGHT

While astronomers use slightly different filters than those used in televisions and commercial cameras, they use a similar technique to combine images taken with different filters to produce color images.

A. Go to this website to look at a color image of a star cluster and nebula that was created from separate red, green, and blue images:

Explore the results of toggling the “on” and “off” switches on this webpage. Set one color to “on” and the other colors to “off” to see what the image taken with that one color filter looks like. Now repeat this step with the other colors.

B. Now look at some other examples of color images taken with the Hubble Space Telescope:

Click on the individual examples, and read the accompanying text to learn more about how those color images were made.

Here is something else to notice: images taken with a particular color filter do not come off the telescope looking like that particular color. For example, an image of a nebula taken with a red filter does not appear red. Instead, the image appears black and white, with brighter white pixels indicating that the nebula emits more red light in those regions, and darker gray or black pixels indicating that the nebula emits little or no red light in the corresponding regions. To create a color composite image, astronomers assign colors to images taken with individual filters, and then combine those color-assigned images.

There are some other things to notice in these examples, too. Some of the example images (of the galaxy ESO 510-G13 and of the planet Mars) are “natural-color” images, created by combining images taken with three fairly wide filters and essentially showing us the objects’ true colors. In other cases (the images of the Cat’s Eye Nebula and the Eagle Nebula), individual images were taken with narrowband filters centered on emission from particular elements found in the gas within these nebulae. Colors were then assigned to these individual images to highlight the emission from those elements, rather than to preserve true color information. In the case of the Cat’s Eye Nebula, all three narrow-band filters would appear reddish if we were to assign “true” colors to them, so a natural color image would not be particularly informative. In still other examples (images of Saturn and of the Egg Nebula), individual images were taken in infrared filters that are centered on wavelengths longer than the human eye can detect. In these cases, it would be impossible to create a natural-color image. Instead, colors are assigned to the individual images to highlight features that stand out in the different infrared filters.

As noted previously, for images of astronomical objects at non-visible wavelengths, we often use color as a tool to enhance an object's detail or to visualize what the human eye could never see. The electronic information from Hubble spans the visible spectrum, but also includes data from infrared and ultraviolet wavelengths. Most of the spectacular Hubble images that you have seen are processed to highlight subtle details or to show scientifically important features in the objects.

In the next activity, you will see how black-and-white images from seven of Hubble’s different wavelength bands can be combined to make a beautiful color image.

COMBINING SEVEN FILTERS TO MAKE AN IMAGE

Go to this website to see a color image made from Hubble Space Telescope images of a galaxy taken in seven different filters.

Toggle through the different filters to see what the individual images that make up the color composite look like. Notice that some features of the galaxy show up very strongly in certain filters and other features show up more strongly in other filters. Noticing this difference can help an astronomer understand what physical processes are going on in the object in the image.

For instance, toward the outer regions of the galaxy are knotty-looking areas that are particularly bright in ultraviolet wavelengths. We know from our understanding of the Planck spectrum that objects that are brightest in the ultraviolet are hotter than objects that are brightest in visible or infrared wavelengths. The stars toward the outer regions of this galaxy must be hotter than stars toward the galaxy’s inner region. Astronomers can then ask why various regions of the galaxy look different at different wavelengths and follow up with further studies that may even indicate how this galaxy was formed and has evolved.

For non-visible bands of light, scientists typically choose a color palette that differentiates between levels of brightness (e.g., brighter colors such as red could be brighter regions) or between energies of the detected photons (e.g., regions with more energetic photons could use bluer colors). However, colors can also represent physical properties that are inferred from modeling the data. For example, comparing the number of photons of different energies can yield the temperature of the emitting material, photons in specific energy bands can yield information on chemical composition, and additional modeling can produce images that show density, mass, or different physical components such as gas, dust, or stars.

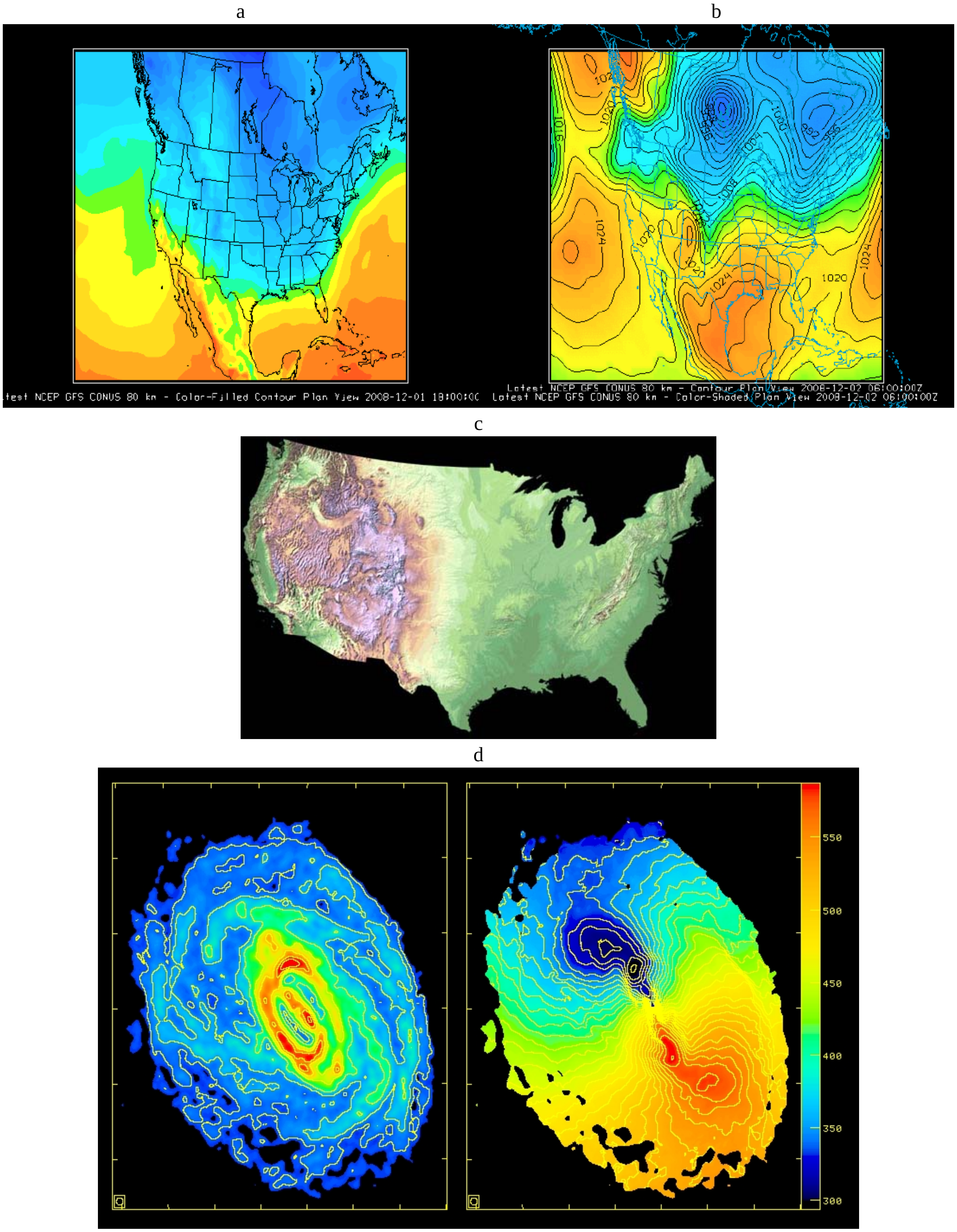


Figure 3.11: False Color Images. Color can be used to represent many things in an image. These images use different colors to depict (a) surface air temperature, (b) air pressure, (c) elevation, and (d) gas density (left) and gas velocity (right). Only the last of this set has an astronomical context, but other astronomical images can also use color to contain information about the image. In addition, sometimes astronomical images are taken in wavelengths that our eyes cannot see, and so “false” colors must be assigned to the different wavelengths shown. Credits: (a and b) Unisys, (c) USGS, (d) radio HI image of the Circinus galaxy by B. Koribalski (ATNF, CSIRO), K. Jones, M. Elmouttie (University of Queensland) and R. Haynes (ATNF, CSIRO).

FALSE-COLOR, NON-OPTICAL WAVELENGTHS

A. Initial thoughts

Take a look at Figure A.3.1:

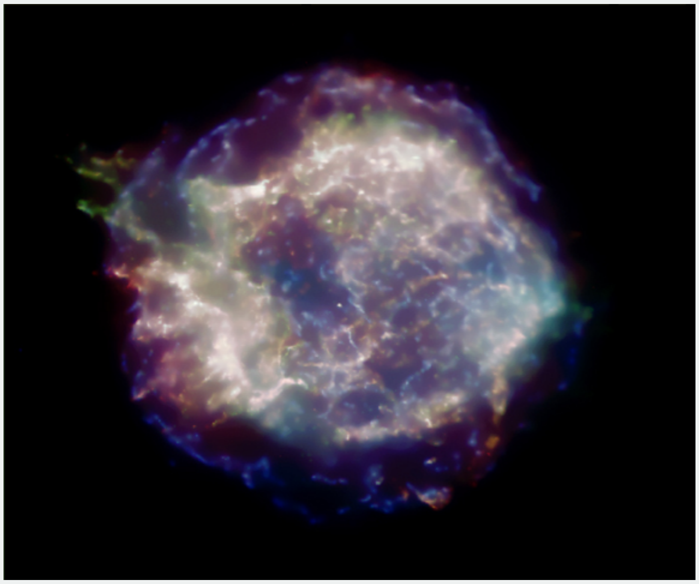


Figure A.3.1: Credit: NASA/CXC/SAO

Now take a look at Image A.3.2:

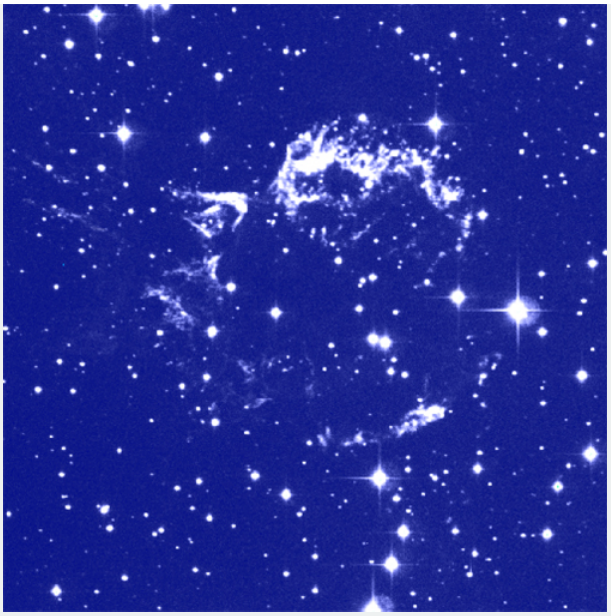


Figure A.3.2: Credit: NASA, Hubble Heritage Team (STScl/AURA)

B. Explanation and final thoughts

Both are images of the same object: Cassiopeia A (Cas A), a supernova remnant.

The second image is taken with only one visible-color filter and is not a composite image. It is what Cas A would look like if your vision was monochromatic (only able to see one color, but it would be a visible color none the less).

The first image was taken by Chandra, an x-ray satellite observatory. There are three colors in this image, red, green, and blue. Each color represents a different wavelength of x-ray light that corresponds to a range of energies. The red in the image corresponds to energies of 0.3 to 1.55 kilovolts. The green ranges from 1.55 to 3.34 kilovolts, while blue, the most energetic, includes energies between 3.34 to 10 kilovolts. Where the colors overlap are where there are greater ranges of those energies, and where it is white in the image is where the emitted energy ranges from 0.3 to 10 kilovolts.

In all of these examples, the images themselves, while often beautiful to behold, are not the primary sources of scientific data. It is the numbers behind the images that are analyzed for scientific content and that are modeled to try to understand what is actually happening.

This page titled 3.3: Imaging Astronomical Objects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

3.4: Determining the Brightness of Astronomical Objects

WHAT DO YOU THINK: WHAT ARE PHOTONS?

Members of the Stargazers Club are talking about their readings for class.

- **Jennifer:** Our book says that photometry is about counting photons—but what does that mean? What's a photon? Is it like a proton?
- **Lily:** Photons are light so I think the number of photons has to do with how bright the object is.
- **Mallory:** I think a digital camera must have some way to count them.

Whether our telescopes are here on Earth or in space, we can only measure the light that hits our detectors. This process tells us how bright an object appears to be, and astronomers call this apparent brightness the flux or intensity. The measured flux is different from the object's true inherent brightness, called its luminosity. The flux is just the energy we measure at a particular point in space in a given time, while the luminosity is the object's total power output, measured in watts. Flux and luminosity are related in a way that you are probably familiar with from everyday life: If something is farther away, it will look dimmer, and if it is closer, it will look brighter. This is true no matter how bright an object is intrinsically: moving farther away from it will always make it appear dimmer. Here, you will learn how to measure the fluxes of sources using a CCD detector.

A CCD detector is composed of an array of rows and columns of individual light-sensitive elements, or pixels. One way to think of the pixels is as tiny buckets that collect photons, sort of the way rain buckets might catch raindrops during a storm. An incoming photon of light strikes the detector and produces electrons that are stored in the detector until they are read out. Built-in electronics organize these electrons as charge packets, and a computer can display and record them as a digital image. The resulting digital image is composed of numbers that correspond to the relative numbers of electrons stored in a given pixel, and thus the number of photons that struck that pixel. These numbers are generally called the "pixel values." Figure 3.12 is a schematic representation of a detector with X columns and Y rows.

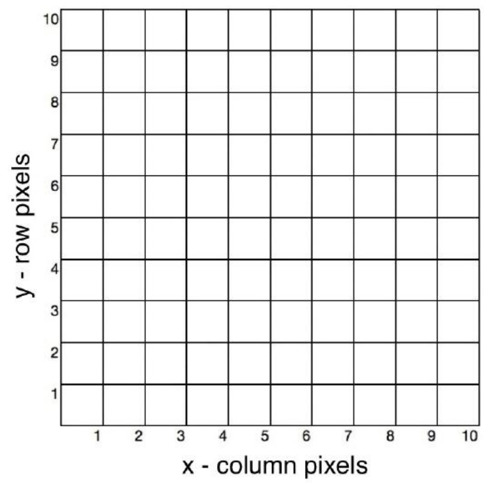


Figure 3.12: A CCD array is a grid of pixels, here numbered starting from the lower left. Each pixel is a mini light detector, storing electric charge in proportion to the number of photons that hit it. It is a bit like having buckets set out in a grid to catch rain: Each bucket will fill in proportion to the number of raindrops that fall into it. The pattern of charge on the array is what produces an image when the charge is read out by the camera circuitry. Credit: NASA/SSU/Kevin McLin.

Even though stars are technically point sources of light, an image of a star will have some physical size due to the blurring by Earth's atmosphere, as well as by the diffraction and interference of the light waves (due to the telescope's optics) that produce the image. A real star image will appear as a so-called seeing disk (see Figures 3.13, 3.14, and 3.15). This pattern will be brighter at the center of the disk and approach the sky background level at the edges of the pattern. Because of this blurring of stellar images, bright stars are represented by larger disks of light than faint stars on an image. That does not mean that the bright stars are necessarily bigger than fainter ones. In fact, they could be much smaller; we cannot tell just from the seeing disk.

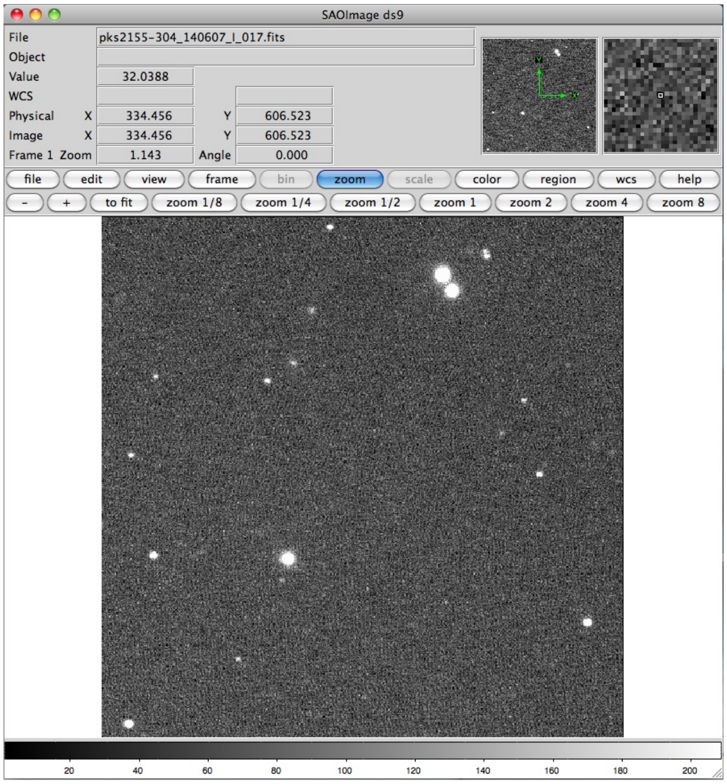


Figure 3.13: An image displayed in a popular software package. The stars are the bright circles and the sky background is dark in this example. The “bigger” stars are actually just brighter than the “small” stars. All stars are unresolved point sources in this image. The stars are actually collections of square-shaped pixels, as can be seen in Figures 3.14 and 3.15. Credit: NASA/SSU/Kevin McLin.

It can be useful to visualize a digital star image in terms of a three-dimensional histogram of pixel values (see Figure 3.14). Here, the height corresponds to the value at a particular pixel and the rows and columns correspond to the location on the detector. In this representation, taller places represent larger pixel values—a brighter part of the image.

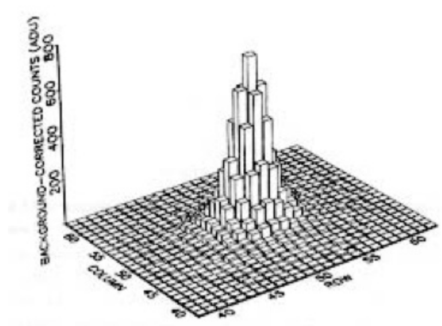


Figure 3.14: This plot displays the image brightness along a vertical axis above the pixel array. Each pixel contains a different number of counts, corresponding to the number of photons that landed in that pixel, and therefore, to the brightness of that part of the image. Notice how in this image, which is a star, the center is quite bright, and the edges gradually fall off to the sky background level. The width of the star is not zero, even though stars are technically point sources. The width of point sources is caused by optical effects in the telescope, as discussed in the text. Credit: NASA/SSU/Tim Graves.

Figure 3.15 shows a highly magnified digital image of a real star. Notice that you can see the individual pixels in the star image, as well as in the sky background around the star. The inner circle represents the region within which nearly all the light from the star has struck the detector. For determining the brightness of an object, this circle is termed the analysis aperture. If we add up all the pixel values in the analysis aperture, we will have a number that represents the brightness of the star. This number is directly related to the number of photons coming from the star that struck the detector during the time the exposure was being made.

https://phys.libretexts.org/@go/page/30470

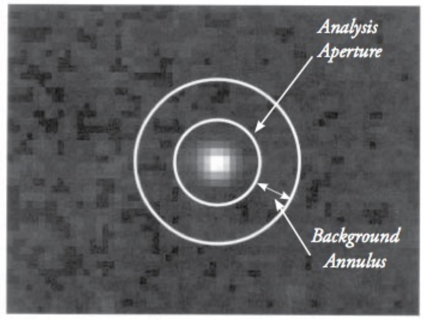


Figure 3.15: This image shows a star surrounded by the apertures used to measure its brightness. The inner aperture is the analysis aperture. It is assumed that all of the star's light falls within this aperture, though that is not absolutely true. The outer circle sets the outer limit of the background annulus. In this region, the sky background is determined. More commonly, a third circle is used, and the annulus between the second and third circles is used to measure the sky background. Notice the pixelation of the image. Credit: NASA/SSU/Tim Graves.

When making a measurement of a star's brightness, we must be careful to correct for any other sources of light. These light sources can contaminate the values of the star brightness. Spurious sources include light from the sky and from the detector itself. Together, these are called background counts. Even under the most ideal circumstances, the sky is never truly dark. There is always at least a small amount of light. In an urban or even a suburban environment, light pollution from artificial light sources will illuminate dust and molecules in the atmosphere. The atmosphere itself emits some light, resulting from interactions with solar wind particles. Another source is the integrated light from faint and distant stars - even Hubble cannot escape this background source. All of these sources from the sky are recorded by the detector and are collectively called the sky brightness.

Other important background sources results from the detector itself. The act of reading out the array introduces "read noise." This is from fluctuations in the count levels measured that are intrinsic to the camera and its electronics. Even in the total absence of light, a detector will produce electrons that get placed into the pixels. These are called the dark current of the detector and are caused by thermal motions (on the atomic level) of its internal components. Cooling the detector can almost completely negate this source of background counts.

To accurately measure the brightness of an astronomical object, the observer must correct for all these background counts (as well as several other sources of noise). The background annulus, formed by the inner and outer circles in Figure 3.15, is a region we can use to estimate the average background sky brightness reaching each pixel. The dark current can be determined by taking an exposure while the camera shutter is kept closed, thus only accumulating dark current since no light reaches the detector. The read noise is a more or less constant value for a given camera and need only be measured one time and then applied whenever the camera is used. These corrections can then be employed to adjust the counts obtained from the analysis aperture so that only the counts from the star are tallied.

GOING FURTHER 3.6: THE MAGNITUDE SYSTEM

In optical light, astronomers measure the brightness of astronomical objects using the magnitude system. Historically, the magnitude system is based on a concept first introduced by the Greek astronomer Hipparchus (c. 190 BCE–120 BCE). In about 129 BCE, Hipparchus produced the first well-known star catalog in the western world. In this catalog, Hipparchus ranked stars by what he called magnitude. He called the brightest stars he could see those of the first magnitude. Stars not so bright he called second magnitude. Using this system, he called the faintest stars he could just barely see sixth magnitude. This basic system has survived to today. Galileo forced us to change the system slightly. In 1610, when he used a telescope to view the sky, Galileo discovered there were fainter stars, and the magnitude scale became open ended. As telescopes became bigger and better, astronomers kept adding more magnitudes to the bottom of the scale. By the middle of the 19th century, astronomers realized it was necessary to define a more rigorous magnitude system. It had been determined that a first-magnitude star was approximately 100 times brighter than a sixth-magnitude star. Accordingly, in 1856, Norman Pogson (1829–1891) proposed that a difference of 5 magnitudes be defined as exactly a factor of 100 to 1 in brightness. Since at the time in the western world, it was believed that all human senses were logarithmic; it seemed perfectly reasonable to define a magnitude difference between two sources in the following manner:

$$m_1 - m_2 = 2.5 \log \frac{F_2}{F_1}$$

https://phys.libretexts.org/@go/page/30470

where the Fs are the fluxes, or amount of light from two objects, labeled 1 and 2. This gives the magnitude difference between the first object and the second. Keep in mind that fainter objects will have larger magnitude numbers than brighter objects. (Very bright objects will even have negative magnitudes—Hipparchus defined magnitudes this way 2,000 years ago, and so it is easier to keep it than to go back and redefine past observations using a new system.) This rule describes how brightness measured by a light-sensitive instrument can be represented as astronomical magnitudes. The magnitude of an object is also known as apparent magnitude (m), because it describes how bright an object appears to us on Earth. As we can see from the equation, it is related to flux. There is also a quantity known as absolute magnitude (M), which describes how bright an object is inherently; it is related to the object's luminosity.

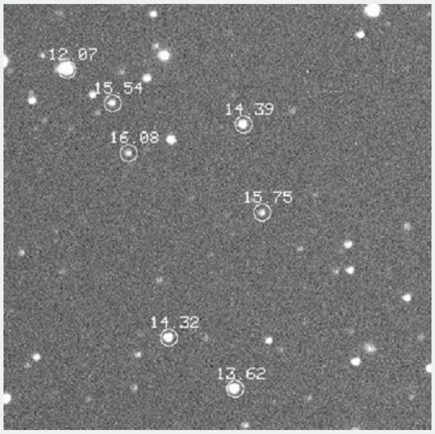


Figure B.11 Stars enclosed in circles have magnitudes labeled. Notice that the fainter stars, which appear to be dimmer, with smaller seeing disks have larger magnitudes. Credit: NASA/SSU/Kevin McLin.

IMAGE ANALYSIS WITH A CCD DETECTOR

In this activity, we will use a model showing the pixels of a CCD, like those found in cell-phone cameras, observatory telescopes, and other imaging devices. Each box represents a single bin or pixel on the CCD. If this were a real CCD, it would be a 225-pixel camera (0.000225 Megapixels).

Within each pixel is a number. This represents the total number of photons that have hit that particular pixel during one exposure. Even on the darkest nights, some amount of background light still can be faintly detected. Additionally, some pixels might be defective, reading higher or lower values than they should. In some cases highly energetic charged particles from space, known as cosmic rays, can strike pixels. These produce enormous numbers of electrons and give a false reading as if a huge number of photons have hit that single pixel.

A. Finding the source and subtracting the background.

Decide where you think the star ("source") is located in the image based on the photon counts of the pixels. The pixels in the image of the star will have higher than average counts. Use the drop-down menu to make sure the "source pixels" button is selected. Click on the pixels you want. When selected, the source pixels are outlined in orange. To deselect a pixel, click it again.

Once you have selected the pixels representing the star, click on the "select background pixels" button in the drop-down menu and select a large number of pixels that seem to have the lower background rate. When selected, the background pixels are outlined in blue. To deselect a pixel, click it again.

Below the CCD array is the formula for determining the star's photon intensity, which is calculated for you. The star's intensity is the total photon count for your selected source pixels minus the average background photon count for each source pixel. (The average background count rate per pixel was determined by adding up all the counts in the background pixels and dividing by the number of background pixels that you selected.)

Play Activity

https://phys.libretexts.org/@go/page/30470

If the average background photon count rate was not subtracted, the star's intensity would be too high because it would include the average number of photons from background sources along with the photons from the star.

Changes in temperature, time of night, and even whether there are car headlights passing nearby can all subtly change the intensity of the background photon count. This is why astronomers take multiple images of the same object, remove the background photons for each, and then average the images together.

Once you are done selecting your source and background data, continue to the false-color activity below.

B. Colorizing CCD images.

In this section, you will build a false-color image of the object above. The color boxes on the right allow you to choose the colors you use and the photon count thresholds for those colors.

You can add as many colors to the image as you like. To do so:

- Click "add color"
- Choose a color
- Set the minimum and maximum value of photon counts in the pixels to be shaded with that color

For example: If your average background photon count was 15 (from the formula in the first part of this activity), you might set the first color from 0 to 15 and make its color black (or white if you wanted to see a negative exposure, which is what astronomers commonly use). Then you could set a second color from 15 to half the value of the highest photon count of all your pixels.

Play Activity

This page titled 3.4: Determining the Brightness of Astronomical Objects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

3.5: Wrapping It Up 3 - Light, Telescopes, and Astronomical Images

In this chapter, you have learned about light and the many things astronomers can learn from it. For this activity, you will use your new knowledge to study images of an astronomical object of your choice in multiple wavelengths of light. Your job is to select an interesting object that has been imaged by telescopes in three different wavebands.

Select one object from the “Sample Astronomical Objects” list below that most interests you. Once you have chosen an object, you will need to visit three of the telescope webpages listed in the “Helpful websites” list and locate an image of your chosen object. Each page will have a search function that will allow you to locate what you need.

Before you begin the activity, take some time to reflect on what you have learned about light, color, and imaging. After studying the images you have chosen, answer the questions that follow.

Sample Astronomical Objects:

- Crab Nebula supernova remnant
- Cassiopeia A supernova remnant
- Center of Milky Way Galaxy (Sgr A)
- M101 (Pinwheel) galaxy
- M87 (Virgo A) galaxy
- M51 (Whirlpool) galaxy
- Perseus A cluster of galaxies
- Hydra A cluster of galaxies

Helpful websites:

- Hubble Space Telescope
- Chandra X-ray Telescope
- Spitzer Space Telescope
- Very Large Array radio telescope
- W. M. Keck Telescope
- Astronomy Picture of the Day

For each of the three images that you have chosen for your object, answer the following questions:

This page titled 3.5: Wrapping It Up 3 - Light, Telescopes, and Astronomical Images is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.


f(x)
s
a
b


f(x)
s
a
b

D. Questions to be graded for accuracy

This picture of the Centaurus A galaxy is a composite of three images in different wavebands.

This image will be used for questions 1-4.

https://phys.libretexts.org/@go/page/31229

This page titled 3.6: Mission Report 3 - Light, Telescopes, and Astronomical Images is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

4: Moving Through Space

Chapter 4 discusses how objects in the Universe move. First, the chapter explores the relationship among distance, speed, and time and the concept of lookback time. Next, you will learn how to measure the Doppler Shift and use it to determine the velocities of moving objects. The motions of astronomical objects – their rotations and orbits – are also discussed.

This page titled 4: Moving Through Space is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

4.0: Moving Through Space Introduction

We are all familiar with the concepts of distance, speed, and time. In fact, we use them every day - even if we are not aware of doing so. For example, in the morning, we are awakened by an alarm clock at a certain hour (time), hurry up to get from our house to school or work (speed), as we travel from one place to the other (distance). Yes, this is only one way in which speed, time, and distance are relevant to our world. Everything in the Universe is affected in some way by these three important factors. In this chapter, we will examine the concepts of distance, speed, and time; how they are related; and several ways that they are measured.

Video Transcript

A New Day Dawns

A new day dawns.

The ancients perceived the Sun as traversing the sky with deific purpose, unaware that it was they and the Earth beneath their feet spinning about the terrestrial axis that are actually in motion, and that it is this revolving motion of the Earth which gives birth to our days.

It is this motion, this axial rotation of the Earth, which led ancient humans to understand and define time, day, night. And as they unknowingly spun around with the planet, they used the position of the Sun and stars to further break down the segments of time: hours, minute, seconds. The night sky appears as a tapestry of lights rotating around the North Star.

Yet, not so long ago, humans pulled back the tapestry to reveal that the stars are not at one fixed distance; that they are not merely pinholes in a celestial sphere. Rather, each is like our Sun, though existing at various distances immense by terrestrial standards. And, of the wanderers - the planets who travel their own paths at their own speeds, indifferent to the progression of the background stars - they too rotate around their axes, and as they do they travel their own elegant elliptical causeways around the Sun marking out their own years with each completed cycle, as does the earth every 365.26 days.

And while we speed around the Sun, traversing 30 kilometers around our circular orbit each second, nearly a billion kilometers each year, the Sun and our solar system with it are orbiting the center of our Galaxy at over 200 kilometers per second. And despite this amazing mind-boggling speed to complete just one orbit around the galaxy, one galactic year, the journey takes us 240 million years to complete.

Such are the motions and times of our universe.

This page titled 4.0: Moving Through Space Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

4.1: Relationship Between Distance, Speed, and Time

Learning Objectives

- You will be able to distinguish among distance, speed, and time
- You will be able to convert between various units of time
- You will be able to convert between various units of speed
- You will be able to perform calculations using the relationship between distance, speed, and time
- You will know what a light-year means and how it is related to a year
- You will be able to determine the lookback time to various astronomical events, including ones occurring at different distances and times

What Do You Think: How Long to Get There

https://phys.libretexts.org/@go/page/30473

4.1.1: Distance, Speed, and Time

Before we can apply the concepts of distance, speed, and time to the Universe, we must understand the specific meaning of each and examine how they are related.

Distance can be defined as how far apart two objects are in space. For example, the distance between Chicago and San Francisco is about 2,100 miles (about 3,400 km). As we discussed in Chapter 1, the distances involved in astronomy are so large that special distance units are often used for convenience. For example, Earth is about 150 million km from the Sun (93 million miles). This distance is called 1 astronomical unit (AU), a unit of measure invented specifically for astronomy.

Other units used specifically by astronomers to measure distance include light-years (ly), light-hours, light-minutes, and light-seconds. The light-minute is how far light travels in 1 minute. For example, Earth is about 8 light-minutes from the Sun (it takes 8 minutes for the light to travel from the Sun to Earth), so an astronomical unit is also 8 light-minutes. We often find it convenient to measure distances in the Solar System in terms of light-minutes, or even light-hours, the distance light travels in 1 hour.

A light-year is the distance light travels in a vacuum (empty space) in 1 year. In SI units, a light year is 9.5×10^{15} m. The light-year is used to measure large distances, those beyond our Solar System. We generally use light-years to measure distances to other stars in our Galaxy, or even to galaxies beyond our own. For example, the distance across the disk of the Milky Way Galaxy is about 100,000 light-years, and the nearest major galaxy to ours (Andromeda, or M31) is 2.5 million light-years away.

Another unit that astronomers use is the parsec (pc). A parsec is equal to 3.26 light-years, and so it is also used to describe large distances. Parsec is usually accompanied by Greek prefixes, such as kilo (10^3) and Mega (10^6), to mean a thousand parsecs or a million parsecs. When abbreviated, these units are written as kpc and Mpc, just as km is a thousand meters and MW is a million watts. Therefore, the distance across the disk of the Milky Way can be described as 100,000 light-years or as 30 kpc (30,000 parsecs).

Time is a measure of how long it takes for an event to occur, or its duration. It is an observed phenomenon that we use to describe past, present, and future events. Intervals of time are usually measured in units that are convenient to the events in our daily lives. These are the familiar measurements of years, months, weeks, days, hours, minutes, and seconds. For shorter times, we can again use Greek prefixes, so that a millisecond (ms) is one thousandth of a second. For longer times, we use the year: it takes 1 year—about 365 days—for Earth to travel around the Sun one time. Humans use the year to measure the age of people and things—including the Universe itself. We can also add the Greek prefixes to the year; for example, a Gyr (Gigayear) is 1 billion years.

Speed is defined as the rate that an object travels through space; it is the distance traveled in a certain amount of time. A familiar example from our daily lives can be seen when we drive to work or school. Our speed is measured by how fast or slowly we drive. Our car is said to have a speed of 60 miles per hour (mph) if it will travel a distance of 60 miles in exactly 1 hour.

Velocity is another term that you will see in the activities below. Velocity can be simply defined as speed in a particular direction. In the above example for speed, we said that a car travels at a speed of 60 mph. If we add a direction, say the car is traveling north at 60 mph, we then have the car's velocity: A car traveling 60 mph north is different from a car traveling 60 mph west. Both travel the same distance, 60 miles, in 1 hour, but if they start at the same spot, they will certainly arrive at different destinations! It is this difference that velocity takes into account.

Mathematically, velocity is often represented by the letter v as seen in the equation:

$$\vec{v} = \frac{\vec{d}}{t}$$

which means that velocity equals displacement (change of position) divided by time. Here, the *v* and *d* have arrows above them to remind us that they include the direction traveled.

Having made the distinction between speed and velocity, we must warn you that people often say velocity when they really mean speed. For instance, when discussing planets orbiting the Sun, or the flight path of rockets and other spacecraft, we might refer to their velocity when we really have no interest in the direction of travel. Fortunately, it is generally easy to tell from the context if the direction as well as speed is important.

In physics and astronomy, the speed of light plays an important role in understanding how our Universe works. Light-speed is the rate that light travels through empty space (a vacuum), which is about 3.0×10^8 m/s or 3.0×10^5 km/s. It is represented by the symbol *c*, and measurements show it to be constant—it does not change, regardless of the state of motion of either its source or the one measuring it.

Using the mathematical relationship between velocity, distance, and time is how we find the equivalent distances of light-minutes and light-years in SI units. Using the previous equation, and ignoring the arrows in this example because we are not concerned about direction, we can determine how far the light year and light minute are in meters:

$$v = \frac{d}{t}$$

Rearranging,

$$d = v \times t$$

Plugging in values,

$$\begin{aligned}1\text{ light-minute } &= (3\times 10^{8} meters/\ \cancel{second})\times(60\ \ \cancel{seconds})\\ &= 1.8\times 10^{10}\, meters.\end{aligned}$$

and

$$\begin{aligned}1\text{ light-year } &= (3\times 10^{8} meters/\ \cancel{second})\times(1\ \cancel{year}\times 3.15\times 10^{7}\ \cancel{second}/\ \cancel{year})\\ &= 9.5\times 10^{15}\, meters\end{aligned}$$

Play Animation **Play Animation**

Keep the definitions of time, distance, and speed, and of the relationships between them, in mind as you work through the activities that are presented in the following sections.

Powers of Ten: Timescales

In this activity, you must drag and drop the tiles from the right-hand side of the screen into the correct open space in the table.

The first column gives a physical description of a time span, the second is the measurement of that time span in convenient units, and the third represents the time span expressed in seconds using scientific notation.

You will know you have dragged the tiles into the correct space when you see a green check mark.

Play Activity

The previous activity gave you a chance to think about some different timescales you might encounter in your daily life or as part of this class. It asked you to compare them and rank them. This is a useful skill to have when you are trying to understand how one process or event might relate to another. However, in the previous activity, you were given all of the times in a common unit, seconds. Normally, things are not so convenient. You will generally have to convert to a common unit before you can make a comparison. The next set of activities gives you additional practice with this useful skill

Unit Conversions: Time

In the next set of activities, you will work with the relationship between distance, speed, and time. These quantities can be described in various ways. For instance, speed can be miles per hour (mi/hr or mph) or kilometers per second (km/s), or even furlongs per fortnight! The important thing is that speed is always the ratio of a distance traveled to the time needed to travel that distance: distance/time. While the units we use to describe a speed are arbitrary, we generally use whatever is most convenient. The next several activities are intended to give you a better understanding of speed and its relation to distance and time.

https://phys.libretexts.org/@go/page/30473

DISTANCE, TIME, AND SPEED IN EVERYDAY LIFE

A. Speed

Worked Example:

1. If you have been driving on the highway for 1 hour and have traveled 30 miles, what is your average speed?

- Find: v
- Given: t = 1 hour, d = 30 miles
- Concept(s): $v = d/t$
- Solution: v = 30 miles/1 hour = 30 miles/hour

Question:

B. Distance

Worked Example:

1. If you have been driving on the highway at 60 miles/hour for 2 hours, how far have you traveled?

- Find: d
- Given: t = 2 hours, v = 60 miles/hour
- Concept(s): $v = d/t \rightarrow d = vt$
- Solution: d = (60 miles/~~hour~~)(2 ~~hours~~) = 120 miles

Question:

https://phys.libretexts.org/@go/page/30473

C. Travel Time

Worked Example:

1. If you have been driving on the highway at 20 miles/hour and have traveled a total of 80 miles, how much time have you been traveling?

- Find: t
- Given: d = 80 miles, v = 20 miles/hour
- Concept(s): v = d/t → t = d/v
- Solution: d = (80 ~~miles~~)/(20 ~~miles~~/hour) = 4 hours

Question:

CC BY-NC-SA

HOW MUCH TIME TO GET THERE

Now the goal is to determine how much time it would take you to get to various places in the Universe, if you were traveling on foot (if it were possible!), in our current fastest spaceship, at the speed of light, or at half of the speed of light.

Assume the following speeds:

- Speed of a person walking: about 3 miles/hour = 5 km/hour
- Speed of our fastest spaceship: 10 miles/second = 16 km/second = 57,600 km/hour
- Speed of light: 186,000 miles/second = 300,000 km/second
- Half of the speed of light: 93,000 miles/second = 150,000 km/second

Drag and drop the travel time tiles for the following places you might like to visit. Their distances are:

- Moon : 240,000 miles
- Nearest star (Proxima Centauri): 4 light-years
- Across the Galactic Disk : 100,000 light-years across
- Nearest galaxy : 2.5 million light-years

There are several worked examples below to get you started. All of the times can be calculated based on the concept that speed = distance

time ($v = d/t$), or equivalently, that time = distance/speed ($t = d/v$).

Worked Examples:

1. How much time would it take to walk to the Moon, if you could?

- Find: t
- Given: Use the facts that the distance d to the Moon is about 240,000 miles, and the typical person can walk at a speed, v, of about 3 miles/hour (without breaks).
- Concept(s): Use the concept that v = d/t → t = d/v.
- Solution: Put in the known numbers: t = d/v = (240,000 ~~miles~~) / (3 ~~miles~~/hour).
- Cancel miles, and use calculator: t = 80,000 hours.
- Convert hours to years: t = 80,000 ~~hours~~ × (1 ~~day~~/24 ~~hours~~) × (1 year/365 ~~days)~~ = 9 years.

2. How much time would it take to get to the nearest star (4 light-years away) traveling at light-speed?

- Find: t
- Given: d = 4 light-years, v = speed of light
- Concept(s): t = d/v

https://phys.libretexts.org/@go/page/30473

- Solution: (4 light-years) / (1 light-year/year) = 4 years.

3. How much time would it take to get to the nearest star (4 light-years away) traveling at half light-speed?

- If you are going at half of the speed of light, that means you are going slower by a factor of 2. That means it will take you 2 times longer than if you were traveling at light-speed.
- 4 years × 2 = 8 years.

Now you are ready to fill in the chart.

Play Activity

4.1.2: Lookback Time - Looking Far Away is Looking Back in Time

What Do You Think: Light-years

The Stargazers Club is discussing a galaxy that is far, far away. Keisha has pulled up an image of it on her phone and tells the group that it is 10 billion light-years away.

- **Indira:** “That means we’re seeing the galaxy as it looked 10 billion years ago, when it was very young. It probably looks a lot different now.”
- **Jason:** “Light-years is the amount of time it takes the light to get to us, right?”
- **Keisha:** “If the galaxy is 10 billion light-years away, then it must be 10 billion years old.”

The distances to stars and galaxies are so large that even light, traveling almost 300,000 km every second, still requires *years* to travel to them. That means when we look at a star, Alpha Centauri for example, we are not seeing it as it is. We are seeing it as it was *4 years ago*. The same is true for the Sun, of course. We do not see it as it is, we see it as it was 8 minutes ago. That has certain implications for our observations of the Universe; if the Sun were to shut off at this moment, we would not know it for 8 minutes. That is because any photons that just left the Sun would not arrive for 8 minutes, and until they got here we would have no way to know that the Sun had gone out.

This concept is so important in astronomy that it is given a special name: lookback time. We would say that the lookback time to the Sun is 8 minutes. To Alpha Centauri, the lookback time is about 4 years. But what about other parts of the Milky Way Galaxy, or other galaxies? Well, the typical lookback time to objects inside our Galaxy is several tens of thousands of years. That is because our Galaxy is about 100,000 light-years across. As a result, light requires up to about 100,000 years to reach one part of the Galaxy from another part. For external galaxies, the lookback times are even bigger.

The nearest large galaxy to the Milky Way is M31, the Andromeda Galaxy. This galaxy is about the same size as ours, but it lies about 2.5 million light-years away. So, if you go out tonight and find M31 (it can be seen from the northern hemisphere in the late summer or fall if you know where to look), the glow you see will have begun its journey 2.5 million years ago. Think about that for a moment. When that light started out, human beings (*Homo sapiens*) did not yet exist as a species. Parts of the Coast Range of California were still pushing upward from the sea, the current volcanoes of the Andes mountains and the Cascades had not yet grown, and the island of Hawaii was still hidden below the Pacific waves.

Still, if we could somehow violate the laws of physics and see M31 as it is right now, we would see very few differences between now and 2.5 million years ago. Galaxies do not change very much over such a time period, unlike Earth and its living creatures. However, if we look at even more distant objects, and that includes all galaxies, then the lookback time is greater. Nearby galaxies are tens of millions of light-years away, and the most distant galaxies are *billions* of light-years distant. We are seeing those galaxies, the nearest of them, as they were tens of millions of years ago. For the most distant galaxies, we see them as they were *billions* of years ago. Even for a galaxy, a billion years is a long time. In fact, the lookback time for the most distant galaxies goes back to nearly the beginning of the Universe. (Did you know that the Universe had a beginning?) In a sense, the finite speed of light turns the entire Universe into a time machine, allowing us to see its history.

Lookback Time and Units

In this activity, we will review measures of distance and time that are typically used by astronomers.

A. Basic definitions

B. Observing an astronomical event

In 1994, scientists were amazed and thrilled to watch pieces of Comet Shoemaker-Levy 9 hit the planet Jupiter. Tremendous explosions resulted, creating plumes many thousands of kilometers high, hot "bubbles" of gas in the atmosphere, and large dark "scars" on the atmosphere that lasted for weeks. At the time, NASA's Galileo satellite was about 1.6 AU from Jupiter, and the Voyager 2 satellite was at a distance of 42 AU from Jupiter (on its way out of the Solar System).

Hints: Recall that 1 AU = 8.3 light-minutes. Also, when speed is the speed of light, if the distance is in light-minutes, the time is in minutes. (The same is true for light-years and years.)

C. Communicating with satellites

In 2003, NASA sent missions to Mars that placed the rovers Spirit and Opportunity on the surface of the planet.

Hints: Recall that 1 AU = 8.3 light-minutes = 150 million km.

1. The closest distance from Earth to Mars is about 55 million km. At this distance, how much time does it take for radio control signals from Earth to reach Spirit? Round to the nearest minute.

LOOKBACK TIME

In this activity, you should see a star field. In the center is the Observer star. Around the Observer star are various other stars that may go supernova (explode). In fact, three of them are going to go supernova, and you need to figure out the order that the Observer star will see them.

When you select the “next” button in the bottom right, the “event order” field at the top will be filled out. The “event order” displays the three stars in the star field that will go supernova. Also displayed is the timeline for when the stars explode. The first star explodes at time zero and starts the clock running.

Find the stars that will go supernova in the star field and hover your cursor over them. This displays the distance to the star from the Observer star in light-years. This distance corresponds to its lookback time in years. For example, Star D is 1.75 light-years from the Observer star, which means that if it were to go supernova, 1.75 years would pass before the Observer would see it. The lookback time is 1.75 years for that star.

Determine the order in which the Observer will see the three supernovae based on the time between each supernova and the distance from the Observer star.

Once you have made your selections in the “observation order” drop-down boxes, click the “next” button again and watch the supernovae.

As each supernova is seen by the Observer star, their order will be indicated at the bottom, over your selections. Once all three supernovae have been observed by the Observer star, the application will let you know if you were correct or not.

Click the “next” button to start another round.

Play Activity

This page titled 4.1: Relationship Between Distance, Speed, and Time is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

4.2: Measuring Motion - the Doppler Shift

Learning Objectives

- You will know that spectra can be represented as ribbon or line plots
- You will know that the motion of sources and observers change the observed wavelengths and frequencies of light

What Do You Think: Doppler Shift

Ever wonder how scientists are able to measure the masses of stars or galaxies? Or how they detect planets around other stars, or how it is possible to determine the distance to a galaxy across the Universe? It turns out that these measurements have something in common with everyday experiences like the apparent change in pitch of a car horn honking as it passes you. We will explore how these very different things are related in the following section.

Imagine you are sitting in a boat on the ocean. The boat will rise and fall periodically because waves in the water lift it as they pass. Now, further imagine that the boat is in a harbor that is sheltered such that only waves from a certain direction and with a constant wavelength (or frequency) are allowed to enter. In that case, the boat will rise and fall with the frequency of the waves. Now imagine that the boat begins to sail directly into the oncoming waves. Will the frequency at which the boat rises and falls remain the same, or will it change?

If you examine the situation, you can probably convince yourself that the frequency will change. Since the boat is moving toward the waves, successive waves do not have to travel quite as far to reach the boat as they would if the boat were stationary; the boat rushes to meet them as they travel toward it. Thus, they reach the boat a bit sooner than they would, and the boat is forced up and down more often than when it was stationary. That is, as perceived by a person on the moving boat, the waves have a higher frequency than when the boat is stationary. The faster the boat moves toward the waves, the higher the frequency will be when measured on the boat.

Now consider the case where the boat moves in the same direction as the waves (but still slower than they move through the water). Each successive wave must catch the receding boat, and so it takes a little bit longer for the waves to reach the boat than it would if the boat were stationary. This causes the boat to go up and down more slowly than it otherwise would. A person on the boat would therefore measure a lower frequency for the waves. Similar to the previous case, the faster the boat moves, the bigger the effect. The faster the boat moves, the slower the frequency of the waves the sailor perceives. Of course, in the case where the boat moves at precisely the same speed as the waves, the waves can never catch the boat. In that case, the boat does not go up and down at all, and the sailor measures zero frequency for the waves.

Next, you will find a set of activities, each depicting a person in a boat and waves, but in different configurations.

Boat Wave Demonstration

A. Source of waves is stationary, observer moves.

In this case, the source of waves is far to the right of the screen, so far in fact that the waves arrive parallel to each other. The observer (the person in the boat) is able to move around the water.

In the bottom right corner, the frequency of the waves is displayed. Two numbers are given. The first is the frequency emitted by the source. The second is the frequency of the waves as measured by the observer.

Move the boat around with your mouse.

Play Activity

1.

2.

3.

B. Student discussion about a moving source.

Play Activity

C. Stationary observer, moving source.

2.

3.

4.

These examples are meant to illustrate how the frequency of any periodic source is shifted if the source and observer are in motion relative to each other. We considered waves in water, but the arguments would be similar for any other periodic signal, such as sound waves (as in Figure 4.1) or light waves. This effect was first explained by the Austrian physicist Christian Doppler in 1842 and it has come to be called the **Doppler shift** .

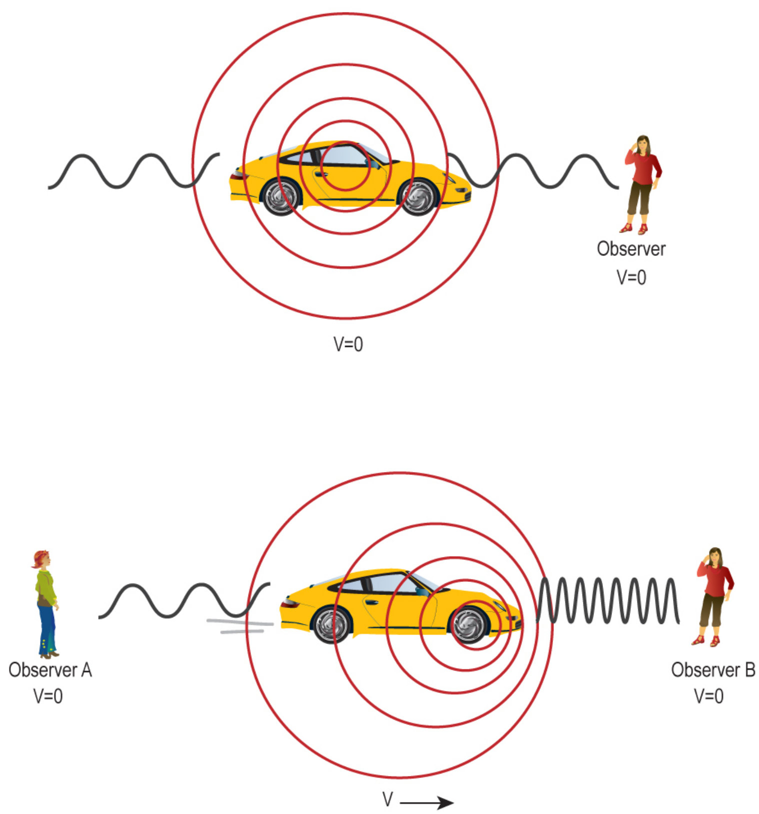


Figure 4.1: Sound waves from a car that is not moving (top) and one that is traveling forward (bottom). The moving car's sound waves are Doppler shifted so that the sound waves are shifted to lower frequencies (longer wavelengths) with respect to Observer A and shifted to higher frequencies (shorter wavelengths) with respect to Observe R B. Credit: NASA/SSU/Aurore Simonnet.

Doppler was working with sound, but the ideas also apply to light. Just as we considered successive water wave crests passing our boat, we can consider successive waves of light passing the boat. The same effect will be noted: The waves will pass by more frequently when we move toward the source, and they will arrive less frequently if we move away from the source.

A slight modification for light is that water waves and sound waves require a medium like air (or water) for the waves to travel upon. Light requires no such medium. As a result, the mathematical treatment of the Doppler shift in sound is slightly different from that for light, but the effect is qualitatively the same for both. Another difference is that light travels much faster than either water waves or sound, so the shift in the frequency of light is not generally large enough to notice without sensitive equipment.

When we observe light from a moving source, we can measure a shift in wavelength for specific absorption or emission lines and compare the observed wavelength to a reference wavelength, the one at which the lines would occur in the laboratory for an unmoving source. This reference wavelength is called the *rest wavelength.*

In the following equation, the Doppler shift (denoted by the letter z) is calculated by comparing (i.e., taking the ratio of) the difference ($\Delta\lambda$) between the observed wavelength (λ_{obs}) and the rest wavelength (λ).

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_{obs} - \lambda}{\lambda}$$

If the observed wavelength is greater than the rest wavelength, the Doppler shift will be a positive number. This indicates that the object is moving away from the observer. If the observed wavelength is longer than the rest wavelength, the shift is called "redshift." On the other hand, if the object is moving toward the observer, z will be negative, and so will the velocity. In this case, we say that the spectrum is "blueshifted." For historical reasons, the letter z is often called the "redshift" because in many astronomical examples, objects are moving away from Earth-bound observers. Also, "redshift" and "blueshift" do not mean exactly shifted to red and blue, they mean shifted to longer or shorter wavelengths, respectively. Figure 4.2 shows the spectrum of hydrogen at rest, redshifted, and blueshifted.

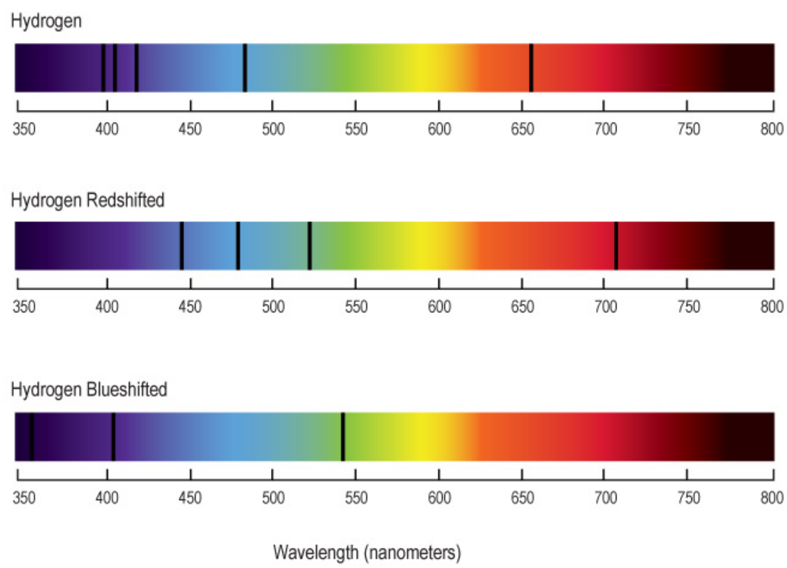


Figure 4.2: Three spectra. The spectrum of hydrogen in the visible portion of the electromagnetic spectrum. The top image shows the spectrum of hydrogen lines as seen in the lab (and therefore stationary). The second spectrum displays the hydrogen lines as seen from an object moving away from an observer. The third spectrum shows the hydrogen lines from an object moving toward an observer. Credit: NASA/SSU/Aurore Simonnet.

After we find the Doppler shift z, we can calculate the velocity v of the moving object by multiplying the Doppler shift by the speed of light c.

$$v = cz$$

If the Doppler shift z is a positive number, then the velocity v will also be positive, meaning moving away. If v is negative, the object is moving toward the observer. We can also see from this equation that the bigger the velocity the greater the redshift and vice versa. This equation is valid as long as the speed is much less than c. For greater speeds, a more complicated variant of the equation must be used, but the two are the same at speeds much lower than the speed of light.

The Doppler shift has some important and useful consequences for the interpretation of the spectra emitted by celestial bodies. These will be explored in the next set of activities.

Doppler Shift of Hydrogen

Hydrogen atoms emit a series ("signature") of spectral lines. When an object containing hydrogen is at rest relative to the observer, its spectrum has a red line (Line 1), a turquoise line (Line 2), a blue line (Line 3), and a series of purple lines. When an object is moving, this whole sequence is shifted in wavelength due to the Doppler effect.

In this activity, we will use the *Spectrum Explorer* tool to explore the motions of objects. Use the slider bar below the spectrum to Doppler shift the series of lines in order to check your predictions below.

Play Activity

1.

2.

3.

Often for quantitative work, it is more convenient to graph the intensity of light at each frequency or wavelength. Two ways of displaying the information in spectra are shown in Figure 4.3. In the next activity, you will practice converting between these two display methods.

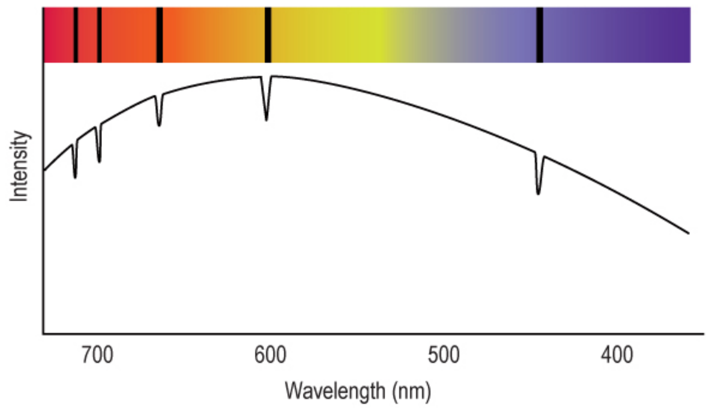


Figure 4.3: Two ways to display information in spectra. Both depict the same colors, just displayed in different ways. The graph at the bottom numerically indicates the intensity of the light at each corresponding point on the “ribbon” spectrum at the top. For example, the dips in intensity on the bottom graph correspond to the dark bars on the top graph. Credit: NASA/SSU/Aurore Simonnet.

Two Ways of Representing a Spectrum

In modern times, astronomers rarely display spectra as a colorful ribbon of light with dark and bright bands. That was how spectra were displayed when they were gathered on photographic plates. With the advent of electronic detectors, which are easy to digitize, it became much easier to display spectra as a plot of intensity vs. wavelength.

This type of display has several advantages. First, it displays the intensity in an easily readable form (on the vertical axis). Inferring the intensity of a source across a ribbon spectrum is nearly impossible to do by eye. Second, very faint lines are more easily seen in line plots because they show up as slight dips or rises in intensity in some region. In ribbon plots, they can be extremely difficult to see. Finally, the strength of lines is more easily understood at a glance because the entire width of a line can be discerned on a graph, whereas this is difficult to do with the old ribbon plots.

To help you understand how ribbon plots correspond to line plots, we have created an interactive activity that allows you to convert a ribbon plot to a line plot. Use the activity as follows:

- Click the mouse on any point in the ribbon plot shown. The intensity of the light at that point will be plotted at the appropriate wavelength on the axes below the ribbon.
- Do this for many points along the ribbon, and see how the plot spectrum is built up. Be especially careful as you plot the regions around emission or absorption lines because the intensity changes very quickly with wavelength in those regions. Astronomers use computer programs to manage this conversion procedure for the spectra they collect at telescopes.
- When you have what you believe to be enough sampled points on the ribbon plot, click the “Connect Points” button. The points will be connected, allowing you to view your completed spectrum.
- Another plot can be displayed in addition to yours by clicking the button labeled “Full Data .” This plot is generated by sampling along the entire spectrum, not using the discrete (and somewhat arbitrary) sampling you used.

In this activity, the wavelength scale is in Angstroms (Å), a unit commonly used by astronomers (1 Å = 10^{-10} m).

Play Activity

1.

2.

3.

4.

The spectrum we have used for this activity is that of an A-type star, between about 3,500 and 4,200 Å. Vega is one example of such a star. We will be displaying spectra in the line format for the remainder of the modules.

Quantitative Doppler Shift

Your goal in this activity is to determine the velocities of several galaxies based on their spectra using the Doppler shift. A small part of each galaxy's spectrum is provided, showing the Hβ (hydrogen beta) spectral line, which has a rest wavelength of 4,860 Å (1 Å = 10^{-10} m). The top graph is that of a reference galaxy that has not been shifted. The lower graph displays the observed spectrum for the selected galaxy.

A galaxy can be chosen by double-clicking on it. You should see that the Hβ line has been shifted to longer wavelengths for each of the galaxies. Double-click the galaxy image when you have finished analyzing it, to get back to the entire selection of galaxies.

Play Activity

For each galaxy:

- Find the observed wavelength , λ_{obs}, for the Hβ line by hovering over the highest point of the Hβ line and reading off the corresponding wavelength.
- Calculate the shift in wavelength, $\Delta \lambda = \lambda_{obs} - \lambda$, in the Hβ line.
- Calculate the redshift , $z = \Delta \lambda / \lambda$, for the Hβ line.
- Use this redshift to find the "velocity " in km/s using: $v = cz$. The speed of light , $c = 3 \times 10^5$ km/s.

Worked Example for Galaxy A:

- The observed wavelength measured from the spectrum is λ_{obs} = 4,950 Å.
- The rest wavelength for Hβ is 4,860 Å, so the shift is Δλ = 4,950 Å - 4,860 Å = 90 Å.
- The redshift is then given by z = Δλ/λ = 90 Å / 4,860 Å = 0.0185
- From the redshift, we can calculate the velocity of the galaxy: v = (3.0e5 km/s)(0.0185) = 5546 km/s. Since the value is positive, we know this is a redshift and the galaxy is moving away from us.

1.

2.

3.

This page titled 4.2: Measuring Motion - the Doppler Shift is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

4.3: Astronomical Objects in Motion

Learning Objectives

- You will be able to distinguish between rotation and orbit
- You will know what orbits around what
- You will know that everything is in motion in some frame of reference
- You will know that acceleration is change in motion

What Do You Think: Motions of Earth

One of the most obvious aspects of our world is that things move, or in other words, that their positions tend to change with time. This is apparent when we look out a window and see birds and insects fly past, cars drive by on a nearby roadway, or people walk across our view. We can even see clouds float across the sky. The one thing that seems as if it does not move is Earth itself. Apparently Earth and objects like buildings that are attached to it do not ever move. However, this apparently obvious truth is not the truth at all.

4.3.1: The Motions of Earth

We realize that the seeming immobility of Earth is a result of our perspective. If we could leave Earth and hover in space close to the Solar System, we would see that Earth travels in orbit around the Sun once every year. It also rotates once every 24 hours. So what seems to be solid and unmoving when we stand on it is actually moving through space while spinning like a top! Earth's rotation and the orbits of the planets around Sun are depicted in Animated Figure 4.4.

a.

Video of the Earth rotating (original link)

b.

Animated Figure 4.4 (a) Earth’s rotation, (b) Planets orbiting the Sun (original link).

Earth’s Rotational and Orbital Speeds

Worked Example:

Earth rotates on its axis once in a day. If you are standing on the equator, how fast (in km/hr and mi/hr) do you have to be moving to complete this trip?

- Find: speed v
- Given: time t, distance d

To answer the question, you must know the distance you travel and the time required.

The time, we know, is t = 24 hours.

What about the distance?

Earth is roughly spherical, and it has a radius of about 6,400 km. Since the circumference of a sphere is given by C = 2πr, just as for a circle, we can write:

$$C = 2\pi(6400\,\mathrm{km}) = 40,200\,\mathrm{km}$$

This is the distance we would travel on the equator as Earth rotates once.

- Concept: v = d/t
- Solve: v = 40,200 km / 24 hr = 1700 km/hr
- Converting to miles per hour: 1700 ~~km~~/hr × (0.62 mi /1 ~~km~~) = about 1,000 miles/hr.

Questions

1.

4.3.2: The Motions of the Sun, Stars, and Galaxies

The Sun, too, is in motion. It spins on its axis, much as Earth does, taking about a month to make one complete rotation at its equator. It also travels in a huge, roughly circular orbit that takes it once around the Galaxy in about 230 million years. This corresponds to a speed of about 500,000 mi/hr. Objects closer to the center of the Galaxy also orbit around the center, but they take less time. And objects farther than the Sun from the center take longer to go around. Our entire Galaxy is rotating. The inner parts complete their orbits in less time than the outer parts (see Figure 4.5 and Animated Figure 4.6). This motion makes the patterns of stars in the sky change very slowly over millions of years. In addition to this average motion, the stars in the Galaxy can also have random velocities in local areas. The relative speed between nearby stars is typically about 10 km/s, or 20,000 mi/hr. This seems very fast, but it would still take a long time to notice the change because the distances separating the stars from one another are so large.

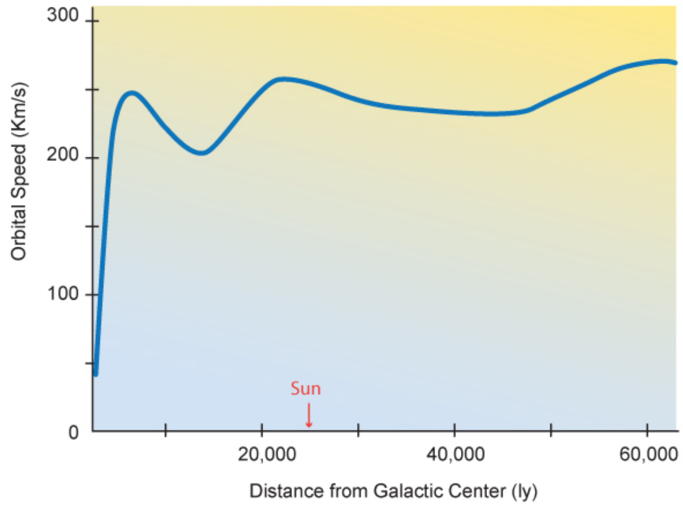


Figure 4.5: The actual measured orbital speeds of stars within a typical galaxy as a function of their distance from the center of the galaxy (in light-years). Credit: NASA/SSU/Aurore Simonnet.

Animated Figure 4.6 Galactic rotation.

The motion does not end with our Galaxy spinning. As was discovered in the early part of the 20th century, all of the other galaxies we see are moving relative to our own, and most of them are moving away from us. So, does this make our Galaxy the center of the Universe, with all motion away from that center? As we will learn in later chapters, it does not. In fact, there is no center of the Universe. But this observation does give us some important insights about motion.

4.3.3: Relative Motion

Whenever we talk about motion, we mean that some object is changing its position relative to another object. On Earth, we generally measure position, and therefore motion, relative to the ground, but this is just for the sake of convenience. As we have already pointed out, Earth itself is in motion. Since we generally want to know where things are on Earth, using Earth as a reference is just fine most of the time. However, if we liked, we could measure motion and position relative to ourselves. In fact, this is what our senses generally do.

As an example, imagine yourself sitting in a car that moves down a straight section of the highway, moving at constant speed. You will notice the landscape rush past at 60 or 70 mph. You do not see yourself move across the ground at that speed; you see the landscape fly past. If not for the little bumps in the road that jostle you in your seat, you might conclude that the landscape really is rushing past and that you are yourself stationary. You might have this impression much more strongly if you are on a train or a boat, where the ride can be much smoother than in a car.

As you drive, you also see other drivers on the road. Most of them will be moving relative to yourself. Drivers traveling in your direction seem to be moving at very low speeds, usually only a few miles per hour. Some of them seem to move forward, others seem to move backward. On the other hand, drivers in the oncoming lanes seem to be moving at 120 mph or faster, either toward or away from you.

So, which is the correct viewpoint? Are the oncoming drivers moving at 60 or 70 mph or at 120 or 140 mph? Are we moving across the ground at 60 mph or is the ground moving past us with that speed in the opposite direction?

That question has no absolute answer. It depends on our point of view. As long as we are clear about which point of view we are using, the answer we give will also be clear. We are free to measure positions and speeds using any reference point we find to be convenient. The only caveat is that, once chosen, we must stick with that single reference point for all of our measurements. If we do not, we cannot compare them in any meaningful way. So, any of the descriptions of motion we used above would be fine as long as we make clear what the motion was measured against (us or the ground). This is one form of relativity, called Galilean relativity: We are free to measure motion (and position) using any convenient point of reference, but we always require some specific reference point to make any such measurement meaningful.

Relative Motion

In this activity, a cart will travel along some tracks. There is a person on the cart who will tell you at what speed she will throw a ball.

We will define the cart's speed to be negative when it is traveling from the right side of the screen toward the left. Likewise, left to right motion is defined so the speed is positive). The ball's velocity with respect to the cart will always be positive because the person will always throw the ball toward the right side of the screen.

Play Activity

Click the "Generate" button, and you will be given the speed of the cart and the speed of the ball in meters per second.

1.

2.

3.

Click on the "Generate" button to start another round. You may want to try a few combinations of negative and positive velocities to make sure you understand the concept.

In the last section, we imagined cars moving on the freeway or trains moving along a track at a straight section of track. This type of motion approximates uniform motion, which is just a shorthand way of saying that the objects move with a constant speed and constant direction. Of course, cars and trains do speed up and slow down, and they change direction, too. But we had imagined what we would see traveling at a constant speed and direction. If either of these conditions is not met, then the motion is not uniform, and its character changes fundamentally.

You might have noticed this effect yourself. Riding in a car feels very different when traveling straight down the road at a constant speed than it does when the car brakes rapidly or goes around a sharp curve. In these latter cases, we have strong departures from uniform motion, and we notice this departure because we feel as though we are thrust either forward or back, or side to side. Such motion is called accelerated. Note that acceleration for a physicist means that an object speeds up, slows down, or changes direction. This is an example of a word that has a slightly different, though still related, meaning in science than it does in our everyday language. Science is full of such words because scientists often require more precise meanings for things than we do in day-to-day conversation.

It turns out that uniform motion is an idealized view of motion. None of the moving objects we have mentioned are really in uniform motion at all. For instance, Earth is orbiting the Sun in a circular path, so it is constantly changing direction. It must be constantly accelerated. Earth is also spinning on its axis, so everything in and upon Earth must also be going around, being accelerated. These accelerations are small, and so we are not generally aware of them. This is why it took people a long time to realize that Earth moves at all. However, any acceleration can in principle be measured. Furthermore, the rate at which direction or speed changes can be measured in absolute terms, without reference to an arbitrary point of reference. In this way, accelerated motion differs profoundly from uniform motion. We will have more to say about accelerated motion when we discuss gravity in later modules.

This page titled 4.3: Astronomical Objects in Motion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

4.4: Wrapping It Up 4 - The Andromeda Shift

4.4.1: Part I. Lookback Time to Andromeda

1.

2.

Play Activity

4.4.2: Part II. Measuring Andromeda's Motion

In this part of the activity, you will make measurements of Andromeda's motion using the Doppler shift. From those measurements you will determine Andromeda's rotation speed and systemic velocity.

1. First you will look at spectra taken from opposite sides of the Andromeda Galaxy (M31), and compare them to a spectrum taken *at rest* to determine which way the galaxy is spinning. By observing the Doppler shift (either redshift or blueshift) of the spectra at either end, you can tell in which direction M31 is spinning. The spectrum being observed is a very small sliver of the galaxy's actual spectrum. It is limited to the region around one of the most prominent lines of galaxy spectra, the Hß line.

- Click on either side of the galaxy to observe the spectrum at that location. Compare this spectrum to the reference spectrum above to determine if this side is redshifted or blueshifted. Redshifted means the material is moving away from you, and blueshifted means it is coming toward you. After measuring one side, click on the opposite side to determine if it is redshifted or blueshifted. Also pay attention to *how much* each side is shifted.
- Click either the "clockwise" and "counter-clockwise" buttons, depending on which way you think Andromeda is spinning.
- You will be informed whether or not you are correct. Resolve any discrepancies with your answer before moving on.

2.

3.

4.

4.4.3: **Part III: Time to collide with the Milky Way**

1.

This page titled 4.4: Wrapping It Up 4 - The Andromeda Shift is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

4.5: Mission Report 4 - The Andromeda Shift

A.

B.

C.

https://phys.libretexts.org/@go/page/31234

D. Questions to be graded for accuracy

Below is a small part of the spectrum for galaxy NGC 6643, showing the Hβ (hydrogen beta) spectral line, with a rest wavelength 4860 Angstroms (1 Angstrom = 10^{-10} m). The top graph is that of a reference galaxy that has not been shifted. The lower graph displays the spectrum for the galaxy NGC 6643.

1

2.

3.

4.

This page titled 4.5: Mission Report 4 - The Andromeda Shift is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

5: Moving Through Time

Chapter 5 explores how objects in the Universe change over time. It examines how the ages of objects can be determined: first through radiometric dating and next through the lifetimes of stars. The evolution of stars, from their formation to their eventual "death," is discussed. The evolution of the Universe as a whole, from its beginning through today, concludes the chapter.

Thumbnail: Artist's impression of a Type Ia supernova, as revealed by spectro-polarimetry observations (CC BY 4.0; ESO via Wikipedia)

This page titled 5: Moving Through Time is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

5.0: Moving Through Time Introduction

It is often said that telescopes are time machines, showing us views of the Universe from earlier times as we study more distant objects. Yet the objects themselves also age.

Stars and galaxies are born, evolve and die, but with lifetimes that are so long, we can only study snapshots that (in almost all cases) seem unchanging. In this chapter, we examine several techniques for determining the ages and lifetimes of stars and then see how the Universe and its constituents evolve through cosmic time.

Video Transcript

The Life Cycle of Stars

Away out there in space there's huge clouds of dust and gas and if one of those clouds of dust and gas is massive enough, it's own gravity causes it to start to collapse so it folds in on itself towards the center of that cloud. t gets denser and denser and hotter and hotter and eventually, the particles of that the gas and the dust are made of are brought so close together that they start to stick together. They start to fuse. That's the energy source of a star. The star switches on and begins to shine. Inside every newborn star, hydrogen atoms are fused together to make helium. This process is called fusion, and it creates the energy of every star. What happens to a star during the rest of its life depends on how massive it is at its birth.

A star like the sun is in a delicate balance between gravity which wants to make the star collapse in on itself and the pressure that pushes outward that comes from the energy produced in these fusion reactions at its core. At some point in the future, the hydrogen runs out. At that point, the core of the star will start to collapse in on itself under its own weight it gets denser, it gets hotter until the point where you can actually start to use the helium atoms themselves as the fuel for the fusion pushing helium atoms together, making carbon and oxygen the next heavier elements on the periodic table. As the star begins to fuse helium, it creates more energy and that causes the outer layers of the star to expand.

One day, our sun will grow so large it will swallow up the inner planets of the solar system out as far as the Earth! It will become a red giant. For the sun, this will be the beginning of the end what happens is that the outer layers of the star get farther and farther from the middle. The force of gravity that they feel is getting weaker, and, actually, the star loses hold of its outer atmosphere. Its outer atmosphere drifts off out into space. It expands to become a planetary nebula. And they're some of the most beautiful objects in the universe. Once the outer layers have drifted away, all that is left of the star is its core. A white dwarf star is the dead, remnant core of a star like the sun at the end of its life It's something that might weight as much as 1/2 the mass of the Sun but it's only about the size of the Earth, so it's an incredibly dense object. It's dead, there's no nuclear fusion going on any more, it's incredibly hot but then over millions of years, it will gradually cool down to become a black dwarf.

Some stars, however, are much more massive than the sun and they lead very different lives. They are able to fuse heavier and heavier elements inside their core the star gets bigger and bigger some grow up to 1000 times the size of our sun until it has fused elements all the way up to iron and once we've formed an iron core, there's no more energy that can be got from fusion that core collapses the rest of the star starts to collapse in after it but then it bounces off. There's a huge shock wave. And in just a second: BANG! The outer parts of the star are blasted off in to space in a huge super-nova explosion! These super-nova *explosions are so powerful that when one of these stars explodes it can actually outshine the whole galaxy of which is part (a galaxy of maybe a 100,000 million stars).*

For these super giant stars all that is left is a super dense core known as a neutron star an object that can have a mass greater than our sun but be less than 20 kilometers across but for the most massive stars of all we think, that when the core collapses the gravity is so strong, it becomes a black hole from which not even light can escape so stars are actually the places in the universe where the elements are created.

After the Big Bang, our universe contained only hydrogen and helium. All the other, heavier elements were therefore fused inside stars. The amazing thing is that virtually everything you see around you was made inside a star billions of years ago before the sun and planets were formed and when that star died and blasted its guts out into space that formed the raw materials from which our sun the planet earth and indeed ourselves were made and, ultimately, that's one of the major reasons I think understanding stars is crucial, because its actually telling us where we came from.

This page titled 5.0: Moving Through Time Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

5.1: Measuring the Ages of Objects - Radiometric Dating

What Do You think: Geologic Dating

There are a variety of independent methods for measuring ages of different objects in the Universe, and these ages provide additional data to incorporate into our models of the Universe. Hopefully, this will give you an understanding of how we know the ages of various objects, so that they do not feel like random facts imposed from authority.

5.1.1: Radioactivity

At the turn of the 19th century, research into the phenomenon of radioactivity not only helped to radically transform our understanding of matter, but it allowed us to determine accurate ages for many materials for the first time.

Radioactivity is a process by which one atom undergoes a change in its nucleus, becoming a different kind of atom as it does so. These changes occur spontaneously, though similar changes can be artificially induced under the right conditions. The processes by which one type of atom spontaneously turns into another type are called nuclear decays.

Radioactivity was first reported in 1896 by Henri Becquerel (1852–1908), who noticed that samples containing uranium salts would expose photographic film, even if the film remained closed away in a box. In 1899, Ernest Rutherford (1871–1937) showed that the activity comes in two types, which he called alpha radiation and beta radiation. They were distinguished by their ability to penetrate through barriers. The alpha-type rays were more easily stopped than rays of the beta-type. A third type of radioactivity, termed gamma radiation, was discovered in 1900 by Paul Villard (1860–1934). A year later, Rutherford and his colleague, Frederick Soddy (1877–1956), showed that these radiation- producing processes are accompanied by the conversion of one type of atom into another type. Also in 1900, Becquerel showed that the beta rays emitted in the beta process are actually electrons. It took Rutherford and Thomas Royds (1884–1955) another six years to discover what the alpha particles are. They are discussed below.

The phenomenon of radioactivity was most thoroughly studied by the Polish/French chemist and physicist Marie Curie (1867–1934). Curie coined the term "radioactive" to describe these nuclei - a term that is unfortunate in that none of the so-called radioactive processes is associated with the emission of radio waves. Curie discovered radioactivity in thorium, and also discovered, along with her husband Pierre Curie (1859–1906), two new radioactive elements, polonium and radium.

It turns out that most naturally occurring atoms are stable; they do not undergo spontaneous decays. Only a small number of atoms are radioactive; this is because, over time, the unstable nuclei decay into stable ones. Nonetheless, radioactive atoms have played a large part in furthering our understanding of the atomic nucleus.

All chemical elements come in several different forms, known as isotopes. These have nuclei with slightly different masses due to a different number of neutrons. For instance, hydrogen, the simplest element, has three isotopes. The first, written ^1H, has a nucleus containing just a single proton. The nucleus of the second form of hydrogen (^2H) contains one proton and one neutron. This form of hydrogen is called deuterium. Finally, the third form of hydrogen (^3H) is called tritium. Its nucleus contains a proton and two neutrons. Each form is denoted by the chemical symbol for hydrogen, with a superscript indicating the total number of protons plus neutrons. What makes each of these isotopes hydrogen is the single proton. Coupled with an electron it forms the hydrogen atom, as distinguished from the hydrogen nucleus. It is the electron that gives hydrogen its chemical properties.

Since it is the electrons that determine the chemical properties of atoms, each atomic isotope is chemically indistinguishable. For example, each isotope of hydrogen has a single electron and forms the same chemical compounds and undergoes the same chemical reactions. However, they do behave slightly differently in some respects. First, because they have different masses: Deuterium has twice the mass of ^1H and tritium has three times the mass. So, as an example, water with deuterium is slightly heavier than water with regular hydrogen. Deuterium's extra weight makes it somewhat less likely to convert into vapor when water boils or evaporates.

Despite the chemical similarities of different isotopes of hydrogen, they behave very differently in nuclear reactions. This is especially true of tritium, which is radioactive. If we wait around for about 12 years, a tritium nucleus has about a 50% chance of turning into a ^3He nucleus. This is a helium nucleus with two protons and one neutron. It is not guaranteed to do so, but there are even odds. If it does not decay in the first 12 years, then it has a 50% chance to decay in the next 12, and so on. In this decay, we would call tritium the parent nucleus and ^3He would be called the daughter nucleus. Both normal hydrogen nuclei (protons) and deuterium nuclei are stable. They will not decay no matter how long we wait (at least, as far as we know), so tritium nuclei are strikingly different in this respect.

Each isotope of hydrogen has a single proton (that is what makes an isotope hydrogen), but each has a different numbers of neutrons. Similarly, helium has several isotopes. The most common has two neutrons to go along with its two protons and is called ^4He. It is the alpha particle of alpha decay, as originally shown by Rutherford and Royds in 1906. Tritium decays into the other stable form of helium, ^3He. There are four other isotopes of helium, but they are all unstable—in other words, radioactive.

In addition to hydrogen and helium, there are 90 additional kinds of atoms that occur naturally. They are listed in the Periodic Table of the Elements. Each one differs in the number of protons in its nucleus, and thus also in the number of electrons orbiting it nucleus. All of these atomic species have multiple isotopes, the nuclei of which differ only in their numbers of neutrons. Generally, most of the isotopes are unstable, but many atoms have more than one stable isotope. Some atomic species, the heaviest ones, have no stable isotopes at all.

It is not possible to predict exactly when a particular unstable nucleus will undergo a nuclear decay. This makes radioactivity a strictly non-Classical process; it can only be understood under the framework of quantum mechanics, just as is true for the distinct patterns of spectral lines emitted by atoms. We can know the probability that a nucleus will undergo a decay in a certain amount of time (for example, tritium has a 50% chance to decay in 12 years), but we cannot say exactly when that decay will occur for an individual nucleus.

The amount of time for half of a sample to decay is known as the isotope's half-life. Again, since radioactive decay is a statistical process, it is not possible to predict exactly which atoms in the sample will decay and when, but we do know that approximately half of the sample will be gone in one half-life. Half of the remainder will decay after another half-life passes, and so on. The greater the number of atoms (N) in the sample, the better our accuracy, because the uncertainty in the number N equals the square root of N. For example, for a sample where we expected 100 atoms to decay after one half-life, the number actually decaying will typically be $100 \pm \sqrt{100} = 100 \pm 10$, or anywhere between 90 and 110. The percent uncertainty would be $\sqrt{N}/N = 10/100 = 10\%$. In a sample with 1000 atoms, the uncertainty will be $\sqrt{1000} = 32$, and we expect the number of decays to be roughly between 968 and 1032. This corresponds to uncertainty of 32/1000, or about 3%. In real samples used in the lab, typically the number of atoms is much, much greater than 100 (or 1000!), and so the percent uncertainty is correspondingly much lower.

If we have a sample of radioactive atoms in a lab, and if we measure the number of decay products it emits in a period of time, we will notice that it slowly drops the longer we watch the sample. And even though it is not possible to predict when any particular atom will decay, we can predict quite accurately how the number of decays in a large number of atoms will decrease over time.

As an example, imagine we had sample of ^{131}I, a form of iodine used to treat certain types of cancers. If we measure the decay rate (by the detection of decay products, beta rays in this case), we would notice that after about 8 days the number of decays would have decreased by half. We could infer that our original sample would have only half of the original number of radioactive atoms of

^{131}I. So if we started with 1000 ^{131}I atoms, after 8 days only 500 would be left. The rest of the iodine would have been converted to an isotope of the element xenon, ^{131}Xe. After another 8 days, half the remaining iodine would have converted in turn, leaving only 250 in the sample. The other 250 would again have converted to xenon, and this pattern would repeat every eight days - see below for further discussion. In this example, iodine is called the parent nucleus, and xenon is the daughter.

In the decay process described above, the iodine nucleus starts out with 53 protons and 65 neutrons. In the decay process, one of the neutrons converts to a proton, an electron, and an antineutrino. The latter two are ejected from the nucleus, which is then left with 54 protons and 64 neutrons. The process is an example of beta decay because the radiation emitted includes an electron. In some nuclei, positrons are emitted instead of electrons, and neutrinos instead of antineutrinos. Positrons are the antimatter particle for electrons. They are identical to electrons but with positive electric charge rather than negative. Neutrinos and antineutrinos are also a matter/antimatter pair. Either way, the process is called **beta decay** (beta-plus or beta-minus, depending on whether a positron or electron is emitted).

A shorthand way of describing the decay of iodine above is to write it as an equation. These are similar to the chemical equations you might have learned in chemistry class, except here we have a nuclear reaction instead of a chemical reaction.

$$^{131}\mathrm{I} \longrightarrow {}^{131}\mathrm{Xe} + \mathrm{e}^{-} + \bar{\nu}_{e^{-}} \tag{5.1.1}$$

Equation 5.1.1 says that in the decay of iodine with atomic mass 131, the nucleus converts to xenon of atomic mass 131. In addition, an electron (e^-) and an electron **antineutrino** $\bar{\nu}_{e^{-}}$ (the bar over the neutrino symbol indicates it is an antineutrino) are emitted. Of course, what is really happening is that a single neutron in the nucleus of the iodine is converting to a proton. Since the proton and neutron have essentially the same atomic mass, and the electron and neutrino are almost massless by comparison, the atomic mass of the entire nucleus remains nearly constant in this process, with the daughter being only slightly lighter than the parent. We could write down the decay of a neutron *(n)* into a proton *(p)* in a manner similar to the equation above. This is the process occurring in any nucleus that decays by beta decay:

$$\mathrm{n} \longrightarrow \mathrm{p}^{+} + \mathrm{e}^{-} + \bar{\nu}_{\mathrm{e}^{-}} \tag{5.1.2}$$

Xenon-131 is a stable nucleus, so it remains in our sample unchanged. As discussed above, if we continue for another 8 days, we will notice that the activity has been reduced further by half, to only one quarter of its original amount. Our sample will then contain only 250 atoms of iodine. The remaining 750 atoms will be xenon. In the next 8 days, the activity will reduce by half again. We will have only 125 atoms of iodine, and the remaining 875 atoms will be xenon. This continues, with half of the ^{131}I converting to ^{131}Xe every 8 days, until all the radioactive iodine has been converted to stable xenon. The total number of ^{131}I and ^{131}Xe nuclei (their sum) is constant. This is illustrated graphically in Figure 5.1.

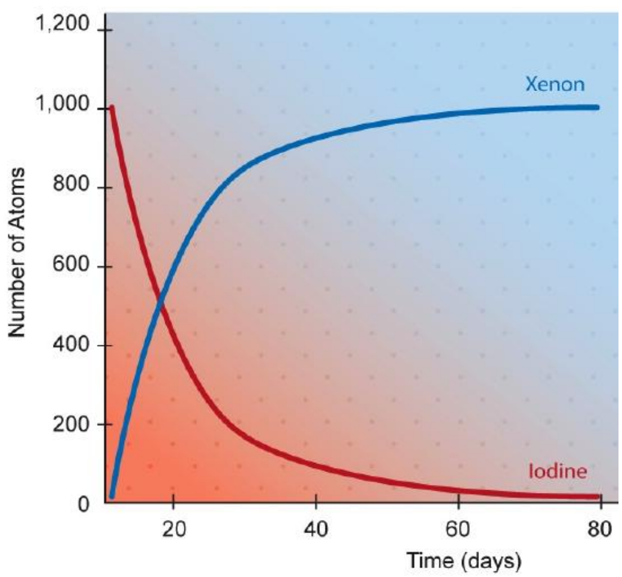


Figure 5.1: The exponential curves show that as ^{131}I decays, the amount of ^{131}Xe increases. The half-life of ^{131}I is about 8 days. Credit: NASA/SSU/Aurore Simonnet.

RADIOACTIVE DECAY

In this activity, you will observe particles that undergo either alpha or beta decay, and you will have to determine which type of decay it is. Make sure you do enough examples to be comfortable identifying both alpha and beta decay.

When you begin the activity, you will see a linear table of elements, with a label on the highlighted element. The legend indicates that protons are represented by red particles, neutrons are gray, and electrons are smaller blue particles. Neutrinos are not shown.

- Click on the “next decay” button at the top of the application and a random element will be selected. Watch closely as the particle will soon decay into another element, and you will need to decide if an alpha or beta decay occurred. In the upper right-hand corner of the activity, it will display the original element’s name and the name of the element into which it decayed.
- Once you have decided on the type of decay you observed, click on your choice between the two buttons on the left, “alpha decay” or “beta decay.”
- Click “next decay” again to view another element’s decay.

Play Activity

5.1.2: Determining Ages

This property of radioactive atoms makes them extremely useful as clocks. If we have a sample of material that contains radioactive elements, then we can use the relative amounts of the parent and daughter nuclei to deduce the number of half-lives that have passed since the sample was formed. Then, since we typically know the half-life in years, we can compute the age of the sample in years. This is how radiometric dating is done.

You may have heard of carbon dating, a method that uses the ratio of ^{14}C to ^{12}C to date organic samples, or items that were once-living things. For example, wood, charcoal, bones and shells can all be dated using the method of carbon dating. The ratio of $^{14}C/^{12}C$ (by number of nuclei) in CO_2 molecules in Earth’s atmosphere is 1.3×10^{-12}, and all living things have this same ratio because they constantly exchange CO_2 with their surroundings. When an organism dies, this constant exchange ceases, and the $^{14}C/^{12}C$ ratio begins to decrease due to decay of ^{14}C. Since the half-life of ^{14}C is 5,730 years, it is a useful element to use to date objects that are between 1,000 and 25,000 years old. We use other radioactive elements to date older or younger objects or objects, such as rocks, that are not made of organic material.

In real samples, the process is somewhat more complicated than we have described. Real samples can be disturbed. Parts can be broken off, for example, or some atoms might evaporate or dissolve away. Or some amount of the daughter or parent species might be introduced by an outside agent. Furthermore, it is likely that some amount of the daughter nuclei is present in the material when it forms. The presence of this extra material can contribute to a spurious result for the age if it depends on just a single isotope.

In practice, a substance might contain several different kinds of radioactive elements, since essentially all atoms have some isotopes that are radioactive. It might also contain different minerals, each with different chemical properties. Each of these atoms or minerals can be analyzed separately, and an age computed from each of them. These ages will generally be close to each other, but not precisely the same. So some weighting of the individual ages based on their known chemical properties and the environment in which the sample was found can be used to determine an estimate of the true age, or to give a plausible range of ages. Uncertainties can be less than a percent or so (based upon the spread in ages from different parts of the sample). In certain instances they can be larger as well.

Some examples of common parent-daughter species are shown in the table below. The table also lists half-lives and typical ages that can be determined for each species. Note that the uranium decay chains, which are used in tandem to determine ages, are complex and involve intermediate unstable nuclei with much shorter half-lives than the primary one shown.

Table 5.1

PARENT	DAUGHTER	HALF-LIFE (YR)	PROCESS	INTERMEDIATES	AGE (YR)
Uranium (^{235}U)	Lead (^{207}Pb)	700 million	Alpha and beta	Th-Pa-Ac-Fr-Ra-At-Rn-Bi-Tl-Po	Millions to billions
Uranium (^{238}U)	Lead (^{206}Pb)	4.5 billion	Alpha and beta	Th-Pa-Ac-Fr-Ra-At-Rn-Bi-Tl-Po	Millions to billions
Potassium (^{40}K)	Argon (^{40}Ar)	1.3 billion	Beta plus (electron capture)	-	Thousands to billions
Carbon (^{14}C)	Nitrogen (^{14}N)	5,730	Beta minus	-	Thousands

Radiometric dating has been used to find absolute ages for rocks on Earth’s surface. For instance, many of Earth’s oldest rocks have been found in Northeastern Canada. They have ages of around 3.8 to 4.3 billion years. Radiometric dating has been used to assign ages to other rock layers in the geologic record, as shown in Figure 5.2. Rock samples from the surface of the Moon have also undergone radiometric analysis, giving ages for the lunar highlands (the light parts of the Moon) of about 4.4 billion years. The lunar maria (the dark areas) are younger, with ages of about 3.8 billion years. In addition, some of the oldest meteorite fragments have also been analyzed, with their ages being about 4.6 billion years. All of these ages point to the origin of the Solar System being slightly more than 4.5 billion years ago.

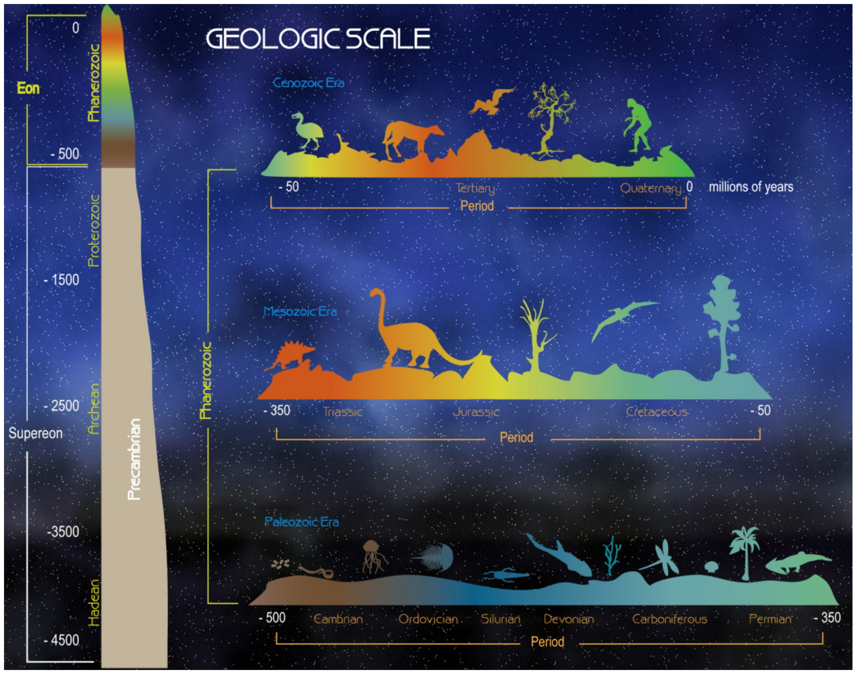


Figure 5.2: The names of the geologic epochs and their corresponding ages in millions of years. The spike-scale on the left shows the major epochs from Earth's beginning to present day, while the colorful Phanerozoic section on the right of the graph corresponds to the tip of the spike-scale. Credit: NASA/SSU/Aurore Simonnet.

We can also place limits on the duration of the Solar System's formation using radiometric methods. Primitive meteorites contain inordinate amounts of the isotope ^{26}Mg; the more common form of magnesium, about 70% of the total, is ^{24}Mg. ^{26}Mg is formed from the decay of ^{26}Al via beta-plus decay. ^{26}Al has a half-life of only about 700,000 years, so the meteorites containing ^{26}Mg must have formed, at most, a few million years after the formation of the ^{26}Al, presumably in a supernova that predated the formation of the Solar System. This indicates that at least some of the building blocks of the Sun and planets were assembled fairly quickly after being synthesized in an earlier, now long-dead star.

Half-life Activity

In this activity, you are given a set of 400 red atoms. These atoms have a half-life of 5 seconds, and they decay into blue atoms.

Play Activity

1. Click the "start" button, and you will see the atoms decay over one half-life. At the end of this cycle, the counter will let you know the number of blue and red atoms that are now in the sample, as well as the percentage of red atoms remaining. There will still be a total of 400 atoms.

2. Predict: about how many red atoms will be present after another (second) half-life.

4. Predict: about how many blue atoms will be present after a third half-life is complete?

5. Click the “start” button to observe the third half-life. How did your predictions compare with your observations? Reconcile any discrepancies between your predictions and your observations and resolve them here.

Radiometric Dating

1. Can carbon dating be used to measure the age of a dinosaur? Explain.

2. A 10.0 gram sample of ^{226}Ra, which has a half-life of 1,660 years, is left to decay. How much of the original sample is left after one half-life?

3. An archeologist finds some bones in an ancient burial site. The ^{14}C content is 1/16 of that of living humans. How long ago did the ancient people live?

Explain your reasoning:

This page titled 5.1: Measuring the Ages of Objects - Radiometric Dating is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

5.2: Measuring Ages - Lifetimes of Stars

? What Do You Think: Burning and Stars

Advances in the understanding of radioactivity at the beginning of the 20th century enabled scientists to calculate the age of Earth accurately. However, the ages derived presented a problem to astronomers: they were not consistent with estimates of the age of the Sun.

Several decades earlier, the physicists William Thomson (Lord Kelvin) and Hermann von Helmholtz had attempted to estimate the Sun's age. They made the assumption that the gravitational potential energy of the Sun was the source of energy that allowed it to shine. According to their model, as the Sun radiates energy from its surface, it should slowly contract, basically falling inward due to its own gravity. The contraction would provide the energy needed to keep the Sun hot and allow it to shine. (This is similar to the source of energy we tap into when we allow water to fall over a dam and then drive a turbine in an electrical generator.) However, the model gave a lifetime that was younger than the age of Earth, as detailed in Going Further 5.2: The Gravitational Lifetime of the Sun.

Another possible power source for the Sun that people once considered was combustion. Could it be that the Sun is just a big burning object, like a big piece of coal in the sky? That is what people did think a long time ago. We can put the idea to a test and see if it works, as you will do yourself in the activity *The Chemical Lifetime of the Sun*. Combustion is a type of chemical reaction, a process of combining a substance with oxygen in a way that produces a lot of energy. The typical energy in chemical bonds is about an electron-volt, or 1 eV (= 1.6×10^{-19} joules in SI units). We are already familiar with these energy units from our explorations of light in the previous chapters. Even if we could assume that oxygen was available in the Sun to run the combustion reactions, the lifetime of the Sun derived from this method is even shorter than that of the gravitational method.

So, if neither gravity nor chemical burning is the source of energy for the Sun to shine, what is? That was a question that could not be answered by Kelvin or Helmholtz or by any other physicist or astronomer of the time. New physics was needed before a satisfactory understanding of the Sun could be found. By the start of the 20th century, scientists realized that some other, unknown form of energy must be powering the Sun. We have briefly touched on it: the physics of the atomic nucleus.

5.2.1: Calculating the Lifetime of the Sun

We can calculate the lifetime of the Sun (or any star) based on an analysis of its energetics. We determine how much total energy it has available, and then we compare that to the rate at which it is using its energy up. Mathematically, this can be expressed by the following equation that relates its power output, or luminsity (L), to its energy production.

$$L = \frac{E}{t}$$

Luminosity (L), or power, in measured in watts and energy (E) in joules. The time (t) is in seconds. Recall that 1 watt = 1 joule/second.

If we rearrange the equation, we get:

$$t = \frac{E}{L}$$

This means that the lifetime of the star depends on the total amount of energy (E) and the rate at which it is using that energy, which is the luminosity (L). If the luminosity is *bigger*, the lifetime will be *shorter*. An analogy for this would be driving your car until it runs out of gas: If you know how much gas is in the tank, and you know the rate at which your car uses gas, you can figure out when your car will run out of gas and stop moving.

To give you a better feel for the units of luminosity and energy, a typical compact fluorescent light bulb uses about 15 joules every second (for a 15-watt bulb), and a typical household might use about a thousand or so joules each second. A typical incandescent bulb is 60 watts. Keep these numbers in mind to compare to the luminosity of stars as we discuss them below.

The amount of energy that the Sun is emitting can be pretty easily measured at Earth's surface. From that, we can deduce its total luminosity. Measurements have shown that there is about 1.4 kW of solar energy striking each square meter of Earth's surface (on the daylight side). This value is an average and varies with cloud cover, time of day, latitude, and things like that. Nonetheless, if we had a square meter of material and held it facing the Sun at the distance of Earth's orbit, it would have about 1.4 kW of solar radiation striking it. Of course, the Sun is not sending energy only through the 1 square meter that we measure. It is sending that amount of energy through every square meter on a sphere that completely surrounds it and has a radius equal to Earth's orbital radius, 1AU. To figure out how many square meters there are surrounding the Sun on a sphere with a radius the size of Earth's orbit, we use geometry. We know that the area of a sphere is given by:

$$A = 4\pi R^2$$

Since the radius of Earth's orbit is about 150 million km, or 1.5×10^{11} m, the area will be: $4\pi(1.5 \times 10^{11}\text{ m})^2 = 2.8 \times 10^{23}\text{ m}^2$. Since each of these square meters has about 1.4 kJ of energy passing through it every second, the Sun must be emitting about $(2.8 \times 10^{23}\text{ m}^2) \times (1400\text{ W/m}^2) = 4 \times 10^{26}$ W from its entire surface (i.e, in all directions). This is known as the solar luminosity—in other words, the Sun produces 4×10^{26} J each second that it shines.

So, is a particular energy source enough to power the Sun for billions of years? We know how much energy the Sun emits every second. We also now know how much energy the Sun contains, both in terms of chemical and gravitational energy. With these numbers, we can estimate how many seconds the Sun can continue to shine.

GOING FURTHER 5.2: THE GRAVITATIONAL LIFETIME OF THE SUN

GOING FURTHER 5.3: WHY THE GRAVITATIONAL LIFETIME CANNOT BE FIXED

The Chemical Lifetime of the Sun

Long ago, people thought the Sun was just a big burning object on fire, like a big piece of coal in the sky. We can put that idea to a physics test and see if it works.

Combustion is a type of oxidation, a process of combining a substance with oxygen. There are other ways to oxidize materials, but combustion produces a lot of energy. The typical energy in chemical bonds is about an electron-volt, or 1 eV. We are already familiar with these energy units from our discussion of light in the previous chapters. However, electron-volts are not convenient for our purposes here. We should use joules; 1 eV = 1.602×10^{-19} J.

Now assume that we have oxygen available in the Sun to run the combustion reactions. We now know that is not true, but certainly no one knew that prior to the turn of the 20th century. In the activity below, we will assume that the Sun is entirely made of oxygen, and when one oxygen atom combusts, an electron-volt of energy is released.

1. The atomic mass of oxygen is 16, and the mass of a proton (or a neutron) is about 1.67×10^{-27} kg. What is the mass of oxygen in kg?

kg

Show your work:

2. Assuming that all the mass of the Sun is in the form of oxygen atoms, and that the mass of the Sun is 2×10^{30} kg, use your answer from part 1 to figure out how many atoms of oxygen could be in the Sun.

atoms

Show your work:

3. Assuming that burning 1 oxygen atom releases 1 eV of energy, how much total energy would be available from combustion of all the atoms of oxygen in the Sun (from question 2)? Express your answer in joules.

joules

Show your work:

4. The Sun's luminosity is 4×10^{26} W. Use this luminosity to compute the time it will take for all of the available combustion energy (from question 3) to be used up. Express your answer in years.

years Show your work:

5. From this estimate of the chemical combustion lifetime of the Sun (from question 4), can combustion be the source of the Sun's energy?

Yes

No

Explain your reasoning:

5.2.2: The Nuclear Lifetime of the Sun

We have seen how certain atomic nuclei undergo decays that convert one nucleus to a different one. It turns out to be possible to make those reactions run the other direction. Of course, the right physical conditions are required for the inverse decays to happen. Still other nuclear processes are possible, ones involving only stable nuclei. All of these processes go under the general name of nuclear reactions, and they play a vital role in the Sun and other stars.

There were many scientific advances between the time the study of radioactivity was undertaken and a workable theory of solar energy production was put forward. One of the most important of these was the realization that the Sun is composed almost entirely of hydrogen and helium. Small amounts of heavier elements are also present, but together they make up only about 2% of the mass of the Sun. This discovery was made by Cecelia Payne-Gaposchkin and presented as her doctoral dissertation in astrophysics at Radcliffe College, now part of Harvard University, in 1925. Payne-Gaposchkin used the then-nascent field of quantum mechanics to reach what was a very startling conclusion for the time. Her thesis is still considered by many to be the greatest Ph.D. dissertation ever done in the field of astrophysics.

Another important discovery bearing on solar energy production had been made 20 years earlier. Published in 1905 under the title *Zur Elektrodynamik bewegter Körper* (On *the Electrodynamics of Moving Bodies*) by Albert Einstein, at the time a clerk in the patent office of Bern, Switzerland, the paper made the astounding claim (one among several) that mass and energy are interconvertible. Many other advances were relevant as well, but we will explore how these two ideas impacted the understanding of the workings of the Sun and other stars.

As a starting point, consider the alpha particle, ^{4}He. We introduced this particle earlier during our discussion of radioactivity. The alpha particle is the most common isotope of helium. It is composed of two protons and two neutrons. The masses of these particles are shown in Table 5.2.

Table 5.2

PARTICLE	PROTON	NEUTRON	2 PROTONS + 2 NEUTRONS	ALPHA (MEASURED)	DIFFERENCE
Mass (10^{-27}kg)	1.672621637	1.674927211	6.695097696	6.64465620	0.050441496

If we add up the masses of the protons and the neutrons that make up the alpha particle, we find that the alpha particle has a mass lower than its constituent parts. This addition has been done in the table, with the difference being given in the last column. The difference is small, only about 0.7%. Still, how can it be that the protons and neutrons have lower mass when collected into an alpha particle than they do when out on their own? This difference is known as the binding energy, and it relates directly to mass-energy interconversion. Some of the mass of the protons and neutrons is converted to energy in creating an alpha particle, and that energy is released to the environment when the alpha particle is formed. The amount of energy is given by Einstein's famous mass-energy equivalence.

$$E = mc^2$$

where E = energy, m = mass, and c = the speed of light.

This energy release presents a possible source of energy for the Sun. As was shown by Cecelia Payne-Gaposchkin, the vast majority of the Sun is hydrogen; most of the rest is helium. Perhaps the Sun derives its energy by converting the hydrogen to helium and extracting energy in the process. The general nuclear reaction would be the conversion, in net, of four protons to an alpha particle.

$$4\,\mathrm{p} \rightarrow {}^4\mathrm{He}$$

This type of nuclear reaction, where lighter elements combine to form heavier elements, is known as **nuclear fusion**. These are distinct from reactions in which heavier elements split apart into lighter ones in nuclear fission.

GOING FURTHER 5.4: NUCLEAR REACTIONS IN THE SUN AND OTHER STARS

Energy of Hydrogen Fusion

1. Using the information in Table 5.2, find the difference in mass (in kg) between the helium nucleus and the four protons.

kg Show your work:

2. How much energy does this mass correspond to in J? Hints: use Einstein's famous equation E=mc^2. The speed of light is about 3 x 10^8 m/s.

J

3. How much energy does this correspond to in eV?

eV

Show your work:

In the previous activity, you should have found that energy could be extracted if we could devise a way to change four protons into a helium nucleus. We will save the discussion of how that might occur for later. For now, you will find the nuclear energy lifetime of the Sun and see if it is consistent with the Sun's lifetime of at least 5 billion years or so - the approximate age of Earth.

Nuclear Energy Lifetime for the Sun

In this activity, you will estimate the amount of time that the Sun could exist if it were being powered by the conversion of hydrogen to helium through nuclear fusion reactions. In the nuclear reactions we are considering, one kind of atom (hydrogen) is being converted into another (helium). Follow these steps to work through the activity:

1. Recall that the mass of the Sun is 2×10^{30} kg. Assuming that all the mass in the Sun is hydrogen, and remembering that hydrogen is just a single proton (of mass 1.67×10^{-27} kg), how many hydrogen atoms are in the Sun?

atoms

2. Recall that it takes 4 protons (hydrogen atoms) for each nuclear reaction. How many reactions will the Sun be able to produce?

reactions

3. Given the number of reactions (from question 2) and the energy (in joules) produced in each reaction (that you calculated in the activity of Energy of Hydrogen Fusion), how much total energy is produced?

J

4. Now you will find the lifetime. Use the total nuclear energy available that you just calculated and the luminosity of the Sun, which is 4×10^{26} W. Express your final answer for the nuclear lifetime in years.

years Show your work:

5. Is nuclear energy generation a viable way for the Sun to produce its energy given what we know about the age of the Earth?

6. Is this still the case, even if only 1/10th of the mass of the Sun ever undergoes nuclear reactions?

In this case, what value do you get for the lifetime of the Sun (in billions of years)?

billion years

5.2.3: Lifetimes of Stars Other Than the Sun

The models of hydrogen fusion are the processes that set the basic timescale for stellar lifetimes. However, there is one other aspect of nuclear fusion that we have not really touched upon: the rate at which the hydrogen fusion takes place depends on the temperature and density of the gas. And not just in a small way. The nuclear reaction rate in the process used by low-mass stars like the Sun (see Going Further 5.4: Nuclear Reactions in the Sun and Other Stars) depends on the temperature to about the fourth power! That means if the temperature doubles, say, from 10 million to 20 million degrees, the nuclear reaction rate goes up by a factor of 2^4. That is a factor of 16, so a star with twice the Sun's temperature would run through its fuel 16 times faster.

The interesting part of all this is that more massive stars must have higher internal pressure to support themselves against their own weight (see Going Further 5.3: Why the Gravitational Lifetime Cannot Be Fixed). More internal pressure means they must be hotter, and hotter means that they run through their fuel faster. To make things even more extreme, the most massive stars do not produce their energy via the same chain of processes as low-mass stars (see Going Further 5.4: Nuclear Reactions in the Sun and Other Stars). In high-mass stars, the temperature dependence is more like T^{15}. For these reactions, doubling the temperature increases the fusion rate by more than a factor of 10,000! As a result, even though massive stars have more fuel, they last much less time than low-mass stars because they go through their mass so much more quickly.

To make an analogy with cars, low-mass stars are like hatchbacks: they have smaller gas tanks but run through it less quickly so they can run for a long time; high-mass stars are like SUVs: they have bigger gas tanks, but run through their fuel so quickly that it is not long before they run out.

LIFETIMES AND LUMINOSITIES OF STARS WITH DIFFERENT MASSES

In this activity, you will explore the relationships among the masses, luminosities, and lifetimes for stars that are fusing hydrogen.

A. Predictions: lifetime and mass

1. How do you think a star's lifetime will depend on its mass?

If star A is twice as massive as star B, it will have a lifetime that is two times longer.

If star A is twice as massive as star B, it will have a lifetime that is half as long.

If star A is twice as massive as star B, it will have a lifetime that is more than two times longer.

If star A is twice as massive as star B, it will have a lifetime that is less than half as long.

Even if star A is twice as massive as star B, mass does not matter; they will have the same lifetime.

Now watch the animation, which shows how much time it takes for stars to end their main sequence lifetimes. (More massive stars are depicted as larger in size.)

Play Activity

2. Resolve any discrepancies with your predictions.

3. What additional information do you need to fully test your predictions?

B. Examining lifetime, mass, and luminosity

Next, click on the stars to further examine their properties, such as mass, luminosity, and lifetime. To look for more stars, click and drag with the mouse inside the window to move around.

Play Activity

1. Now you can fully test your predictions. Choose one star from your data set and another star with approximately twice its mass. How do their lifetimes compare? Resolve any discrepancies with your predictions.

2. How does the lifetime of a star with 5 times the mass of the Sun compare to that of the Sun?

3. Based on your answers to the above questions, how do you think the nuclear fusion rate for a high mass star is greater than, less than, or the same as that of a low mass star?

4. Now we will graph the data to determine the general trends and relationships. You can toggle between "plot" mode and "starfield" mode by clicking the appropriate buttons. To add more stars to the data set, go to starfield mode and click on them (click on a star again to remove it).

a. In the plot mode, use the pull-down menus to select mass on the x-axis and lifetime of the y-axis. What does this plot tell you about how a star's lifetime depends on its mass?

b. Now compare the curve options to your data. What type of relationship fits mass and lifetime the best?

c. In the plot mode, use the pull-down menus to select mass on the x-axis and luminosity on the y-axis. What does this plot tell you about how a star's luminosity depends on its mass?

d. Now compare the curve options to your data. What type of relationship fits mass and luminosity the best?

e. If a star A is twice as massive as star B, how will its luminosity compare?

5. Summarize the relationships among mass, lifetime, luminosity, and nuclear fusion rate for a main sequence star.

There is a mathematical relation that describes the tendency of more massive stars to use their fuel more quickly. It turns out that the luminosity of a star is roughly proportional to its mass raised to the third power.

$$L \sim M^3$$

The luminosity, L, thus increases rapidly with mass, M.

The actual relationship is a bit more complicated, with the exponent being closer to 2.5 for stars with masses similar to the Sun's or lower. For much more massive stars, the exponent is more like 3.5. Still, we can use this mass–luminosity relationship to explore, somewhat crudely, the typical lifetimes of stars.

We know that the lifetime of a star should be roughly given by the ratio of the amount of fuel it has to the rate at which it uses up that fuel. This is how we estimated the lifetime for the Sun previously. So, we can write as follows.

$$t \sim \frac{M}{L}$$

Using the mass–luminosity relation, we can combine these expressions. We then have the relation below.

$$t \sim \frac{M}{M^3} = \frac{1}{M^2}$$

So, we should expect that the *more* massive a star, the *shorter* its lifetime. The effect is dramatic. More massive stars live much shorter lives. A star with a mass 10 times that of the Sun will live roughly 1% as long. On the other hand, a star with a mass only 1/10th the Sun's mass will live about 100 times longer.

We thus expect that any stars that have ever formed over the history of the Universe to still be around if they have masses 1/10th the mass of the Sun. In contrast, stars 10 to 100 times the mass of the Sun last for only a few million years. This is such a short period of time that they should not even be able to leave the sites of star formation where they are born before they run out of fuel.

Though these are only rough estimates, they still give us a general idea of stellar lifetimes. To compute the actual lifetimes for a given star, we have to use sophisticated computer models that take into account not only the star's mass, but also things like composition. The models also use great care to compute nuclear reaction rates, energy transport, and more. However, those details turn out not to make a huge difference in the lifetimes. Our simple estimates still give us good ballpark figures. They also produce approximately the correct trends of stellar lifetime with mass.

Comparing the Lifetimes of Stars With Different Masses

Worked Example:

1. The Sun's lifetime is about 10 billion years. Estimate the lifetime of a star with twice the mass of the Sun.

- Given: The ratio of the masses is 2:1
- Concept: The lifetime goes as the inverse square of the mass: $t \sim 1/M^2$
- Solve: So to find the ratio of the lifetimes, take the inverse square of the ratio of masses: $t \sim 1/(2/1)^2 = ¼$
- So, for a star with twice the mass of the Sun, its lifetime would be one-quarter that of the Sun, or ¼ × 10 billion years = 2.5 billion years.

Questions:

1. Use the relation for stellar lifetimes to estimate the lifetime of a star with ½ the mass of the Sun

years

which is

billion years

5.2.4: Stellar Properties and the H-r Diagram

On a perfectly clear evening far away from city lights, you would only be able to see about 3,000 stars. Within 10 light-years of Earth, there are only 12 stars, including our Sun. Yet, beyond those that we can see with our naked eyes, we have found that there are hundreds of billions of stars in our Galaxy alone.

You might have noticed that not all stars appear the same. Looking up at night, a careful viewing shows that some stars are slightly more red or blue or yellow than others. Early astronomers noticed these color differences too, but a true understanding of stars did not begin until the 19th century. That is when astronomers began to employ the new tools of spectroscopy and photography, and these allowed them to begin to study the stars' properties in detail. The application of photography and spectroscopy to astronomy in the late 19th century marks the birth of the modern science of astrophysics.

Using the new spectral methods, astronomers created a classification scheme for the stars based on patterns of lines seen in their spectra. Stellar spectral classification was first based upon the presence or absence of four prominent spectral absorption features: those from hydrogen, those from calcium and sodium, those from molecules, and those from carbon. This scheme, developed by the Italian astronomer and Catholic priest Angelo Secchi (1818–1878) in the 1860s, was taken up and expanded upon by Edward Pickering (1846–1919) and his "computers," a group of female astronomy researchers at the Harvard College Observatory. These investigations started in the 1880s, and they quickly developed into the most advanced set of stellar observations in the world.

Wilhemina Fleming (1857–1911) was one of the Harvard computers, and she was tasked with making sense of stellar spectra. She subdivided each of Secchi's classes into subclasses and gave them letter designations, A through Q. These were based upon more refined differences in their spectra that had become apparent from the superior data that were being collected by the Harvard observers. Another of the computers, Antonia Maury (1866–1952), then took up the work. She rearranged the ordering of some of Fleming's letter classes and also replaced the letter classification system with one based upon numbers. Her system was quite detailed, but also complicated. It did not find much favor with her colleagues at the Observatory because it was difficult to understand and use. Finally, in 1901, Annie Jump Cannon (1863-1941) took up the challenge. She greatly simplified the spectral classification system of Maury. She also returned to letter classes, like Fleming, but with far fewer class designations in total. She mostly retained the ordering of Maury while reordering a few additional spectral types herself. Each of these revisions of spectral classification was based on the increasingly more detailed measurements of the spectral features of stars. Cannon's basic system of stellar classes is the one still in use today.

There are seven primary categories in the classification system created by Annie Cannon. They are O, B, A, F, G, K, and M. Cannon further divided these into subclasses, so that a B5 star would be midway between a B and an A star, an A2 would be partway from A to F, etc. In 1925, Cecelia Payne-Gaposchkin, whom we have already mentioned, showed that the spectral types are related to temperature, but that discovery depended upon the development of the quantum theory of the atom. Some of the properties of the stellar types are shown in Table 5.3.

Table 5.3: Stellar Properties and Spectral Classification

SPECTRAL CLASS	COLOR	TEMPERATURE (K)	MASS (SOLAR UNITS*)	RADIUS (SOLAR UNITS*)
O	Blue	$>3.0 \times 10^4$	>16	>6.6
B	Blue white	$1.0 \times 10^4 – 3.0 \times 10^4$	2.1 – 16	1.8 – 6.6
A	White, bluish white	$7.5 \times 10^3 – 1.0 \times 10^4$	1.4 – 2.1	1.4 – 1.8
F	White, yellowish white	$6.0 \times 10^3 – 7.5 \times 10^3$	1.0 – 1.4	1.2 – 1.4
G	Yellow	$5.2 \times 10^3 – 6.0 \times 10^3$	0.8 – 1.0	1.0 – 1.2
K	Orange	$3.7 \times 10^3 – 5.2 \times 10^3$	0.5 – 0.8	0.7 – 1.0
M	Red	$<3.7 \times 10^3$	0.1 – 0.5	< 0.7

*The units of mass and radius in the table are solar masses and solar radii. One solar mass is equal to the mass of our Sun (about 2×10^{30} kg), and one solar radius is equal to the radius of our Sun (about 7×10^5 km).

While the computers of Harvard College Observatory busied themselves classifying stars (and doing other things), other astronomers were searching for different patterns in stellar properties. In 1911, Ejnar Hertzsprung (1873–1967), in Denmark, plotted stellar luminosity against stellar color. He measured the “color” of each star using its brightness measured in two filters, centered at different wavelengths of light, and comparing them. Two years later, the American astronomer Henry Norris Russell (1877–1957) independently plotted a star’s brightness against its spectral class. Because both spectral class and color are different ways of characterizing the star’s temperature, these two relationships are really the same thing. Hertzsprung and Russell discovered this relationship independently but nearly simultaneously. And so plots of stellar brightness vs. temperature are now called Hertzsprung-Russell diagrams, or just H-R diagrams, to honor the astronomers who first made them (Figure 5.3).

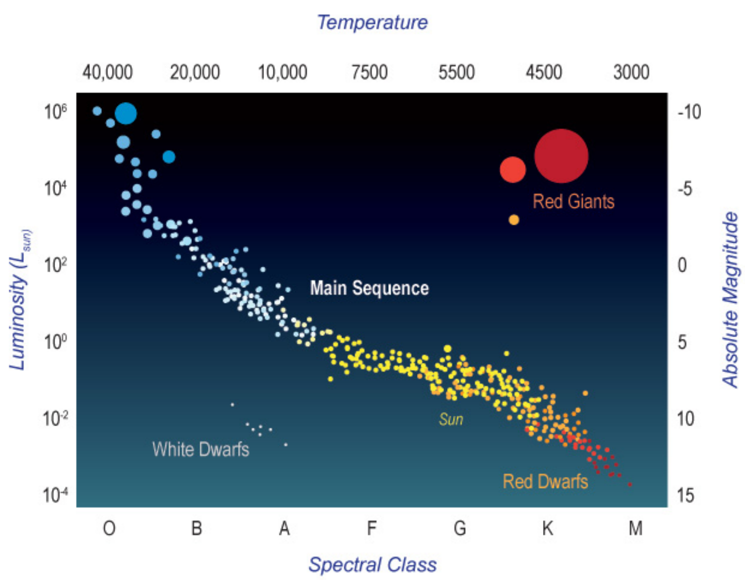


Figure 5.3: The Hertzsprung-Russell diagram shows the relationship between a star’s temperature, spectral class, luminosity, and absolute magnitude. Notice that high temperatures are on the left, corresponding to spectral class O, and lower temperatures are on the right, corresponding to spectral class M. Brighter stars are toward the top of the diagram and dimmer stars are toward the bottom. Credit: NASA/SSU/Aurore Simmonet.

The majority of stars lie along a central diagonal line in the diagram, called the main sequence. Main-sequence stars are in the main part of their lifetimes, fusing hydrogen into helium. During this steady hydrogen-burning phase, the stars are in a state of hydrostatic equilibrium. This means that the gravitational pressure pushing inward and compressing them is balanced by outward thermal pressure. The stars are stable because any tendency of the star to compress will heat its interior, thereby increasing the nuclear fusion rate, releasing more energy, and thus increasing the pressure. The increased pressure then overcomes the compression and expands the star. On the other hand, any tendency of the star to expand will cool the star, thereby slowing the nuclear reaction rate. The reduced reaction rate means lower energy production and further lowering the temperature and, thus, pressure. The reduced pressure then allows gravity to compress the star. So any departure from equilibrium causes effects that return the system to the original steady state: These negative feedback creates a stable equilibrium condition that lasts as long as the star has nuclear fuel to fuse in its core.

In a sense, a star is a giant self-regulating thermostat. It adjusts itself so that it maintains just enough nuclear fusion to create pressure to support itself against gravity, but not so much that it blows itself apart. Again, this is true as long as there is fuel in the core for fusion to occur.

On the H-R diagram, as one goes from the red to the blue-white stars on the main sequence, the stars not only become brighter but they also become hotter and more massive. Color, temperature, mass, and radius are all intrinsically linked for main sequence stars. This is due to the physics involved.

As stars age, they run out of hydrogen to fuse, and this causes them to adjust their internal structure. They evolve off the main sequence and occupy different regions of the H-R diagram. Some of these changes will be described in Section 3.6, on stellar evolution.

Examining clusters of stars plotted on the H-R diagram gives us an additional way to determine the ages of objects in the Universe. We can estimate the ages of star clusters because all stars in the cluster started forming at the same epoch from the same parent cloud. By looking at which stars have ceased hydrogen fusion and which ones have not, we can estimate the age of the cluster.

Using such plots, we have learned that the oldest parts of our galaxy are around 13 billion years old, nearly as old as the Universe itself. We have also learned that there are very young parts of the Galaxy, some even in the process of forming. There are objects with intermediate ages as well. You will determine the ages of several star clusters in the next activity.

Star Clusters and Stellar Lifetimes

The plots shown here are for three different star clusters. Along the horizontal axis is the star's surface temperature, with lower temperatures on the right and higher temperatures on the left. The vertical axis is the star's brightness, with dimmer on the bottom and brighter at the top. Brightness is given in terms of apparent visual magnitude (see Going Further 3.6: The Magnitude System).

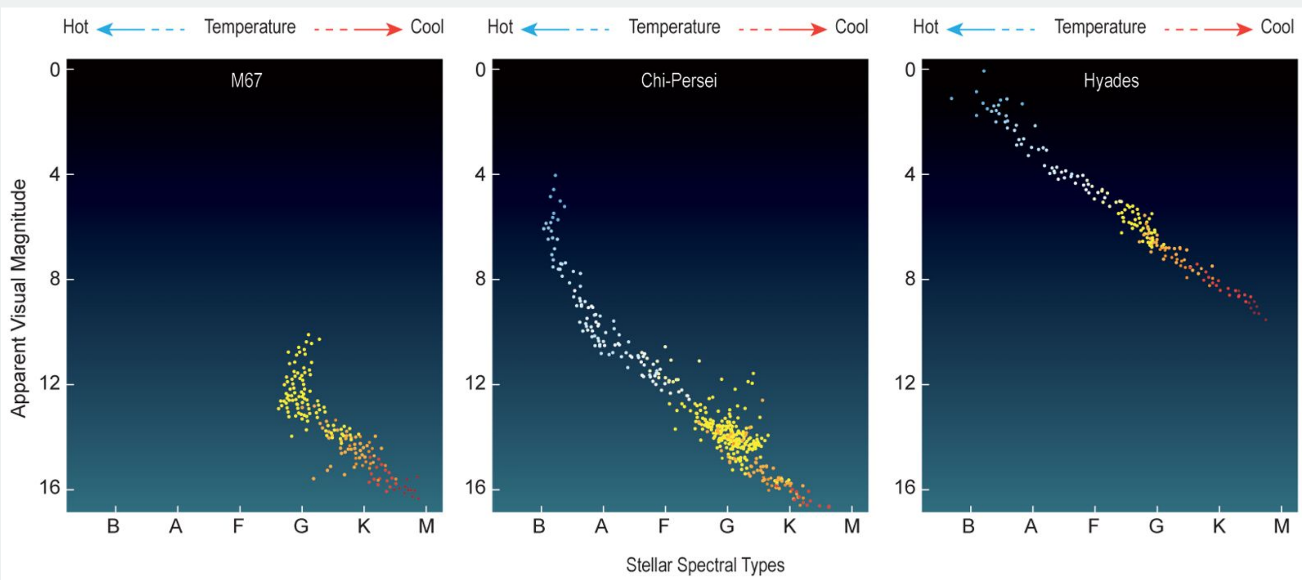


Figure A.1: The Hertzsprung-Russell diagrams for three different open clusters. Credit: NASA/SSU/Aurore Simmonet.

Notice that all of the star clusters plotted show a line of stars running from the lower right to upper left. We will ignore stars that are not along this line for the time being.

1. Do you expect the stars at the left of each graph to have low masses or high masses? Explain your reasoning.

2. Given your answer in question 1, should these stars have long lifetimes or short lifetimes?

3. Assuming that all the stars in a given cluster were formed around the same time, does the plot make sense, given your answers in questions 1 and 2? Explain how it either does or does not.

4. If you can, list the clusters from youngest to oldest. Explain your reasoning.

This page titled 5.2: Measuring Ages - Lifetimes of Stars is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

5.3: Change Over Time - Evolution of Stars

What Do You Think: Change

In this chapter, we have discussed how we measure ages of objects in the night sky. Another thing we want to know about objects like stars and galaxies is whether they evolve or change over time. After all, it would be nice to know if the Sun will change in the near future, and if so, whether it will affect us. Similarly, we would like to know if our Galaxy, the Milky Way, has changed or will be changing. And what about changes in the Universe itself? It turns out that we can learn a lot about how the Sun and the Milky Way have evolved and continue to evolve by comparing them to observations of other stars and galaxies. Similarly, we can learn about the Universe's evolution through our observations of the objects—stars and galaxies—within it.

5.3.1: Formation of Stars

Stars form when clouds of gas and dust collapse due to gravity. In our Galaxy, we observe these large clouds as nebulae, and we often observe star-forming regions within them (Figure 5.4). The clouds are supported by the motions of the gas inside them, but if they become too massive or if they get disturbed (say, by colliding with another gas cloud), the resulting compression will allow gravity will take over in the most dense regions. The gas in those regions will begin to collapse and fragment. Eventually, gravitational compression in the densest fragments can cause heating to the point that nuclear fusion begins. A star is born.

Figure 5.4: The Orion Nebula, as seen by Hubble Space Telescope, is a vibrant star-forming region of space. Credit: NASA/STScI/Hubble Space Telescope

As a star-forming cloud becomes more compact, it spins faster, like an ice-skater tucking in her arms (Animated Figure 5.5). It also flattens into a disk. The center of the disk becomes very hot and dense and forms a star, while at the same time, solid particles condense, growing into larger pieces, eventually forming planets and other objects.

An ice skater spins (video link).

Animated Figure 5.5: When the ice-skater pulls her arms in, she increases the rate at which she spins. When she throws her arms out, her spin rate slows down. A similar thing happens in a star-forming nebula. As the cloud contracts in size due to gravity, it spins faster. Credit: Shutterstock.com

The process to form gas giant planets is thought only to occur in the outer part of the pre-stellar disk where temperatures are cold and the amount of gas available is high. Terrestrial planets are thought to form closer to the star.

First, consider the area of a region of a disk at large radius, shown in green in Figure 5.6. Compare it to the area available in the inner parts of the disk, shown in white and blue. The outer part contains much more area, and thus, there is much more material there to form planets. Furthermore, the warmer temperatures in the inner disk (caused by the proximity to the forming star) prevent the lightest elements from condensing into solids in that region.

https://phys.libretexts.org/@go/page/30480

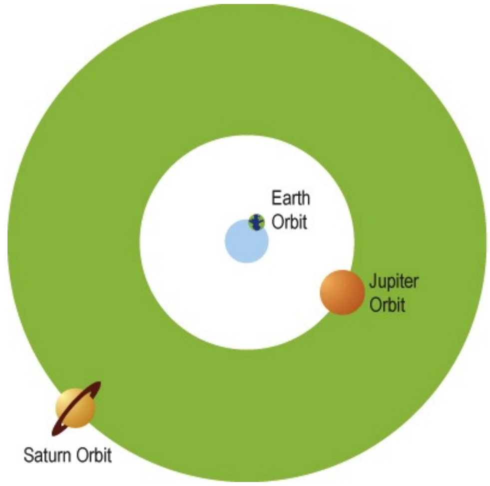


Figure 5.6: More area in the outer Solar System means bigger planets form there. Credit: NASA/SSU/Aurore Simonnet

At higher temperatures, light gases like hydrogen and helium have a lot of energy and are moving too fast to clump together. Only heavier materials, rock-forming silicates and metals, can condense and clump together at the location of the terrestrial planets. The lighter gases only begin to condense in areas with much lower temperatures. This is the outer part of the disk, where temperatures are colder than the freezing point of water. Much colder.

At low temperatures, the gases have less energy and can clump together into solids. As the solids grow, they begin to accumulate the abundantly available hydrogen and helium gas. And as the gases build up, the masses of the forming planets grow. And grow. The resulting stronger gravitational pull from all this mass accumulation allows them to grow ever faster, boosting their gravity further. This positive feedback creates a runaway growth period that only ends when they have gobbled up all the available gas. This is how they become enormous.

Conditions in the inner parts of the star-forming disk do not allow giant planets to form. Conditions are too warm to allow them planets to accumulate large amounts of gas, and there is not very much material available in the inner disk to form planets in the first place. Only, smaller, terrestrial planets can form.

In the inner disk, whatever lighter materials are present, like water, carbon dioxide, methane, ammonia, and others, all remain in the gas phase. Since planets, even gas giants, start forming from the accumulation of solid bodies (rocks), the planets formed in the inner disk tend to be rocky, with only small amounts of light and volatile compounds like water and carbon dioxide. The outer gas giant planets do have rocky cores, but these are buried by the far more abundant lighter gases.

This star/planet formation scenario has a lot of explanatory power. For instance, it easily accounts for the fact that the Sun and planets all rotate in the same direction, because they must all reflect the rotation of the cloud from which they collapsed. In the same way, it naturally describes why all the planets orbit the Sun in the same plane and in the same direction. However, there are some aspects of the Solar System that require other processes to have been at work. For instance, the rotation axis of the planet Uranus is tilted approximately 90 degrees. But we also know that toward the end of the planet-forming process, 4.5 billion years ago, there were many collisions of large bodies. We see evidence of large collisions on the Moon, and these date from that epoch—there is even strong evidence that the Moon itself was formed when a Mars-sized object impacted with the proto-Earth. The large amounts of ejecta blasted into orbit by the collision provided the material that coalesced into the Moon. In addition, the light molecules like water and carbon dioxide that are found in the inner planets could plausibly have been delivered late in their formation, as comets from the outer Solar System impacted with the inner bodies, enriching them with these materials.

The star-formation process is generally understood, though some details must still be worked out. One of the aspects that still puzzles astronomers is how the clouds break up into a number of different stars of different sizes instead of just a single star, or a few very large stars. The reasons probably have to do with turbulent flows within the clouds and angular momentum conservation during the collapse process.

Angular momentum relates to the tendency of spinning objects to continue spinning with a constant spin rate and orientation: It is difficult to either speed up or slow down a spinning object or to change the direction of its axis of rotation. You can test this with a bicycle wheel. Even at low speeds, you will notice the wheel's reluctance to be tipped over, and the faster the wheel spins, the harder it is to tip. This is the same phenomenon that keeps a top from tipping over. You might have also noticed that if you shrink a spinning object, it tends to spin faster, and if you expand it, it will tend to spin more slowly. These are all illustrations of the principle of conservation of angular momentum.

Angular momentum helps support vast clouds of gas against gravitational collapse, and the clouds must overcome this support before they are able to collapse to form stars. However, since total angular momentum must remain constant, a cloud must form stars in a way that allows most of its mass to collapse to relatively tiny objects, while preserving the angular momentum of the whole. Normally, the objects would spin up to very fast rotation rates as they collapsed. These fast rotation rates would halt further collapse at some point. The clouds overcome this through internal collisions in which some of their gas gains energy and angular momentum and moves outward, while the remainder and vast majority of material loses energy and angular momentum and moves inward. As this happens, the clouds tend to flatten into a disk. Most of the material spirals in toward the center of the disk, while all of the angular momentum is contained in a tiny amount of mass that is flung out to large radii through random collisions. Such disks, in which matter is transferred inward as angular momentum is transferred outward, are called accretion disks. They are seen in many different astrophysical settings.

As the disks form, they break up further into smaller rotating vortices. These are really just small disks within the large disks. At the very center of the large disks, stars are formed (Figure 5.7). At the centers of the smaller vortices, gas giant planets like Jupiter and Saturn form. And on still smaller scales, even smaller vortices form the moons around gas giants.

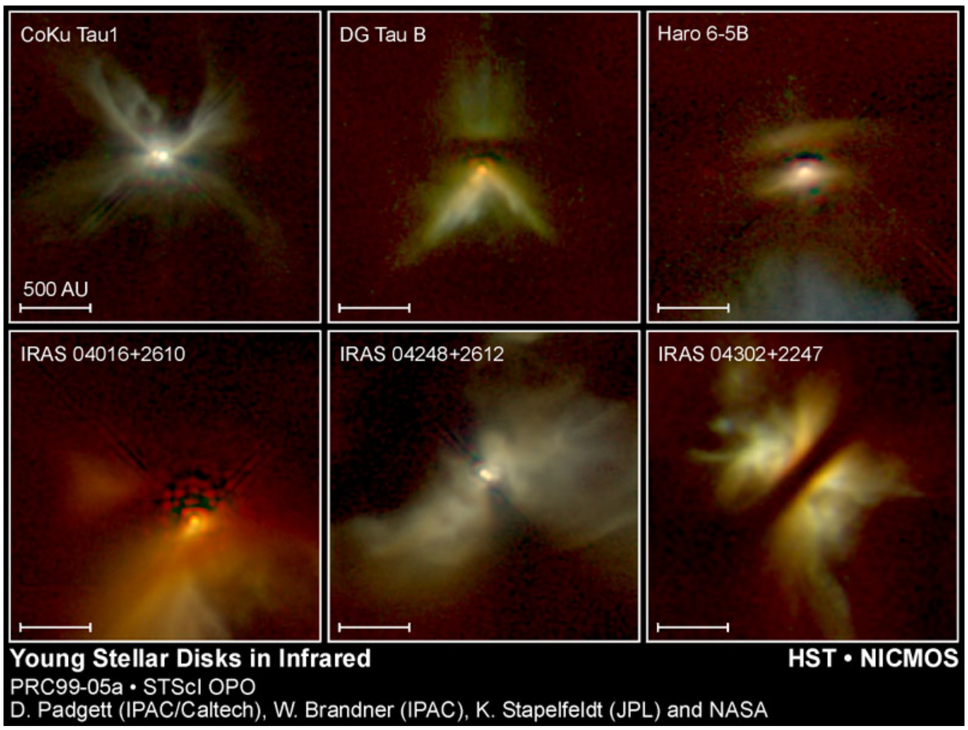


Figure 5.7: Viewed in infrared, six star-forming disks of young stars, some of which may also lead to the creation of companion planets. Credit: NASA/STScI/Hubble Space Telescope

Another result of star formation seems to be the production of high-speed outflows, and even narrow jets of material. These are seen in the later stages of star formation. Molecular CO is seen to be rushing outward from young stellar systems. The outflowing gas has a conical shape and is centered on the rotational axis of the protoplanetary disk. It flows out in both directions perpendicular to the disk. In so-called T-Tauri stars, the outflows are in the form of narrow jets that move at several hundred kilometers per second (Figure 5.8). These jets extend hundreds of light-years from the star, and their termination shocks form nebulosities called Herbig-Haro objects, named after the two astronomers who first studied them.

The causes for outflows from newly formed or forming stars is still not understood. Nor are all the details of the planet-formation process. But, in the past decade, astronomers have begun to catalog hundreds of stars besides the Sun that have planetary companions. We expect our studies of these systems to give us a better picture of how stars and planets form.

We will see the themes of accretion disks and outflowing jets (Figure 5.9) repeated several times in several different contexts throughout the modules.

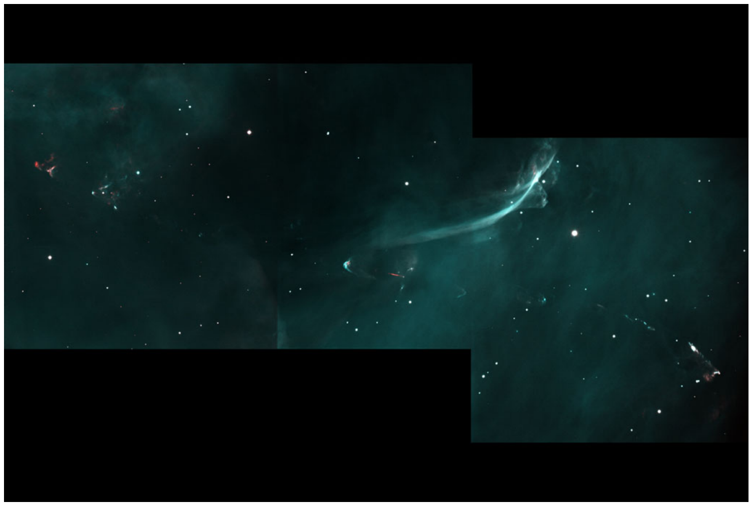


Figure 5.8: Image of outflows from a T-Tauri stars, HH34. Credit: John Bally (CASA, U Colorado, Boulder), Bo Reipurth (IfA, U Hawaii, Hilo), and ESO/NTT

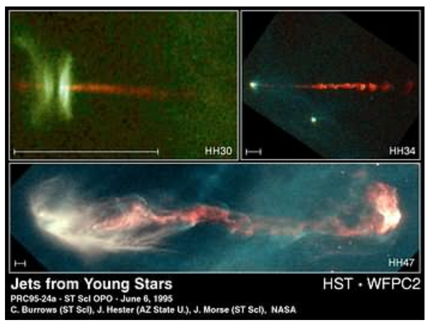


Figure 5.9: As stars form, they can produce jets, expulsions of accreted materials along their magnetic poles, which are perpendicular to their disks. Credit: NASA/STScI/Hubble Space Telescope

Orion Nebula Fly-through

Play this video to take a voyage through the Orion Nebula, a nearby star-forming region. It uses volume visualization to combine images and measured distances to create a three-dimensional fly-through experience.

- As the fly-through approaches the nebula, you will see the hot, bright stars that light up the gases in the nebula. Their radiation hollows out the space in the center.
- Next, the fly-through will pass by some tear-drop-shaped objects; these are young solar systems in the process of being formed.
- Then, we zoom in on one of these objects; you can see the accretion disk and jets.

This video contains no audio. Credit: Greg Bacon (STScI), model based on data by C.R. O'Dell (Vanderbilt University), and The American Museum of Natural History/Rose Center for Earth and Space. Video link.

Now answer the following questions:

1.

2.

5.3.2: Death of Stars Like the Sun

Stars create energy and light through nuclear fusion. At the centers of stars like the Sun, there is so much heat and pressure due to gravity that the nuclei of hydrogen atoms fuse together and create helium nuclei (plus a lot of energy). What happens when the Sun runs out of hydrogen? And what happens in other stars?

In all stars, two major forces are at play: (1) pressure due to gravity, which pushes in toward the center of the star, and (2) pressure due to heat, which pushes outward. As we discussed previously, in main sequence stars, these two forces are in balance, so that the star does not collapse due to too much gravitational pull or fly apart due to too much thermal pressure outward. This "balanced" phase in a star's lifetime can last for billions of years. But, when the star eventually runs out of hydrogen at its center, there is nothing more to fuse together. Without the energy from nuclear fusion, the temperature drops and there is less outward pressure as a result. The center of the star collapses inward due to gravity, causing this inner region to heat up once again. A thin shell of hydrogen gas outside the star's core begins fusion into helium. Pressure related to these fusion reactions pushes the outer layers of the star's gas even further outward, so that the star expands and its outer layers cool down. The star appears red (cool outer layers) and large, and very bright due to the fusion in the hydrogen shell. At this point in its lifetime, the star is called a red giant. Red giant stars are bright and cool, so they can be found in the upper-right portion of the H-R diagram.

During the red giant phase, the very center of the star continues to collapse and heat up so that helium nuclei begin to fuse into carbon. When the helium in the star's core is used up, the core again collapses, causing it to heat up. A thin shell of helium around the core begins fusion, with a shell of fusing hydrogen lying just outside that. Again the star expands, its surface cools, and the star brightens.

However, for a star with a mass like that of the Sun, the core does not continue to contract. The carbon core of the star is already so compressed that the star's gravity cannot push it significantly further inward (it only compresses a little bit and then stops). Instead, the electrons become so compressed that they hold up the carbon core against further gravitational contraction. The hydrogen and helium shells around the carbon core continue fusion but become unstable, and the outer layers of gas around the star are eventually pushed away forming a planetary nebula (Figure 5.10). The small, hot carbon core of the star is called a white dwarf. Over time, white dwarfs cool and fade until we can no longer observe them. White dwarfs are dim, lying below the main sequence, but they are quite hot, so they are found below and to the left of the main sequence (Figure 5.10).

Figure 5.10: At only 2,000 light-years away, the Spirograph Nebula is a planetary nebula spanning less than half a light-year in diameter. At its core is a white dwarf. Credit: NASA/STScI

How do we know the Sun's fate is to turn into a red giant and then a white dwarf, and when will this happen? Astronomers investigate how the Sun will change over time by observing the basic properties of other stars, including their color, brightness, size, and mass. They also determine the stars' ages. They can look at stars that have similar mass to the Sun and see how their properties are similar to or different from the Sun's, and how these things depend on their age. From these observations, astronomers gain clues about what the Sun looked like in the past and what it might look like in the future. Astronomers also compare their observations to our understanding of processes like nuclear fusion that occur at the centers of stars.

On the basis of our current observations and calculations, we know that the Sun began nuclear fusion about 5 billion years ago. We also know that the Sun will continue to fuse hydrogen into helium for another 4 to 5 billion years before it runs out of hydrogen in its core and begins the red giant phase.

5.3.3: Death of More Massive Stars

We have learned that more massive stars run through their fuel more quickly than the Sun. This causes the most massive stars to have much shorter lifetimes than the Sun. How do we know this? As with the Sun, we can learn about how massive stars evolve by observing the color, brightness, size, and mass of many other stars, and by measuring their ages. Graphing the properties of stars on the H-R diagram is a great way to do this. If we compare a group of stars that have the same age (for example, we may look at a cluster of stars that we know all formed at the same time), we find that the most massive stars in the group are no longer present on the Main Sequence, and that many star are found in the upper right corner of the plot. In other words, the massive stars have already begun their red giant phase, whereas the lower-mass stars are still fusing hydrogen, much like the Sun is doing now.

In some cases, when we look at a cluster of stars, we see that the stars that have masses like the Sun's have already reached the red giant phase (bright, red, large), and other stars are already white dwarfs (faint, blue, small). We do not see any high-mass stars in these cases. Where have they gone? The key to understanding this lies in understanding how high-mass stars evolve.

Massive stars generate light and heat through nuclear fusion, just like the Sun. Just like lower-mass stars, when massive stars run out of hydrogen at their cores, they begin a "shell burning" phase, where the gas in a shell around the star's core begins fusion. In this case, as with lower-mass stars, the outer layers of the star are pushed outward, and the star enters the red giant phase. The core of the star collapses due to gravity and begins fusing helium into carbon.

Unlike lower-mass stars, however, very massive stars (about 10 times more massive than the Sun) do not enter a white dwarf phase when they run out of helium for fusion into carbon. Because of their high masses, these stars have enough gravity that the inward pressure and temperature at their cores become large enough to begin carbon fusion into nitrogen. The process of nuclear fusion in the core of the star and in shells around the core continues, and heavier and heavier elements (nitrogen, oxygen, and silicon) are made, assuming the star is massive enough to generate the temperatures needed to fuse these elements. For the most massive stars, the process continues until the silicon in the star's core fuses to form iron.

Fusion does not continue in the iron core. It turns out that the fusion of iron requires energy, it does not release energy (Figure 5.11). This is a catastrophe for any star that reaches this stage.

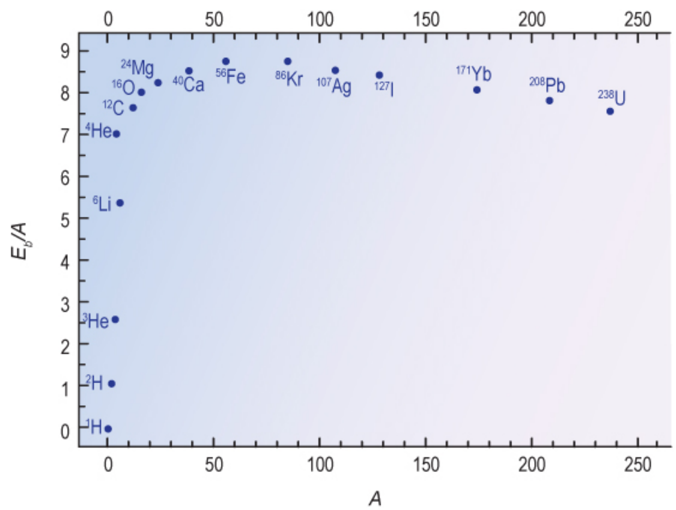


Figure 5.11: Binding energy is plotted vs. atomic number. The binding energy is a measure of how much energy is released when elements undergo fusion (elements lighter than iron) or fission (elements heavier than iron). Stars stop fusing at iron because it is no longer energetically favorable. Credit: NASA/SSU/Aurore Simmonet

What happens at this point? The high pressures and temperatures force the iron to fuse, but this cools the star: remember, iron fusion requires energy, it does not release energy. As the fusion continues, the temperature *drops*. This causes the outward thermal pressure to go away. As a result, there is nothing supporting the stellar layers above the core, or even the core itself. It all comes crashing down.

The force of gravity pushing inward on the star is so great that even the outward push of the electrons in the iron atoms cannot withstand it. Some of the electrons are pushed into the nuclei of the iron atoms, where they combine with protons to form neutrons. This process further lowers the pressure in the core. The stellar core collapses so rapidly that in less than a second, the outer layers of gas fall inward at nearly a tenth of the speed of light. The combination of the electrons with protons (this is inverse beta decay) releases a tremendous number of neutrinos, and these carry away the accumulated binding energy gained by the star over its lifetime. A small fraction, about one percent, of the neutrinos interact with and heat the stellar layers just outside the core. As a result of this near instantaneous heating, the outer layers of the star are rapidly ejected outward in a tremendous explosion called a supernova.

The time required for this entire process to trigger, from the initial onset of iron fusion to the catastrophic heating and expulsion of the layers just above the core, is on the order of a tenth of a second. For these layers to push upward through the overlaying star and eject the remaining layers into space takes several hours; these are typically very big, expanded stars, and they would fill much of the inner part of our solar system. Some of them would fill all of it. In the weeks following the explosion, a supernova is so bright that it can outshine an entire galaxy. It then gradually fades away over many months.

After a supernova occurs, a small remnant of the star's core, made only of neutrons, is left over. This is called a neutron star. A neutron star has a mass of about 1.4 times the mass of the Sun, but is only about 20 km (12.4 miles) in diameter. The neutrons in a neutron star are packed so tightly together that the entire object resembles a huge atomic nucleus. This is incredibly dense—a teaspoonful of neutron star material would weigh over a billion tons!

Surrounding the neutron star is the expanding cloud of gas ejected during the explosion, or in other words, the rest of the star. The expanding cloud is called a supernova remnant (Figure 5.12), and in addition to the outer layers of the star that exploded, it contains any surrounding gas that gets swept up. Thus the remnant includes the elements created in the star during its lifetime and during the supernova explosion as well as the gas in the stars immediate environment.

As the remnant expands, it can collide with adjacent clouds, causing them to collapse and trigger a new generation of stars to be formed. These new stars will incorporate some of the elements that were thrown out into space during the supernova explosion. The elements that are most common in our bodies and in the air we breathe, including carbon, oxygen, nitrogen, and iron, were formed this way, in stars. The iron in your blood was forged in the cores of massive stars billions of years ago and the gold in your jewelry in supernova explosions. This is what the famous astronomer Carl Sagan meant when he said we are all made of "star stuff."

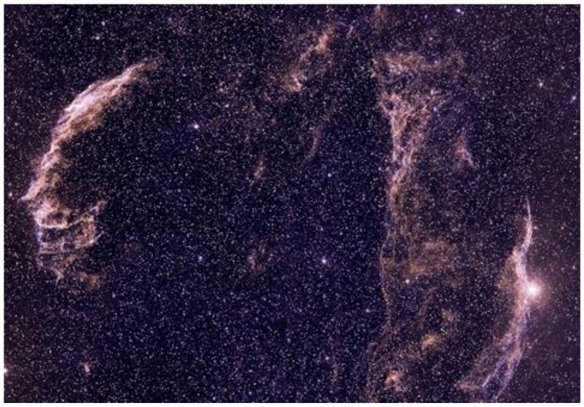


Figure 5.12: The explosive power and legacy of a massive star's death seen here in the Cygnus loop. Credit: ESA and Digitized Sky Survey (Caltech).

If a star is exceptionally massive, more than 20 times the mass of the Sun, then it will have such a strong inward gravitational push, such that even the tightly packed neutrons in a neutron star will not be able to hold themselves up. In that case the star collapses to the theoretical limit. At the center of the budding supernova explosion, a black hole forms. It swallows the stellar core entirely, along with all the neutrinos that are needed to heat the stellar layers above and blast them into space. So in this kind of event, the entire star collapses into its core, a black hole, making a brief gasp of gamma rays before it disappears completely.

A black hole can be described by the point at its very center, called the singularity, which seems to contain the entire mass of the collapsed core, plus whatever additional material from the stellar envelope might have fallen in. Around this is a sphere defined by the event horizon, the radius within which nothing, not even light, can escape the gravitational pull of the black hole. Black holes are incredibly interesting objects, and we will discuss them in much greater detail (including how we can try to observe them) in Chapter 11.

Stellar Evolution

The life cycles for high- and low-mass stars are depicted graphically. Place the tiles for the different stages of stellar evolution in the correct places for each cycle. When you are finished, check your work using the "check" button.

Play Activity

This page titled 5.3: Change Over Time - Evolution of Stars is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

5.4: Evolution of Galaxies and the Universe Itself

What Do You Think: Change in the Universe

In the first section of this chapter, we discussed how astronomers measure distances in terms of the time required for light to travel across them. We introduced the idea of the light-minute, the light-hour, and the light-year as distance measurements. This method works because the speed of light is constant. That means light will always travel a certain distance in a given amount of time. For example, in a time interval of one year, light will always travel one light-year, or about 6 trillion miles. In a different time, say one second, light always travels a light second, which is about 300,000 kilometers. We can use the constancy of the speed of light to measure distances this way for any time period we like.

However, the constancy of light speed means that light is not only a useful tool for measuring distances; light can also be used as a sort of time machine. How can we do this? We will explore this idea by using the Sun as an example.

We know that the Sun is about 8 light-minutes away from Earth. In other words, it takes 8 minutes for light to travel from the Sun to Earth. Therefore, we see the Sun as it was 8 minutes ago. If the Sun were to suddenly stop shining, we would not know about it for 8 minutes!

As another example, the closest star to the Sun is Alpha Centauri. It is the brightest star in the southern constellation Centaurus. Light requires four years to reach Earth from Alpha Centauri, so Alpha Centauri is 4 light-years away from us. In addition, we see the star not as it is, but as it was four years ago. Because stars fdo not change very much on a timescale as short as four years, we do not notice the time delay due to the speed of light when we observe nearby stars.

Viewing other galaxies, however, the situation changes. The nearest galaxy to the Milky Way is M31. It is located in the northern constellation Andromeda. Light takes about 2.5 million years to reach us from M31, so M31 is 2.5 million light-years away from us. But just as above, this also means that we see the galaxy M31 not as it is currently, but as it was 2.5 million years ago. That is a very long time for us, but it is not a very long time for a galaxy. Galaxies change on timescales of hundreds of millions of years, not just a few million years.

What happens if we look at even more distant galaxies? The most distant galaxies we observe are not millions of light-years away, they are billions of light-years away. As a result, as we look at galaxies that are more and more distant, we see them as they were hundreds of millions or billions of years ago. Given those long delay times, we begin to see that even galaxies have changed over time.

Distant galaxies, seen as they were when the Universe was younger, tend to be smaller (in size) and less massive (they contain less stuff) than nearby galaxies. They also tend to be bluer and have stronger indications of ongoing and recent star formation than nearby galaxies have. We also see that the shapes of galaxies has changed over time. Very distant galaxies tend to have odd, clumpy shapes. A much larger fraction of nearby galaxies have regular, disk-like shapes (like the Milky Way or Andromeda) or elliptical shapes.

Does this mean that we live in a special part of the Universe where the galaxies are more massive and have more regular shapes? No, it does not. Remember that when we look at distant galaxies, we are really looking back in time, seeing them as they were when the Universe was much younger than it is now. What we are in fact seeing is evidence that galaxies have grown in size and mass, that they have used up some of their star forming fuel, and in many cases, that their shapes have stabilized over time. What would astronomers living in the most distant galaxies see if they observed the Milky Way? They would not see our Galaxy as the large spiral disk we live in today. Instead they would see a small “proto” galaxy, the one that eventually evolved into the Milky Way. That is because the light they would be observing "now" actually left our Galaxy billions of years ago, when it was much younger than it is now. Over the intervening time the light has been traveling across the immense distance between ourselves and the imagined alien astronomers who are only now able to look out toward us, viewing the ancient light as it arrives at their location.

One of the most difficult concepts that students deal with when studying cosmology is the time-machine effect due to the speed of light. We are used to thinking that light travels instantaneously between objects, but that is not true. Even in our normal experience, it takes time for light to travel from one place to another: light takes 1 nanosecond (a billionth of a second) to travel 30 cm, or about a foot. This amount of time is too short for us notice, though it can be measured by electronic circuits. In the study of cosmology, however, the distances are so large that the light travel time becomes quite noticeable. This effect has some advantages, as we are able to view almost the entire history of the Universe by looking at more and more distant objects.

Some of the most impressive observations obtained thus far are from the Hubble Deep Field and Hubble Ultra Deep Field images taken by the Hubble Space Telescope. In each of these, the telescope was pointed for about 10 days at small patches of sky that contained no bright stars. By exposing the telescope’s cameras to these patches of sky for such a long time, the cameras were able to see “deeply” into the Universe, building up the light from very faint, very distant galaxies. Recently, astronomers estimated that one of the galaxies in the Hubble Ultra Deep Field image is 13.2 billion light-years away, seen as it was when the Universe itself was less than half a billion years old. We can imagine a team of alien astronomers in that galaxy, as it exists "now," peering back in our direction and seeing our galaxy at that time, long before humans or the Sun or Earth even existed. This is the time machine afforded us by the finite, and constant, speed of light.

In the movie at the beginning of the next section, astronomers’ knowledge of the distances of galaxies in the Hubble Ultra Deep Field has been used to create a “fly-through” video of this region of the sky. This visualization allows us to take a virtual trip into the Universe’s past!

Hubble Space Telescope Ultra-deep Field Fly-through

Play this video to take a voyage through the Hubble Ultra Deep Field, into the Universe’s past. These deep field images allow us to peer nearly all the way back to the beginning of the Universe. The galaxies you will see at first are closer to us. As the movie plays, it will be going farther away in distance and farther back in time. All of the images and distances have been measured directly with the Hubble Space Telescope.Observations of galaxies can help us learn about the formation and evolution of the Universe itself. The famous astronomer Edwin Hubble (1889–1953) measured the velocities of nearby galaxies by taking spectra of them and measuring the Doppler shift of their spectral lines. Hubble found that nearly all of the galaxies he studied are moving away from us, and the more distant galaxies are moving away from us at faster velocities than the galaxies closer to the Milky Way. This observation, that more distant galaxies have faster velocities away from us, is called Hubble’s law. In a later chapter we will study the details of Hubble’s measurements and other observations since his time that support his results. But an important thing to take away now is that Hubble’s law indicates that the *Universe is expanding!*

This video contains no audio. Video link.

Answer the following:

1.

2.

When you get to the end where there are no more galaxies visible, it is not because you have reached a *place* in the Universe where there are no galaxies, nor is it because the galaxies at those distances are too faint for us to see. It is because you have traveled back to a time in the history of the Universe, more than 13 billion years ago, *before galaxies had formed!*

3.

The idea that the Universe is expanding might be new to you. Or it may be something you have heard about before. Even if you have heard about the expansion, you could be unfamiliar with the astronomical observations that lend weight to this idea. Either way, you might now be wondering: Why is the Universe expanding? Will it continue to expand, or will it stop at some point? These are good questions, the kinds of questions we will investigate further in later chapters. For now, we will take a little closer look at the Universe's past.

5.4.1: A Quick Overview of Big Bang Cosmology

Imagine that the Universe is not just expanding now, but that it has been expanding throughout its history. That would mean in the past there was less space between galaxies than there is now, and the farther back into the past we looked, the less space between galaxies we would see. We can keep "rewinding" this scenario until we imagine that all the matter in the Universe today was packed into an extremely dense, very tiny volume. That would have been the earliest stage of an ever-expanding Universe.

What caused the Universe's expansion? Astronomers do not know entirely, but they call the model for how the Universe has changed over time since then the Big Bang Theory, or sometimes the Big Bang Model. Some people colloquially call the early moments of the Universe's existence the "big bang." Perhaps this is to emphasize the fact that the Universe sprang into existence a finite time in the past—13.7 billion years ago—but really, the model explains all of the epochs of time: the past, present, and future of the Universe, not just the earliest moments. Therefore, we will use the phrase Big Bang to mean the model or explanation for how the Universe changes over time.

Using our observations, and explaining them in the context of this model, we can reconstruct the history of the Universe. A brief overview is presented below. We will go into more detail in Module 3.

In the beginning, the Universe was extremely hot and incredibly dense. It contained high-energy photons (light) and other subatomic particles. After about 1 minute, the Universe had cooled considerably from its starting point, but it was still incredibly hot (temperature = 1 billion kelvin). At this stage, the density and temperature in the Universe were both high enough that nuclear fusion could occur, just like at the center of a star today. Protons and neutrons combined to form helium nuclei and a tiny amount of lithium (Figure 5.13). This era, when the lightest elements in the periodic table were created, is known as the era of **Big Bang Nucleosynthesis** (BBN). The heavier elements that we are composed of and depend on, such as oxygen, carbon, and nitrogen, have all been produced in stars after this time. They could not be made in the early Universe, because after about three minutes the ambient temperature and density, though still extremely high by earthly standards, were insufficient to sustain fusion.

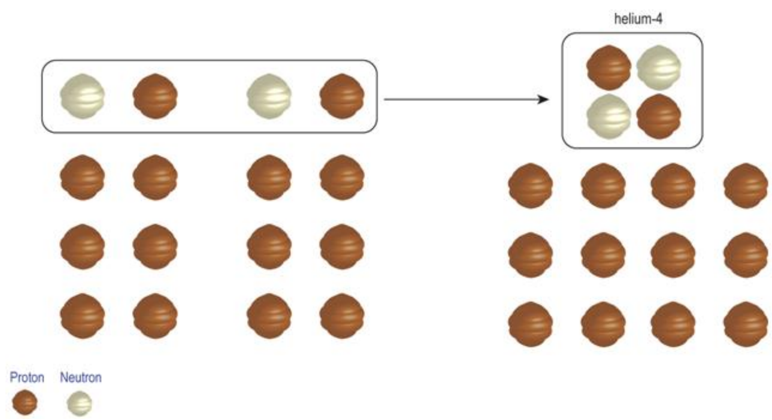


Figure 5.13: When the Universe was a few minutes old, it was the right temperature for nuclear fusion. There were seven protons for every neutron at that time. This led to approximately 25% of the available protons and neutrons forming ^{4}He and 75% forming ^{1}H. Trace amounts of deuterium (^{2}H), ^{3}He, and lithium (^{7}Li) were also created. Protons are represented in the figure as brown particles and neutrons as white particles. Credit: NASA/SSU/Aurore Simmonet

At these early times in the history of the Universe, its temperature was so high that electrons were too energetic to combine with protons or other atomic nuclei. Instead, electrons whizzed around, frequently interacting with photons (light). Because of this, a photon could not travel very far before interacting with an electron, and the Universe was opaque (we would not have been able to see through it). This changed at about 380,000 years after the Universe began, when it had expanded and cooled down to a temperature of 3,000 K. At this point, the cooler temperatures allowed electrons to slow down enough to be captured by atomic nuclei, forming atoms, mostly hydrogen and helium. Because the electrons were now bound in atoms, they no longer interacted with light strongly. Light could then travel freely, and the Universe became transparent.

It is possible to see the light from this era still traveling through space; it is one of the most important lines of observational evidence supporting the Big Bang Model. It was first observed entirely by accident in 1963 by Arno Penzias and Robert Wilson, two astronomers who worked for Bell Labs.

https://phys.libretexts.org/@go/page/30481

While it was still opaque, the Universe itself behaved as a blackbody. If the Universe had expanded and cooled as we think it has according to the Big Bang, physicists predicted that it should contain a background of radiation that preserved this blackbody character. In particular, the relic radiation should have a temperature of a few kelvin above zero. A blackbody of this temperature would emit light most strongly in the microwave region of the electromagnetic spectrum today. This is exactly what Penzias and Wilson detected, and for which they later won a Nobel Prize. The background signal from the early Universe is called the Cosmic Microwave Background (CMB; Figure 5.14).

Astronomers have continued to study the CMB intensely since its discovery some sixty years ago. They have now measured tiny fluctuations in its temperature. These provide information about which regions of the early Universe were slightly more or less dense than average. The more dense regions are those where galaxies or even clusters of galaxies have since formed. Under the pull of gravity, the atoms that formed in the early Universe have combined to form stars, galaxies, and clusters of galaxies. These objects began forming when the Universe was 400 million years old, and we see them still today.

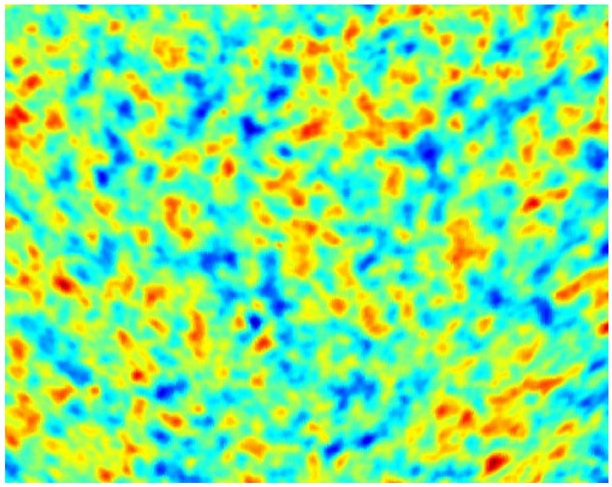


Figure 5.14: An image of the Cosmic Microwave Background (CMB), as measured by the Boomerang Group. Light from the CMB was emitted when the Universe was about 380,000 years old. The blue spots are places that were slightly more dense than average, and the red spots are places that were slightly less dense. Credit: NASA/NSF/Boomerang

This has been only the barest, most general outline of our current understanding of the evolution of the Universe as viewed through the paradigm of the Big Bang Theory of cosmology. We hope you have more questions about this and other topics related to the Big Bang, as well as questions about how the Universe has evolved and what its fate may be. That is the subject of the rest of the modules.

This page titled 5.4: Evolution of Galaxies and the Universe Itself is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

5.5: Wrapping It Up 5 - Cosmic Timeline

Have you ever wondered about the events that created the Universe we live in? Do you want to know when stars first appeared or when galaxies formed? Do you know how old the Universe is? Or when our Solar System formed? Perhaps you have wondered when humankind arrived on the scene.

Explore the following timeline, and complete the questions that follow. This activity will reveal how the Universe evolved from the beginning of time through the present day. The timeline is divided into three major eras, in order from the beginning to now: the early universe (blue), the matter-dominated universe (yellow), and the modern universe (green). To see the events within a particular era, click on that era. To learn more about a particular event, click on the event. Use the scroll bar to view all events in an era.

Play Activity

1.

2.

3.

4.

5.

6.

7.

8.

https://phys.libretexts.org/@go/page/30482

9.

10.

This page titled 5.5: Wrapping It Up 5 - Cosmic Timeline is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

5.6: Mission Report 5 - Cosmic Timeline

A.

C.

D. Questions to be graded for accuracy

1.

2.

3.

4.

This page titled 5.6: Mission Report 5 - Cosmic Timeline is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

6: Measuring Cosmic Distances

Chapter 6 explores the various ways in which humanity has learned to measure the distances to the stars and beyond. Examples of several geometrical methods, the standard rulers, and standard candles are presented. The chapter concludes with a summary of the Cosmic Distance Ladder, which puts together of all the measuring techniques to bridge our understanding of distance in the Universe.

This page titled 6: Measuring Cosmic Distances is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6.0: Measuring Cosmic Distances Introduction

Video Transcript

> *During many celebrations and holidays, such as New Year's Eve around the world or Fourth of July in the United States, fireworks are an integral part of the festivities.*
>
> *But if you cannot see the surrounding landscape, the buildings and houses, and if the night was completely dark, no streetlights or house lights, would you be able to tell how far away they were? Would you need to touch them to tell how far away they were?*
>
> *Looking at the night sky you see thousands of points of light. If you observe them for long enough, you'll notice a few move relative to the others. These are the planets of our solar system, yet they still appear to be only points of light. How can their distances be determined? And what about the distances to stars in our galaxy, or to galaxies outside our own? How can we measure their distances?*
>
> *In this chapter, we will discover the tools, including the mathematics and measurement techniques, used by astronomers to determine cosmic distances.*

The Universe is incomprehensibly vast by earthly standards, and we cannot go to other stars or galaxies with a tape measure or pace out the distances to them. Despite these limitations, one of the most important methods we have to understand what objects are and how they work is understanding their distances. Without a knowledge of cosmic distance, we cannot understand how much energy a faraway object is emitting. In physics (including astrophysics), knowing the energy produced by a system is fundamental to developing models that describe the nature of that system. In this chapter, you will learn how scientists measure the distances to and between objects in the Universe.

Astronomers use three basic types of techniques to measure cosmic distances:

- *Direct Geometric Methods:* These methods use simple geometry, that of similar triangles, to measure the distance to an object. Geometrically determined distances are the most trustworthy because they do not rely on any properties of the objects themselves. As a result, no theoretical uncertainties about the objects can interfere with geometric distances. Unfortunately, geometric methods become difficult to use as objects become more distant, as we shall see.
- *Standard Ruler:* We can use this if we know how big an object is inherently. This method relies on the fact that the farther away an object is, the smaller it looks, in a mathematically predictable way. By measuring an object's apparent size and comparing to its known intrinsic size, we can deduce its distance. This method is not strictly geometrical in nature because it requires us to know something about the object being measured - its size - that we might have know with high precision.
- *Standard Candle:* We can use this method if we know how bright an object is inherently. This method relies on the fact that the farther away an object is, the dimmer it looks, in a mathematically predictable way. By measuring an object's apparent

brightness and comparing to its known intrinsic brightness, we can deduce its distance.

Different techniques work well for different distances. For instance, the geometrical method called parallax can only be used for very nearby stars. On the other hand, Type Ia supernovae, which can be used as standard candles, are so rare that the likelihood of seeing one nearby is quite small, but we see many of them at large distances. For in-between distances, other methods are available. We must use a method (or methods) appropriate to the distance of interest. Aside from the geometrical methods, the techniques used for objects farther away typically rest on calibrations from studies of similar objects that are relatively close and that have known distances from some other method. In this way, we build up a hierarchy of techniques to measure distances. This hierarchy is often called the cosmic distance ladder. To be confident that the ladder is giving us accurate results, we must use multiple independent techniques when possible. If they all give a consistent distance, we can be fairly certain that our measurements are in the right ballpark. We will discuss several of the important distance determination methods in turn.

This page titled 6.0: Measuring Cosmic Distances Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6.1: Geometrical Methods

Learning Objectives

- You will know that the ancient Greeks measure the size of Earth and the Moon as well as the distance between them using geometry
- You will know the principle of parallax and when it applies
- You will be able to give examples from everyday life and from astronomy
- You will be able to perform calculations and understand conceptually how distance is related to parallax angle
- You will be able to convert angles among radians, degrees, arcminutes, and arcseconds
- You will know the principle of the moving cluster method and when it applies

What Do You Think: Geometrical Method

6.1.1: Geometric Methods Used by the Ancient Greeks

One of the first known applications of geometrical methods to an astronomical distance was from ancient Greece. Astronomers there noticed that the shadow of Earth on the Moon during a lunar eclipse was not a straight line, but a curved arc. Using the amount of curvature of the arc, they could deduce the size of Earth's circular shadow, which could be directly compared to the apparent size of the Moon itself. In this way, they were able to deduce the relative sizes of Earth and its moon.

Figure 6.1: When Earth’s shadow falls on the Moon during a lunar eclipse, the size of the diameter of the circle corresponding to Earth’s shadow can be compared to the Moon’s diameter to determine the Moon’s size relative to that of Earth. © Anthony Ayiomamitis. By permission.

Earth’s size had been determined by the Greek scholar Eratosthenes (276–194 BCE), who used the geometry of circles and the lengths of shadows cast by vertical sticks to deduce Earth’s circumference (Figure 6.2). With the results of Eratosthenes in hand, it was possible to assign an absolute size to the Moon. Then, a mathematical estimate known as the small-angle approximation allowed its distance from Earth to be determined. So, more than 2,000 ago, Greek scholars knew that the Moon was about a quarter million miles from Earth, though they did not express the distance in miles, of course.

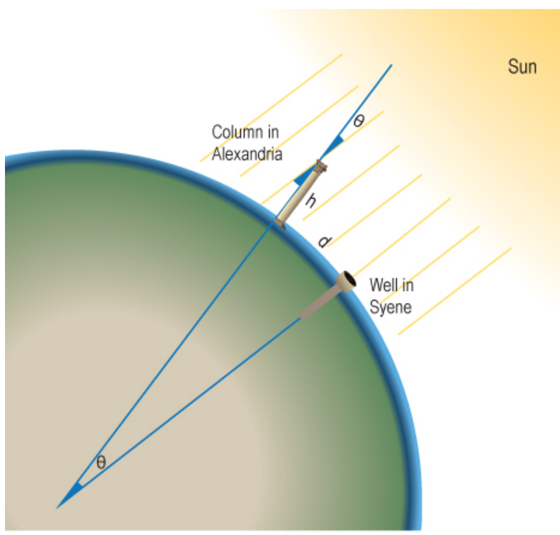


Figure 6.2: In 240 BCE, Eratosthenes measured Earth’s circumference by observing the length of shadows at two different places on Earth, at the time of the summer solstice. At Syene, the Sun was (almost) directly overhead, and the Sun’s rays were vertical, falling directly down a well. At Alexandria, a distance of *d* = 4,400 stades north, a vertical column cast a shadow at an angle θ. By measuring or calculating θ, Eratosthenes could then determine Earth’s circumference because he knew that the proportion of the angle θ to 360 degrees was equal to the proportion of the distance d to the circumference. His value of 252,000 stades is within a few percent of the currently accepted value of 24,907 miles (depending on how one converts stades into miles). Credit: NASA/SSU/Aurore Simonnet

The small-angle approximation, also known as the small-angle formula, can be used to determine the distances to faraway objects.

$$d = \frac{S}{\theta}$$

The distance d to an object, is the ratio of its size, S, and θ (Greek letter “theta”), the angle in radians that the object spans from the location of a viewer (Figure 6.3). For the derivation of this formula, see **Going Further 6.1: Derivation of the Small-Angle Approximation**.

https://phys.libretexts.org/@go/page/31242

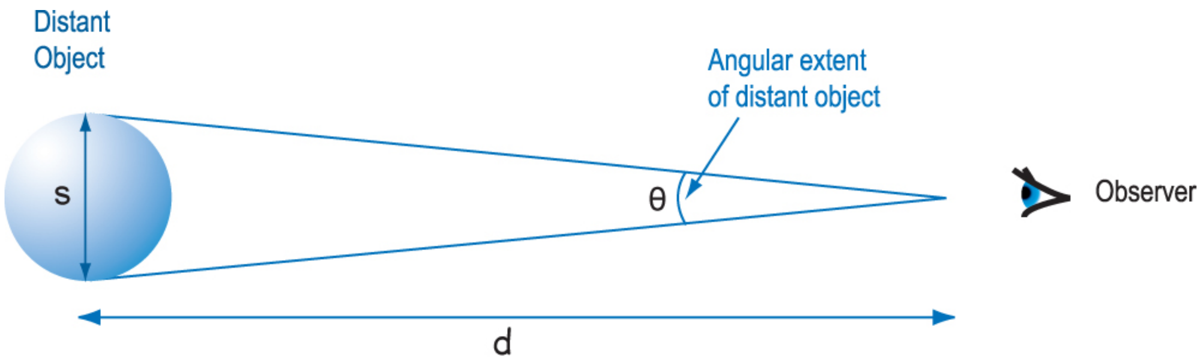


Figure 6.3: When the distance *d* is much larger than S, then the angle *θ* is small, and its value (in radians) is approximately the same as its tangent (or sine). Using this approximation, the size of the object (*S*), the distance to it (*d*), and the angle (*θ*) are related mathematically by a simple relation: *d* = *S/θ*. Credit: NASA/SSU/Aurore Simonnet

Going Further 6.1: Derivation of the Small-angle Approximation

Mathematically, angular diameter, linear size, and distance can be combined in an extremely useful and simple equation called the small-angle approximation. As seen in Figure 6.3, the angular diameter, θ, depends on the distance to the object, d, and its actual linear size, S, according to:

$$\tan\left(\frac{\theta}{2}\right) = \frac{S}{2d}$$

For very small values of \(θ\) measured in radians, $\tan(\theta) = \theta$. Using this approximation, we can simplify the equation relating distance and linear size to:

$$\frac{\theta}{2} = \frac{S}{2d}$$

or more simply

$$\theta = \frac{S}{d}$$

In this **small-angle approximation**, if any two of the quantities are known, the third can be calculated. In astronomy, the angular diameter is usually measured directly, and the equation is used to calculate the distance to or physical size of the object. Since distances to astronomical objects are usually much larger than their linear sizes, this approximation is of great use in all branches and at all levels of astronomy!

The table below shows the relationship between the angle (in degrees and radians) and the tangent of the angle. Also shown is the difference between the angle in radians and the tangent of the angle. The close correspondence between these two quantities is the basis of the small-angle approximation.

Angle (degrees)	Angle (radians)	Tangent (angle)	Difference	% Difference
0.5	0.0087	0.0087	0.0000	0.0025
1.0	0.0175	0.0175	0.0000	0.0102
2.0	0.0349	0.0349	0.0000	0.0406
4.0	0.0698	0.0698	0.0000	0.1628
8.0	0.1396	0.1405	0.0009	0.6550
15.0	0.2618	0.2679	0.0066	2.349
20.0	0.3491	0.3640	0.0149	4.270
25.0	0.4363	0.4663	0.0300	6.870
30.0	0.5236	0.5774	0.0538	10.27

Based on looking at the table, when do you think the approximation breaks down?

6.1.2: Parallax

The first distance measurement technique devised for distances outside of the Solar System involves a phenomenon that you see every day, though you may not have noticed it or realized how you could use it to measure distance. **Parallax** is the apparent shift in position of a foreground object relative to the background. This is how we perceive depth using the small differences in perspective as seen from our two eyes. We typically learn to estimate distances without having to think about it using this technique when we are very young. Astronomical parallax is a more formalized (mathematical) version of our binocular vision. To familiarize yourself with this fundamental distance measurement technique in astronomy, and to connect that measurement technique to everyday things, you should complete the following activity on parallax.

Parallax of Your Thumb

In this activity you will use Figure A.6.1 as a backdrop simulating distant stars as we explore what happens when you view an object from two different perspectives.

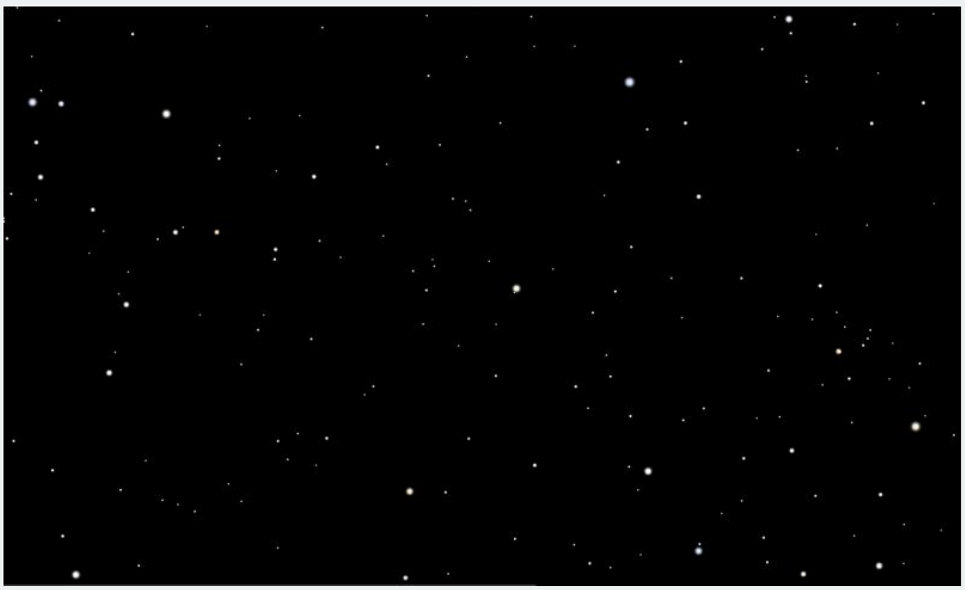


Figure A.6.1: Use this image as a backdrop for the activity. Credit: NASA/SSU/Kevin McLin.

1.

2.

3.

https://phys.libretexts.org/@go/page/31242

Figure 6.4: In the left-hand panel (a), the woman has her arm stretched out and observes a slight parallax shift. In the right-hand panel (b), the woman has her arm bent and observes a greater parallax shift corresponding to the shorter distance to her thumb. Credit: NASA/SSU/Aurore Simonnet

When you observe relatively nearby stars against a backdrop of distant stars, the nearby stars will shift as Earth orbits around the Sun. Figure 6.5 illustrates this type of parallax. The drawing in the right-hand side of the figure shows the geometry of the setup as you look at a relatively close object against a distant background. These might be distant stars in the astronomical context, but the same principle applies to terrestrial viewing as well.

Now bring your attention to Figure 6.5. There are two simulated star fields, labeled 1 and 2. These correspond to viewing the same part of the sky from Earth, with viewing sessions separated by six months. At the left of the figure, View 1 and View 2 illustrate how the setup translates into what you would actually see from the two perspectives. Notice that the two parts of the figure illustrate the important point that the shift in position relative to the background stars represents a change in the angle that our line of site makes with the object relative to some reference line. Thus, parallax is an angle.

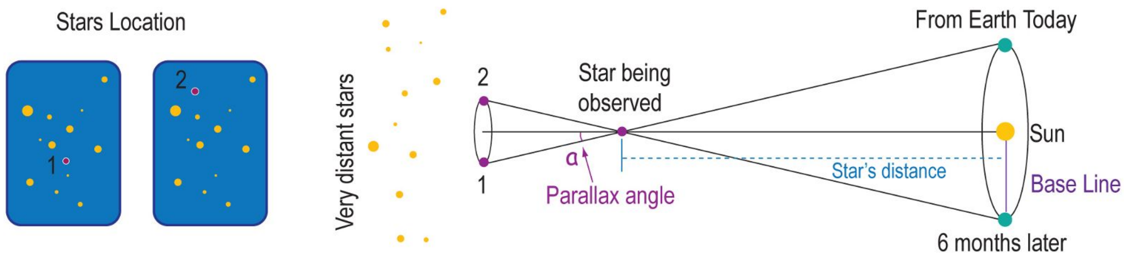


Figure 6.5: Parallax is the apparent shift in position of the foreground object relative to the background. Viewpoints 1 and 2 show the apparent position of the star being observed (in red) when viewed by two telescopes at different points on Earth, or the same telescope at two different points in its orbit around the Sun (January and July, for example). The parallax angle a is also indicated. Credit: NASA/SSU/Aurore Simonnet.

In Figure 6.6 we de-clutter the diagram to concentrate on the geometry. Consider the triangle shown in the figure. We can choose to do our measurements so that one of the angles of the triangle is a right angle (90°). The angle a is the parallax angle, which we can measure by the apparent shift of the nearby star. Note that it is half the apparent angular shift that the star makes between the two measurements, 1 and 2. This is because we have drawn the triangle to have a right angle, a choice made for convenience, not necessity.

An important point to notice about the triangle is that if you make the distance d bigger (keeping the baseline B the same), the parallax angle a gets smaller, while if you make B bigger (keeping d the same), the parallax angle a gets bigger. This matches with the qualitative relationship you found in the activity when you viewed your thumb at different distances from your face.

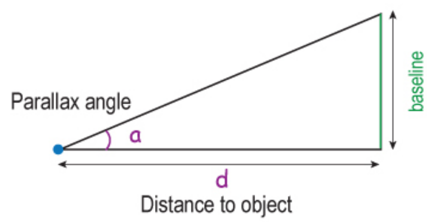


Figure 6.6: Determining distances using parallax. If we know the baseline and measure the parallax angle a, we can calculate the distance d to an object, such as a star. Credit: NASA/SSU/Aurore Simonnet.

Next, we need to calculate the distance quantitatively. We can do this using trigonometry to relate angles to ratios of distances and using the definition of the tangent to write the following relation.

$$\tan(a) = \frac{B}{d}$$

If we know B and measure a, we can calculate the distance d.

$$d = \frac{B}{\tan(a)}$$

This technique for measuring distances is commonly used in surveying for roads or buildings.

In astronomy, the parallax angle is always very tiny because the distances to objects are so huge. The baseline B is the radius of Earth's orbit around the Sun, and even with such a "large" baseline, the largest parallax angles are still less than 1/4,000th of a degree! To describe such small angles the degree is split into small divisions of arcminutes (1/60th of a degree, symbolized by ' or arcmin) and arcseconds (1/60th of an arcminute, symbolized by " or arcsec). The parallax angles for even the nearest stars are less than 1 arcsec. Therefore, the small-angle approximation holds in these cases, and we can rewrite the equation for parallax as:

$$a \approx \frac{B}{d}$$

This formula is consistent with what we found in the activity: the *farther* the distance, the *smaller* the parallax shift. The *bigger* the baseline, the *bigger* the parallax shift. But we emphasize, *this version of the formula only works when the angle is measured in radians*, not degrees, arcseconds, or arcminutes. So, sometimes, we must use the fact that there are pi (π) radians in 180 degrees to convert angles between degrees and radians if we want to use the small-angle formula. And, to re-emphasize, the angle must be small, a condition always satisfied when measuring astronomical parallax to a star.

You have learned about two basic ways to describe the size of an angle. The first uses degrees, arcminutes, and arcseconds. The second uses radians. In the next activity you will learn how to convert between these units. Then you will practice using the parallax formula.

Converting Angles

Worked Examples:

1. Convert 30 degrees to arcseconds.

- Find: The number of arcseconds in 30 degrees
- Given: 30 degrees
- Concept(s): There are 60 arcminutes in 1 degree, and 60 arcseconds in 1 arcminute
- Solution: (30 ~~degrees~~)(60 ~~arcminutes~~/~~degree~~)(60 arcseconds/~~arcminute~~) = 108,000 arcseconds

2. Convert 30 degrees to radians.

- Find: The number of radians in 30 degrees
- Given: 30 degrees
- Concept(s): There are π radians in 180 degrees
- Solution: (30 ~~degrees~~)(π radians)/(180 ~~degrees~~) = 0.52 radians

Questions

1.

2.

3.

https://phys.libretexts.org/@go/page/31242

Sometimes, astronomers use a distance unit called a **parsec** to describe distances. Parsec is short for parallax-second and is abbreviated pc. When the baseline B is the distance between Earth and the Sun (1 AU) and the parallax angle a is measured in arcseconds (*arcsec* for short), then the distance d has units of parsecs. In this case, our formula for distance becomes:

$$d \text{ (in pc)} = 1/a \text{ (in arcsec)}$$

For example, this means that if a star has a parallax angle of 1 arcsec, it is at a distance of 1 parsec; if a star has a parallax angle of 1/2 arcsec, it is at a distance of 2 parsecs. We emphasize: *this version of the formula only works when the angle is measured in arcsec and the distance is measured in parsecs*. Parsecs are a convenient way to keep track of the relative distances of stars. If absolute distances are needed, it is simple enough to convert. For reference, 1 parsec is about 3.26 light-years, and the nearest star to the Sun is about 4 light-years away, or 1.2 pc.

Parsecs and Arcsecs

For this activity, express your answers as whole numbers or fractions, not decimals.

Converting Between Parsecs and Light-Years

1.

https://phys.libretexts.org/@go/page/31242

2.

In ground-based astronomy, we can use parallax to measure distances to some of the nearest stars, out to about 100 pc (several hundred light-years). The method becomes uncertain for stars farther than about 40 pc because atmospheric turbulence interferes with measurement of the tiny parallax angles. For distances beyond about 100 pc, the parallax becomes so small that atmospheric interference prevents measurements of the angles at all, and the method breaks down completely. However, with the development of space-based parallax measurements, the maximum determinable parallax distances have moved out beyond 500 parsecs. They will soon be pushed out much farther still. See Going Further 6.2: The HIPPARCOS Satellite for more information.

Parallax is a relative distance measure. It allows us to say that a star 20 pc away is twice as far as a star that is 10 pc away, but the method does not tell us what a parsec itself is. For that, we must know the size of the astronomical unit because it forms the baseline for our parallax measurements. While the AU is much smaller than the distances we are exploring in this section, it is still much too large for us to determine using familiar terrestrial methods. See Going Further 6.3: Measuring the Astronomical Unit if you want to understand how scientists were able to measure the size of the astronomical unit historically. Today, the astronomical unit is calibrated using radar measurements.

Parallax forms the foundation for the cosmic distance ladder. It is the most direct method for measuring distances to stars, albeit only for those that are nearby. Determining the distance to an object through parallax depends only on being able to see the small shift in its position as Earth orbits the Sun. None of the details of the object's properties or its inner workings matter. However, for distant objects with parallaxes too tiny to measure, we must devise a different measurement tool.

Going Further 6.2: the Hipparcos Satellite

Determination of parallax measurements is difficult from the ground because of atmospheric distortion of images, as we have already mentioned. To avoid the atmosphere, a satellite mission was proposed. This mission was developed by the European Space Agency (ESA) starting in 1980, and the resulting satellite was launched in 1989 from Kourou, French Guiana, aboard a French Arianne rocket. The satellite, called the High Precision Parallax Collecting Satellite, or HIPPARCOS, operated from 1989 until 1993, when it was turned off. The name HIPPARCOS pays homage to the ancient Greek astronomer Hipparchus (c. 190 BCE–c. 120 BCE).

The goals of the mission were to obtain parallax measurements of at least 2 milliarcsecond precision for more than 100,000 stars, as well as high-precision photometry for those stars. This is much higher positional precision than we can attain in ground-based observations. The final HIPPARCOS catalog of stellar parallaxes and proper motions contains 118,218 stars with parallaxes good to 1 milliarcsecond. So, these stars have precisely determined distances out to 1,000 pc (compare this to ~40 pc for ground-based parallaxes).

In addition to the HIPPARCOS catalog, the mission produced a secondary catalog called TYCHO. The TYCHO catalog contains more than 1 million stars with parallax measurements better than 7 milliarcseconds (distances to about 140 pc). An extension of the TYCHO catalog, called TYCHO2, has added an additional 1.5 million stars, bringing the total number of stars to 2.5 million. The TYCHO and TYCHO2 catalogs were based upon data from the satellite's star mapper, the system used to keep the satellite pointing in the proper direction. As a result, they have somewhat lower positional precision than the HIPPARCOS catalog.

The accurate parallaxes obtained by HIPPARCOS have allowed much better calibration of all the nearby rungs on the distance ladder (see Section 6.5) and have improved the precision of the measurements to all more distant objects as a result.

A follow-up high-precision mission, called Gaia, has many of the same goals as HIPPARCOS, though with significantly extended capabilities. In particular, it will obtain positions of a billion stars to 24 microarcsecond precision. This will push stars with precisely known parallaxes out to about 10 kpc. Gaia, also developed by ESA, launched in 2013.

The ideal gas law is easy to remember and apply in solving problems, as long as you get the **proper values a**

Parallax Shift

In this activity, you will observe a field of stars from different positions along the line of sight of Earth's orbit. After observing the parallax shift of each star, rank the stars by increasing distance. To do this:

- Move the Earth icon back and forth along the path of its orbit and observe the motions of the stars in the field. The original position of each star is indicated with a circle.
- Use these relative motions to place the stars in order of increasing distance (from left to right) with the pull-down menus.
- Press "submit" to check your answer.
- Press "reset" to try another set of stars.

Play Activity

Going Further 6.3: Measuring the Astronomical Unit

Historically, determining the astronomical unit was a multistep process. This is typical of cosmic distance determinations, as this chapter shows. The first step to cosmic distances was relating the distances of other planets to the size of Earth's orbit. Venus was the critical planet for this technique.

Figure B.6.1 shows Earth, Venus, and the Sun in a configuration in which Venus is at its largest separation from the Sun, as seen in Earth's sky. This configuration is called greatest elongation. It occurs in two forms, one is when Venus is east of the Sun in the sky, and one is when Venus is west of the Sun. For our purposes, either of these will suffice.

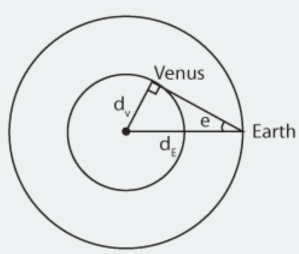


Figure B.6.1. In this figure, the Sun is at the center and the positions of Venus and Earth are labeled. dV indicates the distance from the Sun to Venus, and d_E indicates the distance from the Sun to Earth. The figure is not to scale. Credit: NASA/SSU/Aurore Simonnet.

Notice how, in this planetary configuration, the Sun, Venus, and Earth form a right triangle with Venus situated at the right angle. The hypotenuse of the triangle, d_E, is Earth's orbital distance, i.e., the astronomical unit. The side of the triangle connecting Venus to the Sun, d_V, is the size of the orbit of Venus, of course. So, using some trigonometry, we see that the ratio of Venus's orbital radius to the astronomical unit is the sine of the angle (e) that Venus makes with the Sun:

$$\sin(e) = dV/dE$$

Since we can easily measure the elongation angle by observing Venus as it moves through the sky during the year (it's about 42 degrees), we can determine the size of Venus's orbit in AU. But this is only part of the task. What we want is the size of the AU in meters or some other known unit. To get that, we must view Venus when it is crossing the Sun.

Because Venus's orbit lies between Earth and the Sun, the planet periodically crosses in front of the disk of the Sun as seen from Earth. Such transits, as they are called, do not happen as often as we might think. This is because Venus and Earth orbit in slightly different planes. The orbital plane of Venus is tilted by a bit more than 3 degrees with respect to the ecliptic, the name given to Earth's orbital plane. The Sun, however, is only about half a degree in diameter. That means that Venus must pass between Earth and the Sun while it is within a quarter degree of the ecliptic. This alignment is actually quite rare. Transits come in pairs separated by 8 years, but successive pairs are separated by about 120 years. Usually, Venus passes above or below the Sun, and no transit is visible.

However, occasionally, Venus does pass directly in front of the Sun. When that happens, two observers on Earth, at points A and B in Figure B.6.2 for example, might view the event as shown. Notice that they observe a small angular shift (exaggerated greatly in the figure) of Venus against the Sun's surface. This shift is the key to determining the size of the astronomical unit.

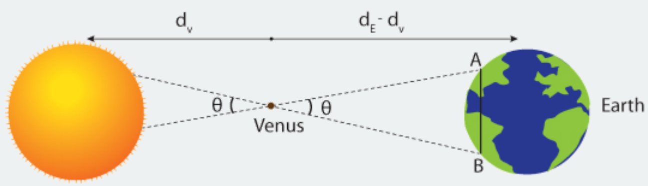


Figure B.6.2. The Sun is at the left-hand side of the figure, and Earth is at the right. Venus is at the point in the middle. The distance from the Sun to Venus is again given by dV, and the distance from Venus to Earth is therefore dE–dV. The distance between two different observers on Earth is given by the line AB. The figure is not to scale. Credit: NASA/SSU/Aurore Simonnet.

In the triangle made between Earth and Venus, the short side, AB, can be measured if the latitude and longitude of each site are known. The angle can be determined by comparing the two observations. Then, using the small-angle approximation, we can deduce the long sides of the triangle. This is the difference in size of the orbit of Venus and Earth's orbit.

$$\theta = \frac{AB}{d_E - d_V}$$

or

$$d_E - d_V = \frac{AB}{\theta}$$

But, we have already determined the ratio of these two distances, so we have two equations with two unknowns, and we can solve for either or both unknown. Using the ratio to eliminate dV, we get:

$$d_E - d_E \sin(e) = \frac{AB}{\theta}$$

or

$$d_E = \frac{AB}{\theta[1 - \sin(e)]}$$

This is how the astronomical unit was originally determined. The method was first employed, without great success, because of the small separation between observing points, in 1639. More successful measurements, the result of large international expeditions, were done in 1761 and 1769, and again in 1874 and 1882. The last two transits of Venus were on June 8, 2004 and June 6, 2012. There will not be another transit of Venus until December 2117, and then 8 years later in 2125.

Much more precise determination of the astronomical unit is now possible by bouncing radio signals off of Venus and using time of flight to measure its distance.

6.1.3: The Moving-cluster Method

There are no star clusters that are close enough to the Sun to have their distances determined through ground-based parallax measurements. Fortunately, it turns out that for one nearby star cluster, the Hyades, parallax is not the only means to measure distance. By using an alternative method to get an accurate distance to this cluster, we can use it to determine the distances to all other star clusters. Below, we explain how this is done.

We generally imagine that the stars are fixed in the sky, but that is not true. The stars appear to be fixed because of their great distances and the relatively short times (our lifetimes) we have to look at them. It would take millions of years for us to notice with our eyes alone that the relative position of stars had changed. With telescopes, these motions can be measured in just a few years or decades. The motion of the stars in the Hyades is what allows us to measure their distance. The technique used is called the moving cluster method.

To understand how this technique works, imagine you are watching the stars in the Hyades move over a very long period. If the cluster is moving away from you (which it is), then it will appear to get smaller and smaller over time. If you could watch it long enough, it would finally vanish into some point in the sky called the convergence point. This is not different from what happens if we watch a car drive away from us on a long, straight highway where our view is unobstructed. Eventually, the car disappears into a point near the horizon. The two sides of the road also seem to meet at this point.

Of course, if the star cluster is moving toward us, then we will see it emerge from the convergence point and grow larger with time (just as a car would if it traveled toward us). This effect is that of perspective, known to all students of drawing and painting.

Play Animation

Animated Figure 6.7: In this animation, the car vanishes into the convergence point as it drives away from the observer. Credit: NASA/SSU/Dominic Nicholson

To understand how the moving cluster method can be used to measure the distance to the Hyades, we have to break down the stellar motions into their components. All motion can be treated this way. In this case it is most convenient to view separately the motion of a star across our line of sight, and its motion along the line of sight. The former is called the star's proper motion and is an angular measure. A star's proper motion is related to its tangential velocity through space, v_T, but is distinct from it. Proper motion is an angular motion, since all we can do is measure the change in the direction to the star over time. We cannot tell if this change in direction is due to the star being nearby and having a relatively small tangential velocity, or if the star is far away, but possesses a larger tangential velocity. Consider how the small-angle approximation applies here, and you will understand why this is so.

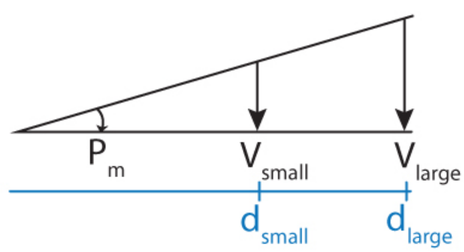


Figure 6.8: In this drawing, the proper motion measured by an observer at the far left is the angle traversed per unit time, P_m. However, the angle that is traversed on the sky is the same for a nearby star with a small tangential velocity, v_{small}, and for a more distant star with a larger tangential velocity, v_{large}. The distance to an object is given by *d*. Credit: NASA/SSU/Aurore Simonnet

The proper motion will tell us where the convergence point lies. Assuming that the random motions of the stars within the cluster are small, proper motions of all the cluster stars will point to the convergence point.

The other component of motion we can measure is called the star's radial velocity, V_R, which is its speed along our line of sight. For that we use the Doppler shift of the star's spectral lines to determine its speed according to the Doppler formula.

$$V_R = \frac{\Delta\lambda}{\lambda}c$$

Here, $\Delta\lambda$ is the shift in the wavelength of some spectral line, λ is the rest wavelength of that line, and c is the speed of light. Of course, the actual velocity of the star through space is a combination of its radial velocity and its tangential velocity. Figure 6.9 shows the relationship between proper motion, radial velocity, and actual velocity. Notice that the radial velocity and tangential velocity form the perpendicular sides of a right triangle, and the total velocity forms the hypotenuse of this triangle. Once we know one of the sides and one of the angles in a right triangle, we can find all the unknown sides using trigonometry. And just a reminder: the radial velocity, V_R, points toward or away from Earth, depending on whether the cluster moves toward us or away from us.

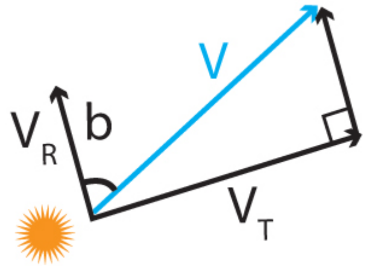


Figure 6.9: This schematic shows the relationship between the radial velocity, V_R, tangential velocity, V_T, and total velocity *V* for a star. *b* is the angle between any star and the convergence point. Credit: NASA/SSU/Aurore Simonnet

Now we use some geometry, and the fact that parallel lines meet at a point infinitely far away, to deduce one of the angles in the right triangle made up of the stellar velocity components. This geometry is depicted in Figure 6.10.

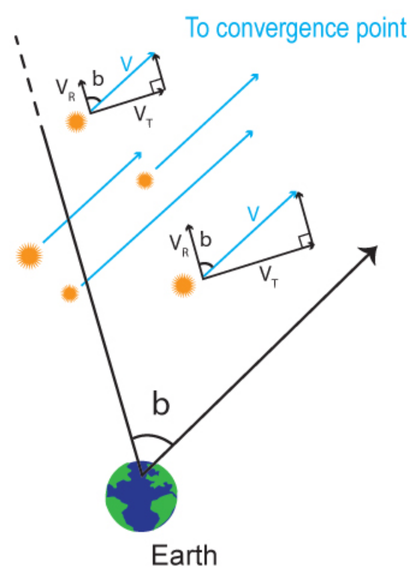


Figure 6.10: All of the stars in the cluster must move toward the convergence point, so their true space velocities (blue arrows) must point toward it. For any star in the cluster, the angle *b* between its true space velocity and its radial velocity is the same as the angle between that star and the convergence point as seen from Earth. Note how the radial velocity and the tangential velocity form a right triangle with the total space velocity. Credit: NASA/SSU/Aurore Simonnet

From Figure 6.10, we see that the angle, b, between any star and the convergence point will be the same as the angle between its radial velocity and its total velocity. This is because the convergence point is located infinitely far away, and from Euclidean geometry, parallel lines will meet at infinity. We can now use this and a little trigonometry to calculate the star's tangential velocity from its radial velocity and the angle b.

$$V_T = v_R \tan(b)$$

This velocity (V_T) relates to the proper motion as in Figure 6.8. We can use the small-angle approximation for the geometry of Figure 6.8 to write the distance as follows.

$$d = \frac{V_T}{P_m}$$

And solving for d, we have the following.

$$d = \frac{V_R tan(b)}{P_m}$$

So, to recap, if we are able to measure the proper motion for the stars in a cluster, we can determine where the cluster's convergence point lies. Knowing the proper motion, the radial velocity, the convergence point, and the angular distance from the cluster to the convergence point gives us another geometrical distance determination method.

This method was the one used to determine the distance to the Hyades before the advent of space-based parallax measurements. An interesting aspect to this is that only the Hyades cluster is close enough for this method to work well. All other star clusters are too distant. As a result, all measurements to objects farther away than the Hyades were once dependent upon an accurate distance to that cluster. Astronomers understandably expended a large effort to get as accurate a measurement of the Hyades distance as possible. Since Hyades distance measurements were only accurate to about 15% from the ground, it meant that nearly all distances had this same uncertainty attached to them. Only the stars close enough for accurate parallax measurements had more precise distance measurements. With the advent of space-based parallax measurements, these distance uncertainties have greatly decreased. The Hipparcos satellite's distance to the center of the Hyades is 46.34 ± 0.27 pc, a precision of only about half a percent.

This page titled 6.1: Geometrical Methods is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6.2: Standard Ruler

Learning Objectives

- You will know the standard ruler principle and when it applies
- You will be able to interpret examples from everyday life and from astronomy
- You will be able to perform calculations and understand conceptually the small-angle formula: which relates distance and angular size

What Do You Think: Standard Ruler

A common way to estimate distances, one we use all the time without even thinking about it, is the essence of the standard ruler technique in astronomy. If you see a friend in the distance, you have a sense of how far away she is because you see her apparent size and compare that to your knowledge of how big she would look if she were right next to you. To do this, your brain is automatically drawing on the principle of perspective: objects appear smaller the farther away they are from you.

Mathematically, this can be expressed as:

$$d = \frac{S}{\theta}$$

where d is the distance, S is the size of the object, and θ is the apparent angular size of the object. You might recognize this as the small-angle approximation. This is also the parallax relationship turned around, because both use the same principle of geometry: given the length of one side of a right triangle and an angle, you can find the length of the other side.

In the standard ruler method, we must compare the observed angular size of an object to its known intrinsic size to calculate its distance. This can be used for everything from figuring out the length of your arm to measuring the distances to galaxies.

Distance to Your Thumb

Worked Example:

1. As an example, hold your thumb out at arm's length. The width of your thumbnail spans about 1 degree. If you know that your thumbnail is 1 cm across, you can figure out how far away it is using the small-angle approximation.

We will assume that a thumb subtends an angle of about 1 degree when held at arm's length and use that to estimate the length of an arm. We will also assume the true width of a thumb is about 1 cm. Then we can use the small angle relationship.

$$\theta = \frac{S}{d}$$

As previously, S is the size (width) of a thumb and d is the distance to it (the length of an arm). We can now solve this for d.

$$\begin{aligned} d &= \frac{S}{\theta} \\ &= \left(\frac{1 \text{ cm}}{1 \text{ deg}}\right)\left(\frac{180 \text{ deg}}{\pi \text{ radians}}\right) \\ &= 57 \text{ cm} \end{aligned}$$

Notice that we had to convert from degrees to radians to use the small-angle approximation. The degrees cancel and radians are not really a unit (being the ratio of two distances), so we are left with cm for the answer.

Questions

Of course, your thumb and your arm might be slightly different from these measurements. The angle subtended by your thumb also might not be exactly a degree, but this illustrates the method of the standard ruler for measuring distances to objects.

1.

2.

https://phys.libretexts.org/@go/page/31243

The same principle that you just used in the activity can be used in astronomy. If we know, for example, how big a galaxy really is, and if we measure its apparent size on the sky, then we can measure the distance to it. The trick is to find objects whose inherent size we know to serve as standard rulers. In astronomy, we can use the small-angle approximation because astronomical objects are typically so far away that they subtend very small angles. Even 1 degree is small enough for this, and most astronomical objects are even smaller; the full moon is only half a degree across, to give a concrete example.

Standard Ruler: Using Your Hands to Measure Angles

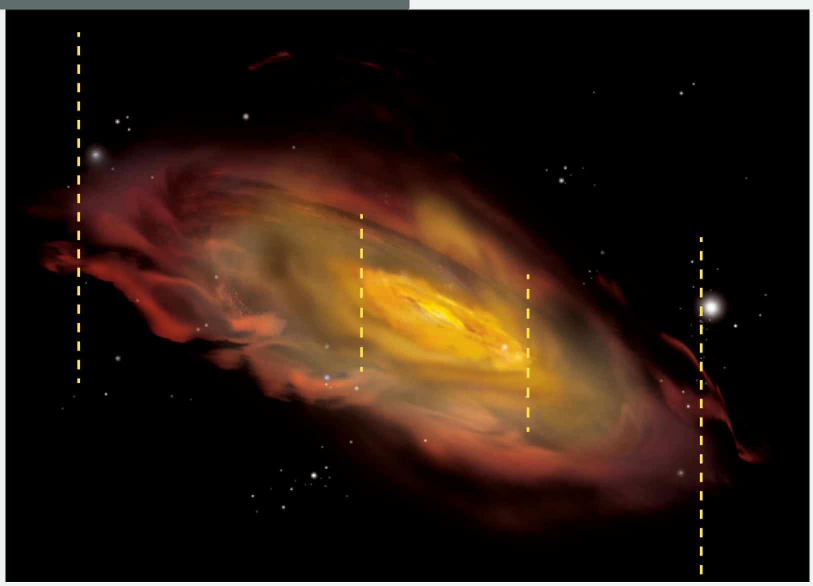


Figure A.6.2: Artist's conception of a galaxy. Credit: NASA/SSU/Aurore Simonnet

A. Measurements.

To see how the standard ruler technique works quantitatively in everyday life, you will be investigating the relationship between size, distance, and apparent angular size. You will need a ruler or a tape measure to do this activity.

1.

https://phys.libretexts.org/@go/page/31243

2.

3.

4. Now use the graphing tool to plot your data from the summary table on a graph, so that distance from the screen is on the x-axis, and angular size in degrees is on the y-axis. Use the default color to plot the data for the disk. When you have plotted the three points for the disk, click the box with the plus underneath the Data Set Control panel to add a second data set and choose a different color for the points for the core. Now add the data points for the core to the graph using this new color.

B. Discussion

1.

ADAPT 6.2.1

2.

ADAPT 6.2.1

3.

ADAPT 6.2.1

4.

ADAPT 6.2.1

5.

ADAPT 6.2.1

Standard Ruler in Astronomy

In this activity, you will see four different spiral galaxies. In the standard ruler method, all of these galaxies are assumed to have the same intrinsic size. Therefore, the relative sizes of the galaxies will provide information about their relative distances. To do the activity:

- Hover over each galaxy to reveal its apparent size.
- Rank the galaxies in order of increasing distance using the selection boxes below the images.
- Click “submit” to check your answer.
- Click “reset” to reshuffle the galaxies, and try again, if desired.

Play Activity

Several types of astronomical objects have been studied in the hope of using them as standard rulers, for example, the diameters of galaxies or the largest ionized gas clouds in a galaxy. Unfortunately, many have not turned out to be very “standard” in size, but instead, have too much variation in their intrinsic size to be useful rulers.

This page titled 6.2: Standard Ruler is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6.3: Standard Candle

What Do You Think: Brightness

6.3.1: Standard Candles and the Inverse Square Law

Recall flux is a measure of the apparent amount of energy from an astronomical object and luminosity is a measure of the object's true total power output. When you look at a distant streetlight at night, how do you know how bright it really is? It could be an extremely bright light shining at a great distance, or it could be a dim light that is much closer. How are you to know which is the case? Your experience might tell you that all streetlights have roughly the same brightness. Under this assumption, when you see a dim streetlight you have an intuitive sense that it appears dimmer because it is far away (Figure 6.11). But by how much? In the next activity, we will explore the relationship between distance and brightness (Figure 6.11).

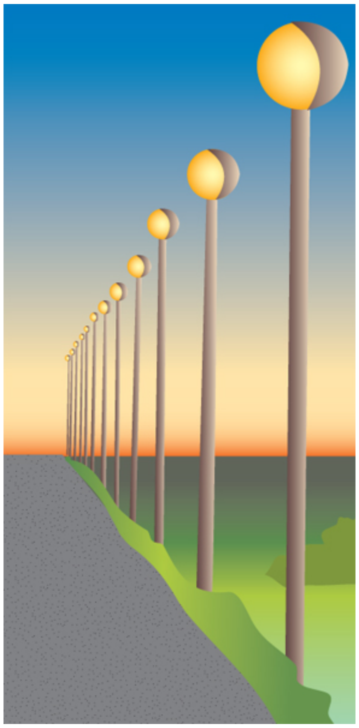


Figure 6.11: Streetlights that are farther away look dimmer. Since all streetlights are roughly the same brightness inherently (same luminosity), we know that the dimmer-looking ones only look dimmer because of their distance. Credit: NASA/SSU/Aurore Simonnet

Flux Vs. Distance

In the following activity, you will collect the flux of light around a star by surrounding the star with spheres of ever-increasing size to determine how much lower the flux becomes with distance.

Play Activity

1.

https://phys.libretexts.org/@go/page/31244

2.

3.

4.

https://phys.libretexts.org/@go/page/31244

5.

The mathematical expression relating the flux of an object to its distance is known as the **inverse square law**.

$$F = \frac{L}{4\pi d^2}$$

In this expression, d is the distance to an object, F is its flux (also known as apparent brightness, or intensity), and L is its luminosity (absolute or intrinsic brightness). This means if an object moves twice as far away, it will look four times dimmer than it did originally. If it moves three times farther away, it will look nine times dimmer, etc. This is illustrated in Animated Figure 6.12.

Play Animation

Animated Figure 6.12: As the lightbulb moves farther away, its apparent brightness decreases by the square of the distance. Here, the size of the bulb is used as a proxy for brightness. Credit: NASA/SSU/Kevin John

The SI units for luminosity are watts, and the units for flux are watts/m^2, which makes sense because the units for distance are meters. Because of the large numbers involved with astronomical luminosity, sometimes it is expressed in terms of the Sun's luminosity, which is 4×10^{26} watts, much brighter than a lightbulb! So, if something is "1 solar luminosity," it is 4×10^{26} watts.

With our detectors, we can measure an object's flux, that is, how bright it appears. The flux is related to its intrinsic brightness and its distance through the inverse square law. The standard candle technique employs the inverse square law to calculate the distance to an object of known luminosity. The key to using the standard candle method is finding similar objects that all have the same luminosity (or at least known luminosities), so that if we know what type of object we are looking at, we can just look up its luminosity, measure its flux, and use the inverse square law to deduce its distance. The search for standard candles in astronomy is a bit like looking for objects that have the equivalent of a lightbulb's wattage stamped on them.

Understanding the Inverse Square Law

1.

2.

3.

USE GRAPH

For more practice, answer the following questions below. Star A and Star B have the same luminosity.

- Saying that something is a fraction as bright is the same thing as saying it is dimmer. For example, if something is 10 times dimmer, it is 1/10th as bright.
- Saying that something is a fraction as far is the same thing as saying it is closer. For example, if something is 10 times closer, it is 1/10th as far.

You can choose to state things either way (choose bright or dim, far or close below) as long as you are clear about your choice.

4.

5.

6.

7.

8.

Using the Inverse Square Law

The inverse square law is a very powerful tool used by astronomers to calculate distances to objects that they observe. The inverse square law equation is written as below.

$$F = \frac{L}{4\pi d^2}$$

Here we say that F is the flux (also known as the observed or apparent brightness), measured in watts/m^2; L is the luminosity (also called the intrinsic brightness), measured in watts; and d is the distance to the object, measured in meters. However, for astronomical distances, meters might not be a very convenient or intuitive unit, so remember that 1 light-year = 9.46×10^{15} meters and 1 parsec = 3.26 light-years.

1.

2.

Standard Candles

Standard candles have known intrinsic luminosities. Typical Type Ia supernovae can be used as standard candles because they all have intrinsic luminosities that are about 10^{44} W at the peak.

In this activity, we are going to compare four "target" supernovae to a "reference" supernova to determine how much farther away the target supernovae are compared to the reference supernova.

The reference supernova is assumed to have a flux of 100 W/m^2. The four target supernovae have fluxes that are lower than the reference supernova and are therefore farther away. To determine how much farther away the target supernovae are:

- Move the mouse over each supernova to reveal its flux.
- Compare this flux to the baseline of 100 W/m^2.
- Use the ratio of the two fluxes to determine the relative distance compared to the "reference" supernova distance according to:
  Relative distance = (100/observed supernova flux)$^{1/2}$
- Choose the correct answer for each supernova. When done, press "submit" to check your answers.

Play Activity

Some of the specific objects and relationships astronomers have used as "standard candles" include: spectrally classified stars on the stellar main sequences, certain types of variable stars, the relationship between a galaxy's rotation speed and its luminosity, and type Ia supernovae. We will discuss each of these in turn below.

Most of the examples we will use in the next sections will describe the luminosities in terms of magnitudes. If you have not read **Going Further 3.6: The Magnitude System**, now would be a good time to do so.

6.3.2: Spectral Classification and Stellar Main Sequences

An important rung of the distance ladder, called main sequence matching (or main sequence fitting), depends on a comparison of the main sequence for one star cluster to that of another. This provides a relative distance between the two clusters, but not an absolute distance. If the absolute distance to one of the clusters is already known through some other means (such as the moving cluster method described in Section 6.2), then the absolute distance to the other cluster can be found through this comparison.

From our discussion of Planck or blackbody radiation, we expect a hotter star to have a higher luminosity for a given size. This is because, for a blackbody, the emitted radiation is proportional to temperature to the fourth power ($F \propto T^4$). We also expect a bigger star to have a higher luminosity, simply because there is more surface area radiating. However, main sequence stars of a given spectral class are nearly the same size. Observationally, we do find that for many stars there is a simple relationship between temperature and luminosity. All main sequence stars of a particular spectral class (which is related to their surface temperature in most cases) have about the same intrinsic luminosity.

Therefore, if we see a main sequence star of a certain spectral type whose distance we wish to measure, we can assume that it has nearly the same luminosity as a nearby main sequence star of the same spectral type. Then we can observe its flux and calculate its distance using the inverse square law. The method works for distances out to tens of thousands of light-years. That is about 10 times farther than the parallax method. However, our ability to use the method still relies on first using parallax on nearby stars to determine the specific relationship between spectral class and luminosity - to calibrate it, in other words.

One challenge to using the spectral type of a star to measure its distance is that astronomers cannot always tell if a star is on the main sequence or not. But there is a foolproof way to get around this problem: compare the main sequences of star clusters. We know that for clusters, the main sequence will stand out in a Hertzsprung-Russell diagram. So plotting the stars in a cluster in a graph of temperature (spectral type) vs. apparent brightness will make the main sequence stars obvious. They can then be compared to a standard HR diagram to determine the distance to the cluster.

Main Sequence Matching for Clusters of Different Distances

This activity shows the main sequences of two star clusters in a Hertzsprung-Russell diagram. Along the horizontal axis is the spectral class, and along the vertical axis is brightness. In the exercise, brightness is given in terms of apparent magnitude. Because all main sequence stars of a given spectral class have the same brightness, the only reason that these two clusters do not lie on top of each other is because they lie at different distances from us. Your job is to slide the lower one up and down until it is superimposed over the upper one. The offset in magnitude will provide a magnitude difference between the two clusters, and this difference can be turned into a difference in distance by using the inverse square law.

The top cluster (shown in blue) is the Hyades, and we have already mentioned that its distance is known from precise Hipparcos satellite parallax measurements. The distance is 46 pc. The other cluster (shown in yellow) lies below the Hyades in the figure. Recall, the greater the apparent magnitude, the fainter the object is.

Play Activity

1.

2.

Use the slider button on the right to slide the cluster main sequence up and down until you have lined it up as best you can with the main sequence of the Hyades cluster. You can read out the difference in magnitudes and the distance on the right. Record these below.

Be careful to use the right side of the main sequence to do the matching and ignore the left side. On the left is where the more massive stars are evolving off the main sequence to become giants, so the matching technique is not valid there. Also, note that you can tell that the Hyades is slightly older than the other cluster because its main sequence turn-off point has moved to lower-mass, longer-lived stars.

3.

4.

This method can be used to determine the distance to any cluster by comparison to the Hyades. All that is required is to create a Hertzsprung-Russell diagram like these, with apparent magnitudes plotted instead of absolute magnitudes.

By matching the main sequences of star clusters, we can compare objects within the disk of our Galaxy to those out into its halo. This is fortunate. Halo objects are much too distant for ground-based parallax measurements. Space-based parallax observations are removing this limitation for halo objects.

6.3.3: Cepheid Variable Stars

Cepheids are a class of extremely bright variable stars. Their luminosity is related to the period of their pulsations: longer-period stars are brighter than shorter-period ones. They are named after the star δ-Cephei (delta Cephei) in the constellation of Cepheus, the first identified example of this particular type of variable star. δ-Cephei can be seen with the naked eye.

This method can be used to measure distances out to about 100 million light-years. Until some Cepheid stars were found close enough to measure their distances using other methods—geometrical methods and main sequence fitting were used historically—

Cepheids gave only relative distances. The distances used to calibrate Cepheid luminosities had large uncertainties at first, and these led to large uncertainties in derived distances from Cepheids. Interestingly, the period–luminosity relation for Cepheids was discovered without knowing either their distances or their luminosities.

In 1912, the astronomer Henrietta Leavitt (1868–1921, Figure 6.13), working at the Harvard College Observatory, discovered 20 Cepheid variable stars in the Small Magellanic Cloud (SMC), a small satellite galaxy of the Milky Way. Leavitt knew they were Cepheids because of their distinctive light curves, which make them easy to distinguish from other kinds of variable stars. They display a rapid rise to maximum brightness, and then a more gradual decline (Figure 6.14).

Figure 6.13: American astronomer Henrietta Leavitt at the Harvard College Observatory. Credit: American Institute of Physics

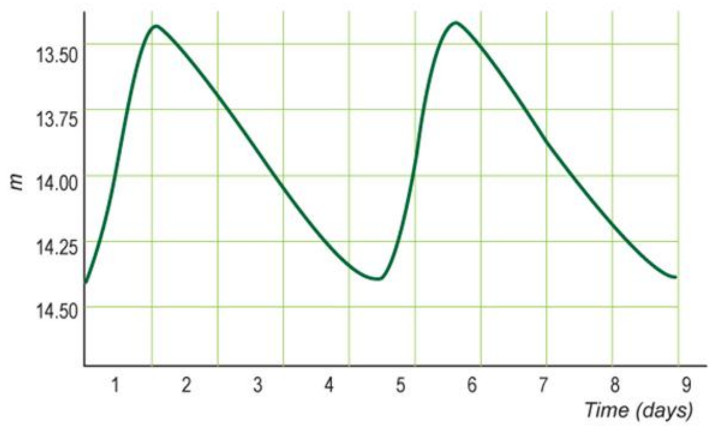


Figure 6.14: Typical Cepheid light curve. A light curve is a graph of brightness vs. time. The light curve for a Cepheid variable star has a characteristic shape, with the brightness rising sharply and then falling off much more gently. The amplitude of the variations is typically 1–2 magnitudes. The period is the time between successive peaks (or troughs) in the curve. Credit: NASA/SSU/Aurore Simonnet

The variations in distance to the individual Cepheid variable stars in the SMC are negligible compared with the much larger distance to the SMC. Essentially, to a good approximation, the stars are all at the same distance. This means that the apparently brightest stars in this group are also intrinsically the brightest. Henrietta Leavitt noticed that the period of a Cepheid variable in the SMC is proportional to its average brightness: brighter Cepheids pulsate with longer periods. This is called the period–luminosity relation for Cepheids. The relation is sometimes also called the Leavitt Relation or Leavitt Law, to honor its discoverer.

By measuring the period of any Cepheid, one can deduce its intrinsic brightness from the Leavitt period–luminosity relation. Then, by measuring the star's apparent brightness, one can calculate its distance using the inverse-square law. This simple procedure makes Cepheid variable stars one of the most important standard candles in the Universe. Not only are they useful distance indicators themselves, they can also be used to calibrate other distance indicators.

Unfortunately, no Cepheids are close enough to allow for an accurate parallax measurement from the ground. To calibrate the Cepheid period–luminosity relation historically, a different calibration, based upon the distances to star clusters, had to be used. The first step was to determine the distance to the Hyades, as we have already discussed. But the Hyades cluster contains no Cepheids, so more distant clusters that did contain Cepheids were compared to the Hyades via main sequence matching. This

https://phys.libretexts.org/@go/page/31244

allowed an absolute calibration for Cepheids to be determined, and this finally allowed an absolute distance for the SMC to be determined. (The actual means by which Cepheids were calibrated was more complicated than this. It had many independent branches, but the distance to the Hyades played an important role for most of them. We are leaving out some of the details for brevity's sake.)

We have now taken several steps up the cosmic distance ladder, and just as with an actual ladder, each successive step has relied on previous steps. Hopefully, you have begun to understand how astronomical distance measurements somewhat resemble climbing a ladder, where each successive step depends on the previous ones.

Thus far, we have moved from Solar System distances out into the Milky Way, and then to "nearby" galaxies. As we look at the last few rungs of the ladder, we will see how they allow us to measure distances farther and farther across the visible universe.

One of the main scientific reasons for building the Hubble Space Telescope (HST) was to measure the distances to galaxies in the Virgo cluster of galaxies. This project was known as the HST Key Project. The measurements relied on observations of Cepheid variables located in Virgo galaxies. Atmospheric distortions combined with the great distance to Virgo prevent astronomers from measuring Virgo's Cepheids from the ground. A telescope in space overcomes this problem.

The Virgo Cluster contains some 2,500 galaxies, among them the spiral Messier 100 (Figure 6.15). From repeated observations of M100, astronomers were able to identify more than a dozen Cepheid variables and measure their brightness as a function of time. From the Hubble measurements of Cepheid variables in M100, the periods of the variables, and hence, their intrinsic luminosities could be determined (see Figures 6.16 and 6.17). Once both the intrinsic and apparent brightnesses were known, the distance to M100 could be calculated. Such observations of M100 and other galaxies in Virgo enabled astronomers to calibrate the period–luminosity relation to higher precision than had been possible in the past. In turn, it allowed calibration of distance measurement techniques suitable for larger distances than with Cepheids alone.

Figure 6.15: The spiral galaxy Messier 100, as imaged by the Hubble Space Telescope. Credit: NASA/STScI

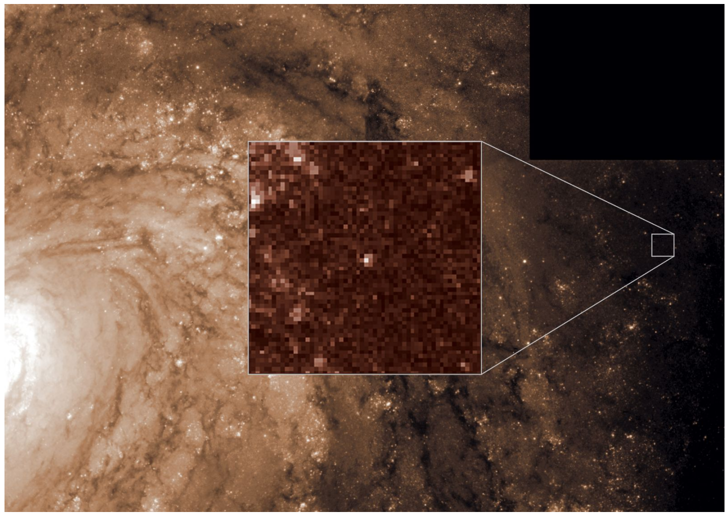


Figure 6.16: Hubble Space Telescope tracks down Cepheid variable stars in M100. Hubble's high-resolution camera detected and picked out one of the Cepheid variable stars used in the next activity. The star is located in a star-forming region in one of the galaxy's spiral arms (the star is at the center of the box). Credit: ESA/ESO

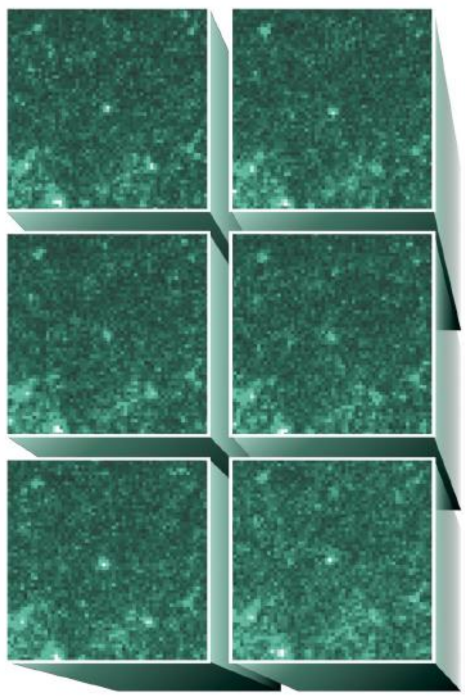


Figure 6.17: Six images taken at different times depicting one of the Cepheid variable stars in the galaxy M100 are shown. The Cepheid is in the center of each frame. It is clear that the Cepheid varies in brightness over time. Credit: ESA/ESO

Measuring the Distance to M100 Using Cepheids

In this activity, we will measure the distance to M100 using Cepheids, retracing the steps of the HST Key Project Team.

A. The Period-Luminosity Relation for Cepheids

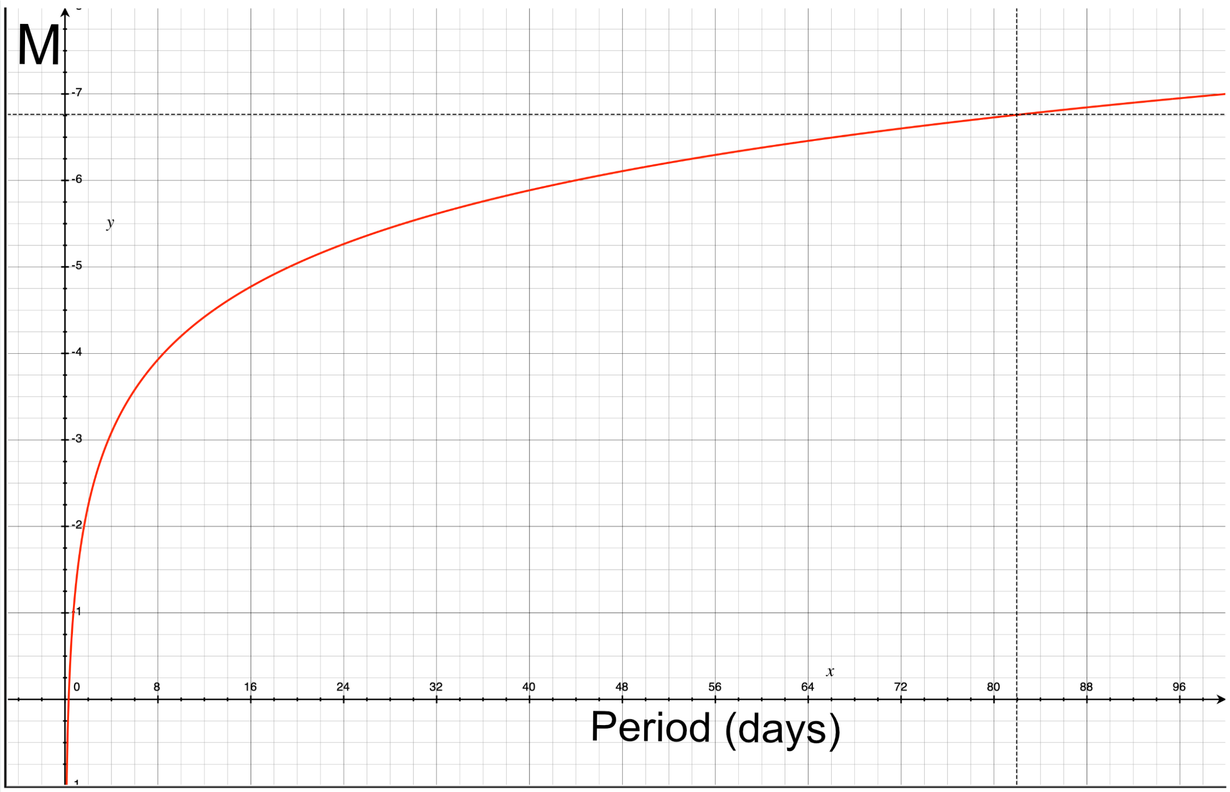
The relation between a Cepheid's period (the time over which its brightness varies) and its absolute magnitude (a measure of its luminosity) is given by the following expression:

$$M = -2.8 log(P) - 1.4 \tag{6.3.1}$$

where M is the absolute magnitude, *log* is the base 10 logarithm, and P is the period in days.

This expression is shown in graphical form in the following figure:

https://phys.libretexts.org/@go/page/31244



Absolute magnitude (plotted along the vertical axis) vs. period in days (running along the horizontal axis) for Classical Cepheid stars is shown by the red curve. This is the Leavitt Period-Luminosity relation from above shown in graphical form.

1.

2.

B. You will now examine Cepheid light curves from HST. (Credit: Ferrarese et al., 1996, ApJ, 464, 568.)

1. From the light curves, determine the periods of five Cepheids and enter the periods that you measured into the data table. The period is the time between successive peaks (or troughs) of the light curve. Click on the curves to read the data. Enter the period into the data table below in section D.

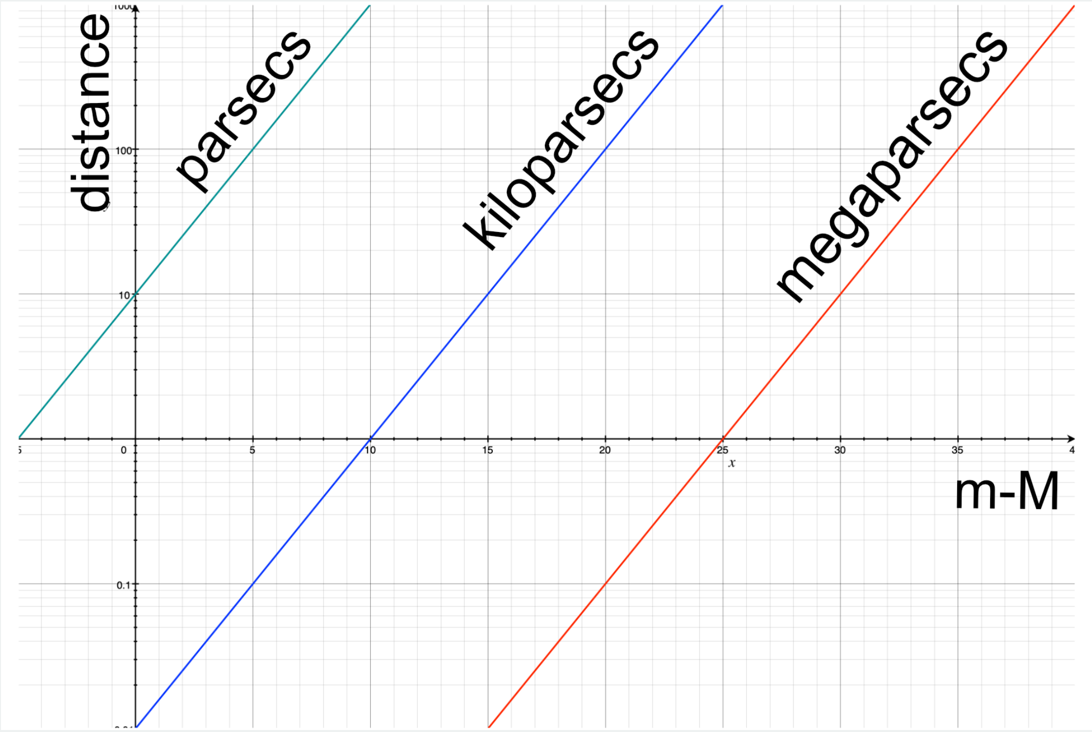
2. Determine the absolute magnitude M of each Cepheid from the period using either the graph above or the expression in part A, and enter it into the data table below in section D.

C. The Inverse Square Law: Absolute and Apparent Magnitude

The absolute magnitude of a star corresponds to the luminosity of a star (related to its inherent brightness), and the apparent magnitude corresponds to the flux (related to how bright it appears to you). If a star is very far away from you, it will appear dim even if it is a very bright star. Thus, the difference between the absolute magnitude and the apparent magnitude is related to the distance. The relation can be written mathematically as:

$$d = 10^{(m-M+5)/5} \tag{6.3.2}$$

where d is the distance in parsecs, m is the average apparent magnitude and M is the absolute magnitude. This relation can be derived from the inverse square law. This expression is shown in graphical form in the following figure.



The figure above shows the distance vs distance modulus, m-M, relation. It has a linear axis for distance modulus and a logarithmic axis for distance. To use the figure, find the appropriate value of distance modulus along the horizontal axis, then scan directly upward until you intercept either the green, blue or red line. Where the line meets your chosen distance modulus you can read the corresponding distance by scanning horizontally over to find the value for distance on the vertical axis: The green line gives the distance in parsecs, the blue line gives it in kiloparsecs and the red line gives the distance in megaparsecs. In the areas where the lines overlap, you can use either unit. For example, for a distance modulus m-M=20, the figure shows that the distance is 100 kiloparsecs (where the blue line intersects m-M=20) or, if you prefer, 0.1 megaparsecs; this is where the red line intersects m-M=20. These are the same distance expressed in different units, and you can use whichever unit you like. This unusual way of plotting the relation allows us to more easily see distances corresponding to a wide range of distance modulus.

1. The average apparent magnitude is halfway between the maximum and minimum brightness in the light curve. Measure the average apparent magnitude m for each Cepheid that you previously studied in part B. Enter your results in the data table in section D.

2. Using the values of the absolute and apparent magnitude for each Cepheid, determine its distance d. Enter your results in the data table in section D.

D. Data Table

E. Discussion

1.

2.

3.

6.3.4: Tully-Fisher Relation

The **luminosity** (L) of a spiral galaxy is proportional to its rotational velocity (v) to the fourth power:

$$L \propto v^4$$

The relationship is called the **Tully-Fisher relation**. By using spectral lines, we can easily determine the rotational velocity, and hence, the luminosity of a galaxy. A similar relation, called the Faber-Jackson relation, holds for the characteristic velocities of the stars in elliptical galaxies and is also measured from the galaxies' spectra. These methods work for greater distances than methods based on stars because galaxies are much brighter than stars and can be seen to much greater distances than can individual stars. In practice, the Tully-Fisher and Faber-Jackson relations can be used out as far as we can measure the spectral lines of a galaxy. This is out to billions of light-years.

The relationship between stellar orbital speed and brightness becomes apparent when astronomers study galaxies in a given galaxy cluster. The galaxies in a cluster are essentially all at the same distance, and so this is reminiscent of how Henrietta Leavitt observed Cepheids in the Small Magellanic Cloud to discover the period–luminosity relation. Without a nearby calibration, only relative distances can be determined, not absolute distances. The HST Key Project for Cepheids in the Virgo Cluster of galaxies provided a precise local calibrator for the Tully-Fisher and Faber-Jackson relations.

Tully-Fisher Relation

In this activity, you will study the Tully-Fisher relationship between rotational velocity and intrinsic brightness for a set of spiral galaxies. The method will bear some similarities to the Cepheid distance method in that there is a relationship between luminosity of an object and an easily measurable quantity. In the case of Cepheids, the relationship is between luminosity and period of pulsations. For spiral galaxies, the relationship is between luminosity and rotation speed of the galaxy.

The rotation speed of a galaxy is determined from its spectrum, specifically the 21-cm emission line from neutral hydrogen gas. If a galaxy is rotating, the gas in the part of the galaxy rotating toward us will show a blueshift, while the gas in the part of the galaxy rotating away from us will show a redshift. However, this motion is superimposed on the net motion of the galaxy in space. This motion is almost always much larger than the rotation speed of the galaxy, and it is away from us for all but a few galaxies. So the net effect is that the entire emission line will be redshifted, but the parts rotating away will have a larger redshift than the parts moving toward us. The center of the galaxy will not be rotating toward or away from us.

The emission line from hydrogen in a galaxy looks strange. That is because the entire galaxy contains hydrogen and the entire galaxy is rotating. We see different components of that rotation depending on where we look. For instance, the outer parts are moving almost straight toward us or straight away from us, so we see the entire motion reflected in the redshift there. But if we look toward the center of the galaxy, we see a lot of gas that is traveling across our line of sight, and only a little that is moving toward or away. This effect causes the emission line to be smeared out in a broad plateau. It is the width of this plateau that gives us the rotation speed of the galaxy.

Play Activity

A. Determining the Tully-Fisher Relation

In the first part of this activity, 10 galaxy spectra are available. You should use at least six of them in order to obtain a good determination of the Tully-Fisher Relation.

1.

2.

3.

B. Using the Tully-Fisher Relation to Estimate the Distance to Galaxies

In this part of the activity, you will use the Tully-Fisher Relation that you created in Part A to determine the distances to three additional galaxies.

1.

This is the general way that the Tully-Fisher relation is used to measure the distances to galaxies. We have omitted some of the complicating details so that you get a better understanding of how the method works.

Now that you have done the Tully-Fisher activity, you may be interested in seeing the original data (Figure 6.18), published in 1977. R.B. Tully and J.R. Fisher are shown in Figure 6.19.

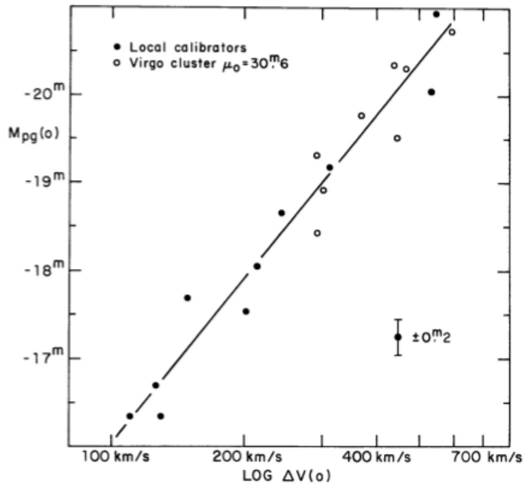


Figure 6.18 The original data published by Tully and Fisher in 1977. The plot shows the photographic magnitude of the galaxy plotted vs. the rotation speed. Credit: Tully, R.B. and Fisher, J.R. 1977, Astronomy and Astrophysics, 54,3, 661

Figure 6.19: (a) R. Brent Tully and (b) J. Richard Fisher today. R. Brent Tully is an astrophysicist at the Institute for Astronomy in Hawaii. He went to graduate school at the University of Maryland before moving to Hawaii, one of the world's best places to do astronomical observations at visible wavelengths. J. Richard Fisher is a radio astronomer, working at the National Radio Astronomy Observatory in Charlottesville, Virginia. Dr. Fisher also went to graduate school at the University of Maryland and is an adjunct professor at the University of Virginia. Credits: (a) www.ifa.hawaii.edu/~tully/ (b) www.astro.virginia.edu/people//faculty/jrf2v/

The Faber-Jackson relation (Figure 6.20) was developed in 1976 by astronomers Sandra Faber (Figure 6.21) and Robert E. Jackson. It relates the luminosity of an elliptical galaxy to the spread in velocities observed near the galaxy's center.

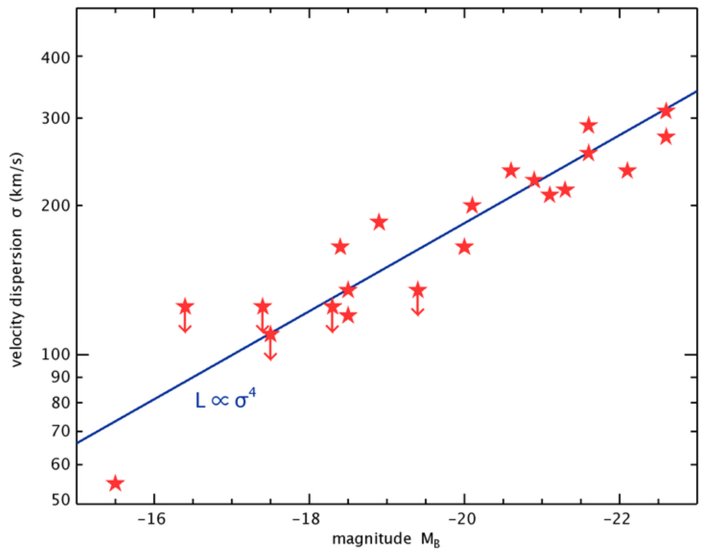


Figure 6.20: Data used in the original Faber-Jackson relation. The plot shows the velocity dispersion vs. the galaxy's brightness in magnitudes. Credit: Faber, S.M. and Jackson, R.E. 1976, Astrophysical Journal, 204,668

Figure 6.21: Sandra Faber is a professor at the University of California, Santa Cruz, and is a prominent cosmologist who has won many awards for her work, including the Heineman Prize, the Harvard Centennial Medal, and the Franklin Institute's Bower Award. She is a member of the National Academy of Sciences and was an early leader in the use of the Hubble Space Telescope for cosmological observations. Robert E. Jackson was Faber's graduate research assistant at UC Santa Cruz, and he helped analyze the data that led to the determination of the relation. Credit: http://astro.ucsc.edu/~dept/faculty/faber.html

6.3.5: Type Ia Supernovae

The shape of the light curve of certain supernovae is related to their luminosity. These supernovae, designated as type Ia, are the result of the thermonuclear explosion of a white dwarf star. This distance method works out to a few billion light-years because the supernovae are so bright that they can be seen at large distances. Supernova distances can be calibrated with the Tully-Fisher and Faber-Jackson relations. Because supernovae are quite rare, we do not have examples in galaxies that are close enough to allow the use of Cepheids for distance calibration. The top panel in Figure 6.22 shows the light curves from many type Ia supernovae, while the bottom panel shows the same supernovae corrected to have the same absolute brightness and decay time. This illustrates how these supernovae can be used as standard candles. A light curve, as we discussed in section 6.3.3, is a plot of brightness vs. time.

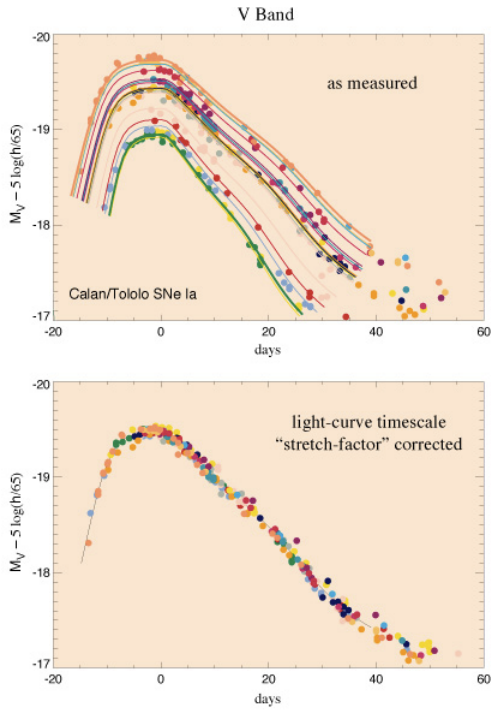


Figure 6.22: The top panel shows the light curves (light emitted as a function of time) of many different type Ia supernovae, which have occurred at different distances. The bottom panel shows how these supernovae can be used as standard candles once their light curves have been corrected by the "stretch-factor." Credit: Courtesy of Alex Kim and the Supernova Cosmology Project

When a star explodes as a supernova, its brightness rises very rapidly, increasing in about three weeks until it rivals the brightness of the galaxy in which it is located. The brightness then begins to turn over, and the star slowly fades over many weeks or months. In the case of Type Ia supernovae, the peak brightness reached is nearly the same for all of them. This is not true for other types of supernovae.

During the 1980s and 1990s, astronomers studying supernovae found that light curves for this kind of supernova had a very useful property. The dimmer the peak brightness of the supernova, the faster it dropped in brightness in the weeks following the peak. In fact, astronomers found a simple linear relationship between the absolute magnitude of a type Ia supernova and its drop in magnitude in the first 15 days after its maximum brightness. This relationship is called the $\Delta m_{15}(B)$ relationship, with *B* indicating that the measurement is to be made using the *B* (blue) filter on the telescope and camera.

The reason this relationship is so useful is that it allows a simple timing measurement to be converted into a brightness measurement, and thus, for type Ia supernovae to be used as standard candles. This should be familiar to you from your study of Cepheid variables.

Examples of supernova light curves are shown in Figure 6.23. The left-hand panel shows the light curves of five supernovae, with absolute magnitude being plotted vs. time—these supernovae occurred in galaxies of known distance, so their absolute magnitudes could be determined from their apparent magnitude and the distance to the galaxy. The right-hand panel shows the same supernovae, but they have been shifted such that their peak brightness coincides. Some of these curves decrease in brightness more quickly than others. But much more interesting, if you compare the two plots, you will see that the faster the drop from peak brightness, the dimmer the supernova. Furthermore, after about 20 days, all of the light curves drop at the same rate, so it is this initial drop from peak brightness that is important for determining the absolute magnitude of the supernovae. There is a linear relationship that gives the peak absolute brightness in terms of the decrease in brightness over the first 15 days after maximum, $\Delta m_{15}(B)$:

$$M_B = -0.748[\Delta m_{15}(B) - 1.1] - 19.258$$

In this equation, M_B is the absolute magnitude in the B (blue) filter.

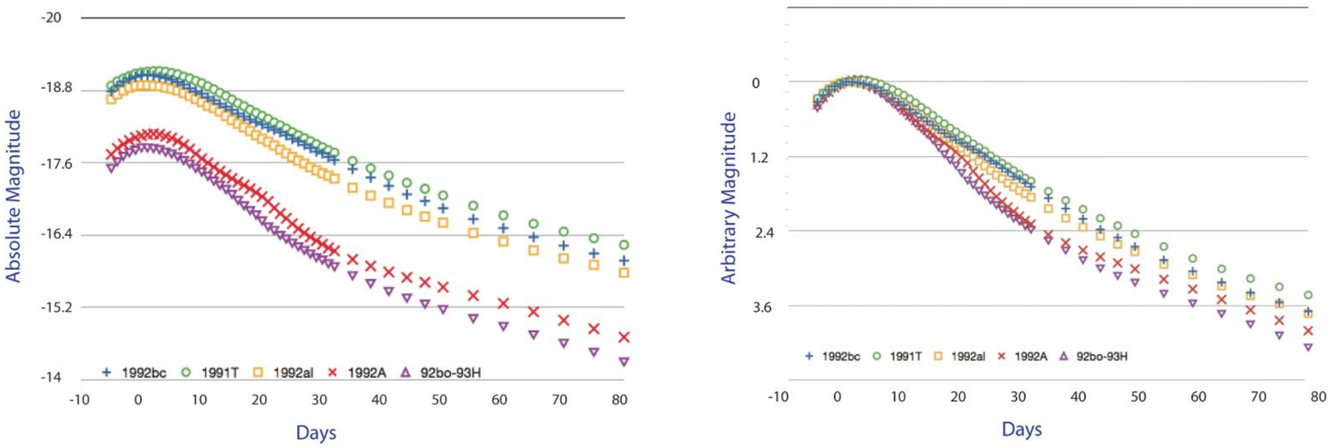


Figure 6.23: Light curves from five selected type Ia supernovae. The left-hand panel shows the absolute brightness for each supernova as a function of time. The right-hand panel shows the results when the light curves are adjusted to have the same value for peak brightness. Note that the dimmer supernovae in the left-hand panel fall faster in the right-hand panel. Credit: NASA/SSU/Aurore Simonnet

Supernovae Type Ia and Distances

In this activity, you will use the relationship between the peak absolute brightness and the decrease in brightness over the first 15 days after maximum to determine the absolute brightness for several supernovae. You will then estimate the distances using the relationship between absolute brightness and the apparent brightness.

Play Activity

1. Plot the light curve for the first supernova, SN 1991T, by clicking on the button labeled with its name. Brightnesses are given in magnitudes and the values for the points can be found by moving your mouse over them.

2. Now click the Calculate Magnitude button. You will see the equation for the absolute magnitude, MB, displayed. Click the “equals” button to compute and display the absolute magnitude. Record it in the data table.

3. Now click on the Calculate Distance button. The expression for the distance to the supernova will display. This will not look like the inverse square law for light, but it is. Recall that we are measuring brightness in magnitudes, and that those are related to the log of the true brightness. That is why the formula might not look the way you expect. It will be filled in with the correct absolute magnitude, which you have just calculated, and the corresponding apparent magnitude, which is read from the point you chose as the peak brightness of the light curve. Hit the “equals” button to display the distance in Mpc. Record the distance in the data table.

4.

5.

6.

7.

This activity outlines the basics of how astronomers use supernovae to determine the distances to galaxies. Only galaxies with a supernova observed in them can be measured this way, but even distant galaxies can be measured because type Ia supernovae are so bright. Supernovae are rare, but there are many, many galaxies in the distant universe. As a result, the type Ia supernova distance method has been extremely useful for building our understanding of the evolution of the Universe. We will return to this topic again when we study the cosmic expansion of the Universe in later chapters.

This page titled 6.3: Standard Candle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6.4: The Cosmic Distance Ladder

The Distance Ladder

In this chapter, we have explored various techniques astronomers use to calculate distances in space. Different techniques work well for different distances, but the techniques for objects farther away typically rest on calibrations from studies of similar objects that are closer. For example, getting the luminosity of a distant star by knowing its spectral class (main sequence fitting) depends on finding the luminosity of many closer stars of the same class by means of some other technique (e.g., parallax or moving-cluster method). As we go farther out into the Universe, distance measurements typically become both more difficult and uncertain. To be confident of our results, it is important that we use multiple independent techniques.

Several of the most important distance techniques have been covered in this chapter. If they give a consistent distance for an object or set of objects (for which each method employed is suitable), then we can be fairly confident that we have a good estimate for the distances we obtain.

Match the Technique to the Distance

Drag and drop the tiles to match the distance–finding techniques with the appropriate astronomical object(s) and distance range.

In the activity, “MW” refers to the Milky Way and "SNe" refers to supernovae.

When you are done, hit the Check Answers button to check your answers. Revise as necessary.

Play Activity

The final correct figure is called the cosmic distance ladder. The ladder shows the individual steps used to construct the astronomical distance scale. As you read down the ladder, each step depends on the steps above, and goes out to farther distances.

This page titled 6.4: The Cosmic Distance Ladder is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6.5: Wrapping It Up 6 - The Supernova of 1885

In the year 1885, what was thought to be a bright nova ("new star") was observed in the center of M31, now known to be the Milky Way Galaxy's closest large galactic neighbor, the Andromeda Galaxy. The nova was originally called S Andromedae (S And for short), using the letter-constellation naming scheme commonly employed for variable stars, of which novae are one class. At the time, the distance to M31 was unknown, so the true brightness of this object was also not known to astronomers of the day. Since that time, we have learned that M31 is actually quite distant, and that the object observed in 1885 was not a nova, but what we now call a supernova—an exploding star, millions of times brighter than the much more common novae.

In the months after its discovery, S And gradually faded from view. However, more than a century later, two American astronomers, Robert Fesen of Dartmouth University and Andrew Hamilton of the University of Colorado, managed to obtain an image of the supernova remnant using the 4-meter telescope at Kitt Peak Observatory in Arizona. This image was followed up with much more detailed images using the Hubble Space Telescope. In addition, the pair was able to measure an absorption spectrum from the remnant, also using Hubble. In this activity, you will use the data acquired with the Hubble Space Telescope to measure the distance to the supernova remnant, and hence, to the center of the nearby galaxy, M31, in Andromeda.

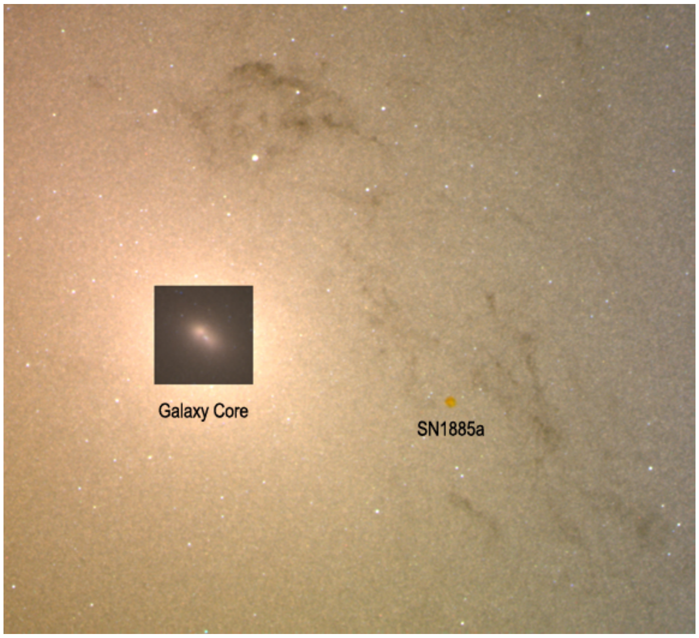


Figure A.6.4: This image shows the center of the galaxy M31. The dark inset is an image of the core, and the orange circle to the lower right is SN 1885a. Credit: Courtesy of Andrew Hamilton, CU Boulder.

If you examine the Hubble image of the core of M31 (see Figure A.6.4), you will notice a faint circle to the right of and below the bright nucleus of the galaxy that appears darker than the surrounding regions. The circle appears quite small. That is the supernova remnant of S And, the star that was seen to explode in 1885, today known as SN 1885a. The remnant appears darker because it is composed of an expanding cloud of cold gas. It absorbs and scatters light emitted from behind it, just as thick clouds of dust or water vapor in Earth's atmosphere absorb and scatter light from the Sun, making them appear darker than the surrounding sky from our vantage on the ground.

To measure the distance, you will use the following procedure:

1. Use the image to measure the angular size of the remnant.
2. Use the spectrum to estimate the expansion velocity of the cloud.
3. Use the age of the remnant and its expansion velocity to determine the physical size of the cloud.
4. Use the small-angle formula to determine a distance from the angular size and physical size.

Play Activity

6.5.1: Part I: Measuring the Angular Size of the Supernova Remnant

Click the "Next" button.

An expanded version of the HST image of the supernova remnant is provided. The scale has been increased so that you can see the individual pixels of the image. Each of these pixels is about 0.049 arcseconds across.

1.

6.5.2: Part II: Measuring the Expansion Velocity of the Remnant

Click “next step.”

In this part, you will use the *Spectral Analysis* tool to examine the absorption spectrum of the supernova remnant. You will notice two prominent absorption features. Line 1, which is the large feature in the center, is mostly caused by Ca II (singly ionized calcium), with a small contribution from Fe I (neutral iron). The one on the right is due to Fe I (neutral iron). Atoms emit or absorb narrow lines in the electromagnetic spectrum when their electrons change from one energy level to another. But the lines in this spectrum are not narrow; they are quite broad.

The lines in this spectrum are broadened by the motion of the absorbing gas. The supernova remnant can be modeled as a spherical cloud of expanding material ejected when the star exploded. The gas on the near side (closest to us) is expanding toward us. Since it is moving toward us, its absorption line is shifted to the blue from our viewpoint. On the other hand, the gas on the far side of the remnant is moving away from us, so its absorption lines are shifted to the red. Since the remnant is composed of gas moving directly toward us, directly away from us, and at all intermediate velocities, we see the intrinsically narrow absorption lines broadened into the features in the spectrum. By using the amount of broadening and the Doppler formula, we can estimate how fast the gas is moving.

By fitting a special kind of curve, called a Gaussian (popularly known as a bell curve), over the data, we can determine the width of the absorption lines, and hence, the expansion rate of the supernova. It is not important that you understand the details of a Gaussian curve, as those are incorporated into the *Spectral Analysis* tool. All you have to do is use your mouse to fit a Gaussian on each of the absorption features in turn.

To use the *Spectral Analysis* tool:

- Click and drag the left button to move the tool into place.
- Click and drag the right button to change the width of the curve.
- Click and drag the slider to change the depth of the curve.

1.

2. Repeat for line 2.

3.

6.5.3: Part III: Determining the Physical Size of the Remnant

The physical size of the supernova remnant can be calculated by multiplying the expansion rate by its age, assuming the expansion speed has remained constant. It is also necessary for you to know that the image you are using was taken by the Hubble Space Telescope in 1998.

1.

2.

6.5.4: Part IV: Determining the Distance to the Remnant

Using the size you just determined, along with the angular size you measured from the image, you will now determine the distance to S And by employing the small-angle approximation.

1.

2.

3.

This page titled 6.5: Wrapping It Up 6 - The Supernova of 1885 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6.6: Mission Report 6 - Distance Measurement

A.

B.

C.

D. Questions to be graded for accuracy

The following questions review various techniques for determining astronomical distances that you have learned in this chapter.

1.

2.

https://phys.libretexts.org/@go/page/31247

3.

4.

This page titled 6.6: Mission Report 6 - Distance Measurement is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

7: Classical Physics- Gravity and Energy

Chapter 7 introduces the concepts of gravity and energy. In the first part of the chapter, you will learn that gravity, the force that causes an apple to fall from a tree, also keeps planets orbiting the Sun and is present throughout the Universe. Through the application of Newton's Laws you will explore properties of surface gravity on other planets and the orbits of planets, stars; you will see that gravity can affect any object in space. The second part of the chapter focuses on the many forms of energy, with special attention given to kinetic energy and gravitational potential energy. You will also learn that energy is conserved. It is passed from one body to another in any interaction, but it is never created nor destroyed.

This page titled 7: Classical Physics- Gravity and Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.0: Classical Physics - Gravity and Energy Introduction

The video shows astronauts moving around inside the International Space Station. Did you notice that they are able to float around and turn in different directions in a manner that is much different from how they would move on Earth? We often say that the astronauts in space are “weightless” or in a “zero-G” environment. But what do these terms really mean? And are they correct?

In this chapter, we consider gravity, the attraction of all matter to other matter. Gravity is the most familiar of the four fundamental forces of nature. It has been so for many thousands of years. However, only relatively recently have we achieved a basic understanding of what gravity is and how it works. In the following pages, we begin to explore the classical theory of gravity.

We will also explore the concept of energy in this chapter. Our focus will be on the energy due to gravity and the energy of motion, as well as the interplay between the two. By understanding gravity and energy we will gain a much deeper understanding of how objects move on Earth, in orbit, and in the Universe more broadly.

This page titled 7.0: Classical Physics - Gravity and Energy Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.1: Gravity on Earth

What Do You Think: Speed of Falling Objects

We all know that we are held to the ground (or floor or our chair) by some physical effect. We can jump off the ground, but we do not rise very far before we reverse direction and fall back down again. We also know that if we step off a high cliff, we will fall. Our downward speed will increase as we fall until some object obstructs our motion. Why is this the case?

People have long wondered why we are joined to the ground, seemingly by an invisible chain. The Greek thinker Aristotle offered the hypothesis that heavy objects fall faster than light objects, and that both fall down because it is natural for them to do so. While this hypothesis seems plausible, it does not explain why objects fall down. Galileo Galilei, the Italian astronomer and physicist, conducted laboratory experiments that showed that all objects under the influence of gravity fall toward the ground at the same rate. His result disproved our intuitive sense of how gravity affects objects: many of us, more than 400 years after Galileo's experiments, tend to think that heavy objects fall faster than light ones do, just as Aristotle did. In fact, they do not, at least not when the effects of friction can be ignored.

Figure 7.1: Galileo's Experimental Apparatus. Galileo used experiments based upon systems similar to this one to study the motion of falling objects. He also studied the motions of pendula, finding that their periods depend only upon their length, not upon the mass of the object suspended. Credit: Museo Galileo, Florence - Photo Franca Principe

Contrary to some popular ideas, Galileo did not study gravity solely by dropping objects from the Leaning Tower of Pisa. He was indeed a professor in that city at the time he conducted his gravity experiments, but he devised more accurate experiments. These included rolling balls of different sizes and weights down inclined tracks (like in figure 7.1), and studying the motions of pendula to which he had attached bobs of different masses and strings of different lengths. Galileo's experiments were not trivial because the speed of rolling objects is affected not only by their mass, per se, but by how that mass is distributed. Two balls of identical weight but different diameters will not roll down a hill with the same speed. Galileo was able to account for this effect.

What Galileo found was that all objects, regardless of their weight, fall under the influence of gravity at a speed that increases by about 9.8 meters per second every second. He did not express his values in terms of meters and seconds, of course. His results mean that if an object starts from rest, after one second it will be moving 9.8 m/s. Then, after another second it will be moving at 19.6 m/s, after another second it will be moving 29.4 m/s, and so on. The increase in the speed of a falling object is called the gravitational acceleration, or sometimes it is called one *g*. Galileo had no idea why things fall toward the ground, nor why they all fall at the same rate. He merely conducted an experiment, and that was his result.

How Fast Do Objects Fall?

Worked Example:

1. Consider what would happen if we dropped a massive textbook from the top of a very tall building. Assume the textbook mass is 5 kg.

How fast would the textbook be falling after 5 seconds?

- Given: acceleration due to the Earth's gravity g = 9.8 m/s^2, t = 5 sec
- Find: v, the textbook's velocity
- Solution: The textbook starts at rest, that is, with an initial velocity of zero. Once you drop it, the textbook falls 9.8 m/s faster every second. So after 5 seconds, the textbook's velocity = (5 s)(9.8 m/s^2) = 49 m/s. Note: the mass of the textbook did not enter into this calculation.

Questions

For more examples, see Math Exploration 7.1.

Math Exploration 7.1

Gravity on the Moon

VIEW TRANSCRIPT

In the video, you should have seen that both the hammer and feather reached the ground at the same time, consistent with Galileo’s findings. On the Moon, gravity is the only force acting on the objects. In the next section, we will take a closer look at how objects’ motions are affected when more than one force acts on them.

This page titled 7.1: Gravity on Earth is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.2: Force, Mass, and Weight

Mass and Weight

To understand gravity, we will have to learn a little bit more about what causes changes in the motion of objects. Clearly, gravity causes such changes, but it is not the only effect that can do so. For instance, we can push on a cart, or pull on it with a cord. In either case, the cart's motion can change. Such a pull or push is called a force by physicists. In physics, force has a very precise mathematical definition:

$$\vec{F}_{tot} = m\vec{a}$$

In this expression, F is force, m is mass, and a is acceleration, and the equation says that total force is the product of mass and acceleration.

There are several things that are important to understand about this equation. The first is the little arrows above the $\vec{F}$ and the $\vec{a}$. They are there to remind us that both force and acceleration have a direction associated with them. Pushing on the back bumper of a car has very different results than pushing on the front bumper - one makes the car accelerate forward, the other makes it accelerate backwards. Quantities that have a size and a direction are called vectors , and the arrows remind us that both force and acceleration are vectors. We will often write the equation without the arrows:

$$F = m\,a.$$

In this case, it means we are not concerned with the direction of the force and acceleration, or sometimes it will mean that their direction is obvious from the context of the problem we are considering. Since the only two vectors in this equation, the F and the a, are on opposite sides of the equals sign, they must point in the same direction.

The second thing we have to consider in this equation is the "tot" subscript on the F. That is there to remind us that we are referring to the total, or net force acting on an object. So, for example, if we push equally hard on the back bumper and on the front bumper of a car in such a way that the two forces are the same size but pointed in opposite directions, then the total force on the car will be zero. In this case, where the total force is zero, the equation says that the acceleration must be zero as well. That means that the car's velocity will remain constant. This is because acceleration is defined to be the change of an object's velocity with respect to time.

On the other hand, if both forces are pointed in the same direction, the total force will be the sum of the two. Then the acceleration will be larger than if only one force was acting because the two forces together, acting in the same direction, are larger than either one alone (Figure 7.2).

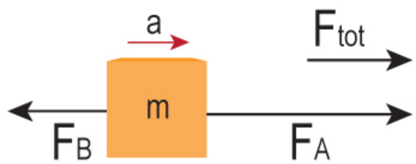


Figure 7.2: Net force and acceleration. The acceleration of an object points in the same direction as the total, or net, force (F_{tot}) acting on that object. The strength of the force, proportional here to the length of the arrow, is equal to the object's mass multiplied by its acceleration. Note that in this image, the net force points in the direction of F_A because F_A is larger than the oppositely directed F_B. Therefore, the acceleration also points in the direction of F_A. Credit: NASA/SSU/Aurore Simonnet

The SI unit of force is the newton (N), named for Isaac Newton, whose discoveries are the subject of this chapter. In SI units, 1 newton = 1 kg m/s^2. The more familiar unit of force in the USA is the pound (lb); 1 N = 0.2248 lb. However, we will use SI units in these modules.

Net Force on an Object

In this activity, you will use what scientists call a **freebody diagram** to see how forces acting on an object in different directions can add together to create a total force on the object.

Play Activity

Worked Example:

1. Consider how two forces add together in one dimension. Assume the first force we are dealing with pushes an object to the right with a force of 3 N. The second force pushes the object to the left with a force of 4 N. What is the total force on the object?

- The results of using the Freebody Diagram tool for this situation are shown in Figure A.7.1.
- The first force pushes 3 N to the right, or 3 N in the positive x direction. There is no push in the y direction. To enter the first force, type 3 in the "x value" box, and 0 in the "y value" box. Now click the "Show" button to enter this force. Notice that the force now appears on the diagram to the right.
- For the second force, there is a push of 4 in the negative x direction, and no push in the y direction. To enter this force, type -4 in the "x value" box, and 0 in the "y value box". Now click the "Show" button to enter this force.
- To add the two forces, click the "Add" button. Toward the bottom of the Input Panel, you should now see the Total Force listed in green. On the right, you can see the two input forces (in red) and the total force (in green) on the diagram.
- The total force is -1 (x-direction) + 0 (y-direction) N. You can see that the total force points toward the left.
- Once you are ready, click "Clear" before going on to the next example.

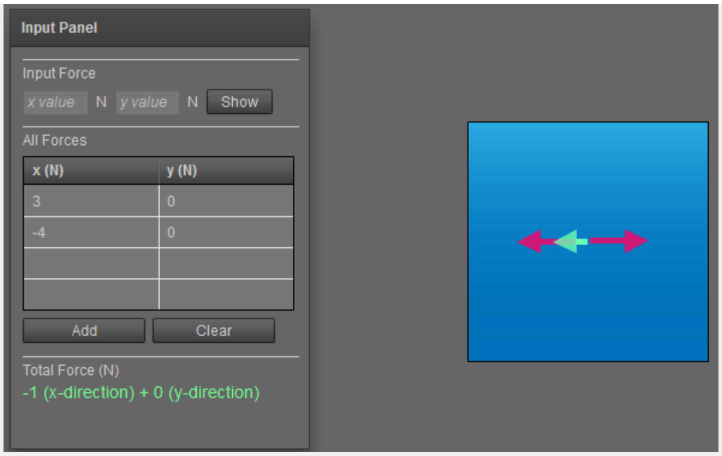


Figure A.7.1: Net force for 2 forces in one dimension. Output from the *Freebody Diagram* tool for 3(x-direction)N + -4(x-direction) N. The red arrows show the individual forces, while the green arrow shows the sum of the two forces. Credit: NASA/SSU

Questions

Now do a few examples on your own, either numerically, or using the interactive *Freebody Diagram tool*.

Another important thing to understand in the equation is the meaning of the symbol *m*, the mass. Mass is a tricky concept to understand at first, but hopefully by the end of this chapter you will understand it. Mass, as defined by the equation above, is a measure of how resistant an object is to changes in its motion. First, we will do a couple of examples to demonstrate how this works. The next thing to be clear about is that mass is not the same as weight. The difference will be addressed at the end of this section.

MASS AND ACCELERATION

Worked Examples:

1. Assume that we want to accelerate a 5 kg object at a rate of 2 m/s^2. We can use the force equation to compute the force we must apply to the object.

- Given: m = 5 kg, a = 2 m/s^2

- Find: F
- Concept: F = ma
- Solution: F = (5 kg) (2 m/s^2) = 10 N

We would therefore have to exert a 10 N force on the object.

2. We can also use the force relation to deduce the mass of an object. If we imagine that when we apply a 100 N force to an object it accelerates at 2 m/s^2, we can calculate its mass:

- Given: F = 100 N, a = 2 m/s^2
- Find: m
- Concept: F = ma, or, rearranging, m = F/a
- Solution: m = F/a = (100 N) / (2 m/s^2) = 50 kg

This defines the mass of an object using its inertia, or its resistance to change in motion. This definition of mass is sometimes referred to as inertial mass. We can also define mass in terms of gravity, as we do below. While philosophically different, the two definitions give the same numerical answer in all cases studied to the best of our ability to measure.

Questions

So the more massive an object is, the more difficult it is to change its velocity. This includes both its direction of motion or its speed. We are all familiar with this from our experiences in the world: It is much less dangerous to run into a person than a freight train. While freight trains do tend to move faster than people, that is not the most vital detail to consider. Even if a train is moving only at walking speed you would probably prefer colliding with a person than with a train. This is because a train has a lot more mass - and thus inertia - than a person has.

ACCELERATION AND MASS

We can use the *Freebody Diagram* again, this time with mass as well as force, to determine both the direction and strength of an object's acceleration.

Play Activity

Worked Examples:

Imagine you and your friend are fighting over a textbook that you both really, really want to read. The textbook has a mass of 5 kg. You pull on the book with a force of 6 N in the x direction, and your friend pulls on the book with a force of 5 N in the negative x direction.

Using the Freebody Diagram tool that allows you to enter the mass, you should get something that looks like Figure A.7.2 below.

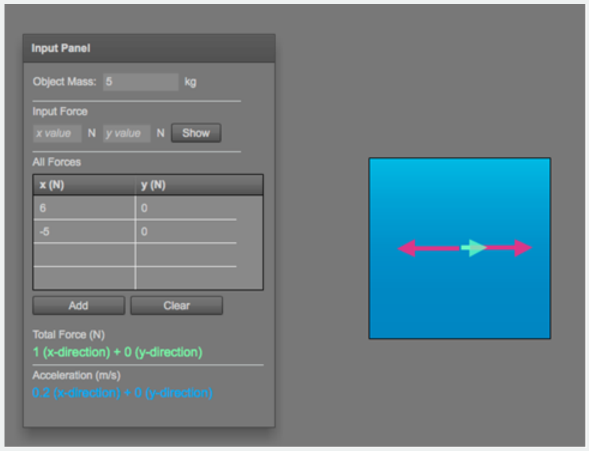


Figure A.7.2: Net force

and acceleration in one dimension. Output from Freebody Diagram for forces of 6 N (x-direction) and -5 N (x-direction) and a mass of 5 kg. The red arrows show the individual forces, while the green arrow shows the net force. The direction of the acceleration is the same as the direction of the net force. Credit: NASA/SSU

1. What is the total force on the textbook (strength and direction)?

We see from the tool that the net force is +6 N – 5 N = +1 N in the x-direction (to the right).

2. What is the total acceleration of the textbook (strength and direction)?

We see from the tool that the acceleration of the textbook is 0.2 m/s^2 in the +x direction. Notice that the direction of the acceleration is the same direction as the net force. The tool calculates the strength of the acceleration from the equation $F = ma$, or $a = F/m$. In this case, $a = 1\ N / 5\ kg = 0.2\ m/s^2$.

3. Will the textbook accelerate toward you or your friend?

It will move toward me.

Questions

Once you decide the matter of the textbook, you and your friend realize that there is a study guide full of answers to all the homework questions in the text, and you begin to fight over that, too! The study guide has a mass of 2 kg. Again, you pull on the guide with a force of 6 newtons in the x direction. Your friend pulls on the study guide with a force of 5 newtons in the negative x direction.

In the activity above, you can see that when two friends are pulling on a fairly massive textbook, the book does not accelerate much. However, when the friends pull with the same force on a less massive study guide, the study guide accelerates more. This demonstrates that under the influence of a force, an object's acceleration is *inversely proportional* to its mass. In other words, an object with a *large* mass will experience a *smaller* acceleration due to a given net force than an object with a *smaller* mass. The object with a *smaller* mass will experience a *larger* acceleration due to a particular net force than the object with the *larger* mass.

Remember that one of the definitions of mass we discussed above refers to the *inertial mass* because it is based on the inertia of an object, or in other words, on its tendency to resist changes in its motion. The equation linking force, mass, and acceleration was first deduced by Isaac Newton, and it is called Newton's second law of motion. Along with his first and third laws of motion, it constitutes the basis of classical mechanics, a branch of physics that describes the motion of bodies through space. We have already encountered Newton's first law, borrowed from Galileo:

Definition: Newton's first law of motion

Objects in motion remain at a constant velocity (straight line, constant speed) unless acted upon by a net force.

We have now encountered Newton's second law in our discussion of force, mass, and acceleration:

Definition: Newton’s second law of motion

The acceleration of an object is directly proportional to the total force on the object, and inversely proportional to its mass. In physics, we write this using the equation $F = ma$, which we introduced and have worked with in its full vector form already.

For more about Newton’s laws see Going Further 7.1: Newton’s Third Law.

Going Further 7.1: Newton’s Third Law

Our daily experiences might lead us to think that forces are always applied by one object on another; for example, a horse pulls a carriage, a person pushes a grocery cart, or a hammer hits a nail. It took Sir Isaac Newton to realize that things are not so simple, and not so one-sided. True, if a hammer strikes a nail, the hammer exerts a force on the nail (thereby driving it into a board). Yet, the nail must also exert a force on the hammer since the hammer’s state of motion is changed, and according to the first law, this requires a net (outside) force. This is the essence of Newton’s third law: For every action there is an equal and opposite reaction. However, it is important to understand that the action and the reaction are acting on different objects.

Try this: Press the side of your hand against the edge of a table. Notice how your hand becomes distorted. Clearly, a force is being exerted on it. You can see the edge of the table pressing into your hand and feel the table exerting a force on your hand. Now press harder. The harder you press, the harder the table pushes back on your hand. Remember this important point: You can only feel the forces being exerted on you, not the forces you exert on something else. So, it is the force the table is exerting on you that you see and feel in your hand.

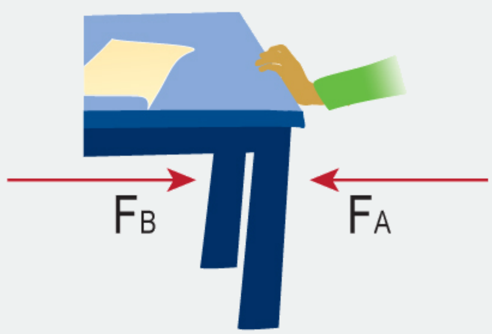


Figure B.7.1. Action–reaction pair illustrating Newton’s third law. Newton’s third law is illustrated in this figure, which shows that the action of the hand pushing on the table is equal in strength and opposite in direction to the reaction of the table pushing on the hand. Credit: NASA/SSU/Aurore Simonnet

Action–reaction pairs like the force of the hand on the table and the force of the table on hand are all around us. If you are reading this while sitting in a chair, you can feel the force the chair is exerting upward on you while you exert a force downward on it. And if you are reading on a computer monitor that sits on the table in front of you, then there is a balanced action–reaction pair between the monitor and the table, as well as between the table and the floor, and so on for all the objects you see around you. Each of these is an action–reaction pair. Such action–reaction pairs are the domain of Newton’s third law.

Not every equal and opposite pair of forces is a third law pair, however. In the third law, two objects are involved with the two forces. For example, a hand and a table, a table and a hand. We can have two forces of equal strength and opposite direction due to Newton’s second law, but these forces act on a single object. For example, if you hold a ball still in your hand, there is an upward force due to your hand on the ball that is exactly equal to the downward force of gravity from Earth on the ball. That is why the ball does not move. These two forces are acting on a single object, the ball, whose net force and acceleration are zero by Newton’s second law.

Newton's third law is also related to the concept of conservation of momentum. Momentum is defined as an object's mass times its velocity; because velocity is a vector, momentum is also a vector. Momentum is also related to force: the change in momentum over time is equal to the net force.

So, for two objects that are a third law pair, the change in momentum of one object will be equal in strength and opposite in direction to the change in momentum of the other. This principle is important in collisions. For example, if two billiard balls collide, they will bounce off of each other and travel in opposite directions. It is also important for propulsion. For example, a squid is a sea creature that takes in water; when it wants to move, it squirts water out of its body in one direction, and it moves in the opposite direction. As another example, if you are standing on a frozen pond having a snowball fight in the winter, if you throw a snowball, the snowball will move forward, but you will also slide backward a little. The velocity of the snowball will be greater than your velocity because you have a greater mass (unless you make a really big snowball).

We will use Newton's laws in our study of objects moving under the influence of gravity. Combined, they provide a basic framework for understanding how objects move, why bridges are able to stand, how water flows in a stream and many other physical phenomena that fill our everyday experience.

What does all of this have to do with gravity? Well, as was noted in Section 7.1, gravity causes an acceleration, *g*, and this acceleration is constant for all objects on Earth, regardless of their mass. If mass and acceleration are involved, then according to Newton's second law, there must be a total, or net, force acting.

In fact, we can use the second law to measure the force exerted by gravity on objects of different masses. We have a special name for the force of gravity acting on a mass; we call it the weight of the object. Substituting the acceleration due to gravity into Newton's second law, we get an expression for the force of gravity on Earth, or the weight of an object:

$$F_g = mg$$

Again, F_g is the gravitational force, or weight, m is the mass, and g is the acceleration due to gravity, 9.8 m/s^2. Now the relationship between mass and weight becomes clear: weight is a force, whereas mass is how much "stuff" there is. The weight and the mass are related by a factor of g, but are not the same.

For more information on why the value of the mass is the same, whether the force is gravity or some other force, see Going Further 7.2: Gravity and Inertia.

7.2.1: GOING FURTHER: GRAVITY AND INERTIA

Gravitational Force on a Mass

Worked Example:

1. If we consider a 5 kg mass, we can use Newton's second law to find the force that gravity exerts upon it.

- Given: m = 5 kg, g = 9.8 m/s^2
- Find: F_g
- Concept: F_g = mg
- Solution: F_g = mg = (5 kg) (9.8 m/s^2) = 49 N

This is equivalent to just over two pounds.

Questions

ADAPT 7.2.1

In the previous example we used Newton's second law to compute the force acting on an object falling in Earth's gravity. This force causes objects to accelerate at one g (9.8 m/s^2). But what if an object is not falling? Does a force still act on it? Yes! But in that case the acceleration is zero, so how can there be a force? Here we have to think carefully about what the second law says: it says that acceleration is caused by a total, or net force. A zero acceleration means only that there is no *net* force acting on the body, not that no forces at all act on it. A picture might help clarify this (Figure 7.3).

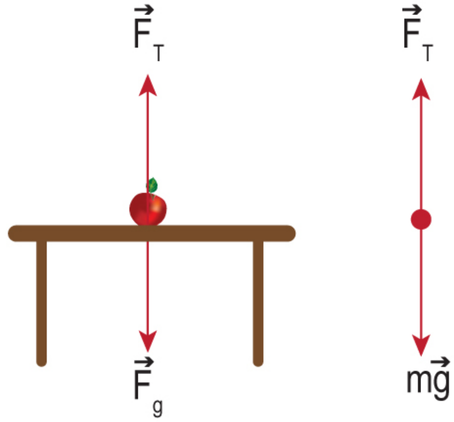


Figure 7.3: Apple sitting on a table. An apple sitting on a table feels the force of gravity, Fg, pulling it downward, but it is not accelerating. If the acceleration of the apple is zero, the net force must be zero. Therefore, there must be a force of equal strength but opposite direction to the gravitational force. This opposing force that is pushing up on the apple is F_T, from the table to the apple, which keeps the apple from falling. The net force on the apple is zero and the apple remains stationary. Credit: NASA/SSU/Aurore Simonnet

Imagine that we have the situation shown in Figure 7.3: An apple sits on a table. In the numerical activity, we saw that the force of gravity is acting on the apple, which should cause it to accelerate downward. However, in this case the apple is not accelerating. Why not? Because the table is in the way. If we apply Newton's second law to this case, we can deduce that the table must be applying an upward force equal in magnitude to the force of gravity. How do we know that the net force is zero? Since the apple does not accelerate, we know the net force is zero. We can draw the apple and the forces acting on it, and since only the apple's weight and the upward force from the table, F_T, are present, they must be exactly equal in strength and opposite in direction.

Let's repeat that again for emphasis: The apple does not remain stationary because there are no forces acting on it. In fact, we have just identified two forces that are acting on the apple. However, one force is from gravity and points downward. The other force is from the table and points upward. The forces are equal in strength and point in opposite directions, so they cancel each other out. The total force (or net force) on the apple is zero.

Lunar Lander

We can put what we have learned so far about Newton's laws and gravity into practice by using the Lunar Lander to simulate landing a spacecraft on the surface of the Moon. To successfully land the spacecraft, you will have to slow it down so that both its horizontal and vertical velocities are very small. You will also have to land the spacecraft upright. Read the Lunar Lander instructions to learn which keys on your keyboard you can use to fire the lander's engine and to rotate the lander. Watch your fuel level!

Some things to think about as you try this activity:

- When you fire the lander's engine, you are giving the lander a "push", or applying a force to it. The direction of the force depends on which direction the lander is pointing. When you stop firing the engine, you stop applying a force to the lander.
- The lander starts with a horizontal velocity . You will have to slow the lander down in the horizontal direction.
- The Moon exerts a gravitational force on the lander, causing it to accelerate downward. You will have to slow the lander down in the vertical direction, too.

Play Activity

This page titled 7.2: Force, Mass, and Weight is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.3: Gravity Is a Universal Force

? What Do You Think: Gravity on the International Space Station

Thus far we have looked at two aspects of gravity. The first is the acceleration due to gravity near the surface of Earth. We know that gravity makes everything fall at the same rate. The second aspect we have considered is the force of gravity, again near the surface of Earth; this force is also called weight. We can now explain to some extent how objects move under the influence of gravity, but we do not really know why they move the way they do. More particularly, we have not learned anything about where gravity itself comes from. Isaac Newton had the first insight into that aspect of gravity.

7.3.1: Newton's Universal Law of Gravitation

One legend suggests that Newton's epiphany on gravity came as he was sitting under an apple tree on the Cambridge University campus where he was a professor. As he sat staring at the Moon, pondering what made it go around Earth, suddenly an apple is supposed to have fallen from the tree onto his head, causing him to exclaim "Eureka!" This story is probably untrue, but it does accurately relate the essence of Newton's insight into gravity: The same force that causes apples to fall to the ground, and keeps people firmly on the ground, also causes the Moon to circle Earth (and Earth to circle the Sun). This was the vital connection to make.

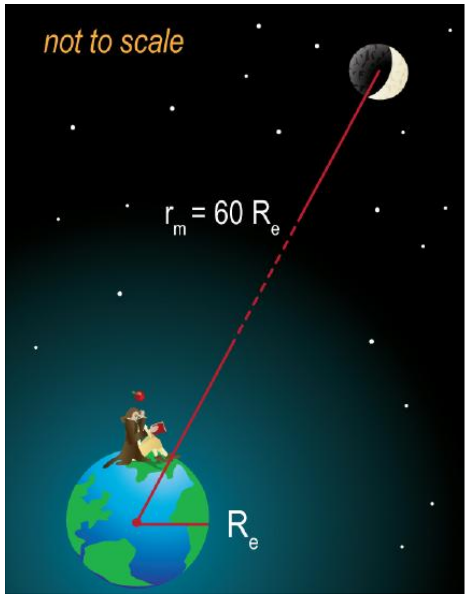


Figure 7.4: Isaac Newton had an insight that the same force that causes an apple to fall to the ground must also keep the Moon in its circular path about Earth. He used this insight and his new laws of motion to deduce the Law of Universal Gravitation. Credit: NASA/SSU/Aurore Simonnet

Newton knew that the Moon circled Earth in 27.5 days. (This is the time required to go 360 degrees around Earth, not the time from new moon to new moon.) Newton also knew that the Moon's orbit around Earth is roughly circular, and that the orbital radius of the Moon is about 60 times larger than Earth's own radius.

To keep an object moving in a circular path, it must be accelerated. Why? Because as it travels along its path, the direction of its velocity is constantly changing, and changing velocity implies an acceleration. If we consider such motion, we realize that a faster object has a larger acceleration than a slower one. Its velocity must change more quickly than the slower object. Furthermore, if an object travels on a very big circle, say like the orbit of the Moon, its acceleration will be smaller than if it travels at the same speed on a smaller circle. The equation that describes the acceleration for circular motion is:

$$a_c = \frac{v^2}{r}$$

where a_c is the acceleration, v is the velocity, and r is the distance between the object and the center of the circle.

With this equation, we see that a larger velocity causes a larger acceleration, while a larger radius for the motion causes a smaller acceleration. You might be surprised that the velocity is squared, but that is how this type of circular motion works.

We know from our earlier discussions that acceleration is a vector and has both size and direction. What is the direction of this acceleration? It is not obvious from this expression, but it turns out to be pointed along the radius of the circle, directly toward its center. For this reason, acceleration of this type is called centripetal (center-seeking) acceleration, hence the subscript "c." It is sensible that the acceleration should point to the center if you recall that the acceleration is just the change in velocity over time. For circular motion like this, where the only thing changing is the direction of the velocity vector, the vector is constantly turning around toward the center of the circle. The velocity vector never turns all the way toward the center of the circle, of course. It continues to point in a direction tangent to the circular path, and the moving object travels farther along the circumference of the circle. See Figure 7.5 for an animated representation of this situation.

Play Animation

Animated Figure 7.5: Object moving in a circular path. This object is moving in a circular path with a constant speed, but a constantly changing direction. Since the direction of its motion is changing, the velocity of the object must also be changing. Thus, there must be an acceleration that points in the direction of its change in motion. This acceleration points toward the center of the circular path, as the animation shows. Credit: NASA/SSU/Kevin John

If you know the formula for centripetal acceleration, as Newton did, it is fairly easy to compute the value of the acceleration in a given situation. For the Moon and Earth, you just plug in the Moon's orbital speed and its distance from Earth. We can use the

Moon's orbital size and period to find the Moon's speed, and then its acceleration, as we show in the next activity. Then you will have a chance to do a similar calculation for Earth's orbit around the Sun.

Worked Example: Calculating Orbital Accelerations

Find the centripetal acceleration of the Moon. The radius of the Moon's orbit has long been known from observations. In modern units it is 384,400 km, or 3.844 x 10^8 meters.

Concept: in order to use the formula $a_c = v^2/r$, we are given r, but we will still need to find the orbital velocity of the Moon, v. We can get this from the circumference around its orbital path and the time period that it takes to go once around.

Step 1. In order to derive the Moon's centripetal acceleration, we first need to find the circumference of its orbit. Given the radius of its orbit, its circumference, C, is:

$$C = 2\pi r = (2\pi)(3.844 \times 10^8 m) = 2.415 \times 10^9 \text{m}. \tag{7.3.1}$$

Step 2. The period, T, for the Moon's orbit around Earth is 27.5 days. Convert this to seconds:

T = (27.5 ~~day~~) (24 ~~hr~~/~~day~~) (60 ~~min~~/~~hr~~) (60 sec/~~min~~) = 2.376\times 10^6 sec

Step 3. Now find the Moon's orbital speed by dividing the distance (circumference) by time (period):

$$v = \frac{C}{T} = \frac{2.415 \times 10^9 \, m}{2.376 \times 10^6 \, sec} = 1,016 \, \text{m/s} \tag{7.3.2}$$

So the Moon moves about 1 km/s.

Step 4. Now we can use the centripetal acceleration formula to get the Moon's orbital acceleration:

$$a_c = \frac{v^2}{r} = \frac{(1,016 \, m/s)^2}{3.844 \times 10^8 \, m} = 2.685 \times 10^{-3} \, \text{m/s}^2 \tag{7.3.3}$$

Questions

Now it is your turn to calculate the orbital acceleration for Earth's motion around the Sun. Follow the steps below to do the calculation:

Newton (and others) deduced (guessed, really) that the reason the acceleration of the Moon toward Earth is so small is because the Moon is very far away. He correctly surmised that gravitational attraction must depend on distance .

How Does the Force of Gravity Depend on Distance?

In this activity, you will make measurements to determine how gravitational force depends on distance.

Play Activity

- The masses are set to 100 kg for this activity.
- Move the slider bar to select different distances between 200 and 600 units. Enter the distance in the x-column and the force into y-column of the graphing tool. For example, for a distance of 200 units, the force is 1250.
- Now use the graphing tool to plot your data and compare your data to the plots provided below in Figure A.7.3.

https://phys.libretexts.org/@go/page/31251

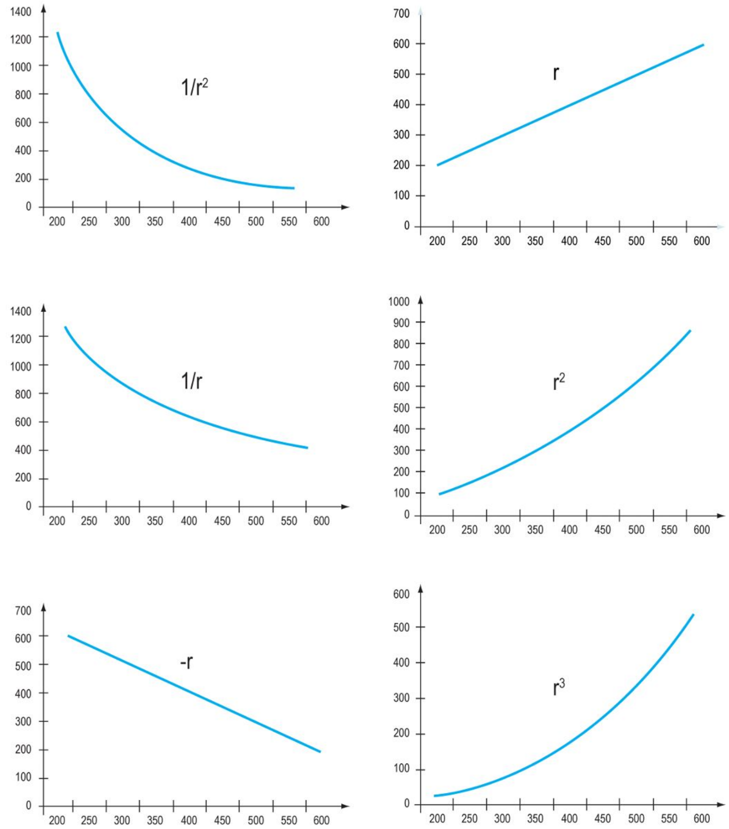


Figure A.7.3: Each of these graphs plots distance on the x-axis, and force on the y-axis. Compare these graphs to the plot you have made in the graphing tool to choose the best match below. Credit: NASA/SSU/Aurore Simonnet

Of course, Newton was not able to make measurements in the laboratory in which he could determine this dependence. Instead, he used mathematical reasoning to infer the relationship, which could then be experimentally verified. Newton's reasoning is traced in Going Further 7.3: Determining the Distance Dependence of Gravity.

Going Further 7.3: Determination of the Distance Dependence of Gravity

Newton reasoned that the easiest dependence to consider is one where the strength of the gravitational acceleration (a_g) decreases with distance (r) raised to some power. In equation form, we use "n" to denote "some power", where n stands for an integer that we still need to determine here. The symbol α means "proportional to."

We expect that the value of n will be negative because only negative values will cause the gravitational acceleration to get smaller with r. By comparing the gravitational acceleration at Earth's surface to the acceleration of the Moon toward Earth, it is possible to determine the value of n.

The r here is understood to be the distance between the centers of the objects involved. So, for example, if we are standing on the surface of Earth, our distance is taken from Earth's center, not its surface. Similarly, the distance of the Moon from Earth is taken to be the distance from Earth's center (not its surface) to the center of the Moon. This is not obvious, and it was one of

Newton's great insights into gravity; he was able to demonstrate its validity mathematically. In the case of the Moon-Earth distance, the distinction between the center and surface is not very important because their separation is much larger than their sizes. However, for we who stand upon Earth's surface, the difference is vital: we are not zero meters from Earth, we are about 6400 km away from its center, at least as far as gravity is concerned.

You should have recognized in the previous interactive activity that the value of *n* turns out to be -2. However, Newton had to derive this, and we will now show how.

We can set up a ratio between the acceleration at Earth's surface and at the Moon's orbit.

$$\frac{a_E}{a_M} = \left(\frac{r_E}{r_M}\right)^n \tag{7.3.4}$$

Then we just have to plug the appropriate numbers into this expression.

$$\frac{9.8\ \mathrm{m/s^2}}{2.685 \times 10^{-3}\ \mathrm{m/s^2}} = \left(\frac{6.638 \times 10^6\ \mathrm{m}}{3.844 \times 10^8\ \mathrm{m}}\right)^n \tag{7.3.5}$$

Note how units on both sides cancel, leaving no units in any resulting expression. We are left with the following…

$$3,650 = (1.657 \times 10^{-2})^n \tag{7.3.6}$$

Now, to solve for *n* it will help if we take the log of both sides. Then we have an expression we can easily solve algebraically. Here are the steps:

$$\log(3,650) = \log(1.657 \times 10^{-2})^n \tag{7.3.7}$$

$$\log(3,650) = n \log(1.657 \times 10^{-2}) \tag{7.3.8}$$

And so...

$$n = \frac{\log(3,650)}{\log(1.657 \times 10^{-2}} \tag{7.3.9}$$

$$n = \frac{3.552}{-1.781} \tag{7.3.10}$$

$$n = -2 \tag{7.3.11}$$

This is essentially the reasoning that Isaac Newton used to deduce the $1/r^2$ dependence of gravity. Such a dependence had been suspected by others, but Newton was the first to be able to demonstrate it in this manner.

Gravity does not only depend on the distance between two objects, it also depends upon their masses. Again using an interactive activity, you will explore this dependence.

How Does the Force of Gravity Depend on Mass?

In this activity, you will make measurements to determine how gravitational force depends on mass.

Play Activity

- Set the distance between the masses to 400 units for this activity.
- We will call the mass of the blue object m_1 and the mass of the red object m_2.

Worked Examples:

1. Move the slider bar such that both masses are 50 units. What is the value of the force on each object in this case?

We see that the force on blue = 78.13 units and the force on red = 78.13 units. Note that the force is the same on both objects.

2. Move the slider bar such that m_1 is twice as much as it was before, i.e. 100 units. The other mass should remain at 50 units. What is the value of the force on each object in this case? By what factor has it changed?

We see that the force on blue = 156.3 units and the force on red = 156.3 units. This means the force is twice as much as it was before. The force on blue is equal to the force on red in this case too.

7.3.1.1: Questions

Now increase m_2 by a factor of 3, i.e. to 150 units while m_1 remains at 100 units.

Again, Newton deduced that the masses of the objects and the distance between them were responsible for the gravitational attraction. He was able to write an expression for the gravitational force between two objects as:

$$F_g = G\frac{m_1 m_2}{r^2}$$

This equation has been given the grandiose title Newton's Universal Law of Gravitation. It describes the gravitational force (F_g) between two objects, one with mass m_1 and the other with mass m_2. The objects are separated (center to center) by a distance r from each other. The geometry is shown in Figure 7.6. The constant of proportionality G is called the gravitational constant. Its value in SI units is G = 6.67384 x 10^{-11} N m^2 kg $^{-2}$. The constant of proportionality was not determined by Newton, but by Henry Cavendish (1731 - 1810) in a famous experiment carried out just before the turn of the 19th century, more than 50 years after Newton's death. At first, Newton's law was a guess, and it required verification by experiments or observations. Now, Newton's gravitational law forms the basis for our classical understanding of gravity and is accurate enough to predict the motions of the planets around the Sun and to send spacecraft, peopled or not, to other bodies in the solar system.

Figure 7.6: Two objects with masses m_1 and m_2, respectively, are separated by a distance r. Each will feel a force from gravity that is proportional to the product of their masses and inversely proportional to their separation squared. Object 1 feels a force directly toward object 2, and object 2 feels a force directly toward object 1. Credit: NASA/SSU/Aurore Simonnet

The masses, m_1 and m_2,could be any two objects in the universe: Earth and the Moon, Earth and the Sun, Earth and you, you and a friend, two tiny particles of interstellar dust , or two galaxies. This is what makes the law universal: it works for everything in the Universe that has mass, no matter how large or small the mass might be.

Sometimes astronomers will be more descriptive in the subscripts they use; instead of calling the objects m_1 and m_2, they might call them M_E and M_S for Earth and Sun for example, or M and m if the objects are different masses. The subscripts are simply a matter of notational preferences.

In the next set of activities, you will practice using Newton's law of gravitation and explore it in more detail.

Working With Newton’s Law of Gravitation

Use Newton’s law of gravitation to answer the following questions:

Exploring the Force and Acceleration Due to Gravity

We can use the Newtonian Gravity tool to take a closer look at how the masses and distances between two objects affect the gravitational forces between them, as well as their acceleration toward each other. In this activity, we have arbitrarily set the G constant equal to 1 in order to avoid dealing with very small numbers. This does not matter at all for understanding gravity. It is equivalent to using a different unit to measure force, rather than using newtons, and units are only a matter of convenience and convention anyway.

Play Activity

A. Force and Acceleration

First, look at the force and acceleration between two objects of the same mass (50 units), and at a given distance (200 units). Use the slider bars to adjust the masses of the objects and the distance between them.

B. The Effect of Distance

Now change the distance between the objects to 500 units, and see how this changes the gravitational force between them, and their acceleration.

C. The Effect of Mass

Now, we will see what happens if we adjust the problem so that the objects have unequal masses. Keep the distance between the objects set to 500 units, and keep the mass of the blue object at 50 units. Adjust the mass of the red object to 150 units.

In the previous activity, you should have noticed a few important points from working with Newton's laws. First, the strength of the gravitational force between two objects is always the same for both of them, even if their masses are different. So, the strength of the force on object 1 from object 2 is the same as the strength of the force on object 2 from object 1. This is because a force is an action that happens between two objects, not a property of a single object. Next, you should have noticed that the directions of the forces go in opposite directions. This ensures that the force is always attractive, which Newtonian gravity certainly is.

Newton's gravitational law depends *only* on the masses of the two objects and their separation. It does not depend on their physical size (their diameters or radii), nor does it depend on whether they spin or not, whether they have magnetic fields, or whether they have atmospheres. Furthermore, the gravitational force between two objects can be felt infinitely far away. It never cuts off, it just becomes smaller and smaller. That means that, in principle, every object in the universe feels a gravitational attraction for every other object, no matter how far. In practical terms, of course, the gravitational interaction between widely separated objects becomes so small as to be negligible.

Even though the strength of the gravitational force felt between two objects is always the same for both of them, you should have noticed that their accelerations are different if their masses are different. We will explore this phenomenon in more detail next.

7.3.2: Applying Newton's Laws

In Math Exploration 7.2, we see with astronomical objects the same phenomenon we noticed with objects on Earth: even though the force is the same, the acceleration depends on the mass. Objects with a larger mass will have smaller accelerations and objects with a smaller mass will have larger accelerations.

Math Exploration 7.2: Gravitational Acceleration between Two Astronomical Objects

Comparing Newton's gravitational law to his second law, we can find an expression for the gravitational acceleration of any body. For example, if we take one of the masses in the equation to be that of Earth, 5.98 x 10^{24} kg, we can find Earth's surface gravity. We can also determine the surface gravity of other objects in the Solar System, as in the following activity.

Surface Gravity of Objects in the Solar System

In this activity, we will use Newton's law of gravitation to derive the value of g, which we have previously measured empirically as 9.8 m/s^2 on the surface of Earth. Then you will have the opportunity to determine the local "g" constant on other planets.

Worked Example:

Consider an object of mass m at the surface of Earth. We can use Newton's gravitational law to find the force Earth's gravity exerts upon the object. In the equation below, M_E and R_E are the mass and radius of Earth, respectively:

$$F_g = G\frac{M_E m}{R_E^2}$$

Since any force can always be expressed in terms of Newton’s Second Law ($F = ma$), we can also write the force in terms of the gravitational acceleration: $F_g = mg$. Setting the two expressions for force equal to each other we get:

$$mg = G\frac{M_E m}{R_E^2}$$

Canceling the common factor of m we get:

$$g = \frac{GM_E}{R_E^2}$$

This is an expression for the gravitational acceleration at Earth’s surface (g). Notice that it is the same for any object, regardless of mass since, just as before, the mass of the object itself has dropped out of the expression. Numerically this is:

g = (6.67384e-11 N m^2 kg^{-2}) (5.972e24 kg) / (6.368e6 m $)^2$ = 9.828 m/s^2

This is the value of the acceleration of gravity (g) at Earth’s surface, as measured originally by Galileo. Newton has given us a theoretical understanding of the origin of the acceleration.

Questions

Now use the table to answer the following questions:

The next activity shows that gravity is really a fairly weak force , and it is only because Earth and other planets have large masses (relative to human scales) that we notice a gravitational force between them. Newton's law of gravitation says that all masses in the Universe attract all other masses in the Universe, no matter how small or far apart. It is just that in everyday life, we do not notice the gravitational forces between small objects because they are smaller than the gravitational forces between large objects. But they are there.

Weight and Gravitational Attraction for Small Objects

Worked Examples: Weight and force for two people

1. We have already seen that the gravitational force of Earth on an object is called that object's weight and that we can calculate the weight using $F_g = mg$ for an object of mass m.

a. A person might have a mass of about 50 kg, though that varies a lot from person to person. What is the weight of such a person?

$$F_g = mg = (50\,kg)(9.8\,m/s^2) = 490\,N \tag{7.3.12}$$

2. But what would be the gravitational force of two such people on each other? We can use Newton's law to answer that question. We will assume that they are a meter apart, which seems like a reasonable distance for two people.

$$F = \frac{(6.67384e-11\,N\,m^2 kg^{-2})(50\,kg)(50\,kg)}{(1\,m)^2} = 1.67e-7\,N \tag{7.3.13}$$

This should make it pretty clear why we are not aware of the gravitational forces we exert upon each other. After all, the electromagnetic force exerted by a strong magnet can be enough to pick up a car - several thousand pounds. Even the magnets we might have on our refrigerators can exert several newtons of force. In comparison, gravity is an anemic force. It generally becomes noticeable only when the masses involved are very large.

Questions

This page titled 7.3: Gravity Is a Universal Force is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.4: Gravity and Orbits

What Do You Think: Orbiting Planets

As you saw in the video that starts the chapter, astronauts in the Space Station are seemingly weightless. We will now explore this phenomenon.

The International Space Station orbits Earth at an altitude of about 380 km. We can calculate the gravitational acceleration at this altitude using Newton's law for the gravitational acceleration: $g = GM_E / r^2$. In this case, the distance r from the center of Earth is the sum of the ISS altitude and Earth's radius.

g = (6.67384 x 10^{-11} N m^2 kg^{-2}) (5.972 x 10^{24} kg) / (6.368 x 10^6 m + 3.80 x 10^5 m $)^2$

g = 8.75 m/s^2

This is almost as large as the acceleration at Earth's surface, and certainly not equal to zero. So why then, do the astronauts appear weightless?

The answer to that question has to do with how orbits work. The astronauts appear weightless because they are in free fall. That means they are falling freely toward Earth. It is a bit like when sky divers jump out of an airplane, except in the case of sky divers the air eventually pushes upward on them as hard as their weight pulls them down, so they are not really falling freely. You should be able to understand why this is if you have ever stuck your hand out of the window of a car moving at high speed. Air exerts a strong drag force on any object moving through it at high speed. This is also similar to the feather example we discussed toward the beginning of this chapter.

In contrast, astronauts do not encounter any drag force, first because they are inside a container, and second, that container is high above most of Earth's atmosphere. Astronauts, and the satellites they inhabit, fall freely toward Earth under the influence of gravity and nothing else. But if astronauts are falling toward Earth, shouldn't they eventually hit it? No, it turns out.

You might be aware that the ISS orbits Earth approximately once every 90 minutes. But it is only a couple hundred miles up, so that means it goes around Earth at an incredibly high speed. In fact, it is moving so fast that, even though it is constantly falling toward Earth, its sideways motion causes it to "miss." On the other hand, it cannot get away from Earth because the gravitational interaction between the space station and Earth bends the path of the ISS, keeping in a circle. It is constantly falling toward Earth, and constantly missing. That is what an orbit is, whether it is the orbit of a satellite (including the Moon) around Earth or the orbit of Earth around the Sun.

Without the gravitational interaction between orbiting bodies, their relative speed would very rapidly separate them. The ISS would go flying off into space, Earth and the other planets would go flying off into space, etc. But the balance between their relative motion and the acceleration caused by their gravitational interaction keeps orbiting objects bound to each other, constantly falling toward each other and constantly missing.

Orbiting Around Planets

Worked Example:

Imagine you could orbit just above Earth's surface. What would your orbital speed have to be, and how long would it take you to go around Earth once? We will assume that Earth has no atmosphere, so we can ignore drag forces. We will also assume that there are no mountains for us to run into; Earth is perfectly smooth for this example.

We know that the gravitational acceleration at Earth's surface is 9.8 m/s^2. For us to be in orbit this would also have to be our centripetal acceleration, just as in our previous examples. Setting these two equal we have:

$$a_c = g \tag{7.4.1}$$

$$\frac{v^2}{R_E} = 9.8 m/s^2$$

Solving for the velocity, v:

$v = [(9.8 \text{ m/s}^2) (6.368e6 \text{ m})]^{1/2} = 7900 \text{ m/s} = 7.9 \text{ km/s}$

Using this velocity, we can find the orbital period, the time required to circle Earth once: It is the Earth's circumference divided by this speed.

$T = C / v = 2\pi R_E / \text{v} = (2\pi) (6368 \cancel{\text{km}}) / (7.9 \cancel{\text{km}}\text{/s}) = 5065 \text{ s}$

Converting this to minutes we get

$T = (5065 \cancel{\text{s}}) / (60 \cancel{\text{s}} / \text{min}) = 84 \text{ min}$

Around the World in 80 Minutes! That is fast. It would be the time required to orbit once around Earth if you could orbit just above the surface. Notice that this time is just a little bit shorter than the time required for the ISS, reflecting the somewhat higher gravity at the surface (and therefore the higher orbital velocity required), as well as the slightly shorter circumference of the proposed orbit. But if you could run 8 km/s (about 5 miles per second), on a perfectly smooth and airless Earth, you could put yourself into orbit just above the ground! You would be moving so fast that, as you fell toward Earth, the curvature of the surface would cause the ground to get out of your way, barely.

Questions

Jupiter's moon Io has an approximately circular orbit at a radius of 4.2×10^8 m from Jupiter's center. Jupiter's mass is 1.9×10^{27} kg.

https://phys.libretexts.org/@go/page/31252

> **Orbit Simulator**
>
> In this activity you will experiment with different combinations of velocities and vary the distance between two objects in order to see what it takes to get into a stable orbit. The orange circle within the simulator represents the Sun. The blue circle represents Earth. When the interactive is started, the distance (given as 1 AU) is scaled to be that between Earth and the Sun (but you can change this parameter). Time has been sped up so that it is about one month per second for the initial orbit.
>
> **Play Activity**

https://phys.libretexts.org/@go/page/31252

You might be asking a question after the last few examples: How do we know the mass of Earth? After all, we cannot put Earth on a scale, measure its weight, and then convert that to a mass using the second law. In fact, the way we are able to determine the mass of Earth, the Moon, the Sun, and many other objects is by using Newton's law of gravitation. If we know the gravitational acceleration of an orbiting body, we can use it to determine the mass of the object causing that acceleration.

DETERMINING MASSES OF ASTRONOMICAL OBJECTS

Worked Example:

In an earlier activity, we found the centripetal acceleration, *a*, of the Moon to be 2.685 x 10^{-3} m/s^2. Using this value, we can find the mass of Earth.

https://phys.libretexts.org/@go/page/31252

In that earlier activity, we used a combination of Newton's law of gravitation and Newton's second law to derive an expression for the Moon's acceleration:

$$a = G\frac{M_E}{r_M^2}$$

where r_M is the distance between Earth and the moon, M_E is the mass of Earth, and a is the acceleration of the moon. Solving our expression for the mass of Earth, we have

$$M_E = \frac{ar_M^2}{G}$$

Now we can plug in the appropriate numbers to find Earth's mass. We are using the distance from Earth to the Moon, r_M, because we are using the value of the gravitational acceleration felt at the Moon's orbital distance.

M_E = (2.685e-3 m/s^2) (3.844e8 m)^2 / (6.67384e-11 N m^2 kg^{-2}) = 5.945e24 kg

This is within 1% of the value we used previously. We could find Earth's mass in several other similar ways. We could use the measured value of the surface gravity and use that to find Earth's mass. Or we could use one of the many artificial satellites that have been launched into orbit above Earth. Any of these methods would allow us to determine Earth's mass.

Question

This activity demonstrates how the masses for all objects in the Universe are determined. Newton's universal law of gravitation, as its name implies, is assumed to be universally applicable. It is therefore used to determine the mass of stars, planets, galaxies and galaxy clusters.

Newton's universal law of gravitation makes a very strong prediction for the motions of the planets in the Solar System -- namely, that their centripetal accelerations should diminish as the square of their distance from the Sun. So Saturn, which is nearly 10 times farther from the Sun than is the Earth, should have a centripetal acceleration only about 1% as large as Earth's. On the other hand, Mercury, at just under half the distance from the Sun as Earth, should have about four times larger centripetal acceleration than Earth does. Remember that centripetal acceleration depends on both the orbital speed of the planet and its orbital radius. If we set the centripetal acceleration of a planet equal to its gravitational acceleration, we will have a relationship between its orbital speed and its distance from the Sun.

$$\frac{v^2}{r} = \frac{GM_S}{r^2}$$

We can cancel one factor of r (the distance from planet to Sun) from each side of this equation and then take the square root, which gives us an expression for the velocity of a planet as a function of its distance from the Sun.

$$v^2 = \frac{GM_S}{r}$$

$$v = \sqrt{\frac{GM_S}{r}}$$

G and M_S are the same for all the planets in the solar system, so only r is changing from one planet to the next. This means that planets orbiting the Sun (or any central massive object like the Sun) should have orbital velocities that decrease with the *square root* of their distance from that object. This relationship is plotted for our Solar System in Figure 7.7 below. A plot of velocity vs. distance is known as a rotation curve.

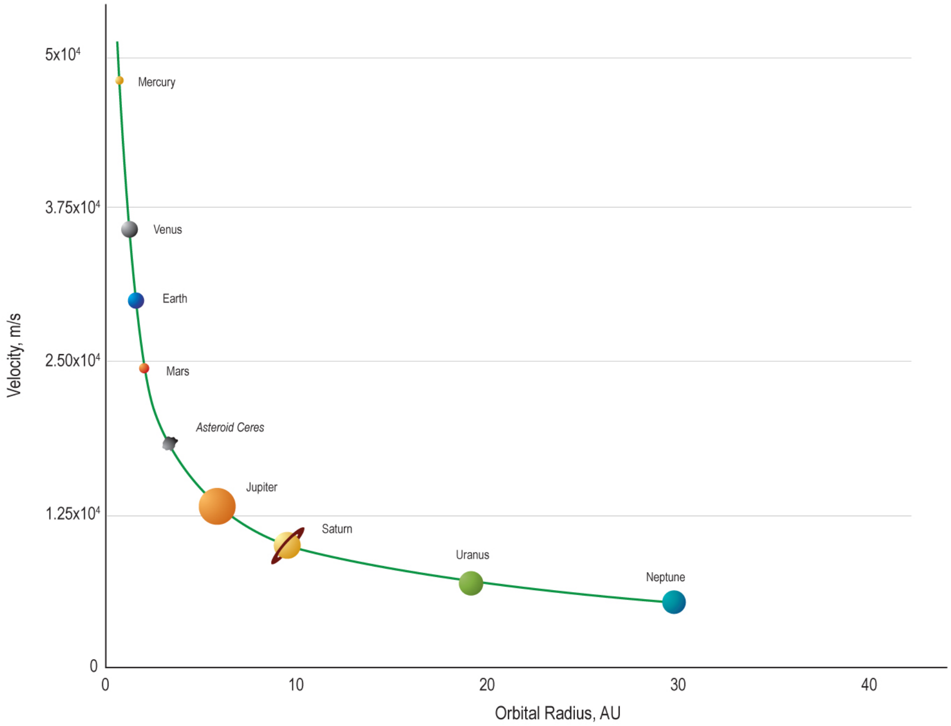


Figure 7.7: The velocities of the planets are plotted as a function of their distance. The rotation curve illustrates that the orbital velocities decrease with the *square root* of their distance from the Sun. Note: the sizes of the planets are not to scale. Credit:NASA/SSU/Aurore Simonnet

All planets in our Solar System follow this relation, but stars in galaxies do not. They follow a different relation that is specific to the details of those systems. However, it is still derived from the laws of gravity.

We can substitute an expression with time for velocity in our previous expressions in order to examine how the orbital time period of planets varies with distance:

$$v=\frac{C}{T}=\frac{2\pi r}{T}=\sqrt{\frac{GM}{r}}$$

where C is the circumference of the orbit, T is the period of the orbit, and M is the mass of the object that is being orbited.

Squaring both sides:

$$\frac{4\pi^2 r^2}{T^2}=\frac{GM}{r}$$

and gathering terms on each side:

$$4\pi^2 r^3=GMT^2$$

or more simply:

$$r^3=\kappa T^2$$

where the constant $\kappa=GM/4\pi^2$. This means that as the orbital distance (*r*) of a planet *increases*, the orbital period (*T*) *increases*. Notice that this expression only depends on the distance to a planet, not its mass. Again, this relationship is true for planets in our Solar System, and the constant κ depends on the mass of the Sun. In other systems, κ will depend on the mass of the object being orbited in that system.

This relationship between the period of an orbit, *T*, and its orbital radius, *r*, in which the square of the orbital period is proportional to the cube of the orbital radius, was discovered by the German astronomer Johannes Kepler (1571 - 1630). For this discovery, Kepler used observational data of planetary motions that had been collected by the Danish astronomer Tycho Brahe (1546 - 1601). Kepler's derivation of this relation (now called Kepler’s third law of planetary motion) was entirely based on these observations. It did not use any theoretical understanding of gravity or motion; at the time there was none.

https://phys.libretexts.org/@go/page/31252

Kepler also discovered that planetary orbits are ellipses, not perfect circles (his first law of planetary motion), and that planets sweep out equal areas in their orbits in equal times (his second law of planetary motion). All of these laws are predictions of Newton's gravitational theory.

Kepler had no knowledge or understanding of the law of gravity because he died more than a decade before Newton was born. His laws were based solely on mathematical fits to observational data. Just as with Galileo's measurement of Earth's surface gravity, Newton's laws provide a theoretical basis for and a unification of Kepler's observational results. Of course, the ability of Newton's laws to explain earlier measurements strongly encouraged Newton and his contemporaries to recognize the validity of his work.

We should note that much of our analysis has been simplified by assuming circular orbits, with the central object not moving. In fact, orbits are elliptical -- or hyperbolic or parabolic, in some cases. The results we have derived still work for the general case of ellipses and other shapes. However a greater level of mathematical sophistication is needed to handle more general orbital shapes. If you wish to see the general treatment of orbits you should take an advanced course in physics.

Finally, it is not quite correct to assume that Earth orbits the Sun, or that the Moon orbits Earth. Instead, we should say that both Earth and the Sun orbit a point between them, as do Earth and the Moon. Recall that the force between Earth and the Sun exerts the same pull on both; Earth pulls on the Sun and the Sun pulls back on Earth just as hard. However, because the Sun is much more massive than Earth, this point for the Earth-Sun interaction is almost at the center of the Sun. Likewise, for the Earth-Moon system it is well within the volume of Earth. But in gravitational interactions, both bodies do move, the more massive one simply moves a lot less. One technique for finding planets around other stars relies on this fact. It uses the tiny back and forth motion of distant stars, evident from a small back and forth shift of their stellar spectral lines, to detect the planets orbiting them.

Orbital Period, Distance, and Velocity

In this activity, you will use the orbit simulator again, but this time to explore how orbital distance affects orbital period and speed.

Play Activity

Keplerian Motion

In this activity, we will use $r^3 = \kappa T^2$ with $\kappa = GM_E/4\pi^2 = 1.01 \times 10^{13}\ m^3/s^2$ to solve problems involving the orbits of satellites around Earth. Using this notation, distance to the satellite from Earth's center (r) should be in meters and the period should be in seconds.

Worked Example:

1. Communications satellites are often put into geosynchronous orbits. This means that from Earth’s surface, they appear to hover over the same spot on the ground. A geosynchronous satellite over a location on Earth’s equator would therefore have an orbital period of 24 hours. Calculate its height above Earth’s surface in meters.

Answer:

- Given: T = 24 ~~hours~~ x (3600 seconds/~~hour~~) = 8.64e4 s
- Find: the distance above Earth’s surface
- Concept: $r^3 = \kappa T^2$
- Solution: r^3 = (1.01e13 m^3/s^2)(8.64e4 s)^2 = 7.54e22 m^3. Taking the cube root gives us r = 4.22e7 m.
- Note that the distance (r) includes Earth’s radius (R_E), so in order to figure out the height above the Earth’s surface we must subtract R_E : Height = $r - R_E$ = 4.22e7 m – 6.4 e6 m = 3.6 e7 m

A schematic of the distances is shown in Figure A.7.4.

https://phys.libretexts.org/@go/page/31252

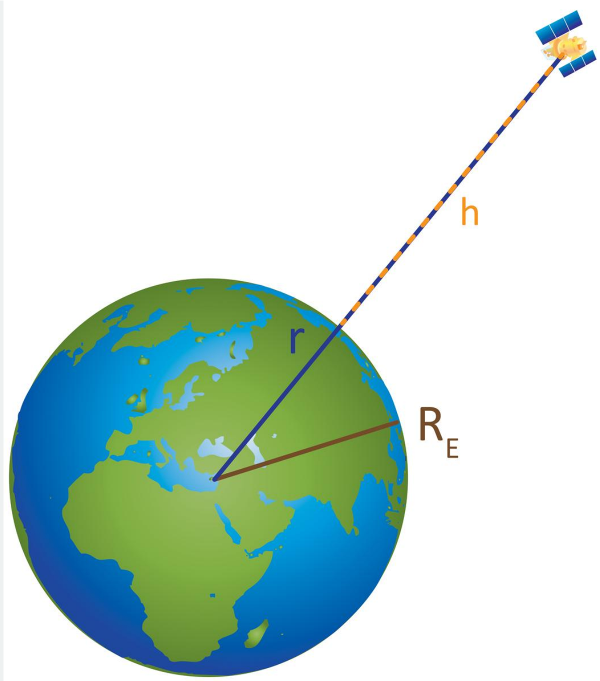


Figure A.7.4: The distance (r) between the center of the Earth and a satellite is a combination of the height of the satellite above Earth's surface (h) and the radius of the Earth (R_E). Credit: NASA/SSU/Aurore Simonnet

Question

This page titled 7.4: Gravity and Orbits is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.5: Forms of Energy

What Do You Think: Energy on the Mountain

We have been describing gravity solely in terms of acceleration or the force of attraction that causes that acceleration. This is correct, but it is an incomplete view of gravity. Sometimes, working with the force of gravity or its acceleration is not convenient. We might only want to understand the general characteristics of a system involving gravity, and be not at all concerned with the details of those characteristics. In such cases it is much easier to work with energy instead of forces and accelerations.

Energy is a word we use in everyday life, and like many terms in these modules, it has a more specific meaning in the context of science. The idea of energy arises in many situations. For example, there are different energy bands of light, there are chemical, gravitational and nuclear energy. Despite the various forms of energy, kinetic, potential, nuclear, chemical and so on, it is always a single number that we can calculate in a given situation. And no matter what type of energy we are dealing with, it can always be characterized by some number of the SI unit joules (J).

In this section, we will look at the concept of energy in some detail. Ideas related to energy will bolster our understanding of how and what changes can occur in the Universe and the objects it contains. The two kinds of energy that we will focus on here will be gravitational potential energy and kinetic energy. We will describe potential energy first.

7.5.1: Potential Energy

Chemical, gravitational, and nuclear energy are all forms of potential energy. For instance, chemical energy is energy that can be stored and released in chemical bonds by changing the configuration of electrons in atoms and molecules. The gravitational energy stored in a proto-star is released as the size of that star slowly decreases. The nuclear energy stored in the configuration of protons and neutrons in hydrogen and helium nuclei can be released when hydrogen is converted to helium, as is happening in the core of the Sun right now. These nuclear reactions are analogous to chemical reactions, but the amount of energy involved in the rearrangement of nuclei is much, much larger than that involved in rearranging the electrons in atoms.

Potential energy always involves the physical arrangement of a system. Let us take an example from gravity to illustrate this point. First, imagine a pinecone hanging in the top of a tall pine tree. Next imagine that same pinecone sitting on the ground beneath the tree (See Figure 7.8). The pinecone has more gravitational potential energy when it is high in the tree than when it is on the ground.

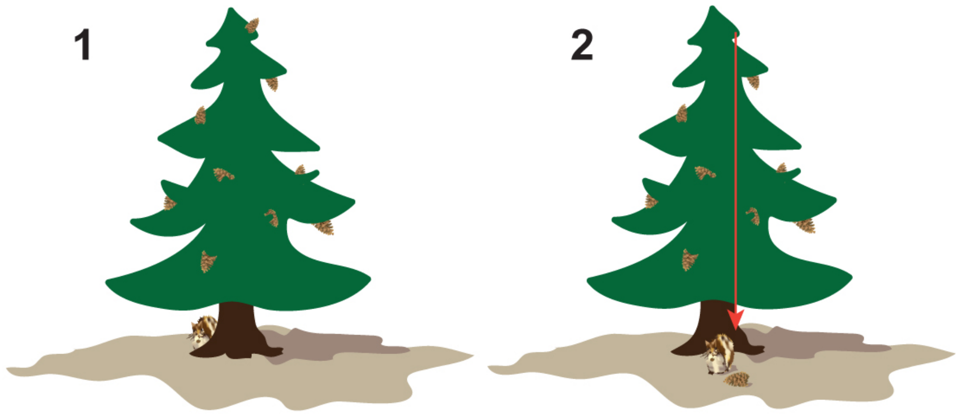


Figure 7.8: A pinecone has more gravitational potential energy when it is at the top of a pine tree than when it sits on the ground below the tree. Credit: NASA/SSU/Aurore Simonnet

There is a simple mathematical expression that describes the gravitational potential energy (*PE*) of an object when the gravitational acceleration can be assumed to be constant, which is certainly the case on Earth's surface:

$$PE = mgh$$

Breaking this expression down, we see that the potential energy is the product of the weight of an object (*mg*) and its height (*h*). From this formula, we see how the units of energy relate to the units of force and distance (or alternatively mass, acceleration, and distance): 1 J = 1 N m or 1 kg $(m/s)^2$.

Some care must be taken here by what height we mean. Is it the height above the ground? The height above sea level? Some other height entirely? It turns out that any of these will do, because potential energy is a *relative* measure of energy. We can only usefully talk about the difference in potential energy between two configurations of a system, not the absolute potential energy of either of them. An activity will help illustrate what this means.

✔ Potential Energy Differences

Assume that a pine tree is 50 meters high and that a pinecone hangs from the topmost point (an unusual place for a pinecone, but it simplifies our example). We will further assume that the pine tree is growing on a hill at a point 1000 meters above sea level. How does the gravitational potential energy of the pinecone change when it falls to the ground? We will assume that the pinecone has a mass equal to 70 grams.

1. First calculate the gravitational potential energy of the pinecone at the top of the tree:

- Given: $m = (70\ \cancel{g}) / (1000\ \cancel{g} / kg) = 0.07\ kg$; $g = 9.8\ m/s^2$; $h = 50\ m + 1000\ m = 1050\ m$
- Find: *PE*
- Concept: From the expression for gravitational PE we have *PE* = *mgh*
- Solution: $PE = (0.07\ kg)(9.8\ m/s^2)(1050\ m) = 720.3\ J$

2. Now calculate the gravitational potential energy the pinecone has at the bottom of the tree:

- Given: $m = (70\ \cancel{g}) / (1000\ \cancel{g} / kg) = 0.07\ kg$; $g = 9.8\ m/s^2$; $h = 1000\ m$
- Find: *PE*
- Concept: *PE* = *mgh*
- Solution: $PE = (0.07\ kg)(9.8\ m/s^2)(1000\ m) = 686\ J$

So, the potential energy the cone has when it is at the top of the tree is 720.3 J and the potential energy the cone has when it sits on the ground below the tree is 686 J. The difference is

$$\Delta PE = 720.3\ J - 686\ K = 34.3\ J$$

We have added the Greek symbol Δ (delta) to indicate that this is a change in the potential energy of the pinecone when it falls to the ground from the top of the tree.

You might notice that the only thing that changes between these two expressions is the height of the pinecone with respect to the tree. We could just leave off the elevation of the tree and use the amount by which the pinecone's elevation changes:

$$\Delta PE = (0.07\,kg)(9.8\,m/s^2)(50\,m) = 34.3\,J$$

This confirms what we have said: potential energy only depends on the configuration of a system, and changes in that configuration cause changes in the potential energy.

Questions

In the previous worked example we could imagine that when the pinecone dropped out of the tree it rolled down the hill, causing it to loose an additional 50 meters of elevation before coming to rest. If that were the case, what would be the total difference in its PE? Do we have to worry about things like the slope of the hill or how bumpy the ground is, or whether there are bushes on the ground that the cone has to break through? No! None of that matters. The only relevant information is the mass of the pinecone and how far it fell in total. In this case, it will lose an additional 34.3 J, and so its total change in PE will be 68.6 J. Half of this amount is from falling out of the tree, and half is from rolling down the hill.

7.5.2: Kinetic Energy

We will return to potential energy shortly, but first we want to explore another kind of energy: kinetic energy. The word kinetic refers to motion, and kinetic energy is the energy that objects have by virtue of the fact that they move. You will not be surprised to find that kinetic energy (*KE*) is also described by a simple mathematical expression:

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

Here, the mass of the particle is m and its speed is v. So you can see immediately that speed makes a big difference in KE. If a particle *doubles* its speed, its energy increases *four-fold*.

UNDERSTANDING KINETIC ENERGY

In this activity you will examine how the mass and speed of an object affect its kinetic energy.

https://phys.libretexts.org/@go/page/31253

Calculating Kinetic Energy

Worked Examples:

1. What is the kinetic energy of a 1000 kg car traveling at 100 km/hr?

- Given: To answer the question we must first convert the speed of the car to m/s.
  v = (100 ~~km~~ / ~~hr~~) (1000 m / 1 ~~km~~) / (3600 s / ~~hr~~) = 27.78 m/s
- Find: kinetic energy *KE*
- Concept: Now we can use the equation for kinetic energy: $KE = ½\, mv^2$
- Solution: $KE = ½$ (1000 kg) (27.78 m/s $)^2$ = 385,802 J = 3.858e5 J

2. Compare this to its energy at 50 km/hr.

At 50 km per hour (kph), the car has half the speed. This factor of ½ in the velocity will be squared when we calculate the energy, so the kinetic energy at that speed will be ¼ of what it was before:

$$(1/2)^2(3.858e5\ J) = (1/4)(3.858e5\ J) = 9.645e4\ J = 96,450\ J \tag{7.5.1}$$

This is why it is considerably safer to drive more slowly.

Questions

https://phys.libretexts.org/@go/page/31253

Just as for potential energy, there is no absolute amount of kinetic energy for a given object. Speed (and velocity) depends on the reference system used to measure it. So, for example, a bird flying 10 m/s will have a different amount of kinetic energy according to someone sitting on a bench watching it than it will as measured by a person riding a bike at 5 m/s. Does the direction of the bike's motion make a difference? Yes, because the relative speed between the bird and the cyclist are different if, say, the two move in the same direction or opposite directions -- or in some intermediate direction. So whenever we talk about energy, either potential or kinetic, we must keep in mind that both of them depend on the reference system we use to measure the speed and position. Observers in different reference systems can get different values for the kinetic and potential energy of particles. The most convenient reference system to use for a given situation will usually be obvious from the conditions of the problem.

7.5.3: Other Forms of Energy

Aside from kinetic and potential energy, there is also energy in radiation, like light. This is something like kinetic energy: light certainly does move. However, light does not have a mass, so we cannot find its kinetic energy using the expression above.

https://phys.libretexts.org/@go/page/31253

Scientists have found that the energy of a particle of light is given by the expression below.

$$E = hf,$$

Here h is called Planck's constant. It is a fundamental constant of nature, like the speed of light, c, and the gravitational constant, G. The other symbol, f, is the frequency of the light.

If you read the chapter on special relativity, you will learn that mass and energy are equivalent: the amount of energy in a given amount of mass is given as $E = mc^2$. This is a sort of potential energy.

So as promised, energy can take on many forms, though they all basically break down into either potential energy or kinetic energy. In the next section we will see more examples that show why energy is such an important and powerful concept.

This page titled 7.5: Forms of Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.6: Conservation of Energy

? Sliding Blocks

As we have already discussed (and you have probably noticed), falling objects speed up as they fall. On the other hand, if you throw an object upward, it slows down, finally stops, and then begins to fall back toward the ground. We already described this motion in terms of the acceleration caused by the force of gravity. There is always a gravitational force downward on objects due to the Earth, even when they are in the air, even when they are traveling upward, even when they are stopped. In fact, it is the gravitational pull that causes them to slow down in the first place. If there were no gravitational force between Earth and, say, a ball, you could throw the ball up and it would just keep going. There would be no forces to slow it down (Newton's first law).

There is also another way to think about why these types of motion occur. In addition to thinking about forces acting on them, we can consider their energy.

7.6.1: Conversion of Kinetic and Potential Energy Near Earth's Surface

In the previous section, we learned that gravitational potential energy increases when an object is placed at an elevated location and that kinetic energy increases as an object moves faster. You might be wondering if the motion of a falling object can be understood in terms of its energy. In fact, in can be. In the example of the pinecone, we imagined that the cone starts out at the top of the tree with a lot of potential energy. It then falls to the ground, and just before it hits the ground we know from experience that it is moving quite fast. How does its potential energy at the top of the tree compare to its kinetic energy just before it impacts the ground? The two are equal.

Energy, it turns out, is one of several quantities in nature that are said to be conserved. This conservation law for energy is one of the most important in science and can be applied in many different situations. (Exceptions only come up in quantum mechanics and are so rare that we need not consider them here, though we will touch upon some of them in later chapters.)

In its simplest form, the law of conservation of energy states that:

Energy can be neither created nor destroyed. It can only be moved from one part of a system to another, or changed from one form to another.

Mathematically, this can be expressed as:

$$E_{final} = E_{initial}$$

where E_{inital} is the total starting energy of a system and E_{final} is the total ending energy of a system. The "forms" of energy referred to are those we have already mentioned: potential, kinetic, and radiation. Conversion of mass to energy can also take place.

Applying this to our pinecone, we can say that its initial energy when it is hanging at the top of the tree, which is entirely gravitational potential energy, *mgh*, is equal to its final energy, entirely kinetic, $\frac{1}{2}mv^2$, just before it hits the ground. The total amount of energy the pinecone has stays the same, but it changes form. As the pinecone falls, the potential energy is converted into kinetic energy.

We can use the law of conservation of energy to trace how energy is transferred and transformed in everyday processes. These include the dropping of pinecones, but also astronomical processes, too.

Choose any everyday task. If you consider it carefully, you will see that the history of the energy involved in it goes back farther than you might at first think. For example, if you walked to class today, where did that energy come from? It came from the chemical bonds in the food you ate for lunch or for breakfast. Where did the food get that energy? From the Sun, directly if it was a plant, and through plants if it was meat. Where did the Sun get its energy? Read Chapter 3 to find out... And so on.

So energy a very useful number. We can calculate how much energy we need for a particular task, and then compare that to the amount of energy we have in a readily available form. We can also trace its origins back in time to see how things in the Universe are (or are not) connected in chains of events, one leading to the next, each acting as a conduit for energy as it moves from the past, through the present, into the future.

As a practical example, we can use the energy method to determine the speed of falling objects. Starting with the law of conservation of energy:

$$E_{final} = E_{initial}$$

And using the case where all of the final energy is kinetic and all of the initial energy is gravitational potential energy, we get:

$$\frac{1}{2}mv^2 = mgh$$

We can solve this equation to find the speed that a pinecone, or any other falling object, is traveling when it impacts the ground:

$$v = \sqrt{2gh}$$

Notice that the mass has again canceled; all objects will have the same final speed regardless of their mass.

In the next activity, we will plug in some numbers from our pinecone example and others to explore the law of conservation of energy.

Potential Energy to Kinetic Energy

Worked Examples:

1. A pinecone of mass 70 g starts out 50 meters above the ground and falls out of a tree. What is its final speed?

- Given: initial height h = 50 m, m = 0.07 kg
- Find: final speed v
- Concept: using conservation of energy and plugging in for *KE* and *PE*, we can use the formula derived above: $v = \sqrt{(2gh)}$
- Solution: $v = [\ (2)\ (9.8\ m/s^2)\ (50\ m)\ ]^{1/2} = (980\ m^2/s^2)^{1/2} = 31\ m/s$

Math Exploration 7.3: Potential Energy to Kinetic Energy

Questions

https://phys.libretexts.org/@go/page/31254

If you have been following our argument closely, you might have a question: Where does the energy go when the pinecone stops on the ground? After all, both the potential and kinetic energy are gone once the pinecone has stopped. If energy cannot be created or destroyed, it must have gone somewhere. And indeed it did.

When the pinecone hits the ground you will notice several things happen. First, it will make a sound. Sound is a vibration in the air (otherwise known as a sound wave), and it has energy. Some of the pinecone's energy goes into the energy of the sound wave. In addition, some dirt, pine needles and other objects will probably be kicked up and flung away from the point of impact. They all have energy, too. Finally, the pinecone itself might break or be deformed in some way, as may the ground itself if a small crater or indentation is created. The ground might also be slightly heated when the pinecone hits, though the heating will likely be too small to notice. This heating and the deformations of the pinecone and ground also require energy. That energy comes from the kinetic energy the pinecone has when it hits. So none of the energy goes missing after the pinecone comes to rest, it just goes into other forms. That is what is expected from the energy conservation law.

Of course, energy conservation works the other way, too. If we throw the pinecone upward at 31 m/s, it will rise to a height of 50 meters before it finally stops. In general, an object thrown upward with a speed v will rise to a height given by $h = ½\, v^2/g$ before it stops and falls back down again.

So how far up can we throw an object? Are there limits, or will it always eventually fall back down? We can use what we have learned to try to predict the answer to this question.

We know that the conservation of energy says that the kinetic energy we give an object thrown straight up must equal the potential energy the object has at the top of its trajectory. As before, we can write this mathematically:

$$\frac{1}{2}mv^2 = mgh$$

According to this expression, as we increase the kinetic energy the object starts with (we throw it upward faster and faster), it will go higher and higher, but it always comes back down after some amount of time. The height h can become arbitrarily large as long as we increase the initial speed of the particle. However, there is a problem with this line of reasoning.

7.6.2: Conversion of Kinetic and Potential Energy Far From Earth's Surface

If you recall, when we introduced the expression for the potential energy we said it was valid as long as the gravitational acceleration, g, was essentially constant. This is true only as long as our distance from Earth's center does not change much. However, as we rise higher and higher above Earth's surface, g does begin to change. We calculated that at the altitude of the

International Space Station, the value of g has dropped from 9.8 m/s^2 to about 8.75 m/s^2, and by the time we get out to the distance of the Moon, g has dropped to only a couple *millimeters per second per second*. Clearly we cannot use the expression valid for a constant g once we have risen a significant distance off Earth's surface.

Not to worry. Physicists know that the expression we have used for potential energy is only an approximation. If the gravitational force changes appreciably along an object's path, due to its changing distance from another massive object, then we must use the correct expression for potential energy. It looks a lot like Newton's law of gravitation, but not quite.

$$PE = -\frac{GMm}{r}$$

Unlike in the expression for the gravitational force, the gravitational potential energy depends on the reciprocal of the distance separating the two objects, not the square of the distance. Also, there are no vector signs here. Unlike force, energy is not a vector. It has no direction. Finally, and this is important, there is a negative sign in front of the potential energy. The negative sign has to do with how we have defined the zero-point for gravitational potential energy. Notice that potential energy is zero when r is infinitely large, otherwise it is *always* negative. This might seem strange at first, but here is a way to think about it: From experience we know that objects tend to move from higher potential energy to lower potential energy. That is why pinecones fall out of trees onto the ground, but they do not fall from the ground up into the trees. With the definition of potential energy as written above, an object at a large elevation will lower its potential energy by moving to a smaller elevation. Notice that as r decreases, the potential energy becomes a larger negative number, or in other words, it becomes *smaller*. (In absolute value it is larger, of course, but negative numbers with large absolute value are smaller than negative numbers with small absolute value.)

So, now we can return to our question, How far up can we throw something so that it will always fall back down? Or to put it another way, how fast must we throw something so that it does not fall back down? In this situation, we have to realize that zero potential energy can no longer be assumed to be at Earth's surface. Now we must define zero potential energy to occur at $r = \infty$. Remember, the zero-point for potential energy is arbitrary, and we are only interested in changes in the potential energy. At Earth's surface the initial energy, call it $E_{initial}$, of the object will be its potential energy there (which is negative) plus whatever kinetic energy we give to it. Assuming the object has mass m and starts out with speed v, we can write its initial energy as below.

$$E_{initial} = \frac{1}{2}mv^2 - \frac{GM_E m}{R_E}$$

The final energy of the particle will be zero because it will be out at $r = \infty$, so its potential energy will be zero, and it will not be moving anymore, so its kinetic energy will also be zero:

$$E_{final} = 0$$

In the following activity, we will work out the details of the answer to our question for Earth, as well as how fast we can throw objects on other planetary bodies before they will no longer fall back down.

How High Can We Throw an Object?

Worked Example:

We want to determine how fast we can launch an object from Earth and have it fall back down. Anything going faster will keep going and anything slower will come back. The speed right on the border is known as the escape velocity.

- The final state of an object in this case will be when it has traveled as far as it can go ($r = \infty$), and has used up all its kinetic energy ($v = 0$) in doing so. In that case, its final energy will be zero: $E_{final} = 0$.
- Since energy must be conserved : $E_{final} = E_{initial}$, its initial energy must also be zero.
- So we can set our expression for $E_{initial}$ above equal to zero:
  $\frac{1}{2}mv^2 - GM_E m / R_E = 0$
- Rearranging terms and solving for the velocity we see that
  $\frac{1}{2} v^2 = GM_E / R_E$
  $v = (2\, GM_E / R_E)^{1/2}$
- This is the escape velocity. For Earth, the value is
  $v = [(2) (6.67384\text{e-}11 \text{ N m}^2 \text{ kg}^{-2}) (5.945\text{e}24 \text{ kg}) / (6.368\text{e}6 \text{ m})]^{1/2}$
  v = 11,160 m/s = 11.16 km/s.

https://phys.libretexts.org/@go/page/31254

To calculate the escape velocity from other astronomical bodies, you would use:

$$v_{escape} = \sqrt{\frac{2GM}{R}}$$

where M is the mass of the object you were launching from and R is its radius.

Questions

If we accelerate an object to its escape velocity, it will never fall back down. It continues to move outward forever, more and more slowly over time. It never comes completely to rest until it has traveled an infinite distance, which of course, it cannot do. In practical terms, this is the speed you must give an object in order for it to escape the gravitational influence of whatever object it is gravitationally bound to.

In the previous activity we found that for Earth, escape velocity is about 11 km/s (about 7 miles per second). Similar calculations lead to values that range from 1.2 km/s for Pluto to close to 60 km/s for Jupiter.

In order to escape the gravity of Earth, an object ejected from its surface must have at least as much kinetic energy as it has potential energy when sitting in the surface. Of course, it could have more kinetic energy. What would happen then? Well, even after moving outward an infinite distance, it would still have some kinetic energy left over. That means the object would still be moving. Under these circumstances we would say that its total energy, potential plus kinetic, is positive:

$$KE + PE > 0$$

On the other hand, if the kinetic energy is less than the potential energy, then the total energy is negative:

\nonumber \]KE+PE<0 \nonumber \]

The boundary case, where the total energy is zero, is the one we used to define escape velocity.

In general, though, objects (or systems of objects) that have negative total energy are considered gravitationally bound. Objects with zero or positive total energy are gravitationally unbound. We don't have to know anything about the details of the motions within a system to make these determinations.

We will explore this idea further in the following activity.

Energy of Binary Systems

Worked Example:

Consider the system consisting of Earth and the Sun: Earth orbits the Sun at a speed of about 30 km/s. Its distance from the Sun is 1.49598 X 10^{11} m. Earth's mass is 5.945 X 10^{24} kg, and the Sun's mass is 1.9884 X 10^{30} kg.

Using these numbers, we can find the total energy of the Earth-orbital system. We ignore the Moon's motion and Earth's spin for this exercise.

- The total energy is a combination of the kinetic and potential energy:
  $E_{tot} = KE + PE = ½M_E v_E^2 - GM_E M_s / r_E$
- First calculate the kinetic energy piece:

KE = (½) (5.945e24 kg) [(30 ~~km~~/s) (1000 m / ~~km~~)]^2 = 2.675e33 J

- Now calculate the potential energy piece:

  PE = -(6.67384e-11 N m^2 kg^{-2}) (5.945e24 kg) (1.9884e30 kg)/ (1.49598e11 m) = -5.273e33 J
- Adding these together we have

  E_{tot} = 2.675e33 J - 5.273e33 J = -2.598e33 J

As we expect, the total energy of Earth as it orbits the Sun, potential plus kinetic, is negative. This means Earth is in a bound orbit around the Sun. If this were not the case, Earth would not be bound to the Sun and would go flying off into space. Also, notice that the magnitude of the potential energy is roughly twice as large as the magnitude of the kinetic energy. This is typical for gravitationally bound dynamical systems.

Questions

A. Calculate the total energy of the Earth-Moon binary system using a similar procedure to that above:

B. Now consider a comet approaching the Sun

We have assumed so far that orbits are circular for simplicity's sake. However, orbits are not usually circular. If they are bound, then orbits are generally ellipses, of which a circle is a special case. Two kinds of unbound orbits exist: parabolic and hyperbolic. But we will concentrate on the bound orbits for the moment. The following activity asks you to answer some questions about the energy of an object orbiting the Sun.

ORBITAL ENERGETICS

Figure A.7.5 shows a possible bound orbit. We are looking directly down upon the orbit from above, so this is not a perspective drawing of a circular orbit. Notice that the Sun is not in the center of the orbit, but off to one side at a point called a focus of the ellipse. The figure shows an orbiting body, perhaps a comet or an asteroid, at several points along its orbit.

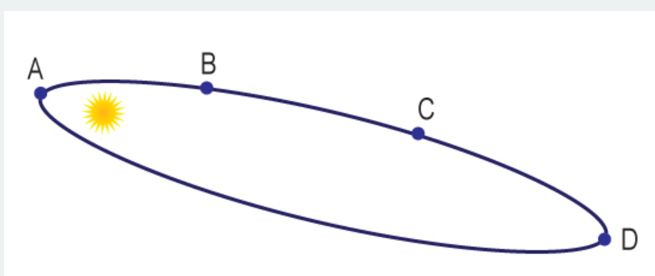


Figure A.7.5: Elliptical orbit of an object around the Sun. The orbiting object is shown at four different points in its orbit around the Sun. Credit: NASA/SSU/Aurore Simonnet

In Figure A.7.5, four different positions of the object are labeled A, B, C, and D.

Questions

This page titled 7.6: Conservation of Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.7: Wrapping It Up 7 - The Galilean Moons of Jupiter

You are tasked with determining the mass of Jupiter by making careful observations of three of the planet's four major moons: Callisto, Europa, and Ganymede. Io's data have already been taken and analyzed as an example.

The mass of an object can be determined through Newton's laws. Earlier in this chapter, we found that:

$$4\pi^2 r^3 = GMT^2$$

where M is the mass of an object being orbited, r is the distance to an orbiting object, and T is the period of the orbiting object. We can re-write this to create an expression for the mass of the object (in this case Jupiter):

$$M = \frac{4\pi^2 r^3}{GT^2}$$

In this activity, you will measure the period and distance of three of Jupiter's moons and use that information to determine the mass of Jupiter.

7.7.1: Part I: Data Acquisition

Play Activity

Click on the moons in the telescope's field of view, which will allow your computer to measure the moon's distance from Jupiter.

- Do this for each of the three moons and then click "next" to retrieve the next day's image.
- Repeat the process for at least 17 consecutive days worth of data (preferably 33 days).

Don't worry if you cannot click on a particular moon during one or more days. You will still gather enough data.

The data used in this activity were obtained from observatory.tamu.edu:8080/Widgets/galilean.html.

7.7.2: Part II: Data Analysis

Now that you have collected 17-33 days worth of data on where the moons of Jupiter are in relation to their parent planet, you can perform data analysis on your results.

- Click on the "Data Analysis" button on the lower right side of the telescope's field of view. This brings up a graph of the data for Io and for the three moons you observed: Europa, Ganymede, and Callisto.
- Click on the "Europa" tab at the top and then click on the "Overlay Sine Wave." The blue line represents the overlay.
- Adjust the Amplitude , Period, and Phase sliders until the overlay matches your data. Repeat this process for your Ganymede and Callisto data.

7.7.3: Part III: Results

Once you have analyzed your data, click on the “Results” button next to the “Data Analysis” button on the lower right. This will bring up the Results window, which you can drag around the field so you can better see your data analysis results.

The Results tool will automatically calculate the mass of Jupiter based on the expression derived from Newton’s laws. To do this:

- Fill out the fields in the Results window with your data, where “amp” indicates the measured amplitude and “per” the measured period .
- Once you have filled out a moon ’s data into the equation, click on the equals sign to get the result.

- Once you have all four moons filled out (Io is already done as an example), click on the equal sign for the “Averaged Jupiter Mass” at the bottom of the Results window.

7.7.4: Part IV: Reporting

7.7.5: Part V. Discussion

We can also use our data to examine the orbital energetics of Jupiter's moons.

*The radius of Jupiter is 7.15e7 m
** The number of seconds in a day is 8.64e4 s/day

This page titled 7.7: Wrapping It Up 7 - The Galilean Moons of Jupiter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

7.8: Mission Report 7 - The Galilean Moons of Jupiter

7.8.1: D. Questions to be graded for accuracy. Show your work!

The James Webb Space Telescope (JWST) is the successor to the Hubble Space Telescope (HST). JWST is currently being developed by NASA scientists and engineers with a goal for a launch date in 2018. JWST will observe the sky in infrared wavelengths, using a much bigger mirror than HST. Unlike HST, which orbits fairly close to Earth (559 km away), JWST will be sent into orbit quite a bit further from Earth, at a distance of about 1.5 million km away from Earth. In fact, it is more helpful to think about JWST orbiting the Sun than orbiting the Earth. While in orbit, JWST will remain farther away from the Sun than the Earth is. JWST will "keep up" with the Earth as it orbits the Sun, so that the Earth will always lie between JWST and the Sun, as in the Figure below.

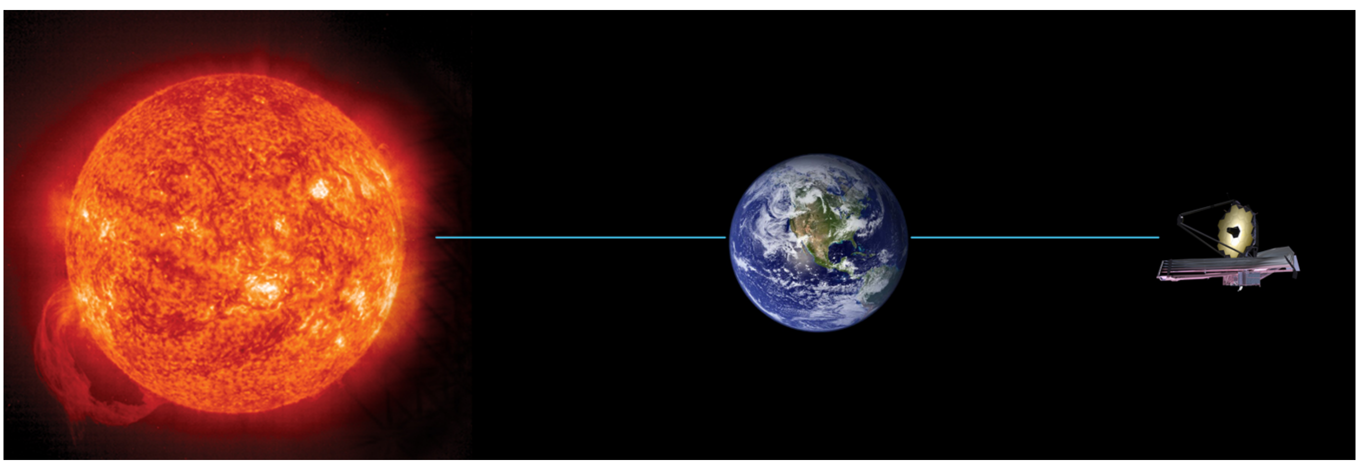


Figure 7.8.1 JWST will orbit the Sun, keeping up with the Earth. Drawing is not to scale. Credit: NASA/SSU/Aurore Simonnet

For a more detailed description of JWST's planned orbit, and a figure showing the Sun, Earth, and JWST, you can visit this website.

1. Force due to gravity.

2. Velocity

3. Kinetic and Potential Energy

4. Total Energy

This page titled 7.8: Mission Report 7 - The Galilean Moons of Jupiter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

8: Dark Matter

Chapter 8 focuses on applying the law of gravity to the motions of astronomical systems to determine how much matter is present and how it is distributed. In the first half of the chapter, you will develop models of rotating systems for various velocity and mass distributions. Next you will explore the evidence for dark matter in galaxies and galaxy clusters by comparing the luminous mass and total gravitational mass. You will also explore models for what dark matter might be and why its nature remains so elusive.

This page titled 8: Dark Matter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.0: Dark Matter Introduction

Video Transcript

> *Small children on a merry-go-round hold on tight as it spins around. Holding on, they resist the forces acting on them because of the circular motion.*
>
> *Just as the children circle around the center of the merry-go-round, water swirls around a drain before disappearing, and tires on a car rotate around the car's axle.*
> *In the vastness of the Universe, among the objects that we can see, we've discovered many similarities and differences. One prominent similarity is the tendency for objects to be spinning and revolving - to have an angular momentum. Planets and stars spin about their axes, moons orbit around their planets, planets orbit around their stars, and stars and gas orbit within their parent galaxy.*
>
> *It was Johannes Kepler who, in the 17th century, discovered the relationship between a planet's distance from the Sun and its orbital period. Building upon that, less than 70 years later, Isaac Newton deduced that the force between a planet and the Sun depends on the distance between them and their masses. According to Newtonian theory, for two objects orbiting around each other gravitational forces determine the distance and speed at which they rotate about their common center of mass.*
> *For centuries, Newtonian gravity was tested against observations of the planets and moons and they almost uniformly agreed. Then, in 1915, Einstein's general relativity supplanted Newton's understanding of gravity and changed almost agreed to totally agreed. According to general relativity, gravity is not a force at all, but is instead a stretching and twisting of space and time.*
>
> *Whether using Newton's view of gravity as a force or Einstein's view of gravity as distortion in spacetime, astronomers have discovered that something is amiss at the largest of size scales: the mass to orbital period relationship that works so perfectly at describing the orbits of planets, that placed astronauts on the moon and robots on Mars, appears to break down when predicting the motions of stars and gas within a galaxy or of the galaxies themselves within their clusters.*
> *In this chapter, we will explore this mystery and the new theory scientists have proposed to explain why the old explanations no longer seem to work at these large scales.*

Previously, we looked at Newton's law of gravitation and how it provides an explanation for the motion of falling objects near Earth's surface, the motion of the Moon and satellites around Earth, and the motion of the planets around the Sun.

In this chapter we step up in scale, applying the law of gravity to the motions of stars and gas in galaxies and galaxy clusters. These larger-scale gravitational studies began early in the 20th century and continue to the present. They have resulted in surprising revelations about our Universe and its composition. Here we begin to explore some of these revelations.

This page titled 8.0: Dark Matter Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.1: Making Models for Rotation

Learning Objectives

- You will be able to use multiple representations to understand a velocity distribution (words, graphs, pictures, equations).
- You will be able to use multiple representations (words, graphs, pictures, equations) to understand an everyday example of a system with a velocity distribution that is constant with radius.
- You will be able to use multiple representations (words, graphs, pictures, equations) to understand everyday examples of systems with velocity distributions that are inversely proportional to radius.

What Do You Think: Models

In this chapter, we will talk a lot about models. Not the kind of models you might have made as a child, of planes or boats or cars. The models we will talk about will be conceptual models. They will involve mathematics and physics. In the dialogue above, models of galaxies are mentioned. What other models can you think of that we use in everyday life? Try to think of several common examples of conceptual and physical models that are commonly used.

This section of the chapter will be focused on using models to describe the velocities of objects as they rotate. Later, we will use models to describe how mass is distributed in objects. We will link our models for mass and velocity using the physics of gravity. Later still, we will compare models to better understand the motions of stars and gas in galaxies, and the motions of galaxies in galaxy clusters. Then we will explore the constituents of the mass of galaxies and galaxy clusters. As you might have guessed from the title of this chapter, dark matter will play a large role in the discussions in this chapter.

8.1.1: Rotation of a Rigid Disk

When we think of rotating objects, some of the first things that come to mind are wheels spinning about their axles, or (if you are old enough) records or CDs. Many objects in nature rotate this way. Because these examples are solid, disk-shaped objects, we will call them rigid disks in this chapter.

We can take a closer look at the example of a wheel. More specifically, look at how the solid material that makes up a wheel rotates about its axle. Suppose the wheel rotates once per second, or in other words, it rotates with a frequency of 1 Hz. A point on the wheel that is close to the axle will go around once each second, and a point on the wheel that is close to the outer edge will also go around once each second. Any point on the wheel makes one full rotation each second.

Now think about the velocities of three different points on the wheel (Figure 8.1.1). Point A, toward the center of the wheel, will go around a small circle each second. Point B, which is farther out from the center than Point A, will go around a larger circle each second. Point C, on the outer edge of the wheel, will go around an even larger circle each second.

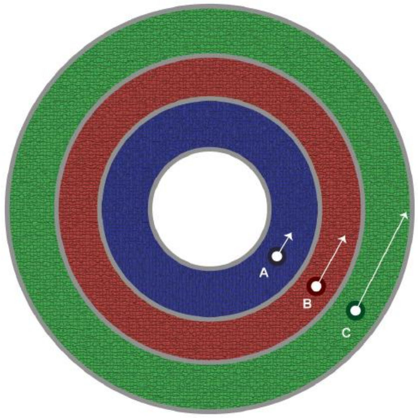


Figure 8.1.1: Rotation of points on a wheel. Points A, B, and C on this wheel all go around a circle in the same amount of time. However, since they are at different distances from the center of the wheel, the sizes of the circles they travel around are different, and so their velocities are different. Credit: NASA/SSU/Aurore Simonnet

Example: Velocities at Different Points on a Rigid Disk

Let's calculate the velocity of Point A on the wheel. Suppose that Point A is 2 m from the center of the wheel and the wheel is rotating with a frequency of 1 Hz.

- Given: radius $r = 2m$, $f = 1Hz$
- Find: velocity v
- Concepts: We can get the velocity from: velocity = distance / time

The size of the circle that Point A will travel around can be calculated using the formula for circumference:

$$C = 2\pi r$$

And if the frequency is 1 Hz, the time to travel once around is 1 second.

Solve

The distance Point A will travel in one second is

$$C = (2)(\pi)(2m) = 12.6\, m.$$

The time to travel once around is 1 second, so the velocity of Point A is v = 12.6 m / 1 s = 12.6 m/s.

Questions

To understand the motion at various points on the wheel, it can be helpful to graph the rotation curve of the wheel. A rotation curve is a graph of velocity versus radius. We do this because it not only helps us better understand the motion of the wheel, but by comparing the rotation curves of different objects (such as wheels, or even the Solar System and galaxies), we can see whether they rotate in the same way or differently. We can also use some physics to begin to examine why they rotate the way they do.

Rotation Curve of a Rigid Disk (Wheel)

We can use the graphing tool to graph the rotation curve of the wheel described above. We want to plot radius on the x-axis, and velocity on the y axis. We can type the following three data points into the graphing tool (we will ignore units for the time being):

- Point A, $r=2$, $v=12.6$
- Point B, $r=10$, $v=62.8$
- Point C, $r=30$, $v=188.5$

The resulting graph looks like the one in Figure $8.1.1A$. See if you can make a graph like this, using the graphing tool.

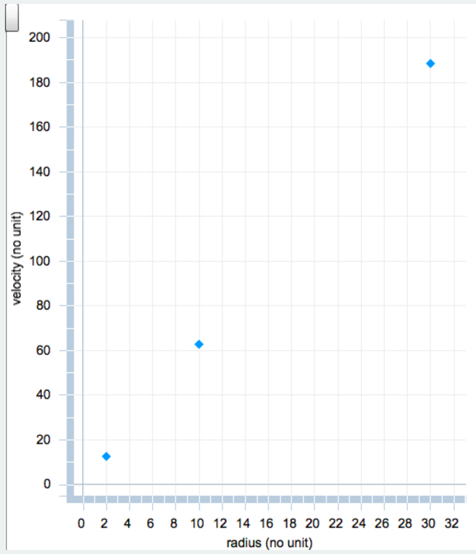
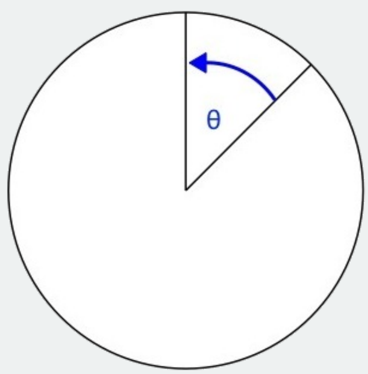


Figure $8.1.1A$: This graph represents the rotation curve of a wheel. Credit: Sonoma State University

The points on the rotation curve for a rigid disk form a straight line. This tells us that the velocities at different points on the wheel are proportional to their distances from the axle. In other words, as the distance from the center of the wheel gets larger, the velocity of the point on the wheel gets larger. We write this mathematically as

$$v \propto r$$

where v corresponds to velocity, and r corresponds to radius, or distance from the wheel's axle. In this equation, the Greek symbol $\propto$ is used to represent proportionality. This relationship between velocity and radius holds for any rigid disk that is rotating about its center. Examples include a CD, the wheel of a bicycle, and an LP record.

> **Going Further 8.1: Linear and Angular Velocity**
>
> Until you do the math, like we have above, it may not seem obvious that a point toward the center of a rotating rigid disk, like a wheel, has a smaller velocity than a point farther from the center of the wheel. Since both points go around a circle in the same amount of time, it seems like they should have the same velocity.
>
> However, as we have calculated, the points have different velocities because they are going around different-sized circles. The inner point goes around a smaller circle, so it travels a smaller distance in the same amount of time that the outer point travels around a larger circle (larger distance).
>
> This is how the calculation works out when we study the linear velocity of the points on the wheel. Linear velocity refers to the linear distance an object has moved (and in what direction) in a given amount of time. This is the kind of velocity we are used to thinking about and have concentrated on so far in these modules:
>
> $$\text{linear velocity} = \frac{\text{distance}}{time} = \frac{d}{t}$$
>
> However, physicists also use angular velocity to describe the motion of objects that turn. Angular velocity is defined as the size of the angle an object rotates through in a given amount of time (Figure $8.1.1B$)
>
> $$\text{angular velocity} = \frac{\text{angle}}{\text{time}} = \frac{\theta}{t}$$
>
> Look at the angular velocities of different points on a rotating wheel—it may help to refer back to Figure $8.1.1$. Point A, close to the center of the wheel, Point B, a bit further out, and Point C, toward the edge, all make one rotation in the same amount of time. This means that each point rotates through an angle of 360 degrees (one full circle) in the same amount of time. In other words, all three points have the same angular velocity.
>
> So, if your intuition was telling you earlier that all three points had the same velocity, you were on the right track—they do all have the same *angular* velocity, and in some cases it is very helpful to study angular velocity. However, for the rest of this chapter, we will continue to focus on studying *linear* velocity.
>
> 
>
>
> Figure $8.1.1B$: Angular velocity. This figure shows the angle a disk rotates through as it spins. Credit: Sonoma State University

8.1.2: Swirling Vortex

Not every rotating thing we can think of in our everyday lives rotates like a rigid disk, or wheel. For example, we can think of water swirling toward a drain, or dust devils and tornadoes— though hopefully you are not experiencing too many tornadoes in your everyday life! These sorts of phenomena are called vortices (plural for vortex).

In a vortex such as water swirling toward a drain, the rotation is actually very different than for a rigid disk like a wheel. Usually, the water closest to the drain will make one full rotation in a shorter amount of time than the water farther from the drain. The two movies below compare rigid disk rotation and rotation that is more similar to water swirling down a drain.

Figure 8.1.2: To the left is a movie showing rigid disk rotation, in which a point on a ring makes one rotation in the same amount of time as a point on any other ring. To the right, we see a movie showing rotation like that of water going down a drain. In this case, points on rings closer to the center make one full rotation in a shorter amount of time than points on rings farther from the center. In other words, a point on an inner ring will “lap” a point on an outer ring. Credit: Wikipedia Creative Commons.

There are two types of velocity models that will result in rotation similar to that of water going down a drain, where material closest to the center of rotation will make a full rotation in a shorter amount of time than material farther from the center of rotation. The first type of velocity model we will talk about may seem the most obvious: it is the case where material closer to the center of rotation has a higher velocity than material farther from the drain.

We can look at an example to understand this better. Suppose a physicist is studying the motion of water as it swirls toward the drain in her lab sink. She fills up the sink and drops three floats in the water at different distances from the drain. She then times how long it takes each float to make one circuit around the drain (Figure 8.1.3).

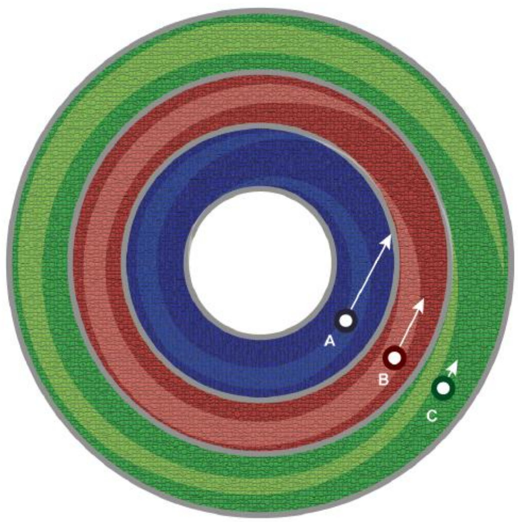


Figure 8.1.3: In this example of water going down a drain, the floats at points A, B, and C each make one full rotation in different amounts of time. Credit: NASA/SSU/Aurore Simonnet

The physicist hypothesizes that she can describe the motion of the water using the following velocity model:

$$v = \frac{0.05m^2/s}{r}$$

where v is the velocity of the water at a given point in m/s, and *r* is the radius (or the distance) between that point and the drain, given in meters. Note that in this model, velocity is *inversely proportional* to radius. In other words, as the distance from the drain *increases*, the velocity of the water *decreases*. In the activity below, we will calculate how it would take each float to circle the drain once.

Worked Example: Velocities Decreasing With Increasing Radius

Assume Float A is 0.1 m away from the drain. Using the physicist’s velocity model, what would we expect the velocity of Float A to be as it rotates around the drain?

- Given: $r = 0.1\,m$
- Find: v, Float A’s velocity
- Concept: $v = (0.5\,m^2/s)/r$
- Solution: $v = (0.05\,m^2/s)/(0.10m)$ or $v = 0.5m/s$

How long would it take the float to circle the drain once?

- Given: $v = 0.5\,m/s$ and $r = 0.1\,m$
- Find: t, the time it takes Float A to circle the drain once (also known as the period)
- Concepts: $v = d/t$, $d = 2/\pi r$

- Solution: $d = 2\pi r = 2\pi(0.1m) = 0.628\, m$
- Using some algebra, we can convert the expression $v = d/t$ to $t = d/v$
- t = (0.628 m) / (0.5 m/s)
- t = 1.26 s

Questions

Rotation Curve of Water Rotating Around Drain

1. Now use the graphing tool to look at the rotation curve that results from the physicist's model of water rotating around the drain of her lab sink. Use the velocities calculated for Floats A, B, and C in our example, and the distances (radii) of these floats from the drain. Be sure to plot the radii on the x-axis, and the velocities on the y-axis. Describe your graph.

2. Your plot should look like Figure $8.1.3A$: below. Note that in the graph, the closer the float is to the drain, the higher is its velocity. The farther the float is from the drain, the lower is its velocity. This is what we expect from the equation the physicist used to model the motion of water in her sink, in which velocity is inversely proportional to radius.

Reconcile any differences between Figure $8.1.3A$: and your graph and describe them here.

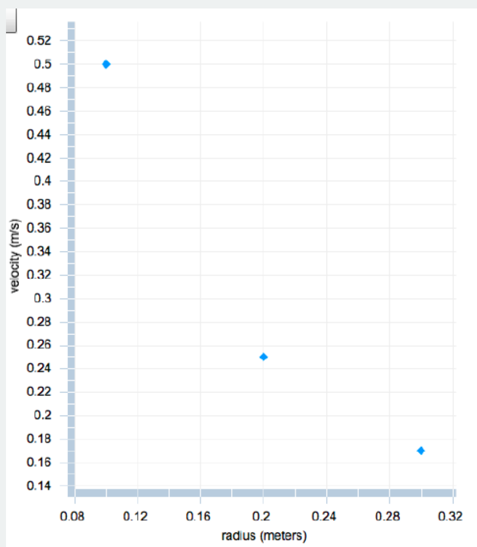


Figure $8.1.3A$: Rotation curve of water rotating around drain. This graph shows the rotation curve that results from a velocity model in which velocity is inversely proportional to radius. Credit: NASA/SSU

8.1.3: Cars Driving Around a Roundabout

There is another, less intuitive velocity model that can result in an object or point toward the center of rotation making a full circle in a shorter amount of time than an object or point farther from the center of rotation (like in the case of water rotating around a drain). This is the case where the velocities at all points, no matter what their distances are from the center, are the same. Imagine that three cars are driving around a roundabout. Each car is in a different lane (Figure $8.1.4$), but they all are traveling at the same speed.

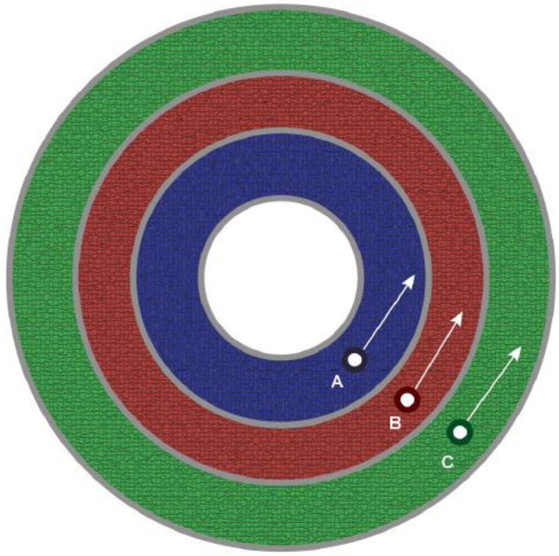


Figure 8.1.4: Cars A, B, and C in this example all have the same velocity. However, they will travel around different-sized circles (depending on their distance from the center of rotation, or which lane they are in). Therefore, it will take each car a different amount of time to go around the roundabout, or make one full rotation. Credit: NASA/SSU/Aurore Simonnet

Mathematically, we can describe the cars' velocities with a model in which the velocity is a constant number, since each of the cars has the same velocity:

$$v = constant$$

In other words, the velocity at any point (in any lane) is the same and does not depend at all on the distance from the center of rotation (the center of the roundabout). We will perform some calculations to see how this velocity model can result in cars at different distances from the center of rotation taking different amounts of time to complete a full circle.

Worked Example: Cars Driving Around Roundabout: Constant Velocity Model

We will describe the cars' motion by a constant velocity model with v = 22 m/s. That means the velocity any car in any lane is 22 m/s, no matter how near or far the car is from the center of the roundabout.

1. Assume Car A is 5 m away from the center of the roundabout. How much time will it take Car A to make one full circle around the roundabout?

- Given: v = 22 m/s and r = 5 m
- Find: t, the time it takes Car A to circle the roundabout once
- Concept: t = d / v, d = 2πr
- Solution: d = 2πr = 2π(f m) = 31.4 m
  → t = (31.4 m) / (22 m/s) = 1.43 s

Questions

At all distances from the center of rotation, the velocity is the same. So the graph looks like a horizontal line (Figure $8.1.4A$). This rotation curve graph may not seem very exciting, but we will see that it can teach us a great deal when we discuss the rotation curves of galaxies!

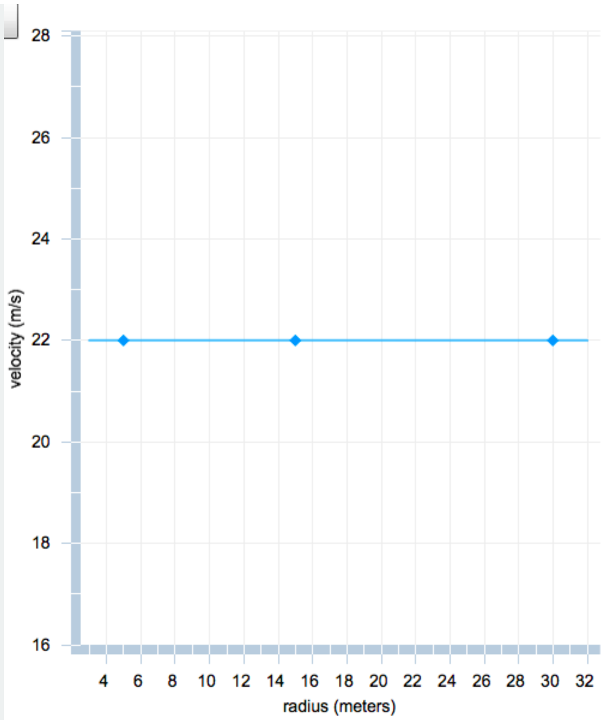


Figure 8.1.4A: Rotation curve for constant velocity model. This graph shows the rotation curve that results from a constant velocity model (here the velocity is 22 m/s). Credit: Sonoma State University

In this section, we have studied several different mathematical relations between velocity and radius. This next interactive will give you the opportunity to connect the equations that you have studied to an animation created using a simple computerized model.

VELOCITIES AND GRAPHS

Play Simulation

The multi-colored rings on the left-hand side of the activity can be arranged to move at different rates by choosing a rotation curve (X, Y, or Z) from the menu above the graph on the right-hand side. Which rotation curve best describes each of the following situations?

This page titled 8.1: Making Models for Rotation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.2: Velocities, Mass, and Gravity - The Solar System

> **?** What Do You Think: Rotation of the Solar System

The systems we have talked about so far in this chapter (wheels and water swirling around a drain) may not seem like they have anything to do with astronomy. But as you will see, these everyday examples will help us understand the motions of planets in the Solar System, the motions of stars and gas in galaxies, and the motions of galaxies in galaxy clusters. Here, we will look at our first astronomical example of rotation, but we will start locally, astronomically speaking— we will look at how the Solar System rotates.

8.2.1: Rotation of the Solar System

We will start our investigation of the Solar System's rotation by listing the average velocities of the planets as they orbit the Sun, and the distances of the planets from the Sun, and see what rotation curve results. Later in this section, we will examine why the Solar System has this kind of rotation curve, using the laws of gravity. Below is a table of average distances of the planets from the Sun, and their average orbital speeds:

Table 8.1: Distances and Velocities for Solar System Objects

PLANET	AVERAGE DISTANCE FROM SUN (AU)	AVERAGE ORBITAL SPEED (KM/S)
Mercury	0.39	47.9
Venus	0.72	35.0
Earth	1.00	29.8
Mars	1.52	24.0
Jupiter	5.20	13.1
Saturn	9.58	9.69
Uranus	19.2	6.81
Neptune	30.1	5.43

We can use the numbers above to look at the rotation curve of the Solar System (Figure 8.6). Does this look like any of the rotation curves we discussed in section 8.1? We will discuss why the rotation curve of the Solar System looks the way it does as we move further into this chapter.

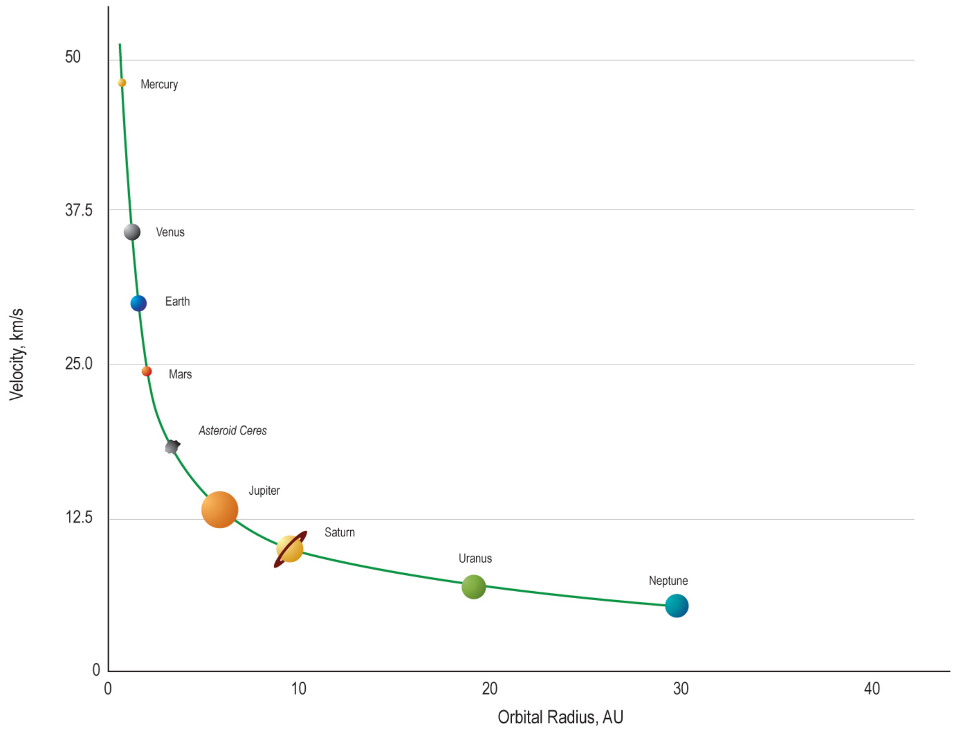


Figure 8.6: The rotation curve of the Solar System shows that the inner planets rotate around the Sun with faster velocities than the outer planets. Credit: NASA/SSU/Aurore Simonnet

Velocities in the Solar System

The dwarf planet Pluto has an average distance of 39.3 AU from the Sun, and an average orbital speed of 4.67 km/s around the Sun. Dwarf planet Eris has an average distance of 68.0 AU from the Sun, and an average orbital speed of 3.44 km/s.

Use Figure 8.6 above, or add these data to your own graph (using the graphing tool), and answer the questions below.

8.2.2: Gravity and the Mass Distribution of the Solar System

By looking at the rotation curve of the Solar System and comparing it to the examples we discussed in Section 8.1, you will notice that the motion of the planets in orbit around the Sun resembles the motion of water swirling around a drain. More specifically, the planets' motion resembles the model we discussed in which velocities decrease with increasing radius. Why do the planets orbit the Sun in this way? Or put more generally, why does the Solar System rotate this way? The reason for this has to do with gravity.

The force of gravity depends on mass, so it will help our understanding to think about how mass is distributed in the Solar System. We know that the most massive objects in the Solar System are the Sun and the planets. But really, the Sun is so massive— far more massive than all the planets combined— that we can ignore the masses of the planets as we model the mass distribution of the Solar System. We can also, of course, ignore the masses of asteroids, comets, and dwarf planets.

Despite having nearly all the mass in the solar system, the sun is relatively tiny in extent; the diameter of the Sun is much, much smaller than the distances between the planets and the Sun. Given these circumstances, we may model the Solar System's mass distribution very simply. To high precision, we can assume that all the mass in the Solar System is concentrated in a point at the position of the center of the Sun. Because we are modeling all the mass as being at a single point, this is called a point mass model.

We can directly use Newton's laws of gravitation and motion to determine the orbital velocity of an object under these assumptions.

$$\frac{GM}{r^2} = \frac{v^2}{r}$$

On the left side of the equation, we have the expression for acceleration due to gravity (from Newton's law of gravitation). On the right, we have the expression for centripetal acceleration. Here, *v* refers to the speed of an orbiting object, *r* refers to the radius of the object's orbit (the distance between the planet and the Sun), and M refers to the total mass enclosed by the orbit. We are assuming circular orbits for simplicity's sake, but this assumption is not necessary. It just makes the math easier to manage.

In our point mass model, *M* is the mass of the Sun. The masses of any planets inside the orbit do not contribute enough to be important, and it is only the mass of the Sun that will be enclosed by the planet's orbit. In other examples that we look at later, we will see that the entire mass of a system is not necessarily enclosed by an object's orbit. Here, however, we can simply plug the mass of the Sun into the equation above.

Continuing with our Solar System example, we cancel one factor of r on the bottom of both sides of the equation. We are then left with the following equation.

$$v^2 = \frac{GM}{r}$$

or

$$v \propto \frac{1}{r^{1/2}}$$

In this case, velocity is proportional to the square root of the distance (r) from the point mass. Velocity will decrease with increasing distance, though in a slightly different way than the example we discussed in Section 8.1.2.

You should complete the numerical activities below to get a better sense for what the proportionality above means. Note that astronomers refer to this relationship between velocity and distance (orbital radius) as Keplerian rotation because it describes the motion of the planets. It is Johannes Kepler who is famous for having derived this expression empirically from observations of the planets in the 17th century.

✔ Keplerian Motion

1. The Earth is 1 AU from the Sun, and Jupiter is 5.2 AU from the Sun. Using the proportionality expression for Keplerian rotation, calculate how much faster we would expect Earth's orbital velocity to be than Jupiter's. How does that compare to the observed ratio between the two planets' velocities?

- Given: r_{Earth} = 1 AU, $r_{Jupiter}$ = 5.2 AU
- Find: $v_{Earth} / v_{Jupiter}$
- Concept: In Math Exploration 8.1, we find that:

$$\frac{v_{Earth}}{v_{Jupiter}} = \frac{r^{1/2}_{Jupiter}}{r^{1/2}_{Earth}}$$

- Solve: Putting in the numbers, we find:

$$\frac{r^{1/2}_{Earth}}{r^{1/2}_{Jupiter}} = \sqrt{\frac{(5.20AU)}{(1AU)}} = \sqrt{5.20} = 2.28$$

The actual ratio between the two planets' observed velocities from Table 8.1 is:

$$\frac{v_{Earth}}{v_{Jupiter}}=\frac{29.8km/s}{13.1km/s}=2.27$$

This is in good agreement with the predicted ratio we calculated based on the Keplerian rotation model.

Math Exploration 8.1: Keplerian Motion

Inverse Square-root Proportionality

Now we can look at the inverse square-root relationship in another mathematical way— by graphing. First, we will assume we have an inverse square-root relationship between v and r like this:

$$v = \sqrt{\frac{5}{r}}$$

Now, use the graphing tool to plot the above six data points (where radius corresponds to the x-coordinate, and velocity corresponds to the y-coordinate). In other words, use the data in the table above to plot a rotation curve.

This page titled 8.2: Velocities, Mass, and Gravity - The Solar System is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.3: Gravity and Models for Different Mass Distributions

What Do You Think: Mass Distributions

In the previous section we learned that the velocities of orbiting objects can be modeled and understood using mathematical formulae and graphs. In this section, we will see that the distribution of mass in astronomical objects can be modeled and understood in the same way. We will examine several example mass models to build our understanding. We will then link these example models to velocity using what we already know about gravity, just as we did in the case of the Solar System.

The mass models we describe here will seem abstract at first. However, they are very helpful for understanding the motions of stars and gas within galaxies, and they can also be used to understand the motions of galaxies within clusters of galaxies.

An important note before we carry on: In the everyday examples of rotation that we discussed in Section 8.1, gravity does not play a major role in causing the objects to rotate the way they do. For instance, in the case of a rigid disk like a wheel, the chemical bonds holding it together keep it rigid. In the case of water swirling around a drain, the physics describing the water's motion has to do with the way molecules interact in a fluid, though sometimes gravity plays a role.

Unlike the terrestrial examples discussed at the outset, gravity plays a major role on the motions of astronomical objects, including planets orbiting stars, stars and gas moving within galaxies, and galaxies moving inside clusters. The everyday examples we discussed at the outset are familiar and serve as a good introduction to thinking about rotation and revolution. They generally involve different physics, but they give us a good reference point for understanding the motion of astronomical objects, which is dominated by gravity.

8.3.1: Gravity and Velocities When Mass is Spread Evenly in Space

When we looked at the motion of the planets in the Solar System, we modeled the mass of the Sun as a point mass. This is because the Sun's mass is entirely enclosed within the orbits of each of the planets, and also because the Sun is small in comparison to the distances between the Sun and planets.

What about systems that cannot be modeled so simply? For example, what if the Sun was not a point mass? What if its mass was evenly spread out in a sphere the size of Earth's orbit? We would still use the following equation to describe the motion of the planets, as we did above.

$$v^2 = \frac{GM}{r}$$

However, we would have to think more carefully when applying this equation. The M in the equation refers to the mass *enclosed* by the planet's orbit. In the case of Earth, Mars, Jupiter, or any of the planets farther out in the Solar System, we could plug in the total mass of the Sun for M (note that this is the case even for Earth, which is orbiting just at the edge of our spread-out Sun). The motion of Earth and outer planets would not be any different than they are now, and they would still follow the Keplerian rotation relation.

Not so for the motions of Mercury and Venus. They *would* be affected because they are closer to the center of the solar system than Earth. Some of the mass of our imagined bloated Sun would lie outside the orbits of these two inner planets. If the mass beyond the orbit of Venus is evenly distributed (as we are imagining), its gravitational effect on Venus will cancel out and have no effect (see Figure 8.7). Only the mass enclosed by Venus' orbit will affect Venus' orbital speed. Similarly, only the mass enclosed by Mercury's orbit will affect Mercury's orbital speed.

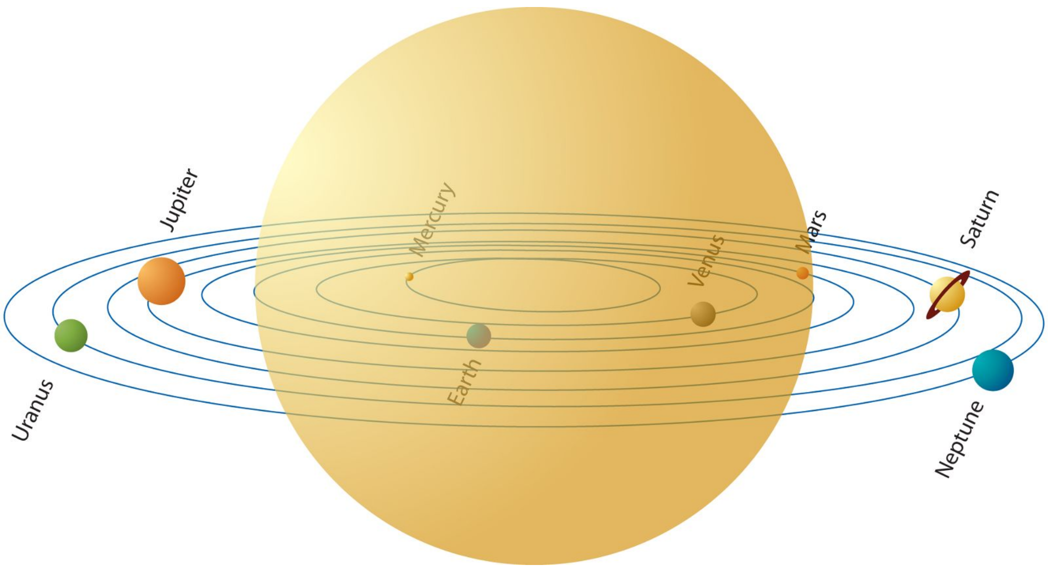


Figure 8.7: In this picture, planets orbit a Sun that is spread out in a sphere the size of Earth's orbit. Mercury's and Venus' orbital speeds are only affected by the mass enclosed within their orbits, respectively. Credit: NASA/SSU/Aurore Simonnet

Worked Example: Planetary Orbits and a Spread-out Sun

We can determine Mercury's orbital speed in the case of the spread-out Sun. We already know that we can use the equation below to solve for orbital speed.

$$v^2 = \frac{GM}{r}$$

We can plug in the value of the gravitational constant for G (6.67e-11 m^3 / kg $s^{2)}$, and the distance between Mercury and the Sun for r (5.79e10 m). But what should we plug in for M? We will need to calculate the amount of the Sun's mass that is enclosed by Mercury's orbit.

To do this, it is helpful to note that when we say an object's mass is evenly spread out, that means it has a constant density. We must also remember that density, ρ, is equal to mass divided by volume, or

$$\rho = \frac{M}{V}$$

(Note: here we use upper-case V for volume, to distinguish from lower-case v, which we have used for velocity.)

To determine the mass of the Sun that is enclosed by Mercury's orbit, we can use the relationship between mass and density written slightly differently:

$$M = \rho V$$

If the mass of the Sun were spread all the way out to Earth, its density would be 1.4e-4 kg/m^3. To learn why, see Math Exploration 8.2.

For V, we will want to plug in the expression for the volume of the Sun that is enclosed by Mercury's orbit:

$$V=\frac{4\pi}{3}r^3_{Mercury}$$

So the full expression for the mass of the Sun that is enclosed by Mercury's orbit is this:

$$\begin{aligned} M &= \left(1.4\mathrm{e}-4\ \mathrm{kg\ m^{-3}}\right)\left(\frac{4\pi}{3}\right)\left(r^3_{Mercury}\right) \\ &= \left(1.4\mathrm{e}-4\ \mathrm{kg\ m^{-3}}\right)\left(\frac{4\pi}{3}\right)(5.79e10\ \mathrm{m})^3 \\ &= 1.15\mathrm{e}29\ \mathrm{kg} \end{aligned}$$

Now, we are ready to plug everything into the equation for orbital speed:

$$\begin{aligned} v^2 &= \frac{GM}{r} \\ &= \left[\frac{(6.67\mathrm{e}-11\ \mathrm{N\ m^2\ kg^{-2}})(1.15\mathrm{e}29\ \mathrm{kg})}{5.79\mathrm{e}10\ \mathrm{m}}\right] \\ &= 1.32\mathrm{e}8\ \mathrm{m^2/s^2} \end{aligned}$$

or

$$v=1.15\times 10^4\ \mathrm{m\ s^{-1}}$$

This is smaller than Mercury's actual orbital speed (4.79×10^4 m/s), which is what we expect. In our imaginary scenario, only a fraction of the Sun's mass is attracting Mercury inward — the fraction enclosed by Mercury's orbit. The rest, outside the orbit, has no net effect. As a result, Mercury does not have to move as fast to overcome the attraction of gravity.

Math Exploration 8.2

Questions

Now we will look at the more general case of orbital motion when mass is evenly spread out in space. How do objects orbit in this case? Again, we can model this kind of system by recognizing the relationship between mass and density. Remember that if an object's mass is spread out evenly, that means it has a constant density. We can work again with our equation for velocity:

$$v^2 = \frac{GM}{r}$$

where we can substitute $M = \rho V$:

$$v^2 = \frac{G\rho V}{r}$$

and we can plug in the volume of a sphere for *V*:

$$v^2 = \left(\frac{4\pi r^3}{3}\right)\left(\frac{G\rho}{r}\right)$$

This leads to the following expression of proportionality.

$$v \propto r$$

So, in a case where the mass is evenly spread out in space, we would expect the velocity to increase with radius.

Velocity Proportional to Radius

What does the expression $v \propto r$ look like in graphical form? To make our example a bit easier to understand, we can take our expression of proportionality and turn it into an equation:

$$v \propto r$$

becomes

$$v = constant \times r$$

In this equation, the constant is a simple number. It does not depend on radius or any other variable.

Imagine we were studying a system that had a constant density, and the velocities of the objects within the system followed an equation like the one above, with the constant set equal to five.

$$v = 5 \times r$$

8.3.2: Gravity and Velocities When Mass is Distributed in Different Ways

In this section of the chapter, we are going to look at the rotation that results, due to gravity, when mass is distributed or spread out in different ways. In Section 8.3.1 above, we looked at what would happen to the velocities of the inner planets if the mass of the Sun was evenly spread out to the radius of Earth's orbit. In that case, we used the concept of density to mathematically describe the fact that the mass of our imaginary Sun was evenly spread. Mathematically, we described this by saying that the density of our spread-out Sun was constant within the radius of Earth's orbit.

Describing the spread-out Sun's *density* as constant (or unchanging at different radii) is not the same as describing the Sun's *mass* as constant. In the case of the spread-out Sun, the constant density leads to a mass that gets larger with radius. We can do the math to see why:

$$\rho = constant$$

and remember that mass (M) is equal to density (ρ) multiplied by volume (V):

$$M = \rho V$$

For a spherical spread-out Sun, we can plug in the volume of a sphere for V, just as we did in Section 8.3.1:

$$M = \frac{4\pi r^3}{3}\rho$$

In other words, as distance (r) from the center of the spread-out Sun increases, the enclosed mass increases quite a bit— the enclosed mass increases as a factor of r^3, or $r \times r \times r$! So, at a larger radius, we are enclosing a lot more mass in this system. As we saw above, this leads to orbital speeds that increase with radius, too.

$$v \propto r$$

Now we will look at some other ways mass could be distributed in a system such as a star or a galaxy.

8.3.2.1: **CASE 1: mass of the system increases directly with radius**

$$M \propto r$$

or

$$M = constant \times r$$

How would objects orbit in this system? Would their velocities be faster at larger radii, or smaller? Go back to looking at how orbital velocities and mass are related, through gravity:

$$v^2 = \frac{GM}{r}$$

We can plug in our expression for mass:

$$v^2 = \frac{G \times constant \times r}{r}$$

Since G is a constant, we can consider G × *constant* = *another constant*. We can also cancel out r in the top and bottom of the fraction above. This leads to:

$$v = constant$$

So for this mass profile, the orbital speeds of the objects in the system are all the same!

8.3.2.2: **CASE 2: Mass of the system does not change with radius**

In other words, imagine the following:

$$M = constant$$

This equation means that at any radius, the enclosed mass is the same. Our point-mass model of the Solar System is an example of a system in which the mass does not change with radius. In that model, at the radius of any planet, the mass enclosed by the planet's orbit is the mass of the Sun.

We can make a mathematical model of the orbital speeds in this type of system, and check whether they would match what we expect in the Solar System. Again, we go back to the equation that relates the gravitational force to the centripetal force:

$$v^2 = \frac{GM}{r}$$

and plug in M = constant:

$$v^2 = \frac{G \times constant}{r}$$

or

$$v \propto \sqrt{\frac{1}{r}}$$

This is the same relationship between velocity and radius (or distance from the Sun) that we find for the Solar System.

Visualizing Mass and Velocity Vs. Distance (Radius)

Below are several graphs that represent relationships, or proportionalities, between radius (on the x-axis) and another variable. In this section of the chapter, we have discussed how mass can be distributed with radius. We have also discussed the resulting relationships (due to gravity) between velocity and radius.

https://phys.libretexts.org/@go/page/31260

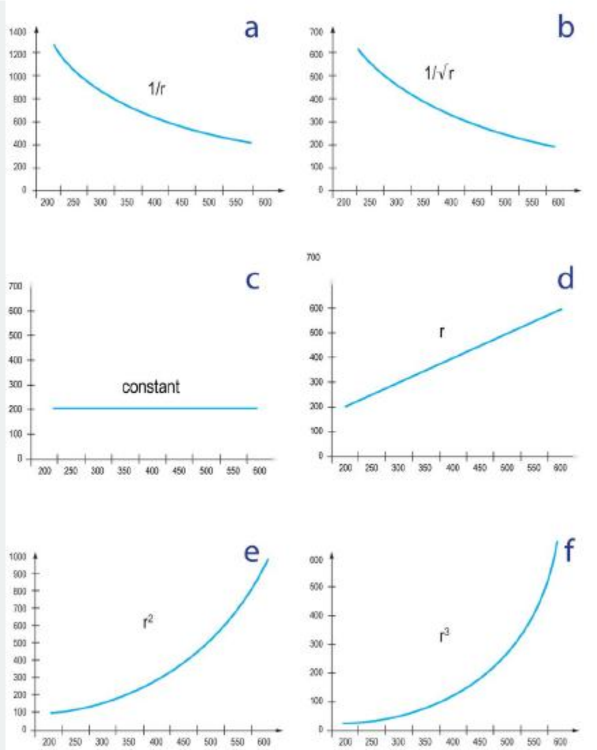


Figure A.8.3: Graphs of proportionalities with distance. Each of these graphs shows distance (radius) on the x-axis, and another variable on the y-axis. Compare these graphs to the relations we have studied in this chapter. Credit: NASA/SSU/Aurore Simonnet

Worked Example:

In Section 8.3.1 and the beginning of 8.3.2, we imagined what would happen if the density of the Sun were constant out to the radius of the Earth's orbit. Within that radius, we showed that a constant density leads to mass increasing with radius according to the following expression of proportionality:

$$M \propto r^3$$

If M is the value plotted on the y-axis, which of the graphs in Figure A.8.3 shows this proportionality?

Answer: graph f

Questions

https://phys.libretexts.org/@go/page/31260

This page titled 8.3: Gravity and Models for Different Mass Distributions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.4: Velocity and Mass Distributions in Galaxies

What Do You Think: Velocities and Rotation

Up to this point, we have explored how velocities of rotating objects can (or not) vary with radius, or distance from the center of the system. We have seen how mass can (or not) vary with radius, too. We have explored ways to model mass and velocity distributions using equations, expressions of proportionality, and graphs. We have also seen how mass distributions and velocity distributions are related according to the effects of gravity. We looked in detail at how mass can be modeled in the Solar System, and how the Solar System rotates (or in other words, how the planets orbit the Sun) due to gravity.

But what does all of this have to do with dark matter? In this section, the connection to dark matter will start to become clear. Here, we will discuss how objects orbit within galaxies and what that tells us about galaxy masses.

8.4.1: Observing Rotation Curves of Disk Galaxies

To begin, we will consider spiral galaxies. Spiral galaxies (like the Milky Way Galaxy) are large systems that typically have three distinct components (see Figure 8.8): The first is the flat disk that contains stars, gas, and dust, and that is most prominent when we look at these galaxies in visible light. You will not be surprised to learn that this is called the **disk component**. The second component is a sphere-shaped collection of stars near the galaxy center. It is called the **bulge**, or sometimes the central bulge. Finally, there is a much larger sphere-shaped collection of stars and star clusters extending out at least as far as the disk, called the **halo**. The halo is much larger in extent than the bulge, but there are not very many stars in it compared to the number of stars in the bulge or disk. The number of stars is so low, in fact, that galaxy halos are nearly invisible. We tend to look right through them.

All of this material is held together by the mutual gravitational attraction of all the material within the galaxy.

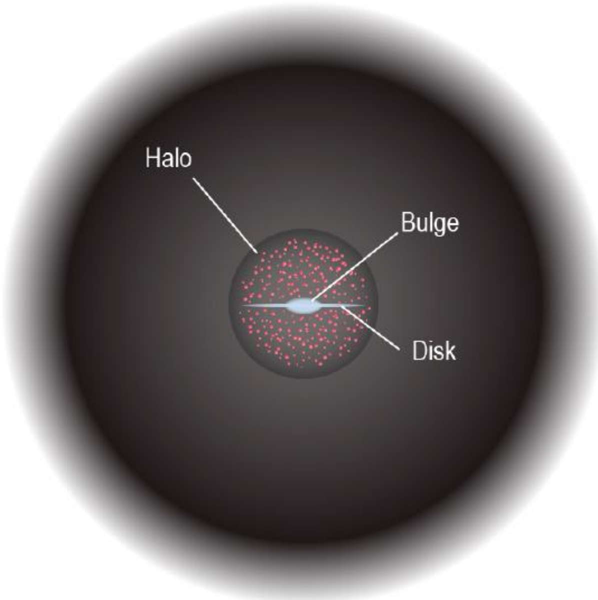


Figure 8.8: Illustration of a spiral, or “disk”, galaxy. Note the bulge, disk, and halo of the galaxy. Credit: NASA/SSU/Aurore Simonnet

How do we measure the orbital speeds of the stars and gas within a spiral galaxy? First note that because the bulges and disks of spiral galaxies are so much brighter than their halos, we mainly concentrate on measuring the orbital speeds of the stars and gas in those components of spiral galaxies.

The stars and gas in the disks of spiral galaxies tend to rotate around the center of the galaxy. If the galaxy is “face-on” from our vantage point, we will not be able to observe its rotation. On the other hand, if the galaxy is “edge-on,” we can measure the rotation through redshift and blueshift. For an edge-on, rotating spiral galaxy, we will observe redshift on the side of the galaxy that is rotating away from us (the light will appear to be “stretched out,” and will appear to have longer, redder wavelengths). We will observe blueshift on the side of the galaxy that is rotating toward us (the light will appear “scrunched” and will appear to have shorter, bluer wavelengths). See Figures 8.9 and 8.10 for more on how we observe rotation in galaxies.

Figure 8.9: (top) We cannot measure rotation in spiral galaxies that are seen face-on. (bottom) We can measure rotation in spiral galaxies that are seen edge-on. One side of the galaxy will appear redshifted, and the other side will appear blueshifted. Note that the color shift in this image has been exaggerated. The real redshift and blueshift we observe in a rotating galaxy is very small and can only be measured by taking a spectrum. Credit: NASA/SSU/Aurore Simonnet

One might think astronomers would study the stars in galaxies to learn about the motion within galaxies. Stars are bright, and there are many of them throughout a galaxy. For very nearby galaxies, we can measure the motions of stars. However, gas is also spread throughout the disk of a spiral galaxy. This gas provides a bright emission spectrum at specific optical wavelengths, as well as bright emission at a particular radio wavelength (21 cm). At these wavelengths, the gas within a spiral galaxy actually emits more light than stars. Also, if we take spectra at different distances out from the center of the galaxy, we will have measurements that can tell us about the motions of the gas at different radii from the galaxy's center. We can do this by placing a slit over the galaxy that blocks out all the light except from a thin strip along the length of the galaxy. This way, we can measure motions only from the light that makes it through the slit, as shown in Figure 8.10.

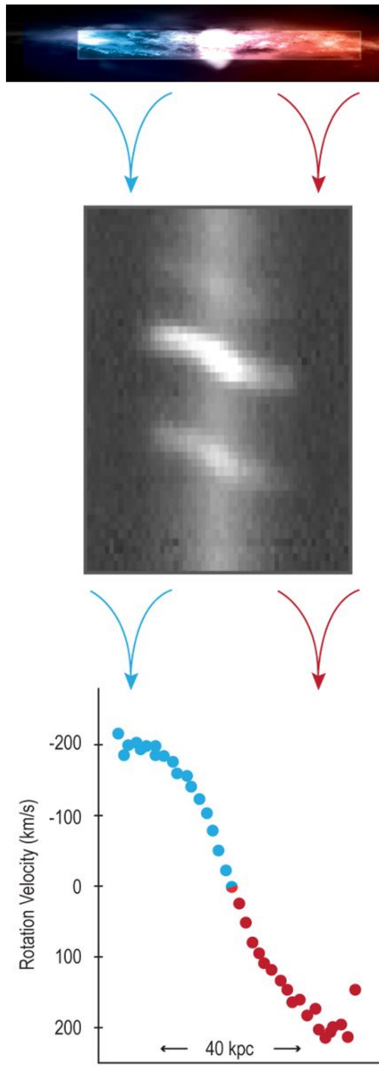


Figure 8.10: Placing the slit from a spectrograph so that it allows only light from a particular region of a galaxy is a way to limit the measurements to a region of interest. At top, an artist's illustration of a galaxy is shown, with a rectangular region highlighted. This rectangular region shows the portion of a galaxy's light that would be observed through a slit. This light would then be dispersed into different wavelengths to show the galaxy's spectrum. The middle of the diagram shows what the resulting spectrum of the galaxy would look like. The top of the spectrum shows the shorter-wavelength (bluer) light from the galaxy, and the bottom of the spectrum shows the longer-wavelength (redder) light from the galaxy. The vertical stripe down the middle is light from stars in the galaxy; the stars emit light at all wavelengths measured here. The wiggly horizontal lines are light from the gas in the galaxy. If the galaxy were not moving, these gas emission lines would be perfectly horizontal, only glowing at one wavelength. However, because the galaxy is rotating, the emission lines are blueshifted on the left and redshifted on the right, resulting in the wiggly appearance. These observed blueshifts and redshifts can be converted into a measured rotation curve for the galaxy, seen at the bottom of the figure. Credit: NASA/SSU/Aurore Simonnet

For nearby galaxies astronomers must take many individual measurements if they want to make a complete rotation curve, or velocity-distance plot. Nearby galaxies are fairly large in the sky, often larger than the field of view of a telescope. As a result of their large apparent size it is not possible to view them in their entirety with a single exposure. Instead, separate observations must be made across the galaxy, each at a different distance from its center. This is a time consuming process.

Perhaps paradoxically, motions are easier to measure in distant galaxies. When a galaxy is distant enough to lie completely within the telescope/camera field of view, then a spectrograph slit can be laid down to coincide with the long axis of the galaxy, allowing a complete map of velocities for the entire length of the system to be collected at once.

Whichever method astronomers use, the end result is the sort of spectrum shown in Figure 8.10 above. It shows the rotation velocity of a disk galaxy versus distance from the center of the galaxy. Such plots are rotation curves, just like the rotation curves you graphed and studied earlier in the chapter.

Rotation Curve of a Spiral Galaxy

In this activity, you will obtain a rotation curve for a spiral galaxy by moving a slider over the image of the galaxy.

To move the slider, click and drag the line under the red "button." Move the slider to each location where you would like to make a measurement, then click the red button. (Alternatively you can drag the slider with the button, and when you release the button, it will trigger a measurement.)

Choose at least five points on each side of the galaxy's center. At each point, the calculations are done to find the rotation speed of the gas in the galaxy. This is done by comparing the measured value of the H-alpha emission line to its laboratory value and applying the Doppler relation to the shift.

Play Simulation

Rotation Curve Matching Activity

Match the images of the galaxies on the left with the rotation curves shown on the right.

For each galaxy image, the relative velocities at three different locations are indicated by the length of the arrows.

Click and drag each image to the square next to the rotation curve that provides the best match. When you have made a correct match, you will see a green check mark in the upper right corner of the galaxy image.

Play Activity

8.4.2: Missing Mass (Dark Matter)

From your work in the last activity you may have noticed that rotation curves for spiral galaxies have two parts. In the first part, close to the center of the galaxy, the rotation velocity increases directly with radius, \(*v* \propto *r*\). This is the rotation curve of a rigid disk (see Section 8.1.1). However, galaxies are not rigid, they are collections of separate objects, and their rotation is determined by how those objects (and their masses) are distributed. The distribution of mass that goes with this curve is the one in which mass is evenly spread out; refer to the models from Section 8.3,.

However, the rotation curve does not follow this model very far. It soon "turns over," or stops increasing, and the velocity is then constant out to the edge of the visible galaxy. In other words, in the outer regions of the galaxy disk, v = constant. Again following from our work in Section 8.3, we know that a constant velocity indicates that in this outer region of the galaxy, the galaxy's mass must be increasing directly with radius. If we double the distance from the center of the galaxy, the mass enclosed within the radius also doubles. This is a surprise.

Why? If you look at a typical spiral galaxy, you will see that it has a bright center, and that its brightness gradually decreases toward its outer regions. This trend of brightness with radius is true for all spiral galaxies. The rapid decrease of the brightness in spiral galaxies creates a puzzle. If we assume that most of the mass of a galaxy is made up of stars, we would assume that most of the galaxy's mass is distributed toward the center of the galaxy, in the parts of the galaxy where those stars are more numerous. At very large radii (large distances from the center of the galaxy), the mass of the galaxy should still increase because the orbit would enclose more stars. However, as we got farther away from the center of the spiral galaxy, the radius would not enclose very many more stars, and the increase in mass (at these large radii) should be small.

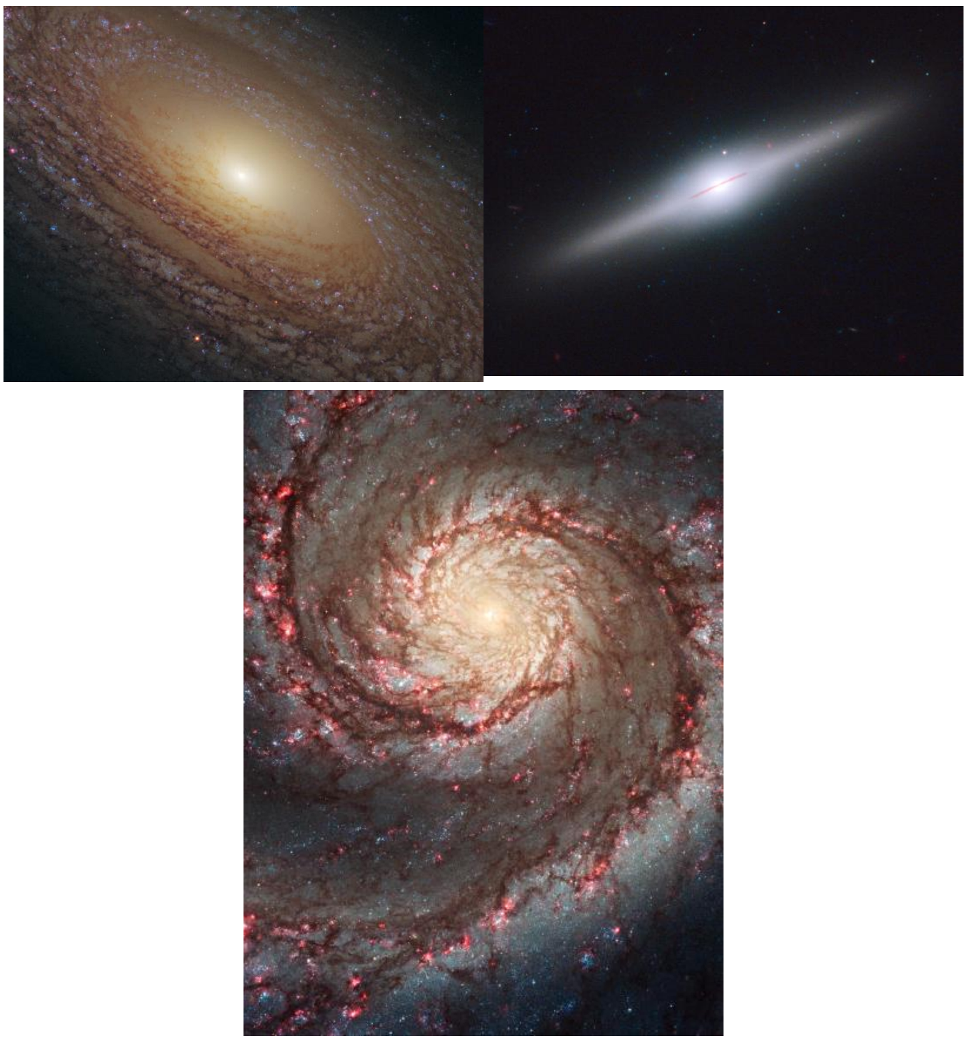


Figure 8.11: Notice that each of the spiral galaxies shown here are brighter toward their centers, and fainter toward their outer regions. If stars make up most of the mass of a spiral galaxy, we would expect the galaxies' masses to increase much more slowly with radius than they actually do. However, spiral galaxy rotation curves indicate that their masses increase directly with radius. Credits: (a) NGC 2841: NASA, ESA, and the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration; (b) ESO243-49: NASA, ESA, and S. Farrell (Sydney Institute for Astronomy, University of Sydney); (c) M51: NASA, ESA, S. Beckwith (STScI), and the Hubble Heritage Team (STScI/AURA)

Instead, the rotation of the galaxies implies that the mass is increasing rapidly with radius. There do not seem to be enough stars present in the outer regions of spiral galaxies to cause the rotation velocities we measure, not on their own. If we take into account only the masses of the stars we see in spiral galaxies, we expect the galaxies to rotate more slowly in their outer regions than what we measure.

An additional check on the rotation is provided by the gas in the galaxies. The gas disk of a galaxy extends out farther than the stellar disk, so it allows astronomers to measure the mass distribution of a galaxy to larger radii than stars alone do. In those far, outer regions of spiral galaxies we *still* see that the rotation speed of the gas is constant. This means that there is still additional mass, increasing linearly with distance from the galaxy center, even in the outer parts of the galaxy where we see no stars at all. This result has puzzled astronomers for a long time.

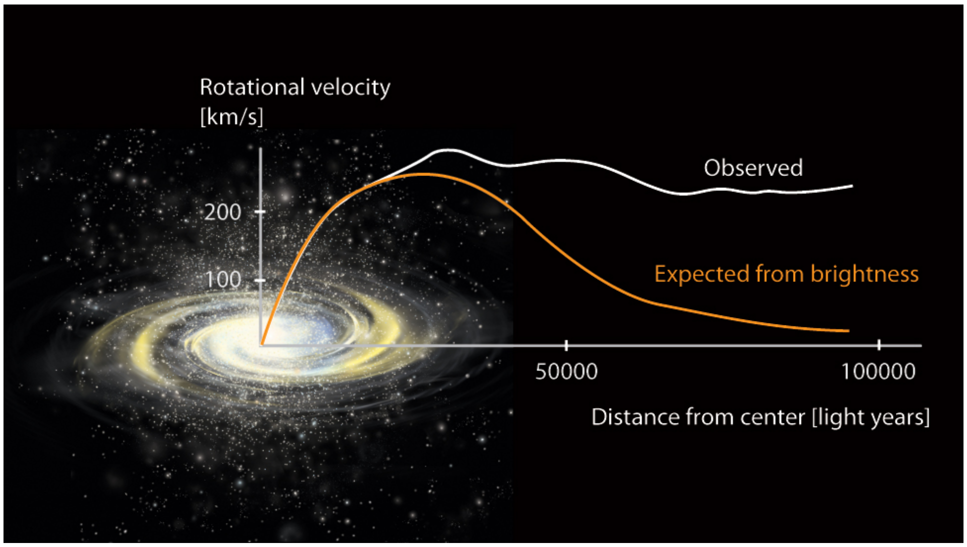


Figure 8.12: This figure shows an artist's rendition of a galaxy, demonstrating that a typical spiral galaxy is much brighter at its center than in its outer regions. A graph is shown on top of the image. In this graph, an orange curve shows the expected rotation curve of the galaxy, as calculated from the galaxy's brightness. In other words, this is the rotation curve we would expect if the galaxy's mass was mostly made up of stars. The white curve shows the observed rotation curve of the galaxy, from directly measuring the motion of the stars and gas in the galaxy. The two curves are very different! Credit: NASA/SSU/Aurore Simonnet

The shapes of galaxy rotation curves seem to be at odds with the distribution of light in the galaxies. The motions of gas and stars suggest that more mass is present than we see in the form of stars or even gas. The discrepancy is about a factor of two in the inner parts of galaxies where the main disk of stars is located. It grows to as much as a factor of 10 in the outer parts where we have only the gas disk to use as a probe.

The rotation curve anomaly was most thoroughly studied by Vera Rubin (b. 1928) and her colleagues in the 1970s and early 1980s. The inconsistency between the visible mass and the mass determined from the motion of stars and gas was originally dubbed the "missing mass" problem. Many explanations have been proposed for what this "missing mass" might be. We will discuss them later in the chapter. For now, we will note that dark matter refers to the matter in a galaxy or other system that does not emit (or absorb) any light, but which we can detect due to its mass (and gravitational effects), and which produces these puzzling rotation curves.

Figure 8.13: Vera Rubin (1928-2016), shown in 1970 measuring spectra. Dr. Rubin conducted the most thorough studies of galaxy rotation curves in the 1970s and early 1980s. Her work strongly suggested that there is much more to galaxies than we can see, and that most of their mass is invisible. Credit: Courtesy of Dr. Rubin

Worked Examples: How Much Dark Matter is there in the Milky Way?

Observations of the rotation speed of gas in the outer regions of the Milky Way Galaxy indicate that the gas is orbiting the center of the Galaxy with a velocity of 220 km/s. Assume the gas is at a distance of 15 kpc from the center of the Galaxy.

1. How many meters from the center of the Galaxy is the gas? Convert from kpc to meters.

- Use the conversion factor: 1 kpc = 3.09e19 m
- Then: r = 15 ~~kpc~~ × 3.09e19 m / 1 ~~kpc~~ = 4.63e20 m

2. What is the total mass of the Galaxy? Use the equation $v^2 = GM/r$ to calculate the total mass. Remember to convert the velocity of the gas to m/s.

- Given: v = 220 km/s = 2.2e5 m/s, r = 4.63e20 m, and we know that G = 6.67e-11 $m^3/kg\ s^2$
- Find: M
- Concept: $v^2 = GM/r$
- Use algebra to solve equation for M: $v^2 = GM/r$ becomes $v^2r = GM$
- Continue to use algebra: $v^2r = GM$ becomes $v^2r/G = M$
- Now plug in the given numbers for v, r, and G to solve for M:
- M = $(2.2e5\ m/s)^2\ (4.63e20\ m) / (6.67e\text{-}11\ m^3/kg\ s^2) = 3.36e41$ kg

Questions

In the above activity, you found that only a little more than half of the Milky Way's mass is made up of stars. That leaves a lot of mass unaccounted for! We must also recognize that the calculation above is a very rough approximation. For example, we made the assumption that every star in the Milky Way has the same mass as the Sun; however, most stars in the Galaxy have smaller masses than the Sun, and some have much larger masses than the Sun. Furthermore, dust within the Galaxy blocks out some of the light of the stars. Scientists who have made much more detailed calculations of the total mass of the Milky Way and its mass in stars have found that only about 10% of the Milky Way's mass is made up of stars and other matter that emits light. That means that roughly 90% of the Milky Way's mass might be dark matter!

Pay attention to the above example, because later in the chapter we will calculate the fraction of dark matter in another galaxy and even in a galaxy cluster.

8.4.3: Stellar Motions in Elliptical Galaxies

So far, we have looked at the rotation of spiral disk galaxies. These galaxies are fairly common in the Universe. The Milky Way Galaxy is a spiral, so we are motivated to study other spirals to learn more about galaxies like our own. However, we can also learn a great deal (and find still more evidence for dark matter) by studying other types of galaxies. In this Section, we will take a closer look at **elliptical galaxies** (see Figure 8.14).

Figure 8.14: Elliptical galaxies have “puffed” up elliptical shapes, and contain mostly old stars and little gas. Credits: (a) NGC 4150: NASA, ESA, R.M. Crockett (University of Oxford, U.K.), S. Kaviraj (Imperial College London and University of Oxford, U.K.), J. Silk (University of Oxford), M. Mutchler (Space Telescope Science Institute, Baltimore), R. O'Connell (University of Virginia, Charlottesville), and the WFC3 Scientific Oversight Committee; (b) M32: Tod R. Lauer/NASA; (c) Dominant Galaxy in Abell 2261: NASA, ESA, M. Postman (STScI), T. Lauer (NOAO), and the CLASH team

Remember that elliptical galaxies contain mostly older stars, and little gas, so we must measure the motions of stars within an elliptical galaxy to determine the stars’ velocities and, in turn, the mass of the galaxy. Elliptical galaxies are also typically found in groups and clusters, indicating that they form when smaller galaxies, such as spirals, collide or strongly interact with each other due to gravity, as shown in Animated Figure 8.15.

Play Animation

Animated Figure 8.15 Galaxy mergers. This simulation shows how an elliptical galaxy is formed when two spiral galaxies collide due to their gravitational pull on each other. Astronomers predict that the Milky Way will collide with the Andromeda galaxy (which is currently 2.5 million light-years away) in about 4 billion years.Credit: Visualization by: NASA, ESA, and F. Summers (STScI) Simulation by: NASA, ESA, G. Besla (Columbia University), and R. van der Marel (STScI)

In part due to the way that elliptical galaxies form, the motions of the stars in ellipticals are more complicated than the motions of stars and gas in spirals. Spiral galaxies rotate, but elliptical galaxies do not. Instead, their stars follow highly elliptical orbits with different orientations. This gives the galaxies their ellipsoidal shapes (like puffed-up grapes), and also their name. While spiral galaxies are supported against gravitational collapse by rotation, elliptical galaxies are supported against collapsing by the fairly random motions of their stars. Elliptical galaxies behave sort of like a gas, which is made out of molecules, except that the stars in elliptical galaxies do not ever collide with each other the way that molecules in a gas do. (Remember that there is a lot of space between stars!)

Animated Figure 8.16 Galaxy cluster interactions. This simulation shows how multiple galaxies interact due to gravity in a galaxy cluster. At the center, a massive elliptical galaxy develops. This movie helps to illustrate how the stars in an elliptical galaxy do not rotate, but instead orbit the galaxy in all different directions. Credit: Copyright © 2008 John Dubinski. Used with permission.

Stars in elliptical galaxies exchange energy through gravitational interactions rather than through direct collisions. Since higher mass galaxies have larger gravitational forces acting, their stars must be traveling at higher speeds on average to resist collapsing in toward the galaxy's center. The speed of stars in an elliptical galaxy is related to the mass of the galaxy by the following equation:

$$k\langle v^2\rangle = \frac{GM}{r}$$

or, solving for mass,

$$M = k\frac{\langle v^2\rangle r}{G}$$

where $\langle v^2\rangle$ denotes the mean-squared velocity, or the square of the spread in the velocities (the velocity dispersion) of the stars in the galaxy. The expression is nearly the same we had for rotating galaxies, aside from an extra constant, k. The value of the constant k depends on how elongated the elliptical galaxy is.

Going Further 8.2: Mean-squared Velocity and Root-mean-squared Velocity

When measuring physical quantities accurately, we need to repeat the measuring procedure many times, and use statistics to help us understand the accuracy of our measurements. In the example above, we measured the velocities of many stars in an elliptical galaxy. This is similar to making repeated measurements, except that we get the information from all the stars at once, in the elliptical galaxy's spectrum. We can use statistics to make sense of our measurements from all the stars in the galaxy, just as we would use statistics to make sense of repeated measurements in an experiment.

We will label each individual star's velocity with a subscript i, where $i = 1, 2 \ldots . n$, for n total measurements. Each velocity measurement is therefore denoted by v_i. The mean velocity, $\langle v\rangle$, is found by adding up all the measured values of velocity, and dividing by n:

$$\langle v\rangle = (v_1 + v_2 + v_3 + \ldots + v_n)/n \tag{8.4.1}$$

In an elliptical galaxy, some of the stars' velocity measurements will be positive and others will be negative, because some of the stars are moving away from us and others are moving towards us. Therefore, we might expect the average velocity to be zero, which does not tell us much. What we really want to measure is the magnitude, or size, of the stars' velocities, not the direction in which the stars are moving. The magnitude of the velocity is what we usually call the speed.

So, in order to find the speed, a more useful thing to calculate is the mean-square velocity. By squaring the velocity before we average it, the negative numbers will become positive, so we do not lose the information about the magnitude of the stellar motions when we add up the measurements. The mean-square velocity, $\langle v^2 \rangle$, is given by:

$$\langle v^2 \rangle = \frac{v_1^2 + v_2^2 + v_3^2 + \ldots + v_n^2}{n} \tag{8.4.2}$$

Then to find the average speed of the stars, we can take the square root of this quantity, called the root-mean-square (RMS):

$$\sqrt{\langle v^2 \rangle} = \sqrt{\frac{v_1^2 + v_2^2 + v_3^2 + \ldots + v_n^2}{n}} \tag{8.4.3}$$

In this example, the RMS value of the velocity provides us with the average speed of the stars in the galaxy.

Astronomers measure the velocity dispersion by taking a spectrum of an elliptical galaxy. Unlike the case for spiral galaxies, where they usually rely on the emission signature from the gas in the galaxy to measure rotation, for elliptical galaxies they look for a signature of motion from the stars. They have to do this because ellipticals typically contain little to no gas. Stars have absorption spectra. Such spectra display dark lines where light is absorbed by the gas in the stars' atmospheres. See Chapter 2, Sections 2.6 and 2.7 for a discussion of absorption spectra.

As we observe the random motions of the stars in an elliptical galaxy, some of the stars will be moving toward us relative to the rest of the galaxy. They will have blue-shifted absorption lines relative to the average for the galaxy. Other stars will be moving away from us, so their absorption lines will be red-shifted. Since all of these stars will be seen through the slit of the spectrograph at the same time, all the spectral features will be overlapping. The overlapping absorption lines are seen as broadened lines, and the width of the lines are related to the mean-square stellar speeds as shown in Figure 8.17. The interactive activity that follows illustrates how this works.

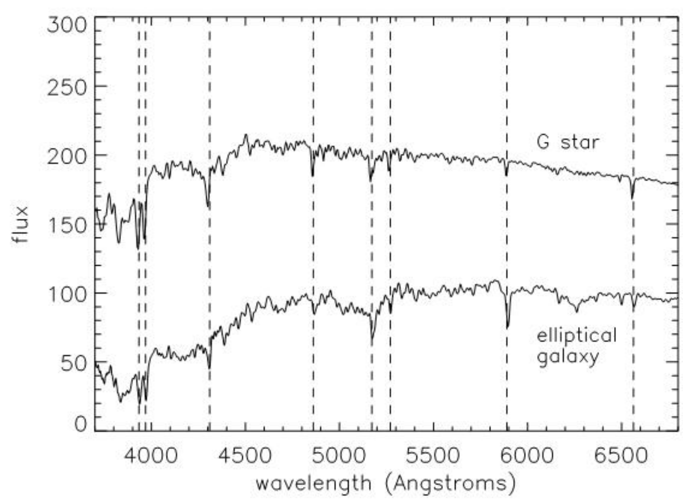


Figure 8.17: This figure shows the spectrum of a single star, above, and the spectrum of an elliptical galaxy, below. Dotted lines highlight absorption lines that are seen in both the star and galaxy spectra. The elliptical spectrum is the sum of the spectra from many stars within the galaxy. Absorption lines in the galaxy spectrum tend to be wider than those in the stellar spectrum, because the absorption in the galaxy is coming from many different stars, some of which are redshifted and some of which are blueshifted due to their movement within the galaxy. The width of an absorption line in the elliptical galaxy's spectrum can be converted to a mean-square stellar velocity measurement. Credit: Gavazzi et al. 2004, A&A, 417, 499 (for spectrum of elliptical galaxy NGC 3379); David Silva (for G star spectrum)

Measuring the Velocity Dispersion of an Elliptical Galaxy

In this activity, you will use an adjustable Gaussian fitting tool to determine the width of an absorption line in the spectra of three different elliptical galaxies.

Play Activity

First, click on one of the galaxy images at the top. The spectrum of the galaxy will appear. Notice that the spectrum has an absorption line, or a big decrease in brightness at certain wavelengths.

In order to see the Gaussian line fitting tool to measure the width of the absorption line, click on the box that says “Show Gaussian” above the spectrum.

The tool has a horizontal adjustment and a vertical adjustment. The horizontal adjustment has two large white dots: the dot on the left can be used to move the tool away from its initial position on the y axis. The dot on the right can be used to change the width of the Gaussian shape. The smaller dot on the vertical axis of the tool can be used to change the depth of the Gaussian shape.

Carefully fit the shape of the absorption line using the tool. The Gaussian fitter will read out the following information: the central (mean) wavelength of your Gaussian fit, and the width (sigma, related to a mean-square measurement) of your Gaussian fit in wavelength and in velocity. Once you are satisfied that you have fit the shape well, read the number under “sigma (velocity)”. This is the velocity dispersion of your galaxy. Record the number in the table below. Repeat for each galaxy.

How Much Dark Matter is There in an Elliptical Galaxy?

Observations of the motions of stars in an elliptical galaxy have provided a velocity dispersion measurement of 200 km/s. Assume the measurements are made within a radius of 20 kpc from the center of the galaxy. You may find the worked example at the end of Section 8.4.2 helpful as you do this calculation.

Questions

https://phys.libretexts.org/@go/page/31261

In the above activity, we see that, just as in the case of the Milky Way, elliptical galaxies contain a significant amount of dark matter. More detailed estimates of the amount of dark matter in ellipticals show that up to 95% of elliptical galaxies' masses may be made up of dark matter. By studying stellar motions in elliptical galaxies, astronomers are able to trace the way their mass is distributed, just as they do in disk galaxies. The results they get are the same as for spirals: most of the mass in elliptical galaxies is invisible, or dark matter. The presence of this unseen matter, as revealed by the motions of matter that is seen, seems to be the rule for galaxies.

This page titled 8.4: Velocity and Mass Distributions in Galaxies is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.5: Velocity and Mass Distributions in Galaxy Clusters

What Do You Think: Galaxy Clusters

8.5.1: Observed Motions of Galaxies in Clusters

The motions of stars and gas in galaxies were not the first evidence that galaxies contain enormous amounts of unseen material, or dark matter. They merely confirmed a mostly forgotten result from decades earlier. The first evidence that dark matter dominated over the visible kind dates back to the 1930s, to studies of the Coma galaxy cluster. Fritz Zwicky (1898 - 1974), an astronomer at Caltech (Figure 8.18), had discovered that galaxies tend to cluster together. His discovery contrasted with the beliefs of most astronomers of the time, who thought that galaxies were evenly sprinkled through space. Of course, this belief was not based upon any evidence one way or another. Until Zwicky, no one had really bothered to look carefully.

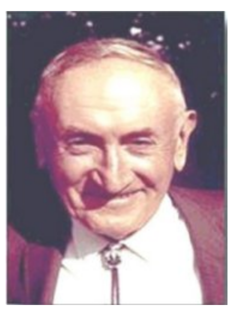


Figure 8.18: Fritz Zwicky was a Swiss astronomer who worked most of his life at CalTech. He was the first to realize that there must be some “missing mass” in clusters of galaxies that was needed to keep the galaxies from escaping from the cluster. Credit: Courtesy Fritz Zwicky.

To follow up his discovery, Zwicky investigated the galaxies in a large galaxy cluster in the Coma Berenices constellation. The galaxy cluster is called the Coma cluster, after the constellation; Figure 8.19. Zwicky measured the amount of starlight in the galaxies (an indication of the number of stars they contain) as well as the total amount of mass in the cluster. This he determined from motions of the galaxies through the cluster. What he found astonished him.

Figure 8.19: The Hubble Space Telescope's Advanced Camera for Surveys viewed a large portion of the Coma cluster, spanning several million light-years across. The entire cluster contains thousands of galaxies in a spherical shape more than 20 million light-years in diameter, and is over 300 million light years away. Credit: NASA, ESA, and the Hubble Heritage Team (STScI/AURA)

Galaxy clusters are a bit like elliptical galaxies, but on a much larger scale. Within elliptical galaxies, stars move in randomly oriented orbits around the center of the galaxy, moving in response to the overall gravity of the galaxy. This includes the gravity from all the other stars in the galaxy, as well as the unseen "dark" matter that makes up most of its mass.

In galaxy clusters, galaxies themselves orbit the center of the cluster in response to the overall gravity of the cluster. We can measure the motions of the galaxies in a galaxy cluster to determine the mass of the entire cluster, similar to the way we measure stellar motions within an elliptical galaxy to determine the galaxy's mass. This is what Fritz Zwicky did to measure the mass of the Coma cluster.

8.5.2: Masses of Galaxy Clusters and Further Evidence of Dark Matter

Although the galaxies in a galaxy cluster have orbits that resemble those of stars in an elliptical galaxy, astronomers use a somewhat different technique to measure the cluster galaxies' motion. In an elliptical galaxy, it is difficult to measure the motion of an individual star, so astronomers look for the signature of the motions of many stars all at once. In a galaxy cluster, however, it is possible to measure the motions of individual galaxies within the cluster. Astronomers can then combine the measurements of the motions of several galaxies in a cluster to get a velocity dispersion measurement of the galaxy cluster as a whole. They then use the velocity dispersion measurement to determine the mass of the galaxy cluster.

To measure the motions of individual galaxies in galaxy clusters, astronomers measure their redshifts, which should not be a surprise. The measured redshift gives them a sense for how fast the galaxies are receding from us. Galaxy clusters as a whole move away from us due to the expansion of the Universe, and because of this we see that all the cluster galaxies have redshifted spectral features. Within a galaxy cluster, a galaxy might be on an orbit that causes its velocity to be slightly larger than that of the cluster as a whole, or on an orbit that causes its velocity to be slightly smaller than that of the larger cluster (see Figure 8.20).

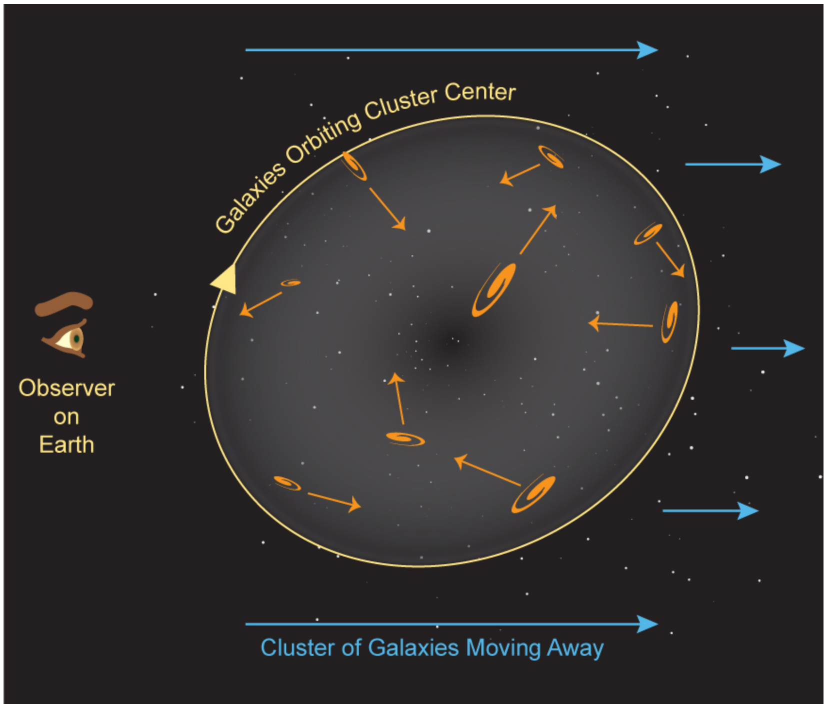


Figure 8.20: This galaxy cluster, as a whole, is moving away from us due to the expansion of the Universe. Within the cluster, some galaxies have even larger velocities away from us due to their orbits within the cluster, and some have slightly smaller velocities away from us. Credit: NASA/SSU/Aurore Simonnet

Remember that when we discussed the motions of stars within elliptical galaxies in Section 8.4.3, we noted that if the stars' velocities were too large, the galaxy's gravity could not hold onto them. In that case they would fly away. Similarly, if a galaxy in a galaxy cluster is moving too fast, the cluster's gravity will not be able to hold onto it, and it would fly out of the cluster. So we know that a galaxy cluster must have enough mass to hold onto any of the galaxies that we observe within it. If it did not, any fast-moving galaxy would most likely have flown away long ago.

Zwicky knew this too, and he used the same expression to relate the mass of galaxy clusters to the galaxy motions as we used for stars in elliptical galaxies in Section 8.4.3:

$$M = k\frac{\langle v^2\rangle r}{G}$$

Remember that in the equation above, the term in brackets is the mean square of the velocity dispersion. In this case, the velocity dispersion of the galaxy cluster is a measurement of the spread of different velocities of the galaxies in the cluster. The constant *k* may depend on the shape of the cluster, among other things.

Zwicky found that galaxies move much faster through a cluster than stars move within galaxies. He found that the typical velocity dispersion for the galaxies in the Coma cluster was well in excess of 1000 km/s. The high velocity dispersion implies a very large mass for the cluster. When Zwicky compared the velocity-determined mass of the Coma cluster to the mass expected from the light from the stars in the galaxies, he found that they were very different. This is much like what Vera Rubin found four decades later when she studied the rotation of spiral galaxies.

In order to hold together, the galaxy clusters required much more mass than was evident from the light of just the stars in their galaxies. Zwicky found that most of the mass of a galaxy cluster lies outside of the galaxies themselves, just as most of the mass of a galaxy is not contained in its stars. The following activity lets you perform the same sorts of measurements that Zwicky did to determine the velocity dispersion for the galaxies in a sample galaxy cluster.

Measuring the Velocity Dispersion of a Galaxy Cluster

In this activity, you will have an opportunity to measure the velocity dispersion of the galaxy cluster Abell 2029.

First, you will need to make a histogram, or bar chart, of the redshifts of the galaxies in the cluster.

https://phys.libretexts.org/@go/page/31262

- The interactive activity has a bunch of galaxy images with numbers printed on them in red. These numbers are actual redshifts of galaxies in cluster Abell 2029 (though the images are simulations).
- To make your bar chart, click on a galaxy image and drag it to the box above the correct redshift range. For example, if the galaxy you clicked on has a redshift of 0.075011, it should go in the box above the redshift range 0.074 - 0.076. Do this for all 16 galaxies pictured.
- Remember that all the galaxies in the cluster are redshifted because the whole cluster is moving away from us due to the expansion of the Universe. However, within the cluster, some galaxies are moving slightly toward us and some are moving slightly farther away, resulting in slightly different redshifts for each galaxy.

Next, you will fit a Gaussian shape to the bar chart to measure the width of the distribution of the cluster galaxies' redshifts.

- Click on the "Next" button to the lower right. You will see that your bar chart gets squeezed a bit to make the Gaussian fitting easier.
- Now click on the box next to "Show Gaussian Tool." This will bring up a Gaussian fitting tool just like the one you used for the elliptical galaxy velocity dispersion measurement, except that the shape is flipped (with the "hill" now pointed upward instead of downward).
- Here are the details of how to use the tool: The tool has a horizontal adjustment and a vertical adjustment. The horizontal adjustment has two large white dots: the dot on the left can be used to move the tool away from its initial position on the y axis. The dot on the right can be used to change the width of the Gaussian shape. The smaller dot on the vertical axis of the tool can be used to change the height of the Gaussian shape.
- Carefully fit the shape of your bar chart for the redshift distribution using the tool. The tool will display the width of your Gaussian shape (sigma) in redshift, and it will automatically convert this number to a velocity dispersion measurement.

Play Activity

The Mass of the Coma Cluster

We can do a calculation to learn more about the mass of the Coma cluster. You may find the worked example at the end of Section 8.4.2 helpful as you do the calculation below. We will use the equation from above, setting $k = 6$, to relate the cluster's velocity dispersion to its mass.

$$M = 6\frac{\langle v^2\rangle r}{G}$$

Astronomers have measured the Coma cluster’s velocity dispersion to be about 1000 km/s within a radius of about 3000 kpc.

Questions

https://phys.libretexts.org/@go/page/31262

As we can see from the numerical activity above, only a fraction of the Coma cluster's mass is comprised of galaxies. This suggests that the Coma cluster contains dark matter. Where is the dark matter located? Well, remember that there is dark matter within galaxies themselves. Galaxies like the Milky Way are composed of 90-95% dark matter. There is also evidence for dark matter between the galaxies in the Coma cluster. In fact, astronomers estimate that the Coma cluster is about 90% dark matter in total (see more on this in Going Further 8.3: What's Between the Coma Cluster's Galaxies?).

Going Further 8.3: What is Between the Coma Cluster's Galaxies?

Is dark matter the only stuff between the galaxies in the Coma Cluster? In fact, it is not. With the advent of space-based observatories sensitive to the x-ray emissions from hot gas, an additional method of measuring the mass of galaxy clusters became available. x-ray observations showed that galaxy clusters contain a tremendous amount of gas that lies outside and between the galaxies within a cluster. In fact, the mass of the hot gas is comparable to, or even in excess of, the mass of all the galaxies in the cluster combined. The gas is often referred to as the intra-cluster medium. It has a low density, but a very high temperature. As a result, it emits x-rays.

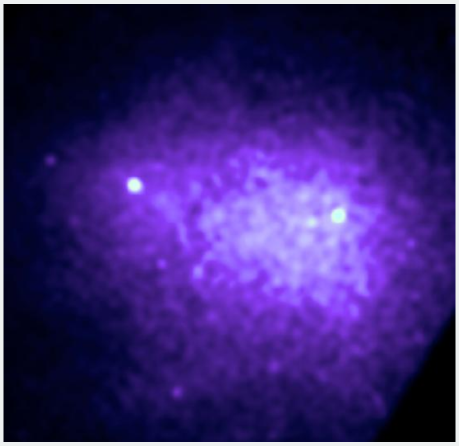


Figure B.8.2: The Coma Cluster in x-ray Wavelengths. This x-ray image from the Chandra Observatory shows the central 1.5 million light-years of the Coma Cluster. The blue haze represents a vast 100-million-degree-Celsius gas cloud which envelops the galaxies in the cluster. Compare this to Figure 8.19, which shows the Hubble Space Telescope visible light image of the cluster. Credit: NASA/CXC/SAO/A.Vikhlinin et al.

The temperature of a gas is related to the average speed of the gas particles, and the higher the temperature, the faster the particles are moving. This is because temperature is a measure of the average kinetic energy of the gas particles. The expression for kinetic energy is ½mv^2 for a particle of mass m and velocity v. So, using the temperature of the gas, we can calculate a velocity measurement for the gas. The velocity of the gas particles is determined by the gravity of the cluster—remember that if the velocity of the gas particles was too high, they would fly out of the cluster. So the cluster must be massive enough and have enough gravity to hold onto the gas. We can use the velocity measurement of the particles in the gas to determine the galaxy cluster's mass. This method is similar to using stellar motions or galaxy motions to determine mass.

Astronomers have found that the mass of the luminous, or light-emitting, matter from the galaxies in the cluster and from the gas in between the galaxies makes up only about 10% of the cluster's total mass. The Coma Cluster contains about 90% dark matter!

Is dark matter the only stuff between the galaxies in the Coma Cluster? In fact, it is not. With the advent of space-based observatories sensitive to the x-ray emissions from hot gas , an additional method of measuring the mass of galaxy clusters became available. X-ray observations showed that galaxy clusters contain a tremendous amount of gas that lies outside and between the galaxies within a cluster. In fact, the mass of the hot gas is comparable to or even exceeds the combined mass of all the galaxies in the cluster. The gas is often referred to as the intra-cluster medium. It has a low density but a very high temperature. As a result, it emits x-rays.

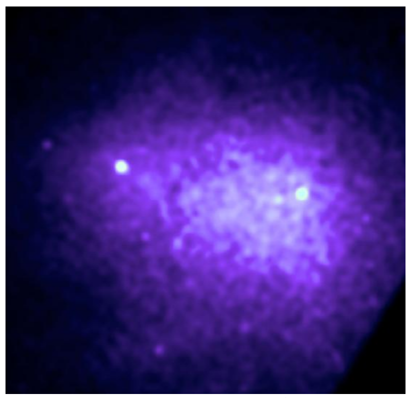


Figure B.8.2: The Coma Cluster in x-ray Wavelengths. This x-ray image from the Chandra Observatory shows the central 1.5 million light-years of the Coma Cluster. The blue haze represents a vast 100-million-degree-Celsius gas cloud which envelops the galaxies in the cluster. Compare this to Figure 8.19, which shows the Hubble Space Telescope visible light image of the cluster. Credit: NASA/CXC/SAO/A.Vikhlinin et al.

The temperature of a gas is related to the average speed of the gas particles, and the higher the temperature, the faster the particles are moving. This is because temperature is a measure of the average kinetic energy of the gas particles. The expression for kinetic energy involved both speed and mass.

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

This is the energy for a particle with mass m and velocity (or speed) v. So, using the temperature of the gas, we can calculate a velocity measurement for the gas. The velocity of the gas particles is determined by the gravity of the cluster—remember that if the velocity of the gas particles was too high, they would fly out of the cluster. So the cluster must be massive enough and have enough gravity to hold onto the gas. If that was not the case, the gas should have long ago dissipated.

So this means that we can use the velocity measurement (the temperature) of the particles in the gas to determine the galaxy cluster 's mass. This method is similar to using stellar motions or galaxy motions to determine mass. In this case, though, we are using the velocities of the particles (protons and electrons) that constitute the bulk of the hot gas.

Astronomers have found that the mass of the luminous, or light-emitting, matter from the galaxies in the cluster and from the gas in between the galaxies makes up only about 10% of the cluster's total mass. The Coma Cluster, like other galaxy clusters, contains about 90% dark matter!

This chapter has given the basic observational evidence for dark matter. The evidence is indirect, but still compelling. Additional evidence will be covered in a chapter on gravitational lenses. There is also indirect evidence for the existence of dark matter that comes from the formation of large scale structure. For the remainder of this chapter we will consider some of the possible forms the dark matter might be expected to take. We will also briefly look at alternative explanations of these observations.

This page titled 8.5: Velocity and Mass Distributions in Galaxy Clusters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.6: Possible Explanations for the Missing Mass in Galaxies and Clusters

What Do You Think: Missing Mass

If we take the observations of the motions of stars and gas in galaxies, and the motions of galaxies in galaxy clusters, at face value, then they imply the existence of additional mass that is unseen: dark matter. The unseen mass could take several forms. Some are fairly simple, others quite exotic. We will consider several possibilities and explore what evidence exists for and against them. Keep in mind that astronomers still do not know the exact nature of the dark matter. While they have made progress in ruling out some of the following possibilities, its composition remains unknown.

8.6.1: Faint Astronomical Objects

The simplest explanation for dark matter would be that it is composed of normal matter that is too faint to see. This matter could take the form of dim, low-mass but otherwise normal stars, for example. Or it could be made up of other very dim objects like planets, white dwarfs that have cooled down, neutron stars, or black holes. If there were many of these sorts of objects floating around in galaxies and galaxy clusters, then they could easily explain the amount of mass we measure. However, they would be extremely difficult, or even impossible, to detect directly with current technology.

One of the early popular ideas to explain the “missing mass,” as dark matter used to be called, was to attribute it to a population of very faint stars. This was an obvious idea for astronomers to consider. They already knew that faint stars vastly outnumber bright stars— the fact that the sky seems to be filled with bright stars is an illusion, the result of bright stars being much easier to see. With moderately large telescopes it is possible to see the brightest stars anywhere in our Galaxy (discounting that they might be behind thick dust clouds). It is even possible to see the very brightest stars in nearby external galaxies. Contrast this with the faintest stars, which are only visible out to a few dozen parsecs, even with fairly large telescopes. Because faint stars are so difficult to see, it is natural to assume that many of them have been missed. As a result, they are a natural candidate for the “dark matter.”

Unfortunately, this idea does not turn out to be viable. When scientists made careful studies looking for faint stars in our galaxy, even with the Hubble Space Telescope, they did not find them, at least not in the numbers required to explain the dark matter. When sensitive cameras using CCDs (charge-coupled devices) first became available in the early 1980s, one of the first ways astronomers used them was to look for faint red stars. Because of their higher sensitivity to light, CCD detectors were much better suited for these kinds of projects than the older photographic films in use until then. The low numbers of faint red stars found in these studies ruled out the most natural and simple explanation for dark matter - faint stars. Instead, it implied that a more exotic explanation was needed.

The next most obvious candidate for dark matter is the burned out remnants of dead stars. These include white dwarfs, neutron stars, and black holes. Because these are all the compact remains of former stars, they have been given (somewhat in jest) the name **MAssive Compact Halo Objects** or MACHOs. Some other objects that are not stellar remnants, like brown dwarfs, are also included in this category. These objects are even more difficult to detect than faint red main sequence stars. In fact, they generally cannot be seen at all from our perch here on Earth. White dwarfs and brown dwarfs are visible if they are not too distant, but neutron stars and black holes are usually invisible. The only way we see them is if they are in particular systems in which they might emit x-rays or radio signatures. Any black holes or neutron stars not in such systems are essentially undetectable. The same is true of most white dwarfs and brown dwarfs, which are too far away to be seen.

So, is there no hope of detecting MACHOs at all? We certainly cannot detect them via their own light, at least not with currently available telescopes and cameras. But there is another way: gravitational lensing. When these compact objects pass in front of a distant background of stars, occasionally one will align exactly with a background star. The background star is magnified by the gravity of the compact object, which acts like a lens (see Figure 8.21). When that happens the background star's brightness appears to increase, and we can detect that increased brightness if we monitor many stars.

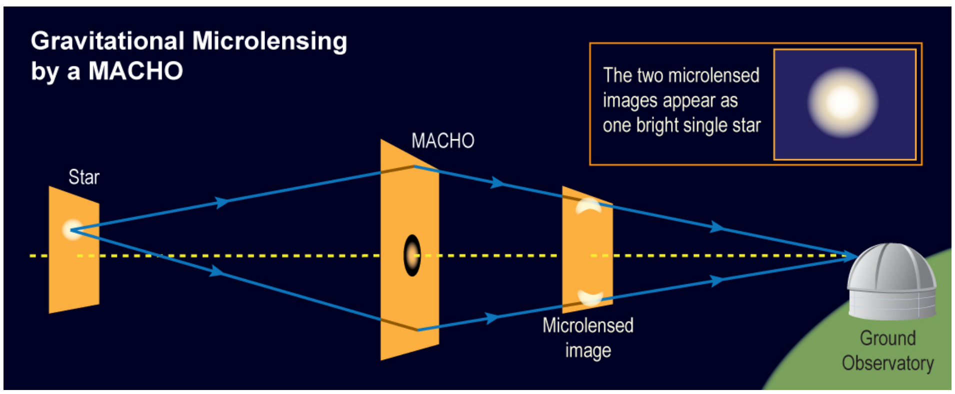


Figure 8.21: Gravitational microlensing. The intense gravitational field of the MACHO acts like a lens to smear and distort the image of the background star. The starlight is bent around the MACHO into two banana-like shapes that are larger than the original stellar image. The telescope cannot resolve the microlensed images, instead seeing a single resulting image that appears brighter than the original star. Credit: NASA/SSU/Aurore Simonnet

You might already have deduced that there are not enough of these compact objects to explain the amount of dark matter in galaxies. Many were detected, and they could comprise some of the dark matter, but only up to half of it. That still leaves a lot of unseen, and unexplained, material in the Galaxy.

✔ White Dwarfs and Brown Dwarfs as Dark Matter?

We will do a couple of calculations to determine the number of undetected white dwarfs or brown dwarfs that would be needed to make up the Milky Way's entire unseen mass, which is about 1.8×10^{42} kg.

Worked Example

How many white dwarfs would it take to account for all of the Milky Way's unseen mass? Assume that a typical white dwarf has a mass of 2×10^{30} kg.

- Given: unseen mass in Milky Way of 1.8e42 kg, mass of white dwarf is 2e30 kg
- Concept: divide the unseen mass by the mass of a white dwarf to find out how many white dwarfs it would take to make up the unseen mass:

    (1.8e42 kg)/(2e30 kg) = 9e11 white dwarfs

This would be 900 billion white dwarfs, which is not realistic given what we know about the Milky Way and its stars.

Question

8.6.2: Dark Matter Particles

Observational searches for faint stars and MACHOS have found that, for the most part, dark matter is not composed of these objects. MACHO surveys would have found non-luminous objects, and surveys for faint red stars would have found the very faint luminous ones. The only possibility that remains is that the dark matter is made of even smaller objects that cannot be easily detected, such as very small particles.

If the dark matter is composed of particles, the particles must have some special properties. First, of course, they must be massive. After all, their combined mass is causing the gravitational effects seen in galaxy rotation curves and the motions of galaxies and hot gas in galaxy clusters. Second, they cannot interact with (absorb, emit, or reflect) light. If they did we would have seen the particles already. This means the dark matter is not made up of atoms or molecules or particles of dust or gas. Nor can it be made up of common sub-atomic particles, like electrons or protons, for example. All of those things interact strongly with light of various wavelengths. Dark matter's lack of interaction with light is the property that makes dark matter dark.

The dark matter particles also must not interact strongly with other kinds of particles. Presumably, if dark matter did interact strongly with known particles, electrons for example, then we would have seen the results of these interactions in particle physics experiments. We have not. Some experiments claim results that *might* be caused by the interaction of dark matter particles within their detectors, but these results are still extremely tentative. Current data simply are not good enough to say that dark matter has absolutely been detected.

Furthermore, if the dark matter particles interacted strongly with each other, then we should have seen the production of known particles like photons, electrons, etc., that would be produced by these interactions. For example, we might expect to detect gamma rays from two massive dark matter particles annihilating each other. The process would be analogous to the annihilation of an electron and positron pair, in which we see x-ray photons with energies of 511 keV, the rest energy of an electron or positron. For the case of dark matter particles annihilating, the energy of the photons would be much larger because of the hypothetically larger mass of the particles involved. These hypothetical particles are often called WIMPs (again, in jest), for **Weakly Interacting Massive Particles**, a term coined in 1985 by physicists Gary Steigman at The Ohio State University and Michael Turner at the University of Chicago.

The Fermi Gamma-ray Space Telescope is being used to search for such gamma rays, but so far nothing conclusive has been found. This could be because the particles are too massive to be detected within the energy range of Fermi, which has a top energy sensitivity of 300 GeV. Or it could be because we have not looked through the data carefully enough, or because the signal is just too weak. Of course, we also might not be seeing a signal from dark matter annihilation because the hypothesized gamma-ray

https://phys.libretexts.org/@go/page/31263

emitting dark matter particles are not there in the first place— perhaps dark matter is made of something else entirely, something that does not annihilate.

What about neutrinos? Neutrinos are well known to interact very weakly. However, they also have extremely small masses, near zero, which makes it difficult to account for so much missing mass. For a while some scientists thought that neutrinos were a good candidate for dark matter. However, as we will see in the chapter on structure formation, they are not suitable to explain the bulk of the dark matter. So far, scientists have seen lots of MACHOs, though not enough to explain the dark matter. No one has ever seen any WIMPs, but particle physicists have several candidates for what these particles might be.

Many scientists were hoping to simply create WIMPs in collisions at the **Large Hadron Collider** (LHC) at CERN, a particle physics laboratory located at the Swiss/French border near Geneva. (CERN is also known as the European Organization for Nuclear Research, or in French, Organisation Européenne pour la Recherche Nucléaire— formerly Conseil Européenne pour la Recherche Nucléaire.) In the collider, protons and antiprotons (the anti-matter counterparts of protons) are accelerated to high energies, as high as 7 TeV (7×10^{12} eV). Since the proton and anti-proton beams travel in opposite directions, when they collide there will be 14 TeV of energy available to create new particles.

Each collision creates a tiny fireball at the collision site, and out of that fireball emerge new particles. Most of them by far are the commonplace photons, electrons, positrons, neutrinos, etc., that we are familiar with already. But scientists have been hoping that WIMPs will be produced from time to time, and that their instruments will record this. At the time of this writing this has not happened. The LHC has been operating at full energy of 14 TeV for more than half a decade. It takes up to two years for scientists to sift through the enormous amount of data produced during each season. Thus far, despite some tantalizing false alarms, nothing convincing has been seen. The search continues.

8.6.3: MOND

If you are like some scientists, you might be questioning the conclusions we have reached so far in this chapter, as well as the assumptions we used to reach them. First, we applied Newton's laws of gravity and motion to stars and gas orbiting in galaxies and galaxy clusters. Those laws make a definite prediction regarding how these objects should move under the influence of a certain amount of gravitating mass. We compared those predictions to the amount of mass we can see and realized that the two do not match. In fact, they are not even close. From this we conclude that there must be mass present that we cannot see— dark matter.

There is an alternative conclusion that we could make instead: perhaps we are seeing all the mass present, but our reliance on Newton's laws is misguided. When we apply Newton's laws to the motions within galaxies and galaxy clusters, we are assuming that the laws are valid there. But does that really have to be the case? Newton developed his laws to describe motions that he could observe on Earth and in its near vicinity. Newton's laws have since been tested within the confines of our Solar System, and they work exceedingly well. For example, we are able to send spacecraft to other planets and planetary systems, insert those spacecraft into orbit and de-orbit them if we wish, land on the surface, etc. But does that mean our physical laws will work when we go outside the Solar System? Here we explore this idea for a moment.

We learned in Chapter 6 that the acceleration experienced by objects traveling in circular orbits, called centripetal acceleration, is provided by gravity. We saw that the general expression for centripetal acceleration is given as

$$a_c = \frac{v^2}{r}$$

where v is the speed of the orbiting object and r is the orbit's radius. We can compare the values of this acceleration for several different physical systems.

The following table shows values for several planets, the Sun as it orbits in the Milky Way galaxy, and a typical galaxy in a cluster. The first column for each object gives the relevant orbital speed, the second column gives the size of its orbit, and the last column gives the resulting centripetal acceleration.

Table 8.2 Comparing Objects in the Solar System, Galaxy, and Galaxy Clusters

OBJECT	V (KM/S)	R (M)	A_C (M/S^2)
Earth	29.79	1.49×10^{11}	5.96×10^{-3}
Neptune	5.43	4.50×10^{12}	6.58×10^{-6}
Sun	220	2.50×10^{20}	1.96×10^{-10}

OBJECT	V (KM/S)	R (M)	A_C (M/S^2)
Cluster galaxy	1000	3.10×10^{23}	3.24×10^{-12}

The accelerations of Solar System objects are clearly much larger than those for stars in galaxies or for galaxies in clusters. (This statement is not true for the Oort Cloud comets, but they are not relevant currently to the arguments presented here because we cannot track them when they are not close to the Sun.) And of course, accelerations are much, much larger for falling bodies near the surface of Earth. Since the sizes of these accelerations are so different, some scientists wonder if the same laws of motion and gravitation can be used to describe the motion of all of them. If the laws of gravity are not the same everywhere, say if gravity works differently at low accelerations than it does at high accelerations, then perhaps our conclusions about the dark matter are in error.

The first papers published that follow this sort of thinking appeared in the early 1980s. They argued that, by replacing Newton's second law with a slightly modified form, the need for dark matter to explain galaxy rotation curves and the motions of galaxies in galaxy clusters could be avoided. The modification was a simple one: change Newton's second law by multiplying it by a function, which we label *f*, as follows:

$$F = ma \rightarrow F = ma \cdot f(a/a_0)$$

In this expression, the function *f* depends on the acceleration, *a*; a_0 is a parameter that we can use to help match the model to the data. The function is defined such that $f = 1$ when $a >> a_0$. Otherwise $f = a$. This modification is similar to the models we created in Section 8.1 for mass distribution. However, in this case we are not modeling how mass is distributed in a galaxy, we are modeling how the force on a particle depends on its acceleration. The value of the parameter a_0 must be about 10^{-10} m/s^2 if the model is to match galaxy rotation curves.

With this definition, when the accelerations are large compared to a_0, as they are in the Solar System, then the usual set of equations devised by Newton is valid and the motions behave in the manner we expect. However, when the accelerations are small, comparable to or less than a_0 (as they are for, e.g., stars orbiting a galaxy) then we must use a modified version of Newton's second law:

$$F = m\frac{a^2}{a_0}$$

Because we have modified Newton's second law, this procedure is referred to as MOdified Newtonian Dynamics, or MOND. If we now set this equal to the force of gravity, assumed to be in the standard form, then we have

$$m\frac{a^2}{a_0} = \frac{GMm}{r^2}$$

We can rewrite this further (after canceling the common factor of *m* from both sides) using the expression for centripetal acceleration to replace *a*:

$$\frac{\left(\frac{v^2}{r}\right)^2}{a_0} = \frac{GM}{r^2}$$

or...

$$\frac{v^4}{a_0 r^2} = \frac{GM}{r^2}$$

Now we can cancel the common factor of r^2 and rearrange to get

$$v = (GMa_0)^{1/4}$$

This is very different from what we got before. It depends only on physical constants (*G* and a_0) and the mass inside the orbit, *M*. If we assume that light traces mass, then we expect the rotation velocity to become constant at some point because the light (and under these assumptions, the mass) drops off rapidly as we move away from the center of the galaxy (as in Figure 8.22).

So under the assumptions of MOND, we find that the rotation velocity is independent of distance from the center of the galaxy, at least for stars orbiting at large distances. That is exactly what is observed for real galaxies. No amount of “hidden,” “missing,” or “dark” matter is required.

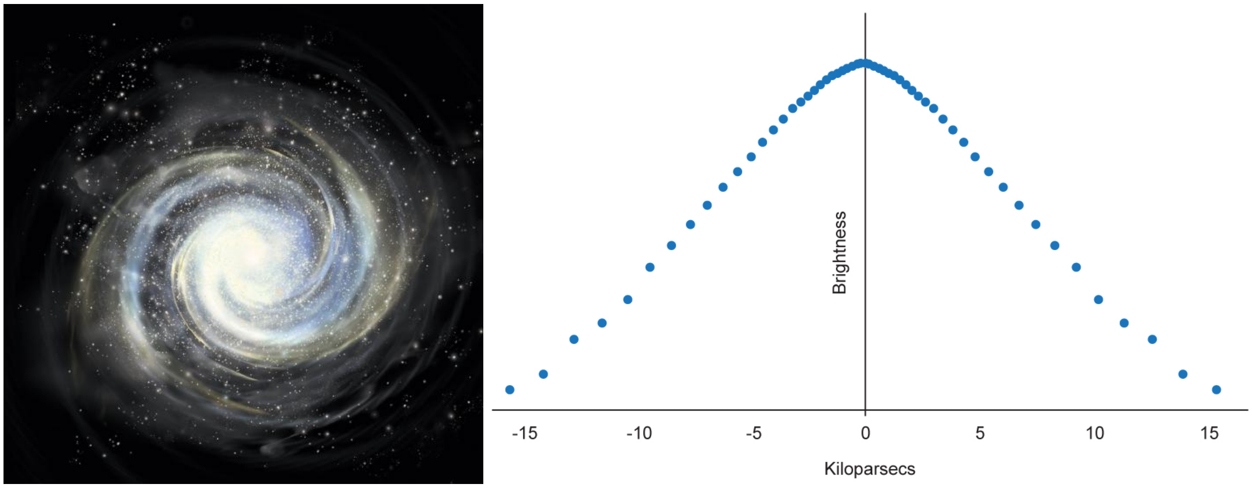


Figure 8.22: Spiral galaxy surface brightness profile. This figure shows (a) an illustration of a spiral galaxy, and (b) a graph of the galaxy’s brightness profile. Credit: NASA/SSU/Aurore Simonnet

So, is this result believable? Can we truly avoid the need for a new kind of matter by making a small modification to the laws of physics, a modification that would go completely unnoticed in any laboratory experiment we could devise? How can we tell whether this or our previous treatment, if either, is the correct one? Have we learned something new about the laws that describe motions of objects in the Universe?

These questions are not easy to answer. Yet, they are the kinds of questions that face scientists every day. How do we decide between two competing ideas when both are able to explain the same data equally well? Given only the information discussed in this chapter it is not possible to refute the ideas of MOND. We tend to have a bias against making modifications to existing physical laws, a bias born of our familiarity with the laws as we first learned them and our many successful applications of those laws, even in a regime far removed from the one that is relevant here. These successes can give us a certain amount of confidence that the laws are valid in all circumstances. But are such confidences warranted? As we will see in the next chapter, they are not, but they are not given up easily either.

The only way that scientists are able to decide among competing ideas is to collect more information about the Universe. At some point, we hope, one of the ideas will no longer be compatible with the improved view revealed by our new measurements. This has occurred with MOND over and over again, at least as we have described it here.

We shall see that other gravitational effects, from Einstein’s general relativity, not just from the Newtonian viewpoint, can be used to distinguish between the ideas of MOND and the need for dark matter. To jump ahead, the results have always been consistent with the dark matter idea, not with MOND… However, MOND continues to evolve; each time an old version is shown to be inconsistent with observations, its supporters modify its equations in such a way that the inconsistencies are removed. We will revisit these ideas in the next couple of chapters and hopefully develop some ideas of our own about whether the MOND approach is reasonable.

Throughout our explorations, you should keep in mind that Newton, when he devised his laws of motion and gravitation, was also searching for ways to express experimental results as simply as possible using mathematics. His method was exactly the one employed by the champions of MOND now.

This page titled 8.6: Possible Explanations for the Missing Mass in Galaxies and Clusters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.7: Wrapping It Up 8 - What Is the Matter With NGC 3198?

In this activity we will be studying the spiral galaxy NGC 3198 (Figure A.8.4). We will use measurements of the rotation curve of the galaxy and of the brightness of the galaxy in order to determine whether it contains dark matter, and if so, how much.

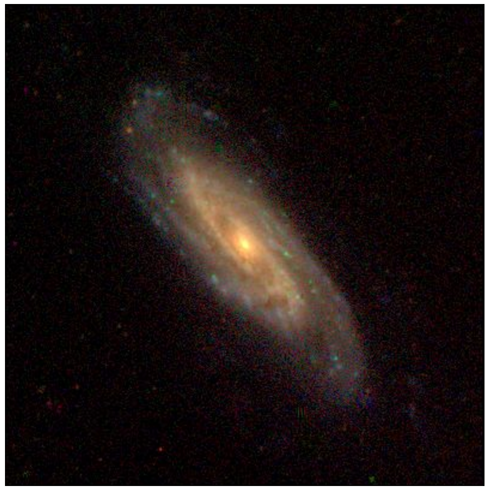


Figure A.8.4: Image of the spiral galaxy NGC 3198. Credit: Frei, Guhathakurta, Gunn, and Tyson, 1996, *Astronomical Journal*, Vol. 111, p. 174

8.7.0.1: **PART I. MODELING A ROTATING GALAXY**

In this part of the exercise, you will use a set of rotating rings to model how the stars and gas in this galaxy rotate. To do this, open up the interactive galaxy modeler:

Play Activity

In the upper-right hand corner, you will see a graph of the galaxy's rotation curve.

- The graph only shows the velocities on the side of the galaxy that is moving away from us.
- You can click on the data points in the graph to see the velocities of the stars and gas in NGC 3198 at different radii, or distances, from the galaxy's center.

Next, model the galaxy's rotation using the slider bars at the bottom right.

- Use the left-most slider bar to match the velocity of the data point closest to the center of NGC 3198.
- Use the next slider bar to match the velocity of the next data point on the graph, and so on, out to the right-most slider bar, which you will use to match the velocity of the right-most data point on the graph (farthest from the center of NGC 3198).
- Once you have matched the sliders to the velocities in the graph, click on "Update" at the bottom.
- The rotating rings on the left will now simulate the velocities you have input with the sliders.
- Congratulations! You have now modeled the rotation of NGC 3198.

8.7.1: PART II. MEASURING NGC 3198'S TOTAL GRAVITATIONAL MASS

Now you will fill out the missing data in different columns of Table A.8.1. Go back to the rotation curve of NGC 3198. Remember that you can click on the data points to read how fast the stars and gas in the galaxy are rotating at different radii from the galaxy's center.

1. Click on and read the data points in the rotation curve graph. Fill in the velocity and radius information in Table A.8.1.

8.7.2: PART III. MEASURING NGC 3198’S LUMINOUS MASS

In the first part of this activity, you measured NGC 3198’s total mass at several different radii. Now, you are going to take a closer look at the luminosity (absolute brightness) of the galaxy and think about what the brightness measured at different radii tell us about the amount of light-emitting matter (like stars, gas , and dust) that is in NGC 3198.

The luminosity graph shows the total luminosity of NGC 3198 encircled within several radii. For instance, the data point at R = 2.5 kpc shows the total luminosity of all the stars and gas and dust within a circle of radius 2.5 kpc from the center of the galaxy. The data point at R = 10 kpc shows the total luminosity of all the stars and gas and dust within a circle of radius 10 kpc from the center of the galaxy. And so on.

Now take a look at the graph of NGC 3198’s luminosity.

- You can do this by clicking on “Show/Hide Luminosity Profile” above the rotation curve.
- You should now be able to see graph of the galaxy’s rotation curve, and below that, the graph of the galaxy’s luminosity versus radius.

8.7.3: PART IV. DARK AND LUMINOUS MATTER IN NGC 3198

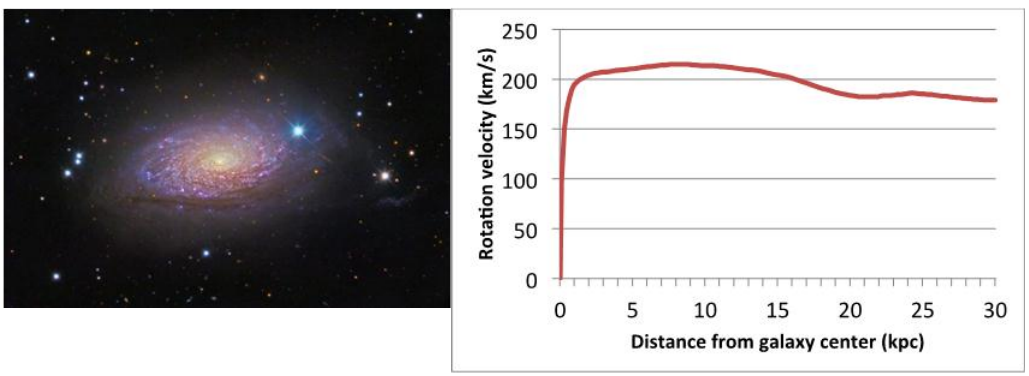
Now we can compare the amount of luminous matter in NGC 3198 (measured in Part III) to the total amount of matter in NGC 3198 (measured in Part II). This will give us insight into whether NGC 3198 has any dark matter, and if so, how much.

This page titled 8.7: Wrapping It Up 8 - What Is the Matter With NGC 3198? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

8.8: Mission Report 8 - What Is the Matter With NGC 3198?

D. Questions to be graded for accuracy. Show your work!

In this section, we will examine the rotation curve of the “Sunflower Galaxy,” also known at NGC 5055. The image of the galaxy and the galaxy’s rotation curve, are shown in the figure below.



NGC 5055, the Sunflower Galaxy, and its rotation curve. Credit: NGC 5055 image: Processed by Juan Conejero / Image Acquisition by Jim Misti and Steve Mazlin; NGC 5055 rotation curve: adapted from data published in Y. Sofue, Y. Tutui, M. Honma, A. Tomita, T. Takamiya, J. Koda and Y. Takeda, 1999, Astrophysical Journal Vol. 523, pp136-146

1. Understanding velocity distributions

This page titled 8.8: Mission Report 8 - What Is the Matter With NGC 3198? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

9: Special Relativity

Chapter 9 focuses on Special Relativity, including the effects of time dilation and length contraction, the geometry of spacetime, and the $E = mc^2$ formula. On the foundation that the speed of light and the laws of physics are the same for all observers, you will learn that moving clocks run slower and moving rulers are shorter. In addition, you will become familiar with spacetime diagrams and the relationship between mass and energy. Several practical and fanciful applications will be explored.

This page titled 9: Special Relativity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.0: Special Relativity Introduction

Although it is not possible to drive through a city at nearly the speed of light, the video you have just viewed shows a computer simulation of a relativistic trip through Tubingen, Germany (Credit: Ute Kraus and Marc Borchers). However, not all of the effects of travel near the speed of light have been shown in this video. For example, the video does not show changes in color due to the Doppler shifting of the light. In this chapter, you will be introduced to the effects that are most commonly measured when objects travel at extremely high, constant speeds: time dilation and length contraction. More exotic effects, such as the warping of the buildings shown in the video, can be explored through resources available at the Spacetime Emporium.

Since human forms of transportation, such as bicycles, automobiles, and trains, do not really travel at the extremely high speeds needed to experience relativistic effects, how do we know these effects really exist? We can observe what happens when much smaller, rapidly moving particles known as cosmic rays move through Earth's atmosphere.

Cosmic rays are energetic particles (such as protons, atomic nuclei, etc.) that originate from outer space. When high-energy cosmic rays strike Earth's upper atmosphere, they interact strongly with the atoms there. Their energy is converted into a shower of particles that then cascade down into the lower atmosphere. Most of these particles are converted to energy at high altitudes, many kilometers above Earth's surface. However, some of the particles do make it to the surface, where they can be detected. Some of the particles produced are muons, particles that are similar to but more massive than electrons.

Muons are unstable and have a 50% probability of decaying after only about 2 millionths of a second. In that time, given that the muons are traveling near the speed of light, they should be able to travel about 650 meters before they decay. Of course, only half the muons decay in this time. The other half have a 50% probability of decaying after an additional 2 microseconds (and an additional 650 m), and so on. So, we expect that a small number of the muons will travel much farther than 650 m. However, the muons are produced at altitudes of 10 km or more. Given their half-life and the altitude at which they are produced, fewer than 20 out of every million muons should reach sea level. Nonetheless, many, many more than that are detected. How can this be?

The answer to this question is the subject of this chapter. It has to do with the unexpected character exhibited by space and time when objects travel close to the speed of light.

This page titled 9.0: Special Relativity Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.1: The Principles of Special Relativity

Learning Objectives

- You will know that the speed of light is the same for all observers.
- You will know that the laws of physics are the same for all observers in reference frames moving at constant velocity with respect to each other.

What Do You Think: Properties of Light

We have already seen how several unexplained experimental results led to the revolutionary ideas of quantum physics at the end of the 19th century, overturning longstanding notions about matter and energy. At around the same time, an apparent logical inconsistency in the theory of electromagnetism spurred a similar overturning of our notions about space and time. The classical theory of electricity and magnetism led to this revolution. The theory predicts the existence of waves of electromagnetic fields—electromagnetic waves. The waves themselves are not the problem, their speed is.

In the classical theory of electricity and magnetism, the speed of electromagnetic waves (i.e., the speed of light) is the product of two physical constants. One of the constants relates to the strength of electric fields; think static electricity. The other relates to the strength of magnetic fields, like the ones that stick magnets to a refrigerator door. These constants do not relate to the state of motion of either the source of the waves or to the motion of someone observing them; they are constants. So both moving and stationary observers should measure them to have the same value. And clearly, physical constants should not change if the source of electromagnetic waves, a light bulb, say, is moving fast, slow or not at all. This implies that all electromagnetic waves have the same speed, regardless of the state of motion of their source or of a person or device measuring the waves. This prediction relates to waves traveling through a vacuum. When the waves travel through a material, their interaction with that material can cause them to slow down in ways the depend in detail on the material and the frequency of the waves in question.

The constancy of the speed of electromagnetic waves (in vacuum), regardless of the motion of the source or the observer of the waves, is extremely unusual. This property is different from the way most things move. Some examples regarding the relative velocities of objects for different observers might help to clarify what we mean.

First, imagine you are standing on the side of a railroad track. You observe a train moving past, and a woman standing on the train throwing a ball in the direction of motion of the train. The woman measures a certain speed of the ball *relative to herself*. You will measure a different speed, *relative to yourself*. You must add the velocity of the train to whatever velocity the woman measures for the ball in order to get the speed you measure.

As another example, imagine you measure the speed of a horse galloping alongside a moving train. You will measure a different speed for the horse depending on whether you are riding on the train or standing beside the tracks. In particular, if the horse moves along the ground at the same speed that the train moves along the ground, and in the same direction, then someone on the train does not perceive that the horse is moving at all. The horse has zero velocity for such an observer.

Again, keep in mind that we are only talking about relative speeds here, not speeds along the ground. You are probably used to measuring all speeds relative to the ground, but this is done for convenience, so that everyone has a common reference point. But there is nothing special about the ground, so you can just as well measure all speeds relative to yourself or any other reference.

According to the theory of electromagnetism, electromagnetic waves do not share the property described above with horses and balls. If the ball or the horse were forms of light, both the observers on the ground and on the train would measure the same speed for them: 300,000 km/s. Strange.

Of course, it is possible that the theory of electromagnetism is not an adequate explanation of nature. Perhaps the form of the equations describing electromagnetic fields depends on the state of motion of the observer or the source of the fields. In that case, the equations might be modified depending on whether the source or observer of the waves is moving or stationary. Any such modification might predict that the speed of waves is different for different observers. However, no other laws governing physics require such modification, and physicists tend to think that laws that predict how nature behaves should be the same whether one is riding along in an automobile or sitting on the side of the road.

In the 1890s, a teenage Albert Einstein imagined what it might be like to move as fast as a light wave. He understood that the laws of electricity and magnetism predicted waves, and he also understood that if you could move along with a light wave, just as fast as the light itself, then the wave would seem to freeze. It would no longer oscillate, but would remain stationary. The electric fields might make a vertical sine wave, sort of like an artistically sculpted picket fence. The magnetic field would then make a horizontal sine wave, in phase with the electric field but rotated by 90 degrees from it. This configuration of the fields is shown in Figure 9.1. The motion of the wave is aligned with the spine of the wave, perpendicular to both the electric and magnetic fields.

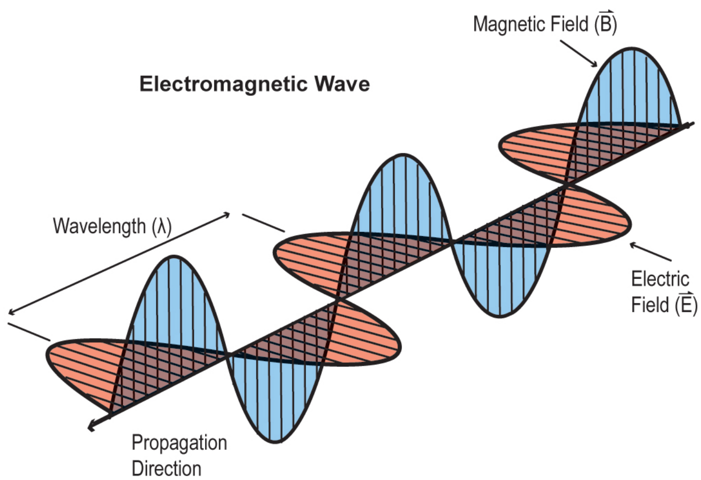


Figure 9.1: A frozen electromagnetic wave as seen by an observer moving along with the wave at the speed of light. Credit: NASA/SSU/Aurore Simonnet

Such a wave is not a wave at all, because it does not oscillate in time. Einstein knew that this idea was inconsistent with electromagnetic theory, and so he assumed that the situation he imagined was impossible. He also assumed that the laws of physics must be the same for all observers, regardless of their state of motion. He elevated the second of these two assumptions to a physical principle, called the *principle of special relativity*. Then he set about the task of seeing where they led. His work on this topic was published in 1905.

The relativity principle is often stated in the form of two separate postulates. Here they are in succinct form:

1. The laws of physics are the same for all observers who move at constant speed and in a constant direction, without any acceleration.
2. The speed of light is constant, regardless of the state of motion of the source or the observer.

The second of these postulates is actually contained within the first (because the laws of electromagnetism predict a constant speed of light), but we state it explicitly for clarity, just as Einstein did in his 1905 paper. The observers who move at constant speed and in a constant direction are often called inertial observers.

To date, all of the electromagnetic waves we have observed in a vacuum travel at this same speed, lending evidence to this idea. The prediction of classical electromagnetism is borne out.

For the rest of this chapter, we will explore some of the interesting consequences of the relativity principle. Again, Einstein simply assumed that it must be true, since to assume otherwise led to predictions that are at odds with experimental results. His resulting theory, The Special Theory of Relativity, though inconsistent with our common-sense notions of the world, has been shown to be entirely consistent with all experimental tests thus far - with one exception, that of certain quantum entangled states. Though these states are a fascinating aspect of the interplay of relativity and quantum mechanics, they are too far afield of our topic to explore here.

This page titled 9.1: The Principles of Special Relativity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.2: Time Dilation

Learning Objectives

- You will know that moving clocks are slower.
- You will be able to correctly use the time dilation formula to compare times in different reference frames.

What Do You Think: Constant Speed of Light

9.2.1: A Little Thought-Experiment

The first result of relativity that we will explore is its implication for time. The primary tool we will use in our explorations will be what Einstein called a Gedankenexperiment, or "thought-experiment." Einstein used many of these experiments to guide his thinking about physics. Only later did he flesh out the details through mathematical calculations. His imagined trip alongside a light wave is a kind of Gedankenexperiment.

For our next thought-experiment, we will imagine a special kind of clock. This clock is very simple, consisting of only three parts: two mirrors that are 100% reflecting and perfectly flat, and a photon that bounces back and forth between the mirrors, as depicted in Figure 9.2.

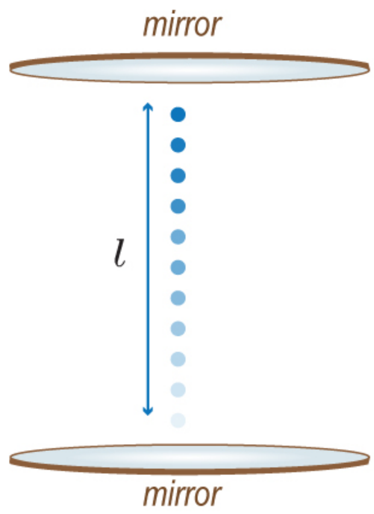


Figure 9.2: An ideal clock can be imagined from a photon bouncing back and forth between two parallel mirrors. This is called a light clock. Credit: NASA/SSU/Aurore Simonnet

This clock keeps time by registering a “tick” each time the photon bounces off the bottom mirror (we could use the top mirror just as well). If the distance between the mirrors remains constant and the speed of light is also constant, then this will be a perfect clock, keeping perfect time. It is not important whether we lack the technology to actually construct such a clock. This is a *thought*-experiment, so this experiment takes place in the ideal laboratory of our mind.

If we could observe such a stationary clock in our lab, we would note that it registers a tick after a time span (Δt) of:

$$\Delta t=\frac{2 l}{c}$$

Here c is the speed of light, as usual. The numerator on the right side of the equation, $2l$, is the distance the photon travels between ticks. We use the Greek letter Δ (“delta”) in addition to the letter *t* to represent the time span, for reasons of convention: In physics and astronomy, *Δ* often implies an amount that a quantity has changed—in this case, an interval of time. This is distinct from an initial or final value (starting or ending time), which would be represented by just the letter *t* itself.

The equation above is the result we get for a clock that is not moving relative to us. We say it is in our rest frame because we measure the clock to be at rest. The term "frame" is shorthand for frame of reference, or simply, reference frame. So far, this result is completely consistent with what we expect from classical physics.

Stationary Light Clock

Worked Example:

How big would our light clock have to be for the time between ticks to equal 1 second?

- Given: $c=300,000 \text{ km/s}$, $\Delta t=1 \text{ sec}$
- Find: l
- Concept: $2 l=c(\Delta t)$
- Solution:

$$\begin{aligned} 2 l &=(300,000 \text{ km}/ \cancel{s})(1 \ \cancel{s}) \\ &=300,000 \text{ km} \end{aligned}$$

Therefore $l=150,000 \text{ km}$.

We can also answer this question by realizing that the total distance the light travels must be one light-second of travel time, so the size of the light clock, l, must be half a light-second since the clock is traversed by the photon once on the way out and once on the way back. So, we can think of the size as half a light-second, or as 150,000 km.

Question:

Now we will look at what happens if the clock moves past us with some constant velocity, v. Figure 9.3 shows the situation that we observe if the clock moves relative to us.

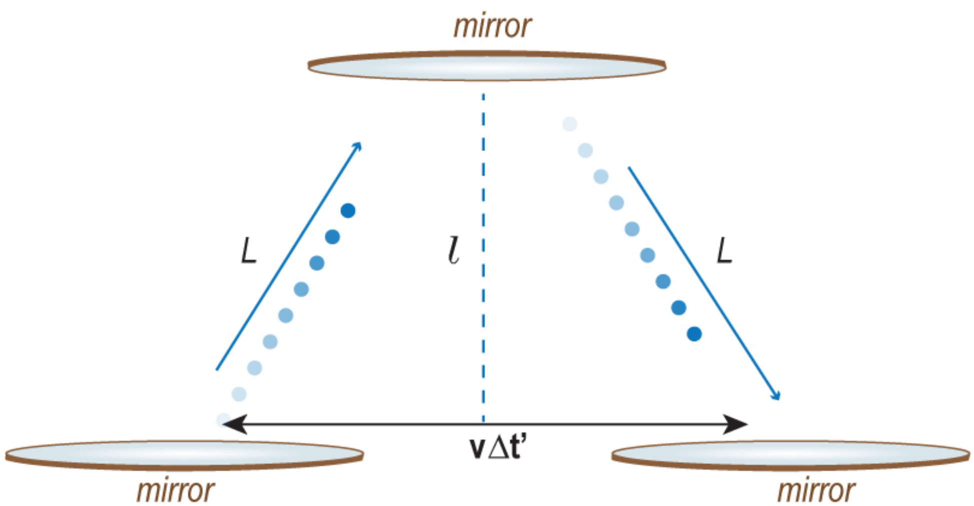


Figure 9.3: A moving ideal clock with a photon bouncing back and forth between two parallel mirrors. The clock is moving to the right with speed v. Credit: NASA/SSU/Aurore Simonnet

In the time that the light travels from the bottom mirror to the top and back again ($\Delta t'$), the clock has moved a distance (d) given by:

$$d = v\Delta t'$$

Here, we have put a prime on the time interval to distinguish it from the stationary case because, as we shall see, the two are not the same.

The light travels directly from one mirror to the other. This is the distance labeled L in the figure. Notice that the path traveled by the light (L) is related to the distance the clock moves (d) and the distance between the two mirrors (l) because they form a (right) triangle. We can write *d*, *L*, and *l* in terms of *v*, the time intervals and *c*, and eventually derive an expression for *Δt′* in terms of *Δt*, *v*, and *c*. The derivation is shown in Going Further 9.1: Moving Clocks. The expression is:

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \frac{v^2}{c^2}}}$$

According to this equation, and thus, to the special theory of relativity, the time interval measured when the clock is stationary and the time interval measured when the clock is moving are not the same; the two observers have measured *different* time intervals for the same two events—successive arrivals of a photon at the bottom mirror.

Some care is required here to understand exactly what this result is saying. The time interval Δt is that needed for a photon to travel from the bottom mirror to the top one and back again *in a frame in which the clock is at rest*. Let us call any frame in which an observer is at rest the *rest frame of the observer.* If a clock is in an observer's rest frame (at rest relative to that observer), then that observer will measure a time interval Δt between successive ticks of that clock. Now imagine the case when the clock is not in the observer's rest frame, but rather, is moving with some constant speed v in a constant direction (uniform motion) relative to the observer. In that case, the observer would measure a time interval $\Delta t'$ between successive ticks of the clock.

Clearly, from the expression above, Δt and $\Delta t'$ are not the same when the relative speed between the two frames is not zero. Since the velocity between the clock and any observer will always be less than c, we can reason that the term multiplying Δt is always greater than or equal to one. That means that a "moving" observer measures the round-trip time for the photon in the clock to always be larger than the round-trip time for the photon measured by an observer in the clock's rest frame: keep in mind that "moving" in this case means moving relative to the clock. The clock therefore ticks more slowly when it moves relative to the person observing it. This means, and this is the tricky part, a clock *ticking slower* will record *less time passing* than a clock ticking more rapidly. Since clocks that are moving relative to an observer always tick slower than clocks at rest relative to that observer, they record less time passing than clocks in the observer's rest frame.

However, it is important to consider more carefully what we mean by "at rest" and "moving." Consider two observers moving uniformly relative to each other. Each can have a clock in her own rest frame, and each will consider herself to be at rest and the other observer to be in motion. Since, according to special relativity, the laws of physics must be the same for both observers, there is no experiment that either can do to show that she is "really" at rest, while the other observer is "really" in motion. Special relativity implies that the claims of both observers to be "at rest" are equally valid. That means that the time distortion effects are completely symmetric. Each observer sees her own clock, in her own rest frame, ticking more rapidly than the clock of her "moving" counterpart.

So, which clock is really ticking slower? That is the wrong question to ask. The "relativity" part of special relativity is there because motion, and also the measurement of space and time, is relative, not absolute. If an observer in uniform motion notices another observer, also in uniform motion *relative to herself*, then she will measure time for that observer to pass more slowly *relative to her own time*. This statement is true for all observers in uniform motion, and there is no absolute measure of either space or time. That is a critical result of special relativity to keep in mind.

Going Further 9.1

For the case of the moving light-clock, the path traveled by the light (L) is related to the distance the clock moves (d) and the distance between the two mirrors (l) by the Pythagorean theorem:

$$L^2 = l^2 + \left(\frac{d}{2}\right)^2$$

We have written d/2 because the right triangle is only half of the light's path.

We can write these distances in terms of v, the time intervals, and c, to derive an expression for Δt' in terms of Δt, v, and c.

Writing the distance moved by the clock in the time $\Delta t'$ we have $d = v\Delta t'$, which we can substitute as below.

$$L^2 = l^2 + \left(\frac{v\Delta t'}{2}\right)^2$$

Traveling the distance L takes only half the time between ticks because the light has to traverse it once on the way up and then again on the way down to make one tick. That is what the factor of two represents in the expression for d: we have only traveled half of the full time interval needed for a clock tick by traveling up to the top mirror. Since we assume that the speed of light is constant (according to the second principle of relativity), we can write a relation between L, c, and Δt' like the one we wrote for l, c, and Δt:

$$\Delta t' = \frac{2L}{c}$$

This looks similar to our expression for l (as measured when the clock is at rest), but it is for the case of the moving clock. Solving these for L and l and substituting them into the equation for L2, we get:

$$\left(\frac{c\Delta t'}{2}\right)^2 = \left(\frac{c\Delta t}{2}\right)^2 + \left(\frac{v\Delta t'}{2}\right)^2$$

Now rearranging by collecting the common time terms we have the following.

$$c^2\Delta t'^2\left(1 - \frac{v^2}{c^2}\right) = c^2\Delta t^2$$

Finally, we may cancel the common term of c^2, rearrange and take the square root.

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \frac{v^2}{c^2}}}$$

9.2.2: The Relativistic Gamma Factor

Examining the relationship between *Δt'* and *Δt*, we see that the term inside the radical (square root symbol) is always between zero and one. It is zero when the velocity of the "moving" clock is *c*, the speed of light (which, according to special relativity, can never happen; see Section 9.7). The term equals 1 when the velocity of the "moving" clock is zero (i.e., when the moving clock is not moving!). At all other velocities, it is some number between zero and one. The square root of a number between zero and one is also between zero and one, and its reciprocal is always greater than or equal to one. This factor plays a central role in relativistic physics. Since it comes up over and over again, it has been given its own name: gamma (γ).

$$\gamma \equiv \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

The symbols used in this expression are what we expect: *c* is the speed of light and *v* is the relative speed between two reference frames of interest. The "equal sign" with three bars instead of two is actually a "defined as" sign. This expression is the definition of the gamma factor in relativity.

A graph of the gamma factor vs. *v/c* is shown in Figure 9.4. An interactive version of this graph will be available as you work through the activities below, in case you prefer to use the graph instead of the equation to determine *γ* or *v* as needed.

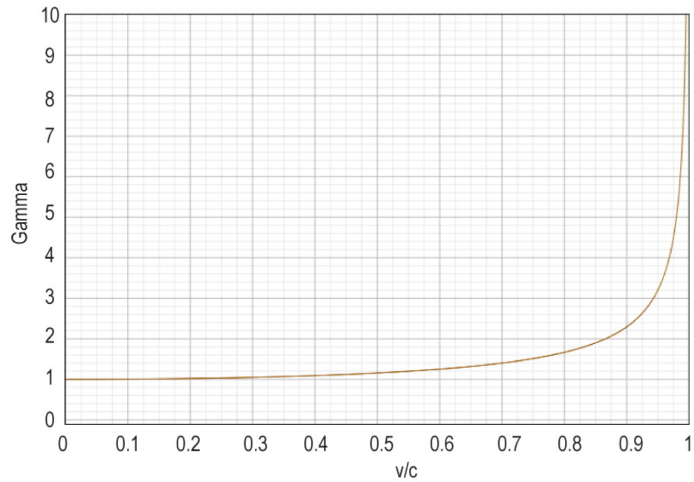


Figure 9.4: The gamma factor (γ) is plotted vs. *v/c*. For velocities much lower than that of light, gamma is very nearly one. In that case the time difference in the two frames is small. Only as *v* approaches *c* does gamma begin to grow, slowly at first, and then quite abruptly after about 0.9c. Credit: NASA/SSU/Aurore Simonnet/Kevin McLin

We can rewrite the equation for time intervals between successive clock ticks on two clocks in relative motion in terms of the gamma factor.

$$\Delta t' = \gamma \Delta t$$

As before, $\Delta t'$ is the length of time it takes for one tick on a moving clock, and Δt is the length of time it takes for one tick on a clock at rest.

Since gamma is always greater than or equal to one, the length of time it takes for one tick on a moving clock is always greater than the length of time it takes for one tick on a clock at rest. This effect is called **time dilation**. Clocks tick more slowly for the moving clock. As we have already discussed, this result is true for both observers: two observers in relative uniform motion each sees her own clock (the one that is not moving for her) ticking more rapidly (time passing faster) and the other clock (moving relative to her) ticking more slowly. If the relative velocity between the observers is zero, then they measure the same time. In this case, they are in the same rest frame. Remember that the motion referred to here is relative motion. Any observer always has her own clock, say a wristwatch, that is *not moving* relative to herself.

Time dilation allows us to understand the problem we first encountered at the opening of the chapter: how muons manage to make it all the way down to the ground before decaying. According to an observer on the ground, the muons can travel to the ground because time for them is passing much more slowly than it passes for that observer, who is at rest on Earth. This idea opens our minds to a new way of thinking about time, as you will see in the next set of activities.

Time Dilation

Worked Examples:

The fastest spacecraft that humans have launched so far is New Horizons, which visited Pluto in 2015. It traveled at a speed of about 16 km/s. If a spaceship in the future could go 1,000 times faster than New Horizons, what would the gamma factor be for that speed?

- Given: v = 16,000 km/s
- Find: γ
- Concept: $\gamma = 1/(1 - (v)^2/c^2)^{0.5}$
- Solution: $\gamma = 1/(1 - (16{,}000\ \cancel{\text{km/s}})^2/(300{,}000\ \cancel{\text{km/s}})^2)^{0.5} = 1.0014$
- Note: here we have used the exponent of 0.5 to denote the square root

Alternatively, you could find *v/c* and then use the clickable gamma plot tool:

USE GRAPH

v/c = (16,000)/(300,000) = 0.053

From reading the graph, $\gamma = 1.00125$ (at 0.05c)

2. If one tick on a clock in mission control takes 1 second, how much time would it take for a tick on a clock in the ultra-fast (16,000 km/s) spacecraft (from the perspective of someone at mission control)?

The observer at mission control thinks that she is at rest and that the spacecraft is moving.

- Given: $\gamma = 1.0014$, $\Delta t = 1$ s
- Find: $\Delta t'$
- Concept: $\Delta t' = (\Delta t)(\gamma)$
- Solution: $\Delta t' = (1\ s)(1.0014) = 1.0014$ s. This means that time on the moving spacecraft is slower (ticks take longer) compared to the time on the ground.

3. If 1 second passes on a clock in mission control, how much time would pass on a clock in the spacecraft (from the perspective of someone at mission control)?

The observer at mission control thinks that she is at rest and that the spacecraft is moving. Since the moving clocks tick more slowly, less time will pass on the spaceship than in mission control, at least, according to an observer sitting in mission control. The two are related to each other by gamma. If we call the time that passes at mission control Δt_m and the time on the ship Δt_s, then we have:

- Given: $\gamma = 1.0014$, $\Delta t_m = 1$ s
- Find: Δt_s
- Concept: The time that passes on the spaceship (Δt_s) will be *smaller* than Δt_m by a factor γ (the time dilation equation is for the time of ticks, not for the time that passes).
- Solution: $\Delta t_s = (1\ s)/(1.0014) = 0.9986$ s This means that time on the moving spacecraft is slower (less time has passed) compared to the time on the ground.

Questions:

Time Dilation and Reference Frames

In this activity, you will change the speed at which clocks move and see how the elapsed time intervals are affected. A small window with a clock will open when you load the activity.

Play Activity

A. Ticking time

A light clock ticks every time a bouncing photon returns to the first mirror. These ticks define the length of a second for the clock, so one second is the round trip travel time for a photon. The duration of this second is adjustable with time dilation, which in turn depends on the relative motion between the clock and observer.

After you have understood how time dilation of the clock works, click on the OK button to load the main activity window. Note that the OK button will not be enabled until you adjust the time dilation with the slider.

If you want to go back to the opening example clock at any point, you can reload the page.

B. Moving light clocks

When the main page loads, you will see windows for two clocks. Each can have its speed (relative to you, the observer) adjusted independently.

Adjust the slider bar on the upper clock so that it will be traveling at 10% of the speed light (i.e. 0.1c or 10% c). Then play the animation.

Now repeat with v = 0.2c, but this time, set the velocity for the top clock to zero, and adjust the slider bar for the bottom clock.

Predictions, observations, and explanations. You may find the clickable gamma graph useful.

USE GRAPH

You might protest that our light clock thought-experiment with mirrors is not an accurate depiction of the real world. For instance, if the mirror moved through space so quickly that it was no longer in the photon's path by the time the photon traveled from the bottom mirror to the top mirror, then the photon would not be reflected. You might conclude that this method would be a way to distinguish the moving mirror case from the stationary mirror case. However, this is not correct. In the case that the photon missed the mirror, it would do so for both observers, not just one. And, we could avoid this problem entirely by imagining two infinitely long mirrors. Everything in our argument above would still be valid, and we would arrive at the same conclusions as before.

The proper way to view the situation is as Albert Einstein did: It is not possible to say which observer is "at rest." Only the observers' relative motion can be determined. If one of them sees the photon bounce off one of the mirrors, then the other observer will see that too. What they will disagree about is the time at which each bounce occurs.

They will also disagree about where the photon is when it bounces, as we will see in Section 9.3. We now revisit the example from the opening of this chapter: muons produced in Earth's upper atmosphere.

✓ Muons From Cosmic Rays, as Seen From the Ground

Worked Example:

We now revisit the example from the opening of this chapter: muons produced in Earth's upper atmosphere. If 10 million muons are produced at an altitude of 10 km, how many of them will reach the ground?

1. First, we will examine this problem without using special relativity.

If we take the speed of the muons to be slightly slower than the speed of light (0.99c), then we can calculate the time required for the muons to reach the ground:

- Given: d = 10 km, v = (0.99)(300,000 km/s) = 2.97×10^5 km/s, N = 10^7 muons
- Find: t
- Concept: v = d/t

- Solve: $t = (10\,\text{km})/(2.97\times10^5\text{km/s}) = 3.37\times10^{-5}\text{ s} = 33.7\text{ microseconds}$

But, this is a very long time for a muon. Muons have a half-life for decay of only 2.2 microseconds. This trip to Earth's surface takes more than 15 half-lives (33.7/2.2 = 15.3). The probability of a muon surviving so long is:

$P = (0.5)^{15.3} = 2.5\times10^{-5}$

Given this incredibly low probability, we should expect to see only about $(10^7$ muons) × (2.5×10^{-5}) = 250 muons reach the ground. However, many more muons do reach the ground.

2. We will do the calculation again, but this time, we will include the effects of special relativity.

We know that the muon lifetime will be increased by a factor of gamma, so we must calculate what this factor will be. You can either use the equation or the clickable graph to determine gamma. Here, we show you how to calculate gamma using the equation. You should verify that you get the same thing using the clickable graph.

- Given: v = 0.99c
- Find: γ
- Concept: $\gamma = 1/(1-(v^2/c^2))^{0.5}$
- Solution: $\gamma = \dfrac{1}{\sqrt{1-(0.99c)^2/c^2}}$ = 7.1
- Again, the exponent of 0.5 means take the square root

USE GRAPH

So, now if we calculate the probability of the muons making it to the ground, but with a lifetime 7.1 times bigger, we have only about 2 half-lives (trip time / half-life = 33.7/(7.1 × 2.2) = 2.1):

$P = (0.5)^{2.1} = 0.22$

This result is a much larger probability, and many more muons reach the ground: (0.22) × (10^7 muons) = 2.2×10^6 or 2.2 million muons. This result is many more than would be seen if relativity did not affect the process, and it agrees with the fraction actually measured.

Questions:

This page titled 9.2: Time Dilation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.3: Length Contraction

Learning Objectives

- You will know that moving rulers are shorter.
- You will be able to correctly use the length contraction formula to compare lengths in different reference frames.

What Do You Think: Distances and Time

You might be puzzled by an apparent contradiction from the example of the muon in the last section. It makes a certain amount of sense that a muon would reach the ground as measured by a ground-based observer. After all, the muon's clock is seen to tick more slowly, and so its lifetime as seen by the ground-based observer appears to be longer. But an observer moving along with the muon would see a clock sitting on the ground moving more slowly than the clock riding along with the muon, so should the muon decay even more rapidly in that frame? The answer is "no," of course, but we have to consider another aspect of relativity for this example to make sense.

It is not just clocks that read differently for different observers moving relative to one another. Meter sticks also give a different measure. We should not be surprised by this fact. After all, velocity is a ratio of distance to time, and if the velocity of light is to be constant between two frames in relative motion, while time is not, then the lengths they measure must not be the same, either. If it was, the ratio of distance traveled by a photon over a time interval could not be the same for both. However, not all lengths are affected, just the ones along the direction of relative motion.

This effect is known as length contraction. If we consider its effect on a ruler of length L, it can be expressed mathematically as below.

$$L' = \frac{L}{\gamma}$$

For this expression, L' is the length in the frame moving with speed v, L is the length measured by an observer at rest with respect to the ruler, and γ is the gamma factor, as defined in the previous section. The ruler is taken to be aligned with the direction of motion, since that is the only direction for which length contraction happens. Since gamma is always bigger than one, L' is always *less than* L, so *moving rulers are shorter* than they are at rest.

LENGTH CONTRACTION

USE GRAPH

Worked Examples:

1. An astronaut whose height on Earth is 1.7 m is flying in a spacecraft at 0.8c. What is her height as measured by her fellow astronauts in the ship?

From the perspective (frame) of the ship, the astronaut is at rest. Therefore, her height is her usual height of 1.7 m

2. What is the height of the astronaut as measured by an observer at mission control?

From the perspective (frame) of mission control, the astronaut is moving, so she will be shorter in height:

- Given: L = 1.7 m, v/c = 0.8
- Find: L'
- Concept: L' = L/ γ
- Solution: from the clickable graph, the γ corresponding to v/c = 0.8 is γ = 1.67, so the astronaut will be L' = 1.7 m / 1.67 = 1.02 m tall, which is indeed shorter.

Questions:

Muons From Cosmic Rays, as Seen by an Observer Riding With the Muon

Worked Examples:

1. We will look at the case of the muon again, this time from the reference frame of the muon instead of from the reference frame of the ground.

As the muon travels toward the ground at 0.99c, from its point of view, the distance from the muon to the ground is not as far away as it appears to be when measured by an observer on the ground. If the height of the muon is L as measured from the ground, then it will be shorter in the muon's frame (L') by the gamma

factor.

First, calculate the distance to the ground. In the muon's frame:

- Given: L = 10 km, γ = 7.1

https://phys.libretexts.org/@go/page/31337

- Find: L'
- Concept: *L*' = *L*/γ
- Solution: L' = (10 km)/(7.1) = 1.4 km

Next, calculate the time (t_{muon}) required to travel this distance:

- Given: v = speed of the muon = 0.99c, L' = 1.4 km
- Find: t_{muon}
- Concept: v = L'/t_{muon}
- Solution: t_{muon} = (1.4 km) / [(0.99) × (3 × 10^5 km/s)] = 4.7×10^{-6} s

Comparing this number to the muon's lifetime in its own reference frame, which is 2.2 micorseconds $(2.2 \times 10^{-6}\text{ s})$, we see that in terms of half-lives, it is $4.7 \times 10^{-6}\text{ s}/(2.2 \times 10^{-6}\text{ s})$ = 2.1 half-lives, just as we found before.

So, the results observed in the two frames are entirely consistent. The same numbers of muons reach the ground. The different observers only disagree about why the muons can make the journey successfully. The observer on the ground thinks it is because the muons' clocks tick more slowly. The observer moving with the muons says it is because the ground is not very distant from where the muons are produced.

Questions:

LENGTH CONTRACTION OF SPACESHIPS

In this activity, you will use an example of a rocket ship and a space station flying past each other to explore the idea of length contraction.

You can use the slider bar to adjust the relative speed of the ship and station.

The changes in length are shown accurately, but the speeds of the objects moving across the screen are not relativistic (otherwise, the motion would be too fast to see).

Play Activity

USE GRAPH

Worked Examples:

A space station and a rocket ship are both 100 m long at rest. Imagine that they are flying past each other at a velocity of 0.5c (50% of the speed of light).

1. From the point of view of an astronaut on the space station, what is the length of the rocket ship as it flies by?

- To figure this out, slide the speed control to 0.5c, and play the animation.
- Notice that from the frame of reference of the space station, the rocket ship appears to be shorter.
- Using the clickable graph for v = 0.5c, you should find that gamma = 1.15, so the rocket ship is about L' = L/γ = 100 m/1.15 or about 87 m long.

2. From the point of view of an astronaut on the rocket ship, what is the length of the space station?

- Play the animation again, this time noticing the perspective from the frame of reference of the rocket ship.
- From the rocket ship, the space station flies by and appears shorter, again only about 87 m long.

Questions:

This page titled 9.3: Length Contraction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.4: The Geometry of Special Relativity - Spacetime

Learning Objectives

- You will recall the mathematical concepts of coordinate systems and the Pythagorean theorem.
- You will be able to draw/interpret spacetime diagrams for a variety of scenarios.

What Do You Think: Spacetime

The examples in the previous sections show how space and time are different for different observers depending on their relative motion. One way to understand this concept is to consider the geometry related to special relativity. It has some similarities to the geometry you may have already learned about in school, but it also has some stark differences. The differences result from the fact that the speed of light is the same for all observers, no matter their state of motion. In this section, we will introduce the idea of spacetime and its peculiar geometry.

9.4.1: Geometrical Refresher

The geometry that most of us learn in school is that of a plane. It is sometimes called **plane geometry**, or **Euclidean geometry** after Euclid, the mathematician in Ancient Greece who codified its properties. In plane geometry, we learn that the angles of a triangle must all add up to 180 degrees, parallel lines never meet, and the ratio of the circumference of a circle to its diameter is the number π. Another relation that holds true in plane geometry is the Pythagorean theorem , which describes the relation between the three sides of a right triangle. The Pythagorean theorem is illustrated in Figure 9.5 and can be written as:

$$a^2 + b^2 = c^2$$

where a and b are two sides of a right triangle and c is the longest side, called the *hypotenuse*. It is the diagonal side in the diagram below.

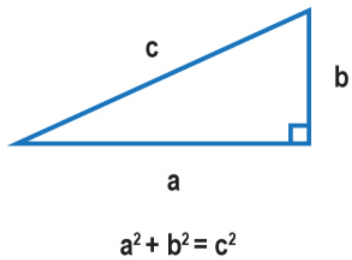


Figure 9.5: The Pythagorean theorem describes the relationship between sides a, b, and c of the right triangle shown. Credit: NASA/SSU/Aurore Simonnet

When we talk about plane geometry, we often talk about points on a plane and the relationships between them. For instance, in Figure 9.5, three points make the vertices of the triangle. Those points can be described by their (x,y) coordinates in a coordinate system, as shown in Figure 9.6.

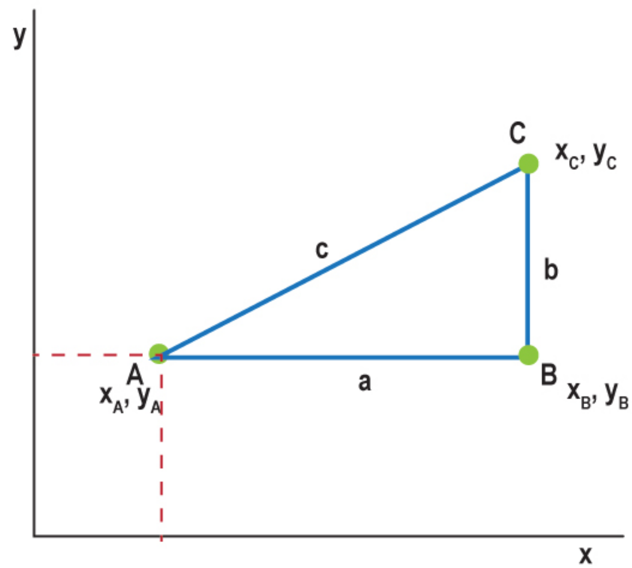


Figure 9.6: This is the same right triangle as in the previous figure, now with the (x,y) coordinates of the vertices labeled. The vertices are points *A*, *B*, and *C*. They have coordinates (x_A,y_A), (x_B,y_B), and (x_C,y_C), respectively. The various relationships between coordinates and lengths are also shown. Credit: NASA/SSU/Aurore Simonnet

Once we have chosen a coordinate system, we can write expressions relating points on the plane using those coordinates. So for Figure 9.6, we can use the coordinate system to write expressions for the length of each side of the triangle:

$$a = x_B - x_A$$

$$b = y_C - y_B$$

$$c^2 = (x_B - x_A)^2 + (y_C - y_B)^2$$

These expressions are true because the length of the side of the triangle labeled a is the difference in the x-coordinates of points *A* and *B*. Similarly, the length of side b is the difference in the y-coordinates of points *B* and *C*. The length of side c is given by the Pythagorean theorem. In addition, points *A* and *B* both have the same y-coordinate, and points *B* and *C* have the same x-coordinate.

Compare the situation just outlined to the one shown in Figure 9.7. We have not changed the triangle ABC at all. It still has the same sides and the same angles, and it has the same relationships between the points *A*, *B*, and *C*. However, we have rotated the coordinate system we are using to describe the positions of things, and so the coordinates of each point have changed.

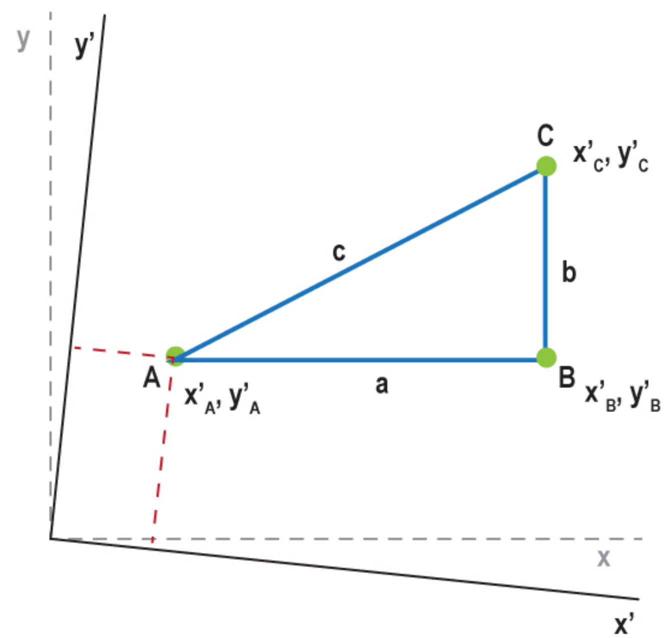


Figure 9.7: Same right triangle, but now with slightly rotated coordinates. Credit: NASA/SSU/Aurore Simonnet

The mathematical expressions for the lengths of the sides *a*, *b*, and *c* in these rotated coordinates are given in Math Exploration 9.1.

Math Exploration 9.1

$$a^2 = (x'_B - x'_A)^2 + (y'_B - y'_A)^2 \tag{9.4.1}$$

$$b^2 = (x'_C - x'_B)^2 + (y'_C - y'_B)^2 \tag{9.4.2}$$

$$c^2 = (x'_B - x'_A)^2 + (y'_C - y'_B)^2 \tag{9.4.3}$$

On first glance, these look different from variations on the Pythagorean Theorem. This is because in the rotated coordinate system none of the points have identical x- or y-coordinates. As a result, the expressions for the lengths of sides *a* and *b* are more complicated. However, in both cases, we are really just using the Pythagorean Theorem to compute the distance between any two points, but it simplifies our task a lot when our triangle is described in a convenient set of coordinates.

The important thing to take away from these examples is that the triangle itself has not changed in the least just because we have used a different coordinate system to describe it. It has the same points at its vertices, and its sides and angles are all the same. We can encapsulate this idea in a Principle of Invariance for Euclidean geometry:

The distance between two points in a Euclidean space is independent of the coordinate system used to describe that space.

This statement means that we are free to choose any coordinate system we like when we want to describe relationships between objects in a Euclidean space. For instance, it would be much simpler to use the un-rotated coordinate system to describe the triangle in the examples above, but we are free to use the rotated one if we wish. These notions of lengths remaining the same in different coordinate systems, called invariance, will be carried over into spacetime. It will be the most important idea to remember when trying to understand its geometry.

To describe the distance between any two points, which is invariant, we can write the complicated expressions above in a more compact and general way so that they are easier to understand. If we use the following notation:

$$\Delta x = x_1 - x_2$$

$$\Delta y = y_1 - y_2$$

for any two points 1 and 2, then we can write a much simpler expression for the distance, *d*, between them:

$$d^2 = \Delta x^2 + \Delta y^2$$

This is illustrated in Figure 9.8. Again we are using Greek Δ's because we are talking about an interval. In this case, we mean changes in x and y.

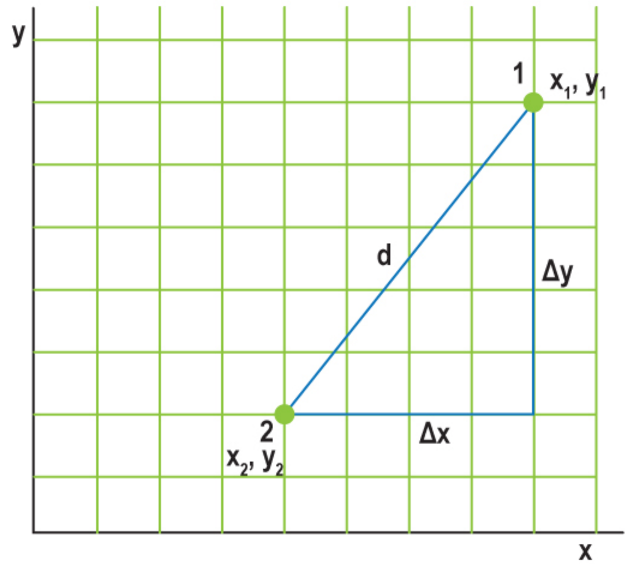


Figure 9.8: Coordinate system with two points, 1 and 2, with a triangle shown to illustrate how the Pythagorean heorem is used to calculate their separation. Credit: NASA/SSU/Aurore Simonnet

✔ DISTANCES AND COORDINATES

Worked Examples

1. Find the distance between the points (4, 5) and (8, 3).

- Given: $(x_1, y_1) = (4, 5)$ and $(x_2, y_2) = (8, 3)$
- Find: d
- Concept: $d^2 = \Delta x^2 + \Delta y^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$
- Solution:

Use the Pythagorean Theorem to find the distance squared. Then take the square root to find the distance.

$d^2 = (4 - 8)^2 + (5 - 3)^2$
$d^2 = (-4)^2 + (2)^2 = 16 + 4 = 20$
$d = (20)^{1/2} = 4.47$

The *Graphing Tool*, which is available in the tool bar, can help you to visualize where the points are in the coordinate system as you work through these problems.

For this example, our two points would be displayed as in Figure A.9.1.

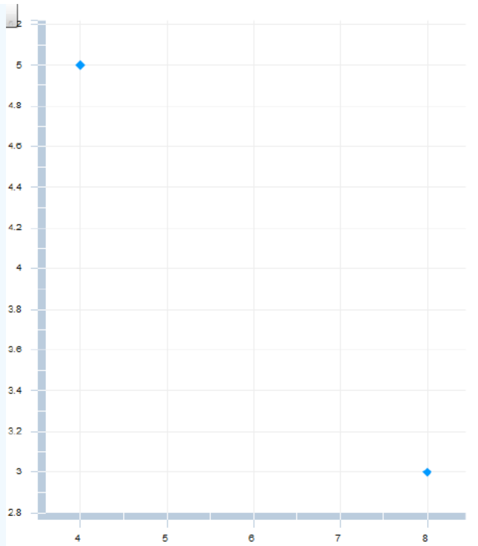


Figure A.9.1: The points (4,5) and (8,3) are plotted using the *Graphing Tool*. Credit: NASA/SSU

For another example see Math Exploration 9.2.

Math Exploration 9.2

Questions:

https://phys.libretexts.org/@go/page/31338

9.4.2: Spacetime Diagrams

When Albert Einstein published his first paper on special relativity in 1905, he did not yet fully grasp its geometrical nature. He came to appreciate these aspects of the theory over the next several years while working with mathematician Hermann Minkowski (1864–1909). Minkowski developed the idea of spacetime , in which events occur in a four-dimensional space that includes both (the three dimensions of) space and (one more dimension of) time—hence, the name.

Spacetime is dynamic. You can plot the wanderings of a particle by tracing its path in this spacetime; i.e., you can plot its position at every moment in time. Such a path is called the worldline of the particle, and it traces out events in the life of the particle, both past and future. Events are just that. They might be the collision of the particle with another particle, or the moment the particle decays into other particles. In any case, the points in spacetime are called events, and a particle's worldline is the string of events that comprise its existence, each tagged with a time and place of occurrence.

As a concrete example, you might describe some of the events in the worldline of your day with the following table:

EVENT	TIME	LOCATION	DESCRIPTION
1	6:00 AM	Bedroom	Get out of bed
2	6:05 AM	Bathroom	Take shower
3	6:30 AM	Kitchen	Have breakfast
4	7:00 AM	Front door	Leave for school
5	7:10 AM	Home bus stop	Board bus
6	7:45 AM	School bus stop	Arrive at school

Table 9.1: A sampling of events that might occur in your day. Other events might happen that we have not shown. These events might include getting dressed, brushing your teeth, sitting down on the bus, etc. A complete description of your day would include every event that occurred, including its time and location. In spacetime, these events link together in a continuous path.

An event in spacetime has four coordinates, three in space and one in time: (t,x,y,z). However, spacetime plots are usually simplified by using only two axes: The vertical axis is time and the horizontal axis is space, as shown in Figure 9.9. We ignore the other two dimensions of space for the sake of convenience. Plotting all three spatial axes plus time would make the diagram too difficult to interpret. We cannot even imagine four perpendicular dimensions, let alone draw them on a two-dimensional piece of paper in such a way as to make sense of them. However, since we are free to choose a convenient coordinate system (as per the discussion above), we can take the direction of any particle's motion to be only in the x-direction. We need only remember that the effects we encounter for the x-direction will also occur in y- and z- directions if there is motion directed along them.

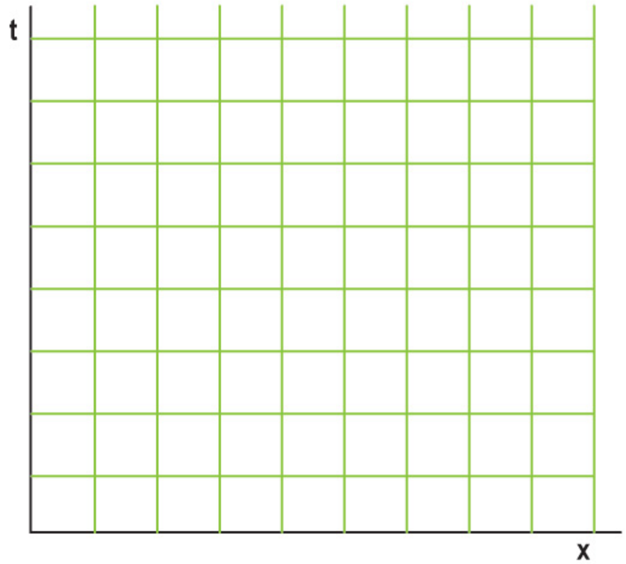
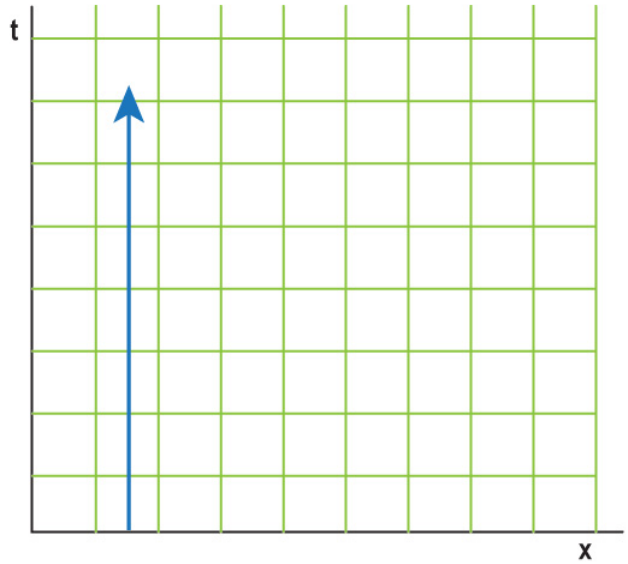


Figure 9.9: In a spacetime diagram, the vertical axis is time. The horizontal axis is space, with only one of the axes shown for simplicity's sake. All motion is taken to be only in this direction, again for the sake of simplicity. Credit: NASA/SSU/Aurore Simonnet

For a particle at rest, the plot is pretty boring. The particle has a constant position (x), so it is carried along a vertical path as time passes: Its worldline is a vertical line that intersects the x-axis at the position of the particle. An example is shown at left in Figure 9.10. Moving objects are only slightly more interesting. Their worldlines make an angle with the vertical, as the example on the right of Figure 9.10 demonstrates.

The slope of the worldline for a particle is related to its velocity. Specifically, since slope is rise over run, and velocity is distance over time, the slope of a particle's worldline will be the reciprocal of its velocity, $1/v$. This idea makes sense because a zero-velocity particle has a vertical worldline, i.e., one with infinite slope. As the particle moves faster, its worldline tips over more and more toward the horizontal because it travels farther in x in a given amount of time.

https://phys.libretexts.org/@go/page/31338



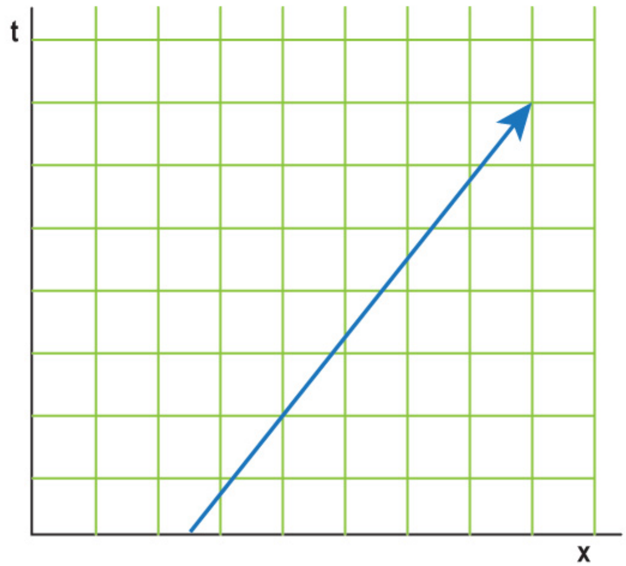


Figure 9.10: Worldlines on a spacetime diagram for a particle at rest (left) and for a particle moving at constant velocity (right). Credit: NASA/SSU/Aurore Simonnet

GRAPHING EVERYDAY SITUATIONS

In this activity, you will represent everyday physical situations in words, as data in a table, and as worldlines on a spacetime diagram. Assume the TV is at the origin of your spacetime graph. You will be using the *Graphing Tool* found in the toolbar.

1. Situation 1

2. Situation 2

3. Situation 3

So far, spacetime looks similar to Euclidean space, except that one of the axes is time. This difference is actually a pretty big one, if you think about it. Generally, in space, all directions use the same units of measure, feet or meters, for example. We never measure northward distances in inches and eastward distances in kilometers. We could do that, but it would not be very convenient because we could not easily compare the distance , say, between Paris and London, to that between Paris and Vienna (see Figure 9.11). The same is true in spacetime; things would be much simpler if we could use the same unit of measure along both axes.

https://phys.libretexts.org/@go/page/31338

Figure 9.11: A map of Europe showing Paris, London, and Vienna, with cardinal directions indicated. Distance from Paris to London: 13,296,371 in. Distance from Paris to Vienna: 1,037 km. Credit: NASA/SSU/Aurore Simonnet

Of course, one axis in spacetime is space and the other is time. Those are fundamentally different things. Still, it is easier to measure them using the same or at least similar units, and we have already been introduced to this idea. In previous chapters, we have used light-years. This is a unit of distance, not time; it is the distance light travels in a year. For example, light will travel 10 light-years in 10 years and 20 light-years in 20 years. Therefore, on spacetime diagrams, we will measure time in years (or hours, minutes, or seconds), and we will measure distances in light-years (or light-hours, light-minutes or light-seconds). This convention makes the two axes have related numerical values, greatly simplifying the comparison of distances along them.

Units of Time and Distance on Spacetime Diagrams

Worked Examples:

1. If we say that two events are 2 light-seconds apart, how much time does it take for light to travel between them?

If we say that two events are 2 light-seconds apart, it means that it takes light 2 seconds to travel from one to the other.

Since light travels 300,000 km/s, each light-second of distance is a very long way by terrestrial standards. If we want to convert back to meters or kilometers for some reason, we can always use the speed of light to do so, but it is more convenient just to use light-seconds or light-years to describe large distances.

2. If two events happen 2 seconds apart, how much time passes between them?

2 seconds. If we are talking about a time, not a distance, then the units have their usual meaning. An event that happens 2 seconds earlier than another event must happen before it.

Questions:

Given that we are using related units to measure both distance and time, we can immediately make an important inference about spacetime diagrams. In the units on our diagrams, light travels at a speed of one light-year per year. This means that its velocity is 1. If you are having difficulty understanding velocity in these units, it might be easier if you consider that our velocities are always in terms of the speed of light , *c*. So, in this way, they can be between 0 and 1. For example, a velocity of half of the speed of light would be 0.5. Since a photon has a velocity of 1, and the reciprocal of 1 is also 1, the worldline of a photon makes a 45-degree angle with the t- and x-axes. As we shall see, all massive particles must travel slower than light, so their worldlines all make an angle smaller than 45 degrees with the t-axis.

This is our first rule for creating and interpreting spacetime diagrams:

Rule 1. The worldlines of particles moving at the speed of light make a 45-degree angle with the t- and x-axes in a spacetime diagram. Worldlines of particles moving slower than light make an angle smaller than 45 degrees with the t-axis.

In the next activity, we will introduce you to the *Spacetime Diagram Tool* and practice drawing worldlines for particles traveling at various speeds.

SPACETIME WORLDLINES

Play Activity

Worked Example:

1. Use the *Spacetime Diagram Tool* to create worldlines for a particle moving with a speed of 0.5c.

To do this, adjust the "% speed of light" slider bar until it says 50%. The orange line will be the worldline for your particle. It should look like Figure A.9.3:

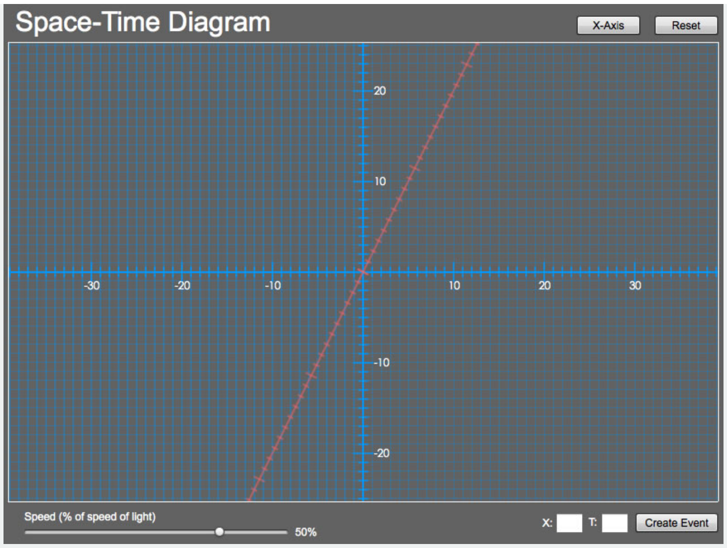


Figure A.9.3: Worldline for a particle moving at 0.5c in the *Spacetime Diagram Tool*. Credit: NASA/SSU

Questions:

A great power of spacetime diagrams is that they allow us to compare measurements of space and time in one reference frame to space and time in a different reference frame that is moving uniformly with respect to the first. To understand how this comparison is done, imagine two frames moving with constant velocity (that is, constant speed and direction) relative to one another.

For an observer at rest in a frame with coordinates t and x, the spacetime axes are drawn perpendicular, as we have done In Figure 9.12. We can then plot the spacetime axes, t' and x', of a moving frame on the same diagram. To do this, we plot the t'-axis as a line with slope *1/v*, where *v* is the relative velocity of the two frames. Why? First, we are free to take the origins of the two systems to coincide at $t = 0$ and $t' = 0$ (convenient coordinate systems again!). Then we can draw the worldline of the origin of the moving system in the system at rest.

Just as before, the worldline of a point moving at constant velocity makes a line with slope $1/v$, and this result is exactly what the motion of the origin of the "primed" system will do in the spacetime diagram of the "unprimed" one. The worldline of the origin of the primed system must run along t' by definition, because $x' = 0$ for the origin and along the t'-axis. This reasoning gives us the location of our t'- axis in the (t,x) frame. Getting the x'-axis is a bit more complicated, but the x'-axis is a line with slope v. The details are described in Going Further 9.2: Finding the Location of the x'-Axis.

It turns out that the x'-axis makes the same angle with the x-axis as the t'-axis makes with t-axis. So, in the unprimed frame, the primed frame gets squeezed in toward the diagonal. The faster the relative velocity between the two frames, the more squeezed the primed frame becomes.

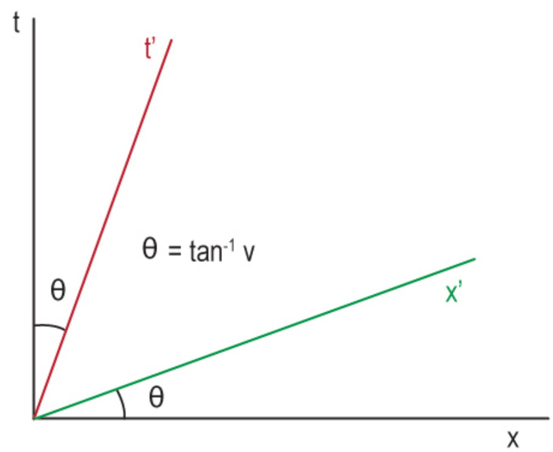


Figure 9.12: Axes (x,t) for a frame at rest with respect to an observer and, on the same diagram, axes (x',t') for a frame moving with velocity v. The angle between the t- and t'-axes and the x- and x'-axes are the same. They are the arctangent of the relative velocity between the two coordinate systems. Credit: NASA/SSU/Aurore Simonnet

Going Futher 9.2: Finding the Location of the X'-axis

To find the position of the x'-axis in the spacetime diagram, conduct the following thought experiment. Imagine that a flash of light is emitted from the point $x' = 0$ at some time $t' = T < 0$, or in other words, a flash of light is emitted from the origin of the primed system before it reaches the origin of the unprimed system. The flash is reflected and arrives back at $x' = 0$ at some time $t' = T > 0$. Because the return trip for the light is just the reverse of the outward trip, the time interval for the return trip ends T seconds after $t' = 0$. The geometry is shown in Figure B.9.1.

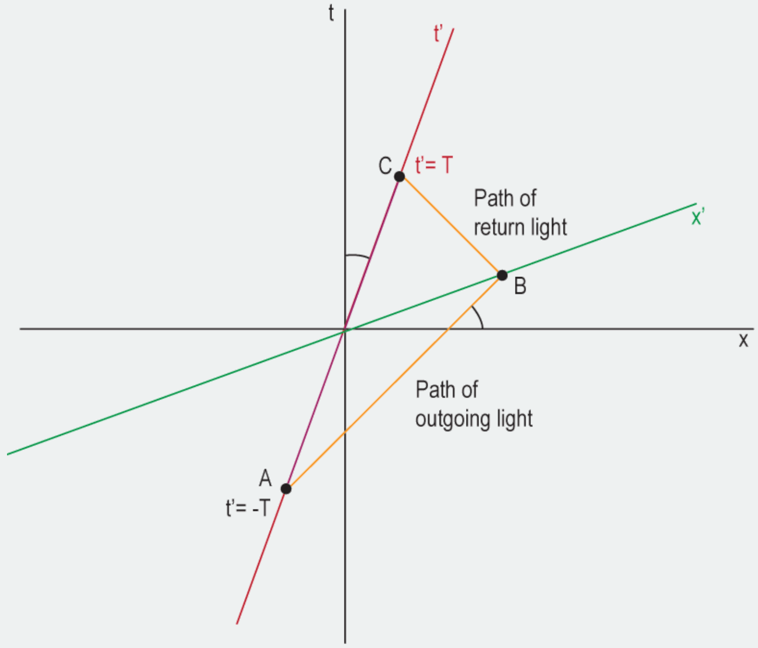


Figure B.9.1. A flash of light is emitted at event A, T seconds before t' = 0. It returns to its starting point at event C, T seconds after t' = 0. Event B, marking the reflection of the light, must lie on the x'-axis. Credit: NASA/SSU/Aurore Simonnet

Under these conditions, the reflection of the light must occur at t' = 0, and so that event, labeled event B, must lie on the x'-axis. Drawing a line from the origin through the reflection point locates the x'-axis.

Now we are ready for the second of our rules for creating spacetime diagrams:

Rule 2. To draw the time axis of a moving frame in the spacetime diagram of a frame at rest, draw a line with slope 1/v. To draw the spatial axis of the moving frame, draw a line with slope v.

In the next activity, we will use the *Spacetime Diagram Tool* to practice drawing axes for moving frames of reference.

SPACETIME AXES

Play Activity

Worked Example:

1. Use the *Spacetime Diagram Tool* to create axes for a reference frame moving with a speed of 0.5c.

- To create them, adjust the “% speed of light ” slider bar until it says 50%.
- The orange line will be the t'-axis.
- Click the “x'-axis” button to display the x'-axis (green) for the moving reference frame.
- Your diagram should look like Figure A.9.4.

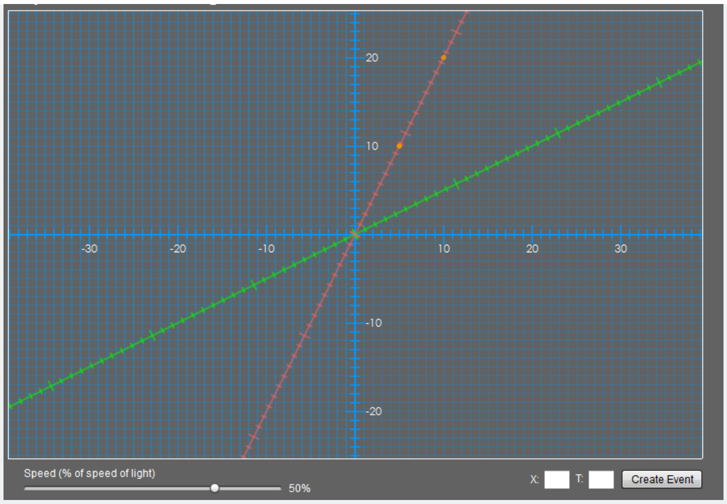


Figure A.9.4: Axes for a reference frame moving at 0.5c, as drawn by the *Spacetime Diagram Tool*. Credit: NASA/SSU

Questions

We are almost ready to use our spacetime diagrams to begin exploring some applications of relativity and what relativity says about the world. However, we still need to know how to compare the scales on the primed and unprimed frames. It turns out that they are not the same. To understand why, we have to employ the invariance principle for spacetime. This principle is analogous to the invariance principle we encountered for Euclidean space, where points have an existence independent of the coordinates we choose to describe them. Any coordinate systems we use to describe the world do not affect the distance between points, nor the orientation of points to each other. Spacetime is the same as Euclidean space in this regard, but there is one critical difference. To understand this difference, recall that the invariance of Euclidean space is encapsulated by the Pythagorean theorem :

$$d^2 = \Delta x^2 + \Delta y^2$$

In spacetime, we have an analogous expression:

$$s^2 = \Delta t^2 - \Delta x^2$$

Here we have used the letter *s* instead of *d* to remind ourselves that we are not measuring a distance between *points* in space, we are measuring the separation between *events* in spacetime—an important distinction. And, just as the distance between two points does not depend on the Euclidean coordinate system used to describe them, so the separation of events in spacetime is independent of the coordinates used. Events have an existence all their own in spacetime, independent of the spacetime coordinates. This is our third rule for spacetime diagrams, but it is far more important than that—it is the essence of special relativity.

Rule 3. The spacetime interval between two events, as defined above, is independent of the coordinate system used to describe t and x. It is said to be invariant.

Notice the minus sign in the spacetime expression. This minus sign seems innocent enough, but it is responsible for the profound differences between the familiar properties of Euclidean space and the counterintuitive properties of spacetime. Also, because of the minus sign, it is possible for the spacetime interval, s^2, to be negative. But, that means that the spacetime separation, s, could be the square root of a negative number, or imaginary. What does that mean? We are not really sure, but we do not have to worry about it. It is really only the spacetime interval, s^2, that is important for our understanding of spacetime.

Going Further 9.3: Examining the Differences Between Euclidean Space and Spacetime

Consider one of the differences between Euclidean space and spacetime: The points in Euclidean space that are all equidistant from a given point make a circle around it, or a sphere, if we consider all three dimensions. This result makes sense. After all, the Pythagorean theorem is the same as the equation for a circle:

$$r^2 = y^2 + x^2 \tag{9.4.4}$$

However, the equation for the spacetime interval has the same form as the equation for a hyperbola:

$$r^2 = y^2 - x^2$$

Apparently, spacetime intervals satisfy this equation, i.e., events that are equidistant from each other lie along hyperbolae! This is definitely not Euclidean. Figure B.9.2 compares Euclidean distances (on left) to spacetime intervals (on right). Note that we have used (x,t) coordinates for the spacetime plot instead of (x,y), in keeping with our convention that spacetime uses time as the vertical coordinate, while normal space uses *y*. As before, a spacetime interval is *s*, whereas *d* is used for a distance through Euclidean space.

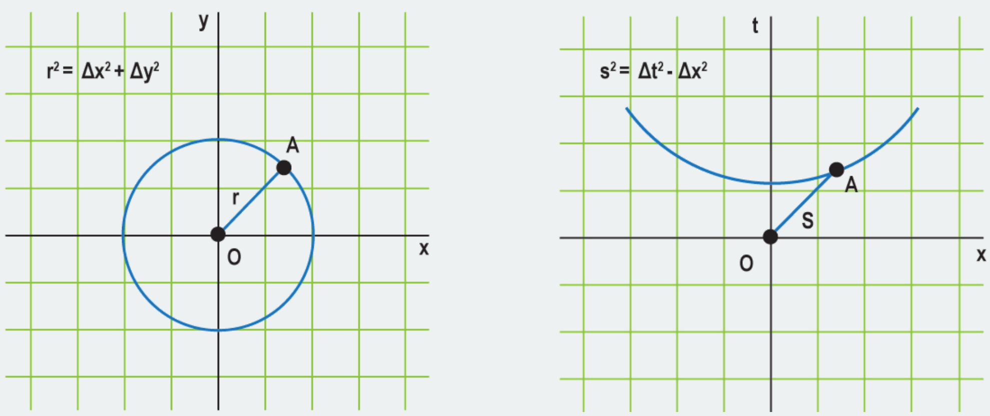


Figure B.9.2. These graphs show points equidistant from the origin. In Euclidean space, at left, the points lie on a circle, with the origin at its center. In spacetime, events equidistant from the origin lie along a hyperbola, as shown at right. To be clear, all points along the hyperbola at right have the same spacetime separation from the origin, just as all points along the circle at left lie the same distance from the origin. Credit: NASA/SSU/Aurore Simonnet

While the points along a hyperbola in Euclidean space are not all the same distance from the origin, they are the same distance from the origin in spacetime. The fact that we are forced to make our plots on the Euclidean space of a flat computer screen or piece of paper obscures the true nature of the spacetime plot. Do not let the limitations of the presentation medium obstruct your understanding of how spacetime differs from Euclidean space in this vital way. All of the points along a given hyperbola in the spacetime plot have the same spacetime separation from the origin.

We can find the correct scale for the (t',x') axis for a moving frame (the space between adjacent ticks on the axis), by plotting these hyperbolae for integer values of s. The intersection of each hyperbola with the t'-axis gives the coordinate where $t' = s$. Why? Because along the t'-axis, the x'-coordinate is zero. The situation is the same as along the t-axis, where the x-coordinate is zero. To take a specific example, consider a hyperbola for $s = 2$. The intersection of that hyperbola with the t-axis is where $t = 2$, because along the t-axis, $x = 0$. So, we have $s^2 = 2^2 = t^2 - x^2 = t^2 - 0^2$, or $t = 2$. Along the t'-axis, $x' = 0$, so t'^2 –

> $0^2 = 2^{2}$, or in other words, $t' = 2$. To find other values of t', you must simply find the intersection of the axes with the hyperbola for a given value of s. The spacings between ticks along the x'-axis are the same as those along the t'-axis.

The final thing we need to understand about spacetime diagrams has to do with reading the coordinates of an event in two different reference frames.

As an example, notice the event labeled A in Figure 9.13. In the unprimed system (the system in which the observer is at rest), the event occurs at position $x = 8$ and time $t = 10$. How can we tell? If you have ever read a graph, you know that you read off the x-coordinate by drawing a line through event A parallel to the t-axis. Lines parallel to the t-axis all have the same value of x, so by reading off the x-intercept of this line, you will determine the x-coordinate of event A. Similarly, to read the t-coordinate of A, draw a line through the event parallel to the x-axis.

In the primed system, which moves with velocity 0.5c relative to the unprimed, the coordinates of the event are $x' = 3.5$ and $t' = 6.9$. How can we read these coordinates? Read off the x'-coordinate by drawing a line through the event A and parallel to the t'-axis. Similarly, to read the t'-coordinate of A, draw a line through the event and parallel to the x'-axis. Lines parallel to the x'-axis all have the same t'-coordinate, which is to say they are simultaneous in the primed frame.

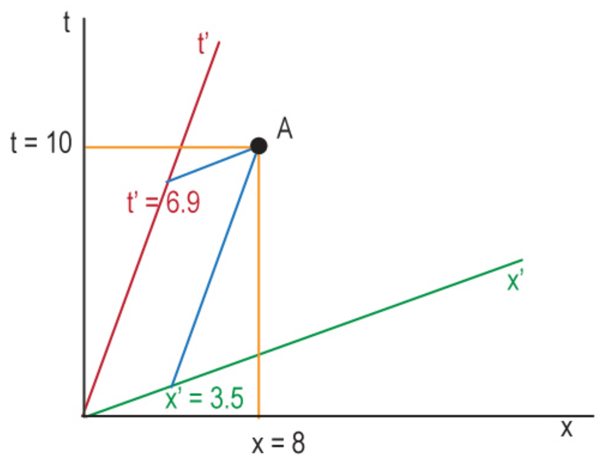


Figure 9.13: Spacetime diagram showing how to find the (x,t) and (x',t') coordinates of an event A. The yellow lines are parallel to the t- and x-axes. They can be used to read off the x and t coordinates for A. The red line is the t'-axis and the green line is the x'-axis. The blue lines are parallel to the t' and x' axes, so they can be used to read off the x'- and t'-coordinates for A. Credit: NASA/SSU/Aurore Simonnet

This is our last rule for building and reading spacetime diagrams:

Rule 4. The spatial coordinate of an event is found by drawing a line parallel to the time axis and finding its intercept with the space axis. The time coordinate of an event is found by drawing a line through the event parallel to the spatial axis and finding its intercept with the time axis. This works for both moving (primed) and rest (unprimed) reference frames.

Those are all the things you will need to know in order to interpret spacetime diagrams. Now, we will use what we have learned to explore how events will be described by observers in two different frames.

READING EVENT COORDINATES

In this activity, you will use the *Spacetime Diagram Tool* to create a graph similar to Figure 9.13.

Play Activity

Hit the Reset button on the tool, if you have not already.

The Spacetime Separation Between Events

Worked Examples:

1. Calculate the spacetime interval

(s^2) between event A in Figure 9.13 and the origin in both the (a) unprimed and (b) primed reference frames, and (c) verify that s^2 is invariant.

(a) Answer:

- Given: $x_1 = 8$, $x_2 = 0$, $t_1 = 10$, $t_2 = 0$
- Find: s^2
- Concepts: $s^2 = \Delta t^2 - \Delta x^2$, $\Delta t = t_1 - t_2$, $\Delta x = x_1 - x_2$
- Solution: $\Delta x = 8 - 0 = 8$, $\Delta t = 10 - 0 = 10$, $s^2 = 10^2 - 8^2 = 100 - 64 = 36.0$

(b) Answer:

- Given: $x_1 = 3.5$, $x_2 = 0$, $t_1 = 6.9$, $t_2 = 0$

https://phys.libretexts.org/@go/page/31338

- Find: s^2
- Concepts: $s^2 = \Delta t^2 - \Delta x^2$, $\Delta t = t_1 - t_2$, $\Delta x = x_1 - x_2$
- Solution: $\Delta x = 3.5 - 0 = 3.5$, $\Delta t = 6.9 - 0 = 6.9$, $s^2 = 6.9^2 - 3.5^2 = 47.6 - 12.3 = 35.3$

(c) The spacetime interval, s^2, is the same in both cases (within rounding error), which means it is invariant.

Questions:

Understanding Event Coordinates

Recreate the diagram in Figure A.9.5 using the *Spacetime Diagram Tool*, and answer the following questions.

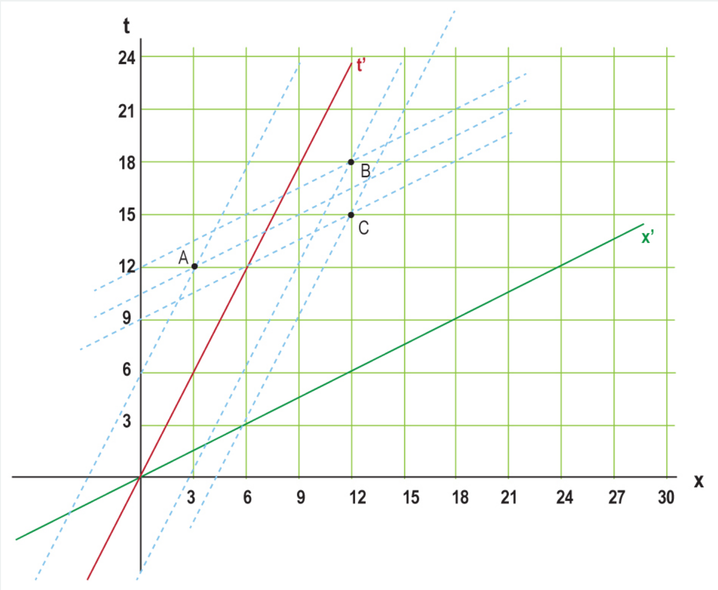


Figure A.9.5: Schematic shows events A, B, and C, with lines indicating how the coordinates of the three events are to be read in two different reference frames. To read the coordinates, the events must be plotted using the *Spacetime Diagram Tool*. Credit: NASA/SSU/Aurore Simonnet

Play Activity

In this section, we have introduced spacetime diagrams as a tool to help us understand the relationships between different events in spacetime and how observers in different frames view these events. In particular, the activities have illustrated that time and space coordinates of events can mix in unexpected ways when viewed from different frames of reference. For example, in the last activity, we saw that both the time and space parts of an event in one frame feed into the time part (or space part, for that matter) of the event as viewed in a different frame. That is why the time differences for the three events A, B, and C were not given by the simple time dilation equation. The next section puts these new tools and ideas to work to help us understand some of the weird results of special relativity.

This page titled 9.4: The Geometry of Special Relativity - Spacetime is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.5: Applications of Spacetime

Learning Objectives

- You will explore aspects of the feasibility of interstellar travel.
- You will understand that events that are simultaneous in one frame are not necessarily simultaneous in another.
- You will understand that events are real, that the same events happen in both frames, but that observers in different frames disagree about the spacetime coordinates of the events.

What Do You Think: Interstellar Travel

In this section, we will see several applications of the ideas introduced in the previous sections. These examples illustrate how our thinking must be modified when we are dealing with objects traveling close to the speed of light, such that the relativistic gamma factor differs appreciably from 1.

9.5.1: Interstellar Travel

Movies and on television often depict journeys between the stars. As we discussed in previous chapters, the stars are extraordinarily far away. Even the closest star to the Sun is about 4 light-years distant. We can use special relativity to investigate the practicality of such a journey and decide if the depictions of space travel from Hollywood are anything near realistic, and whether such trips are at all feasible.

Investigating Interstellar Space Travel

In this activity, you will get a chance to calculate some realistic situations involving space travel.

Questions:

VACATION TO SIRIUS?

In the distant future, when space travel is advanced, your descendants have 3 weeks of vacation time and want to visit the star Sirius, in honor of their favorite *Harry Potter* character. Sirius is 8.6 light-years away.

You might find the gamma graph helpful when thinking about how time is measured differently for observers moving at high relative velocities.

USE GRAPH

9.5.2: The Twin Paradox

One of the apparent inconsistencies in special relativity arises when thinking about extensions to interstellar trips, such as the ones described above. In the previous activity, we took an imaginary trip to Sirius at half light-speed. Imagine that after arriving at Sirius, we turn our spaceship around and return to Earth. This leg of the trip will be identical to the first because there is nothing special about the direction we travel. Due to time dilation, we will again measure a shorter time on our own wristwatch (15 years) than our friends back on Earth, who will again measure 17.2 years for our return trip. However, if the total time that they measure for the round trip is 34.4 years, and the time we measure is smaller (30 years), then we will be younger than we would have been had we foregone the trip and remained on Earth.

This scenario is often called the Twin Paradox because it can be devised in such a way that the space traveler has a twin sibling who remains on Earth. Upon returning to Earth, the traveling twin is younger than the twin who remained. But, that is not the paradox. The paradox is that, since the effects of special relativity are symmetric, both twins can equally claim that they themselves were at rest the whole time, while the other twin was in motion. Thus, they can both equally claim to be the older twin after the trip has been completed. However, this paradox, like most, is the result of not thinking carefully about the situation. Can you think of a reason why special relativity might not work in this case?

To complete a round trip from Earth to Sirius and back, it is necessary to violate one of the conditions of special relativity, namely, that all observations are made in reference frames that move uniformly. This means that neither frame may accelerate. However, that is not the case for the scenario described. For the traveling twin to reverse direction and return to Earth, the spaceship must slow down, stop, and then accelerate back up to half light-speed. We ignored the acceleration parts in our initial phrasing of the trip, but of course, this is a cheat. We cannot change our state of motion from rest to moving without undergoing some acceleration.

Nor can we complete a round trip to another star without accelerating, by virtue of the simple fact that we must turn around to return to our starting point. Once we accelerate, we have stepped out of the realm of special relativity, at least as far as we have been thinking about it. More than that, we have broken the symmetry that exists between nonaccelerating frames.

You most certainly have some experience with this sort of broken symmetry. Think of a time when you were riding in a car on a curving highway. You could feel yourself being pushed around as the car (hopefully) followed the highway. Similarly, if you have ever been forced to stop rapidly, you will recall feeling the car slowing beneath you as you continue moving forward toward the dashboard, perhaps held back by your seat belt. In both situations, it is quite clear that you do not "remain at rest" in the frame of the car. You are slammed back and forth, you fly forward or backward.

Contrast this scenario with the case of two cars moving down a straight, smooth freeway. One car travels at 60 mph, while the other travels past it at 65 mph. If the road is smooth enough, passengers in either car might have the illusion of standing still, while the other car glides slowly past.

The same is true for our imagined space traveler and Earthbound twin. The twin on Earth can claim to have always been in an unaccelerated frame; we will ignore gravity for the moment, and Earth's motion. The traveling twin can absolutely not claim to have been always at rest. This is because, at some point, the spaceship had to turn around; that twin had to change from one inertial frame to another. So, in this case, there is an absolute way to tell which of them was not in uniform motion the entire time, and the symmetry is broken.

The traveling twin will always turn out to have aged less than the stationary twin. We need general relativity to examine how clocks tick in accelerating frames. If we looked in detail we would see that the difference between the clocks can be made extremely large, arbitrarily so. In that case, the traveling twin might age only a few months, whereas on Earth, millions of years could have passed for the twin who stayed behind.

If you think carefully about the expression for the spacetime interval $s^2 = \Delta t^2 - \Delta x^2$ you can probably convince yourself that the time is always greatest for the stationary observer. What is Δx for an observer who is always at rest (an observer who is not moving!) with respect to some reference frame? How does this compare to Δx for an observer who is not at rest relative to that frame? What does this imply about the time part (Δt) of the interval for each observer?

9.5.3: Simultaneous, Says Who?

We will explore one more example of how relativity requires us to abandon our usual ideas of space and time. This also takes the form of a paradox, which is resolved only by considering carefully the situation we will illustrate.

Imagine that you are watching a train approach a barn in a rail yard. Like many barns, this one has a door at each end. Further, imagine that the train tracks lead directly through the barn, passing in one door and then out the opposite one. The illustration in Figure 9.14 shows the action as the train approaches the barn. We will assume for this problem that the barn is 100 meters long.

Figure 9.14: In the figure, a train is approaching a barn at a high rate of speed. The tracks pass through the barn, so in a few moments the train will enter the barn and then pass out the opposite side. Credit: NASA/SSU/Aurore Simonnet

Imagine that this is a special high-speed train that can move at speeds close to the speed of light. In this case, we can imagine that the train moves relative to an observer on the ground (and the barn) at 0.7c. Under these circumstances, we know that the train will appear to be shrunk along the direction of its motion by an amount given by the relativistic gamma factor, with $v = 0.7c$. The gamma factor in this case is 1.4 (check this number if you like). The length of the train as seen in the rest frame of the barn will be 1.4 times shorter than in its own frame.

Now further assume that the left door of the barn, the one nearest the oncoming train, is initially open and that the other door is closed. As the train passes into the barn, a rail-yard worker manages to close the left door just as the back of the train passes into the barn. We will call the moment when the back of the train is fully within the barn Event A.

At the *same time*, another worker opens the right-side door to let the train pass out just as the front of the train reaches the right side of the barn. We will call the act of the front of the train reaching the rightmost door Event B. The barn is apparently just long enough to contain the moving train, because for just an instant, the train is inside the barn with both doors closed, and the ends of the train coincide with the left and right of the barn. This situation is shown in Figure 9.15.

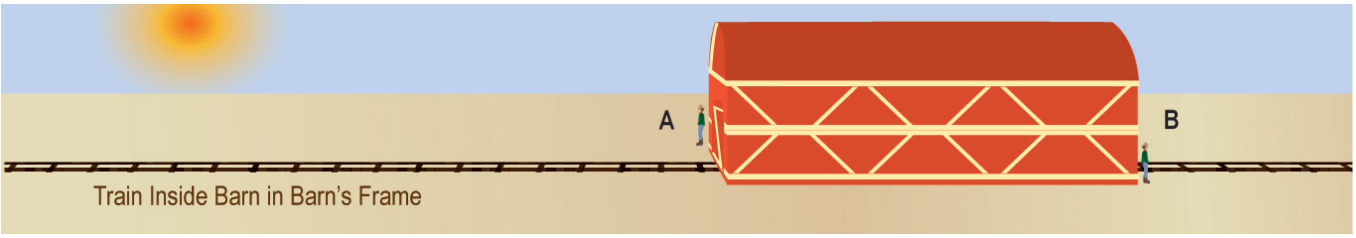


Figure 9.15: In the reference frame of the barn, the train appears to be instantaneously completely within the barn. This view shows the moment right after the rail-yard worker at the left has closed the barn door (Event A), and just before the worker on the right opens the door to let the train pass out (Event B). Credit: NASA/SSU/Aurore Simonnet

Now imagine these events as seen by the passengers on the train. Since the train appears to be 100 m long in the barn's frame, we can deduce that it must be 140 m long in its own frame; recall that it must be a factor of gamma longer. According to an observer in that reference frame, the train is not moving at all. Instead, the barn is rushing toward him at 0.7c. The barn will thus be shorter by a gamma factor of 1.4 than in its own frame. The barn's rest length is 100 m, so its length must be 100 m / 1.4 = 71 meters in the frame of the train. If this calculation is true, how can the train possibly be inside the barn with both doors closed? After all, in the train's frame, the train is 140 meters long, but the barn is only about 71 m long. Figure 9.16 summarizes the problem of a barn that is too short to contain the train from the viewpoint of the train.

Figure 9.16: In the reference frame of the train, the barn is much too short to contain the train. In fact, the train is about twice as long as the barn is deep. How can relativity reconcile these two points of view? Credit: NASA/SSU/Aurore Simonnet

Clearly, this is a paradox. In one frame, the train fits completely within the barn, while in the other frame, it cannot. But, both observers must agree on events. If an event happens in one frame, it must happen in all frames. What is the resolution to this puzzle? To answer that question, we can use the *Spacetime Diagram Tool*.

SIMULTANEITY AND THE TRAIN IN THE BARN

Play Activity

1. To solve the mystery of the train in the barn, reset the *Spacetime Diagram Tool* and create two events, one for Event A, when the back of the train coincides with the left side of the barn, and another for Event B, when the front of the train coincides with the right side of the barn.

- In the rest frame of the barn, these events occur at the same time, since the train fits entirely within the barn. Call this time t = 0. It is convenient to locate Event A at the origin.
- If you make each tick-mark of measure on the graph be 10 meters rather than one meter , then Event B will be located at the x = 10 point along the x-axis. (The axes only extend out to about 30, so it works better to rescale in this way.)

2. Create axes for the frame of the train, which is moving at 0.7 c.

https://phys.libretexts.org/@go/page/31339

From the activity, we see that the solution to the paradox is to realize that two events that appear to be simultaneous to one observer do not necessarily appear to be simultaneous to an observer in another frame. In our example, from the point of view of passengers on the train, the front of the train reaches the right side of the barn, and the rail-yard worker flings open the door (Event B), before the back of the train reaches the left side of the barn and the other worker slams shut that door (Event A). The resolution of the paradox comes in realizing that, in the train's rest frame, only after the right door has opened and the front of the train has moved out of the barn does the back of the train enter the barn, allowing the worker to close the left-side door of the barn (Figure 9.17).

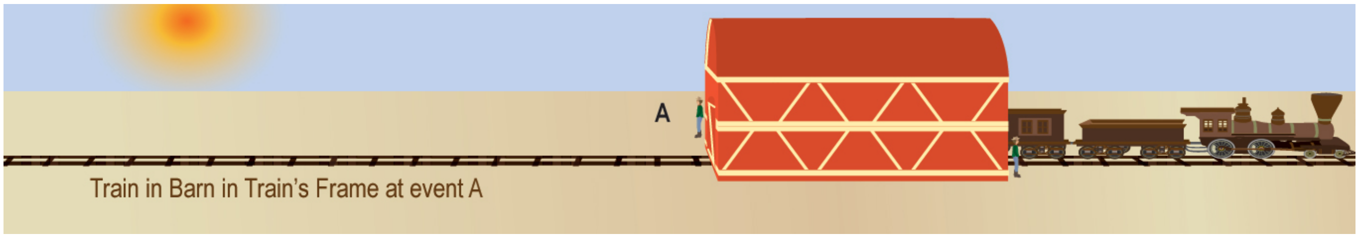


Figure 9.17: In the reference frame of the train, the back of the train has not reached the left of the barn when the front of the train reaches the right of the barn. Thus, event B occurs before event A in the reference frame of the train. Credit: NASA/SSU/Aurore Simonnet

The idea of absolute time, time that is shared by everyone everywhere regardless of their state of motion, is probably the most difficult bias that we must shed if we are to understand special relativity. We have a lifetime of experience built upon the apparent absolute nature of time. This experience is based on an illusion. It only works because we are not accustomed to dealing with objects that move close to the speed of light. The passage of time, the simultaneity of two events, the lengths of objects—all of these things are malleable, and our perceptions of them depend vitally on our state of motion.

This page titled 9.5: Applications of Spacetime is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.6: Mass and Energy

Learning Objectives

- You will be able to calculate rest energy and mass.
- You will understand that mass and energy are different aspects of the same underlying quantity, and that is the total of mass and energy that is truly conserved.
- You will know that a tiny amount of matter can release a great deal of energy.
- You will understand that it is possible to produce a pair of charged particles from an energetic photon, and that the annihilation of a particle-antiparticle pair releases energy.
- You will know that the total energy is greater than the rest energy if a particle is moving

What Do You Think: Mass and Energy

The last major topic in special relativity that we will discuss is that of mass–energy equivalence. This most famous of equations in physics was not even included in Einstein's original 1905 paper on special relativity. He did derive an expression for the increase of the mass of a particle as its velocity increased, suggesting that the mass of a particle should be multiplied by the relativistic gamma factor to obtain its true mass. Only later did Einstein derive his famous $E = mc^2$. But, what does this equation mean?

Quite simply, the mass–energy equivalence relation means just what it says: Mass and energy are equivalent and can be converted one into the other. Do not be distracted by the factor of c^2 in the equation; it is just a constant that depends on the system of units we are using. There is nothing special about a system of units; we just have to be consistent and use just one set of units when carrying out a calculation. The essence of the equation is that energy (E) is proportional to mass (m), and that the two can be interconverted, one to the other.

However, the amount of energy equivalent to given mass can be understood in terms of the units. The speed of light is a large number in standard SI units. When we square it we get an even larger number: $(3 \times 10^8 \text{ m/s})^2 = 9 \times 10^{16} \text{ m}^2/\text{s}^2$. So, converting a single kilogram of mass would release an enormous amount of energy, almost 10^{17} joules. That is enough energy to power a typical house, which uses a kilowatt or so, for several million years.

Generally, the actual reactions that happen in nature are not 100% efficient at converting mass to energy, so only a small fraction of mass is actually converted. For instance, in nuclear reactions like the ones converting hydrogen to helium in main sequence stars, just under 1% of the mass is converted to energy. Typical chemical reactions, like burning a match, are about a million times less efficient than that. Only when a particle meets its antiparticle, like when a positron meets an electron, is all of the mass converted to energy.

Of course, the mass–energy equivalence does not only allow for mass to be converted into energy. It also means that energy can be converted into mass. This is what happens in particle accelerators like the one at CERN near Geneva, Switzerland (Figure 9.18). In those machines, protons and antiprotons are accelerated to extremely high energies and speeds very close to the speed of light. When the particles collide inside the machines, their enormous energies are converted into the showers of particles that are then detected. It is a common misconception that the physicists are somehow “smashing atoms” so that they can see what is inside of them. The atoms in the collisions are certainly destroyed, but the particles that come out are not made of the insides of the destroyed atoms. They are made of the energy contained in the *motion* of those atoms (and from the particles’ masses that have been converted into some of the energy). The next few activities will help to illustrate this concept.

Figure 9.18: Aerial view of CERN, near Geneva, Switzerland. The large yellow ring outlines the location of the Large Hadron Collider (LHC). Credit: CERN/Maximilien Brice

REST ENERGY OF PROTONS AND NEUTRONS

Worked Example:

1. The mass of a proton is about 1.673×10^{-27} kg. Assuming the proton is at rest, how much energy does this correspond to in SI units (joules)?

The energy contained in this mass is given by Einstein’s famous equation. We will use a subscript zero to remind ourselves that the particle is assumed to be at rest.

- Given: $m = 1.673 \times 10^{-27}\ \mathrm{kg}, \mathrm{c} = 3 \times 10^{8}\ \mathrm{m/s}$
- Find: E_0
- Concept: $E_0 = mc^2$
- Solution: $E_0 = (1.673 \times 10^{-27}\ \mathrm{kg})(3 \times 10^{8}\ \mathrm{m/s})^2 = 1.506 \times 10^{-10}\ \mathrm{J}$

This seems like quite a small amount of energy. However, typical objects contain many, many protons. For instance, a typical human body has roughly 10^{28} protons in it (as well as about the same number of neutrons and electrons!). So, converting a typical object into pure energy would release tremendous amounts of energy. Converting all of the protons in a human body into energy would release $(10^{28})(1.506 \times 10^{-10}\ \mathrm{J}) = 1.506 \times 10^{18}\ \mathrm{J}$!

Questions:

In the activity above, we used the subscript zero for the energy when a particle is assumed to be at rest. This is called its rest energy because it is the amount of energy contained in the mass of a particle observed to be at rest. If the particle is not at rest, then its energy will be different. We explore this case in the next activity.

Total Energy of a Proton in the LHC

In this activity, we will explore how the rest energy of a proton in the Large Hadron Collider (LHC) at CERN compares to its total energy.

Protons and antiprotons are circulated at extremely high speeds in the LHC. The total energies of each of the protons and antiprotons accelerated are currently (in 2021) about 10^{-6} J.

https://phys.libretexts.org/@go/page/31340

From these two activities, we can see that most of the energy of the protons in the LHC is not in their rest energy (due to their mass), but must instead be in their motion. We will explore this idea in more detail.

For moving particles, we have to include the relativistic gamma factor when computing their energy. If an object is moving, its total energy is:

$$E = \gamma m c^2$$

or

$$E = \gamma E_0$$

It is this gamma factor that boosts the total energy above and beyond the rest energy value.

ENERGIES, VELOCITIES, AND THE GAMMA FACTOR

Worked Examples:

1. The ratio of an object's total energy to its rest energy is its gamma factor. (In other words, $E/E_0 = \gamma$, as follows from the discussion above.) If a proton were measured to have a total energy of 3×10^{-10} J, what would its gamma factor be?

- Given: $E = 3 \times 10^{-10}$ J, $E_0 = 1.506 \times 10^{-10}$ J
- Find γ
- Concept:$\gamma = E/E_0$
- Solution: \(\gamma = (3\times 10^{-10}\, {\rm J}) / (1.506\times 10^{-10}\, {\rm J}) = 1.99

2. What would the speed of this proton be in terms of the speed of light?

We can use the clickable gamma vs. v/c graph to see that v/c = 0.86. To see the math worked out using the equation for gamma, see Math Exploration 9.3.

USE GRAPH

To see the math worked out using the equation for gamma, see Math Exploration 9.3.

Math Exploration 9.3

Questions:

The most energetic cosmic rays ever detected have total energies around 10^{22} eV.

From the activity, you should have noticed that the velocities of the protons in the LHC and of the cosmic rays are nearly the speed of light. Notice how, at this point, increasing gamma cannot increase the speed of the particles significantly. What's more, no matter how big gamma gets, the speed will never exceed c. However, if the speed increases even by a tiny amount toward the speed of light, then gamma increases by a large amount, and therefore, so does the energy of the particles. It is the increase in gamma, not the increase in velocity directly, that is responsible for the gain in energy. To put this more directly, when a particle is moving at near the speed of light, a minuscule increase in speed results in an enormous increase in gamma, and therefore a correspondingly enormous increase in energy.

You have probably heard that special relativity prohibits particles from moving at or above the speed of light. The last examples and activities give part of the reason why. As a particle with mass approaches the speed of light, its gamma factor, and thus its energy, increase without bound. The particle energy, mostly due to its motion, must come from somewhere. In the LHC, the energy is provided by electricity that is used to create strong fields that bump the particles up in energy as they race around their circular track. We do not have access to boundless energy to put into these particles, we have only whatever power can be delivered by the power plants in Europe. Yet, as the speed of the particles nears the speed of light, their energies increase without limit, and for a massive particle to attain the speed of light we would require an infinite amount of energy to put into it. Clearly, this is more energy than we can ever generate, so there is never enough energy available to accelerate a particle to the speed of light. As a result, massive particles are forbidden from attaining this speed.

On the other hand, some particles have no mass. Photons are one example. The only reason photons have any energy at all is by virtue of the fact that they move. For massless particles like photons, they must *always* travel at the speed of light for any observer in any frame of reference, but that is what we assumed as a postulate of special relativity.

So, particles with nonzero mass cannot travel at or above the speed of light because they would need infinite energy to pass through the point where $v = c$. Could it be possible for particles to travel faster than light if they *always* traveled at such high speeds? In that case, they would never actually have the speed of light, where they would have infinite energy. If you look at the expression for gamma, you will see that if v is larger than c, then the argument of the radical in the denominator will be less than zero. This will cause gamma to be an imaginary number. So, particles that travel faster than light must have an imaginary mass and imaginary energy, whatever that means. Such particles have been suggested, and they are commonly called tachyons. No evidence for tachyons has ever been found, but they are one of the strange ideas that emerge from the theory of special relativity.

This page titled 9.6: Mass and Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.7: Faster Than Light?

Learning Objectives

- You will understand the fundamental reasons why particles cannot travel faster than the speed of light.

What Do You Think: Traveling Faster Than Light?

The discussion in Section 9.6 has, we hope, helped you to understand why no particles can be accelerated from our frame of reference to a speed equal to (or in excess of) the speed of light: Sufficient energy is not available to move any material particles to such a speed. However, other even more fundamental reasons explain why particles cannot exceed the speed of light. We will perform another thought-experiment to understand why this is so.

We are familiar with the idea of causality in our daily world. This is just the notion that some events are the causes of other events. For instance, you flip on a light switch and then the light begins to stream from the bulb; you push a button on your iPod, and the music begins playing; a glass of milk slips from your hand, and then it falls to the floor and shatters. Conversely, we never see milk collect itself from a puddle on the floor and arrange itself into a glass, as the glass itself spontaneously assembles from scattered shards.

These sorts of events never happen in the reverse order. There is nothing in the laws of physics, not the ones we have discussed so far, that would prevent these "backwards" events from happening. Yet, they never do. To reverse the order of causally connected events is easily seen to be completely absurd. With that in mind, we will examine the causal relationships between two events as observed in two different reference frames in uniform motion.

Causal Relationships

Imagine the following thought-experiment: We flip a switch, and a signal is conveyed to a bulb at a speed 5/3 the speed of light—nearly 70% faster than light could transmit the signal. We will compare this situation with signals moving at the speed of light and slower than the speed of light.

Play Activity

1. To begin, use the *Spacetime Diagram Tool* to plot several events.

- It will be easier to understand this example if the first event (call it A), flipping on the switch, occurs at the origin.

- Now plot an event, assuming that the signal from the switch travels to the bulb at 3/5 the speed of light. This signal can travel 3 light-seconds of distance in 5 seconds, so an event at x = 9, t = 15 can represent this case. Call this Event B.
- Next, plot an event to represent the case where the signal travels at the speed of light. The signal will travel 3 units of space in 3 units of time, so an event at x = 9, t = 9 can represent it. Call this Event C.
- The final event to plot should occur at a point with x = 15 and t = 9; at the speed v = 5/3c, the signal would travel 15 light-seconds of distance in 9 seconds of time. This is farther than light could travel in the same amount of time. We will call this Event D.

For an observer at rest, Event B occurs 15 seconds after the switch is flipped. The other two events occur only 9 seconds after. But, what about for a moving observer? Imagine that an observer comes flying past, moving in the x-direction, and passes you just as you flip the switch. Then Event A will also occur at the origin for the moving observer, and so the switch is flipped at t' = 0. We can explore when the other events occur.

In the last activity, you should have noticed that if you increase the speed of the observer enough, you can make the flip of the switch and the lighting up of the bulb simultaneous. A real light switch seems simultaneous to us because the delay is so short. But, in the case we are studying, the two events actually occur at exactly the same time. If you keep going even faster, you seem to be able to make the light turn on before the switch is flipped!

Another example might be even more instructive. Imagine that, rather than turning on a light, our switch connects to a valve for a container of air (Figure 9.19). Imagine also that the valve separates this container from a second container that is completely empty. In the rest frame of the containers, opening the valve will allow the air in the first container to flow into the empty container. Afterward, there will be air in both containers. (For dramatic effect, you can imagine that someone's life depends on getting the air into the second container.)

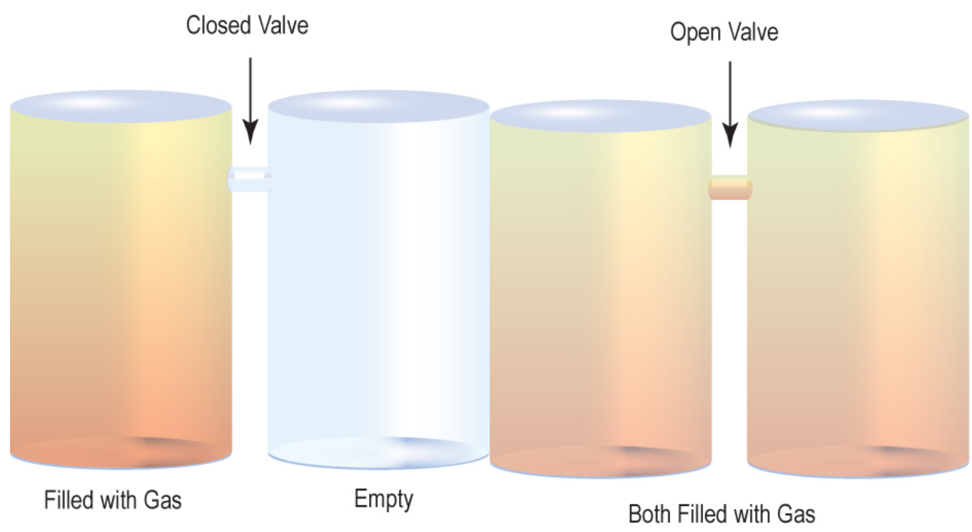


Figure 9.19: (a) One container is full of air and the other is empty. The containers are separated by a valve, which is closed to begin with. (b) After the valve is opened, air flows from the full container to the empty one until they contain the same amount of air. This process increases the entropy of the air, and so is extremely unlikely to spontaneously reverse. Credit: NASA/SSU/Aurore Simonnet

If this scenario played out as with the light bulb above, with the signal traveling from the switch to the valve at $5/3c$, then we could witness the following strange sequence of events: In a moving frame of reference, the air at first would be occupying both containers. At some point the air would spontaneously empty from one container into the other. After that, the valve separating the containers would close, leaving one container empty and the other one full.

We know from experience that this scenario is completely absurd. It is like watching a movie backward; it does not make sense. We never see the air in a room spontaneously flow into a tiny volume up in one corner. This tendency of gases or liquids to expand and occupy more volume, cover more of the floor, etc., rather than contract into a tiny volume is part of a more general trend in nature, the increase of disorder or more technically, the increase of entropy. In fact, so basic is this tendency that it has its own law, called the Second Law of Thermodynamics. This law states that the total entropy for any process must always increase. To learn more, read Going Further 9.4: Entropy, Disorder, and the Second Law of Thermodynamics.

So, the reason that objects or signals cannot travel faster than light is not strictly the result of special relativity. It is special relativity (the speed of light is constant in all frames of reference in uniform motion) along with the second law of thermodynamics (the total entropy must increase for any process). Sending signals faster than the speed of light would force us to abandon at least one of these basic principles. Since they are so well grounded in experimental results, it is much simpler to assume that no signals can travel faster than light. The fact that none has ever been seen to do so suggests that this is the correct choice, but like all things in science, it is subject to modification if contradictory information becomes available.

Going Further 9.4 Entropy, Disorder, and the Second Law of Thermodynamics

Understanding why entropy must always increase is not too difficult. There are many more ways to arrange items so that they are in disarray than there are to arrange them in a nice orderly way. So, given free reign, things tend to get disorderly. As an example, consider a deck of cards.

It is possible to arrange a deck of cards so that all the suits are together and the cards are all in order, from ace to king. But, the number of ways to do so is pretty small. We could arrange them such that the spades were on top, followed by the hearts, then the clubs, and then the diamonds. Or, we could start with the clubs, then have the diamonds, clubs, and spades, and so on. Since this system is so simple, it is not difficult to calculate how many ways we can arrange the deck such that all the suits are collected together and the cards are ordered. We have four suits, so there are four choices for the suit we want to use first. For each of those choices, there are three remaining choices of which suit to place second. So, that is $4 \times 3 = 12$ choices of which suits to put in first and second place. For each of those choices, we have two remaining choices for the third position because there are only two suits left to be placed, so we have $4 \times 3 \times 2 = 24$ possibilities for the placement of the first three suits.

The final suit does not give us any more choices. We have already placed three of the four suits, so only one remains. Thus, the total number of possible ways we can stack a deck of cards such that all the suits are collected and the individual suits are

https://phys.libretexts.org/@go/page/31341

ordered from ace to king is $4 \times 3 \times 2 \times 1 = 24$.

You can come up with other ways to order the cards, perhaps collecting all the aces together, and all the twos, threes, etc. You could go through a calculation similar to the one above to figure out how many ways the cards can be collected that way. Do you think you would have more ways, or fewer? If you like to play poker, then you might already be familiar with these sorts of calculations.

But, what if you want to arrange the cards in no particular order at all? How many ways are there to do that? Well, for the first card, we can choose any of the 52 in the deck (we are ignoring the jokers). For the second, we have one less, so there are 51 possibilities from which to choose. So, for the first two cards, we have $52 \times 51 = 2652$. We already have many more ways to place the first two cards than we had to place all four suits in the first example. How many choices do we have for placing a third card? Fifty, because that is how many cards are left to place. You can probably see the pattern here. To place all the cards, we will have a number of possibilities given by $52 \times 51 \times 50 \times 49 \times 48 \times \ldots \times 5 \times 4 \times 3 \times 2 \times 1$. We simply take one off the previous number of possibilities and multiply by that, repeating this pattern until we get down to one. Such a pattern, as you might already know, has a name. It is called the factorial function. We could shorten our calculation for the cards by saying that the number of ways of arranging the 52 cards in the deck is 52!, where the exclamation point means to take the factorial of 52. This is a huge number. It is almost 10^{68}.

Now imagine that instead of placing cards from a deck of cards, we are placing gas molecules in a box. Each has a position and a velocity. We can choose the x, y, and z components of the position and the x, y, and z components of the velocity independently. How many different ways are there to do that if the box contains a mole (6.02×20^{23}) of atoms? We will not compute it, but you probably see that the number is enormous, much larger than the ways we can place 52 cards.

As an example, we could imaging placing the molecules at random locations, with each of them moving in a random direction and with a speed taken from some random distribution of speeds. This would be a very disordered arrangement of molecules. How many ways are there to do that, do you imagine?

On the other hand, we could arrange the particles to be located on a regular grid, all with equal spacing between them. We could arrange for them to all move in the same direction and all have exactly the same speed. This would be a highly ordered arrangement. How many ways would there be to create a situation like this? It is still a lot given the large numbers or particles we are dealing with, but it is a much smaller number of possibilities than the previous case.

The analysis for molecules has to be slightly modified because, unlike cards in a deck, molecules are identical: every oxygen molecule is like every other oxygen molecule, every nitrogen is like every other nitrogen, and so on for carbon dioxide, etc. We have to take this notion into account, but it does not change the basic analysis. There are vastly more ways to arrange things such that they are disordered rather than ordered.

The second law of thermodynamics makes a very strong statement about this idea, saying that any change in the Universe always increases the total entropy of the Universe - the total amount of "disorderliness." Of course, physicists have a very precise mathematical expression to define what is meant by entropy, but we do not have to consider the details to that level. We should mention that an overall increase in “disorder” or “disarray” in the universe does not prohibit the growth of locally complex and orderly structures. It just means that the local increase in order is made up for, and then some, by disorder someplace else.

This page titled 9.7: Faster Than Light? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.8: Wrapping It Up 9 - A Trip to Alpha Centauri

Learning Objectives

- You will be able to put everything together to demonstrate your understanding of special relativity, using spacetime diagrams.

> *"The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality." —H. Minkowski, 1908*

Imagine you want to travel to Alpha Centauri, the nearest star to the Sun. The star is 4.2 light-years away, so it takes light from the Sun 4.2 years to reach Alpha Centauri, and vice versa. Figure A.9.6 shows the Sun and Alpha Centauri, along with some other distant stars.

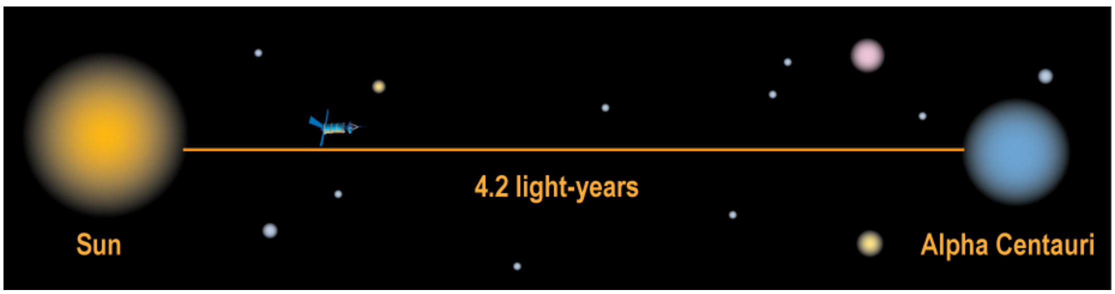


Figure A.9.6: The Sun and Alpha Centauri, with a few distant stars in the background. The distance between the two is 4.2 light-years. The digram is not to scale. Credit: NASA/SSU/Aurore Simonnet

Imagine that you have a spaceship that can travel at half of the speed of light. Traveling at that speed, it would require 8.4 years to reach Alpha Centauri. We will use the *Spacetime DiagramTool* to explore how you would experience the trip if you were a passenger on the ship. You will also need the clickable gamma graph.

Play Activity

USE GRAPH

9.8.1: Part I: The Spacetime Interval between Two Events

The first event is your departure from the Sun/Earth, and the second is your arrival at Alpha Centauri. Each of these events can be represented by a single point in a spacetime diagram.

1. Use the velocity slider bar to create a set of t'- and x'-axes for a frame traveling at 0.5c, the speed of your spaceship.

https://phys.libretexts.org/@go/page/31342

In your frame of reference, you should notice that the entire spacetime interval is made up of the time you measure for the star to arrive. In the reference frame of the Earth, there are both space and time components to the trip.

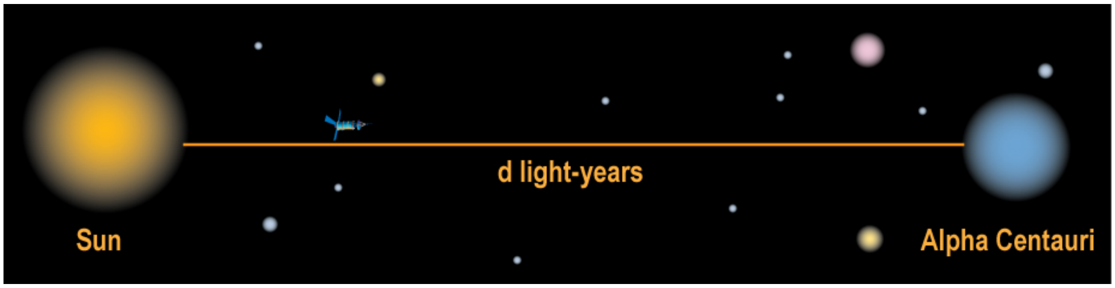


Figure A.9.7: The Sun and Alpha Centauri, with an unknown distance between the stars, due to the length contraction perceived by the traveler on the spaceship. Credit: NASA/SSU/Aurore Simonnet

9.8.2: Part II: Energy of Trip

We cannot make the trip described in this activity with our current technology. In this part, we will explore part of the reason for this limitation. We imagined that our spaceship could travel at half of the speed of light. The relativistic gamma factor for this speed was 1.15.

This page titled 9.8: Wrapping It Up 9 - A Trip to Alpha Centauri is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

9.9: Mission Report 9 - A Trip to Alpha Centauri

D. Questions to be graded for accuracy:

https://phys.libretexts.org/@go/page/31343


f(x)
s
a
b


f(x)
s
a
b

This page titled 9.9: Mission Report 9 - A Trip to Alpha Centauri is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

10: General Relativity

Chapter 10 presents Einstein's General Theory of Relativity, in which the force of gravity is a natural consequence of curved spacetime due to the presence of mass and energy. You will investigate properties of curved spacetime including time dilation, gravitational redshifts, and gravitational lensing. You will also build a mathematical construct for describing curvature and apply it to two- and three-dimensional models.

This page titled 10: General Relativity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.0: General Relativity Introduction

Video Transcript

Close your eyes, and imagine yourself in a comfortable room. Like a small recording studio, the room has no windows and is soundproofed so there is no way to see or hear the outside. There is an exit, however, and just before you are about to leave a voice over a loudspeaker says, "Be careful; watch your step. You are about to exit a moving vehicle."

How could you tell if you were in a moving vehicle, as opposed to in a building? You might wait to see if you feel yourself slow down, or if you feel the vehicle hitting potholes or going around curves in the road. But, if you are moving at a constant speed over a smooth, straight road, what could you do to determine if you were moving?

What if, instead of saying that you were in a moving vehicle, the voice over the loudspeaker said you were in a rocket in space accelerating at 1G? The acceleration due to the Earth's gravity is 1G as well. Are you really in space? How could you tell if the voice over the loudspeaker was telling the truth or not without leaving the room?

Albert Einstein had similar thoughts that led him to develop the general theory of relativity. But this theory doesn't just apply to hypothetical situations of windowless rooms on accelerating rockets; it is our most accurate theory of gravity and is important in our everyday existence.

If the general theory of relativity had not been developed our present-day world would be very different. For example, GPS systems rely on the results of general relativity. If GPS satellites used Newtonian gravity as the basis of their calculations, the positions produced by the satellites would be incorrect and they could not provide accurate directions. General relativity also provides a vital perspective for our understanding of objects like black holes and even the evolution of the Universe. Without it, our understanding of nature would be far less complete.

Imagine you exist in a small room, about the size of an elevator. The room has no windows or any other way to see outside. Now imagine that this room is inside a rocket ship, far out in space and away from any massive objects. The rocket is accelerating at 1g, or 9.8 m/s^2. If you did not know beforehand, is there any way that you would be able to determine that you were on a rocket ship and not, say, sitting in a windowless room in a building on Earth? Is there any sort of experiment that you could do that would distinguish these two cases for you? The answer to that question might surprise you, and it will lead us to a remarkable new view of gravity, just as it did Albert Einstein more than 100 years ago.

10.0: General Relativity Introduction is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

10.1: Einstein's Equivalence Principle

What Do You Think: Floating in Space

The thought experiment outlined above, with the windowless room on a rocket ship (or maybe not on a rocket ship!) is one of the examples Einstein used when thinking about extending his special theory of relativity. Recall that special relativity deals with unaccelerated motion. Its natural extension, therefore, is to try to describe what happens when motion (or frames of reference) are accelerated. Such a theory would be more general than one limited to uniform motion. It might therefore be called the general theory of relativity — and so it is.

Think about the room on the rocket ship a little more. If the rocket ship is not accelerating, and if it is in space far away from any sources of gravity, then we know that it will be in one of the so-called inertial frames, the kind described by special relativity. In inertial frames, objects set in motion obey Newton's first law; they move in a constant direction at a constant speed. So we could imagine floating in the room, with objects like pens, notebooks, and cups of tea floating alongside us. We could also push off of the floor of the ship and travel all the way to the ceiling at a constant speed, only stopping when we ran into the confining surface. We could then push off the ceiling to propel ourselves back to the floor— in such a situation does it even make sense to have a "floor" and "ceiling?" It would be more appropriate to talk about six walls. There would be no basis on which we could distinguish between any of them. Without an acceleration, any direction is as good as any other, so we could also bounce back and forth between any of the walls that we wished, as in Figure $10.1.1$.

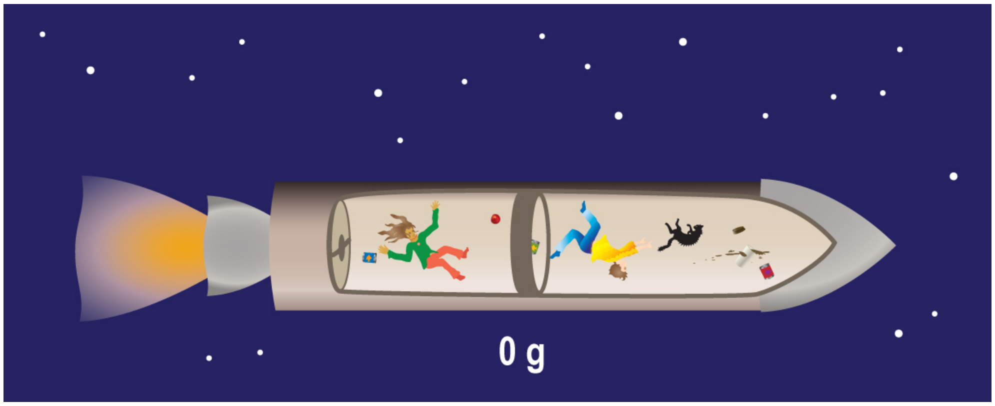


Figure 10.1.1: If we were in a region far from any massive objects, where gravity was essentially zero, we would be in a frame of reference described by special relativity. Credit: NASA/SSU/Aurore Simonnet

Now imagine the rocket engines are turned on and the rocket begins to accelerate forward. Since we and our fellow floating objects are not connected to the rocket ship, we do not move. However, we do see the walls of the room begin to move. In particular, we see the wall of the room toward the back of the rocket begin to accelerate toward us, while the opposite wall toward the front accelerates away. It makes sense to talk about a floor and a ceiling under these circumstances. In order to avoid a painful collision, we will imagine that the acceleration starts off very slowly, perhaps only 0.1g. Then, only after we (and the other objects on the ship) are on the floor does the rocket attain its full acceleration of 1g.

At this point we find ourselves pinned to the floor. If we jump up toward the ceiling, the rocket quickly accelerates to catch us, and we are stuck on the floor again. Furthermore, if we momentarily let go of something, such as a pen, it will hurtle down to the floor as well. Or does the floor race up to meet the object? Can we tell? That is the question that Einstein pondered. His answer was simple: no. We cannot tell if we are falling toward the floor of a windowless room on Earth, or if the floor is racing up to meet us as on an accelerating rocket. From our vantage point, within a small windowless room, both situations look exactly the same. There is no way for us to differentiate between them.

We can turn this situation around. Imagine that we are located in an elevator suspended by a cable inside a building. Just as we did not before realize that we were on a rocket, now we do not realize we are in an elevator. We only know that we are in a small room, and that we are pinned to the floor by gravity (or something we call gravity). Of course, if we let go of a cup of tea it drops to the floor, making a mess. This is all very familiar stuff.

But what if at the same instant that we drop our cup, some mischievous person cuts the cable supporting the elevator? Now we do not see the cup drop to the floor. Why? Because as the cup falls, the elevator itself falls at the same rate. Both accelerate toward the ground at 1g. In addition, we no longer feel ourselves pinned to the floor. We can push off the floor and go floating up to the ceiling if we like. In fact, we will find ourselves in a situation indistinguishable from that of the room on the rocket ship before the engines were turned on. This situation is called **free fall**.

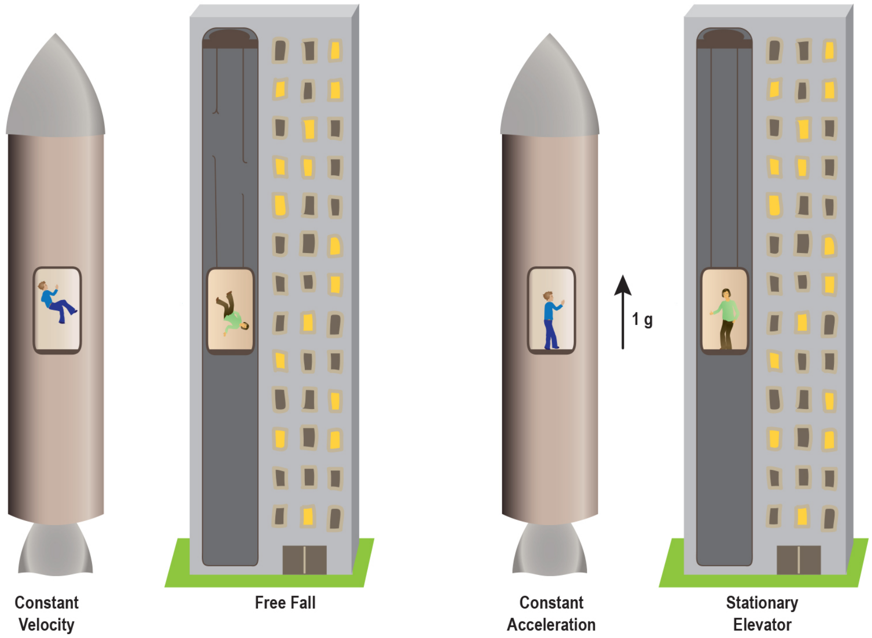


Figure 10.1.2: Accelerating in a rocket ship at 1g feels the same as being in Earth's gravitational acceleration. Credit: NASA/SSU/Aurore Simonnet

Because the case of an accelerated frame is indistinguishable, at least on small size scales, from that of the presence of a gravitational field, Einstein elevated the equivalence of the two to a principle, called the weak equivalence principle. It states, basically, that:

> *For small enough size scales, it is not possible to distinguish between the presence of a gravitational acceleration and an accelerated reference frame.*

In the statement of the weak equivalence principle we limit ourselves to small size scales— or "small enough." What does this mean? What happens when this condition is not met? To understand why we limit ourselves to small sizes, consider how gravity changes as we move through space. From Newtonian physics we know that the strength of the gravitational acceleration varies in proportion to the square of the distance to the source of the gravity. So in principle, the strength of Earth's gravity is a little bit larger at your feet than at your head because your head is a little bit farther from Earth's center than your feet are. Given that Earth has a radius of about 6400 km, the extra meter or two between your feet and head do not create a big change in the gravitational acceleration. It's so small that we can pretty much always ignore it. But in principle, if you jump off a high dive feet first, your feet should fall toward the ground slightly faster than your head does. (This has dramatic consequences when the gravitational acceleration is strong.)

The variation of the gravitational acceleration with distance is called the tidal effect. It causes the ocean tides on Earth, for example, because the variation in the gravity caused by the Moon (and Sun to a lesser extent) is noticeable across the distance of Earth's diameter. The Moon's gravity is a bit stronger on the side of Earth facing the Moon than it is on the opposite side, and this causes bulges in the oceans (and to a lesser extent on the ground as well) as Earth rotates. The tidal effect is always present when there is a gravitational acceleration, because the strength of the acceleration depends on how far we are from the source of the gravity. Given high enough precision, it can be measured.

Contrast this with an accelerating rocket ship. The floor of the room on the accelerating rocket will accelerate toward your feet and head no matter how tall you are; there is no tidal effect in an accelerating reference frame, not on any size scale. (There would be a time delay imposed by special relativity. It takes a little bit of time for the front of the ship to know that the back of the ship is being accelerated forward by the rocket engines, but that is not the same as the tidal effect of gravity.) The lack of a tidal effect can distinguish an accelerating reference frame from one in a gravitational acceleration. As a result, the weak equivalence principle holds whenever the precision of your measuring techniques is too low to discern a tidal effect. An equivalent statement of the weak equivalence principle is that:

> *For small enough size scales, the laws of physics for freely falling observers reduce to those of special relativity.*

This means that, for very small size scales, the effects of gravity (tides) become too small to notice, and the physics of unaccelerated frames becomes completely adequate for describing how things move on these small, local scales. How small this size is depends only on the precision of the measurements you can make.

You might be thinking that the existence of a weak equivalence principle implies the existence of a strong equivalence principle. If so, you are correct. The weak equivalence principle has to do with motions of objects under the influence of gravity alone, whereas the strong equivalence principle includes the other forces of nature, like electromagnetism and the nuclear forces. We will not worry about the strong equivalence principle, as it merely says that all the laws of physics must be the same for all observers, accelerated or not. This is a reasonable assumption to make for any description of the Universe, at least until such time as it is ever shown to be false.

This page titled 10.1: Einstein's Equivalence Principle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.2: Gravity and Curvature

? Curving Space and Time

10.2.1: Gravity Curves Space

In this section we will explore some of the consequences of the weak equivalence principle. Let us again consider a rocket ship far away from any source of gravity. Earlier, we imagined the motions of material objects. Now we will consider what happens when light passes through the rocket ship under these same conditions.

First imagine that a small window is opened in the side of the rocket, just enough to let in a beam of light through the side of the ship. We know that the light beam will travel in a straight line through the ship. Upon reaching the other side it will pass out of the ship, as long as we provide a strategically placed window that allows it to do so. The light will then continue in a straight line in its original direction.

We can imagine three different cases when light shines through the ship, and we consider them in turn. In each case, the light enters through one side of the ship and passes out the opposite side. We consider each case from two different perspectives. In the first, we consider the point of view of an occupant of the rocket ship. In the second, we look at the situation from the point of view of an observer outside the rocket who at rest with respect to the light source.

10.2.1.1: Case One: The ship is stationary with respect to the light source.

This is the case we just considered. The light will travel straight through the ship and pass out directly across from where it entered. Figure 10.3 shows how this might look to observers both on the outside of the ship and inside it.

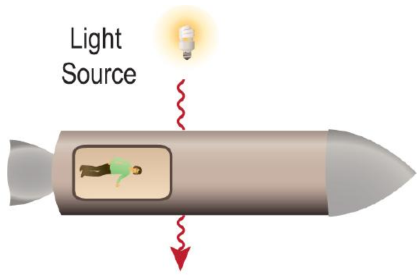


Figure 10.2.1: In the case that the ship and the light source are at rest with respect to one another, a light beam sent perpendicularly through a wall will travel straight across the ship, exiting directly opposite of the point it entered. Credit: NASA/SSU/Aurore Simonnet

10.2.1.2: Case Two: The ship moves at a constant speed, perpendicular to the direction of the light as seen by an observer at rest with respect to the light's source.

In this situation, the light will travel in a straight line, but it will exit the ship at a point displaced from its entry point because the ship moves some distance while the light is traveling across the ship. Figure 10.2.2 shows what this situation might look like to our two observers.

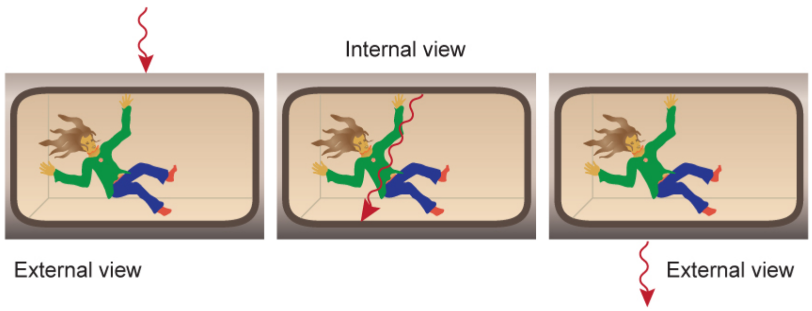


Figure 10.2.2: In the case that the ship and the light source move at a constant velocity with respect to one another, a light beam sent perpendicularly through a wall will travel at an angle as viewed by an occupant of the ship. It will exit at a point displaced in a direction opposite to that of the ship's motion. On the other hand, an outside observer thinks it is the ship that has been displaced, while the light itself has again traveled in a straight line. Credit: NASA/SSU/Aurore Simonnet

10.2.1.3: Case Three: The ship moves at a constant acceleration, perpendicular to the direction of the light according to the same observer in Case Two.

In this case the situation is markedly different for the two observers. Our outside observer still sees the light travel in a straight line as the ship accelerates forward. But the observer on the inside of the ship sees a very different path for the light. For that observer the light beam seems to curve toward the back of the ship as shown in Figure 10.2.3. In any given time interval, a second, say, the light travels a constant distance across the ship, but the ship moves an ever-increasing distance forward as its speed increases. As a result, the light makes a parabolic path relative to the ship and its occupants. Since light travels in straight lines in freely falling frames, it appears to an observer in an accelerated frame that "straight" lines are curved. We will return to this idea below.

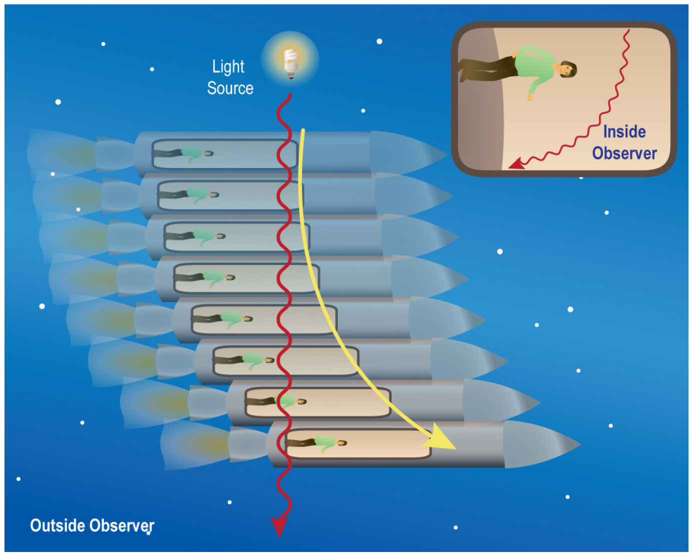


Figure 10.2.3: In this case, the ship moves at an ever increasing speed, while the light moves across it at a constant speed. The result is that the ship's occupant sees the light follow a parabolic path, not a straight one (shown as the red curve in the inset). To an outside observer though, the light still travels in a straight line (red line), while the ship moves ever faster in the forward direction (yellow curve). Credit: NASA/SSU/Aurore Simonnet

A Seagull Flies Across a Ship as Seen by Two Different Observers

Relativity does not require us to consider spaceships and light beams. We can use other analogies, too. For example, imagine a ship at sea that moves forward at a constant speed of 5 m/s. A seagull glides past at 1 m/s in a direction perpendicular to the motion of the ship, as shown in Figure A.10.1.

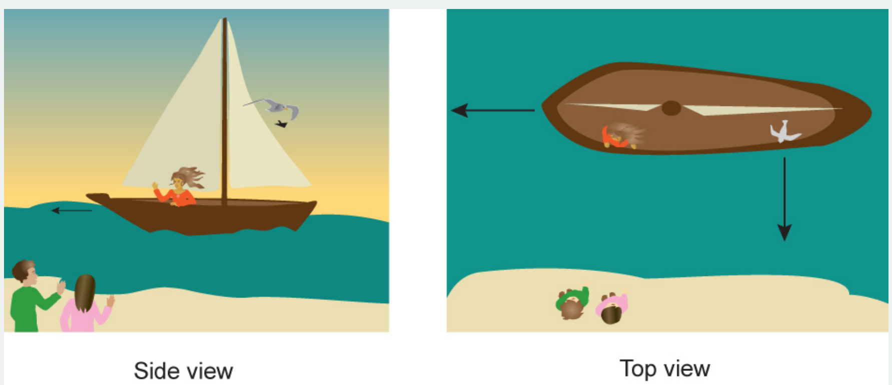


Figure 10.2.4: A seagull flies over a ship moving at a constant velocity. An observer on the ship is watching the seagull. A second observer watches from the shore. The panel on the left shows the view from the side, and the panel on the right shows the view from above. Credit: NASA/SSU/Aurore Simonnet

1.

2.

3.

Accelerating Ship and a Seagull

Consider the ship in the previous activity, but now imagine that it starts from rest and then accelerates at a constant rate of 1 m/s/s (Figure A.10.2).

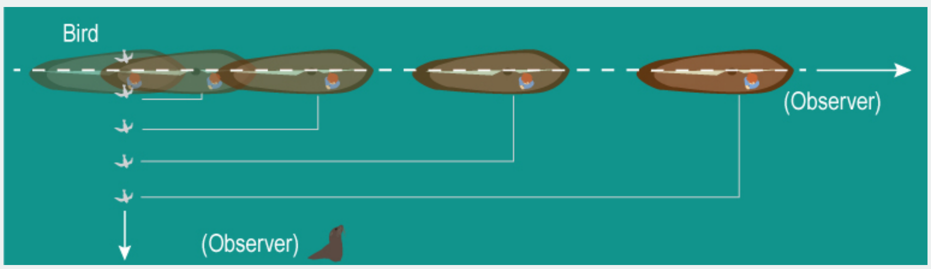


Figure 10.2.5: A seagull flies over a ship that is accelerating. An observer on the ship is watching the seagull. A second observer watches from nearby the shore. Credit: NASA/SSU/Aurore Simonnet

Worked Examples:

1. If the same seagull glides past at the same speed, 1 m/s, how much time will it take to fly across the boat?

Since the seagull is still flying directly from one side of the boat to the other, it must traverse 10 meters to get all the way across. It therefore will take 10 seconds for the bird to cross the boat, just as before.

2. Why can we still ignore the motion of the boat when we are finding the time required for the bird to cross the boat?

We can ignore the motion of the boat because the seagull is still flying directly from one side of the boat to the other. Therefore it must still traverse 10 meters to get all the way across.

3. In Math Exploration 10.1, we show how to calculate the distance the accelerating boat travels.

Math Exploration 10.1

Questions

1.

2.

3.

4.

Going Further 10.1: Accelerated Motion

We have seen that the distance traveled by an object moving with a constant velocity is given by the following equation.

$$d = vt$$

We cannot generally use this expression if the velocity is changing. However, we can use a similar relation if we know the average speed a particle moves. For example, if you traveled 30 km/hr for half an hour, and then you sped up to 90 km/hr for an additional half hour, you would have traveled a total of 60 km in one hour:

$$(30\ \mathrm{km/hr})(0.5\ \mathrm{hr}) + (90\ \mathrm{km/hr})(0.5\ \mathrm{hr}) = 15\mathrm{km} + 45\mathrm{km} = 60\mathrm{km}$$

Since you completed the trip in one hour, and you traveled a total distance of 60 km, your average speed must have been 60 km/hr, though you did not spend any of the trip actually traveling 60 km/hr. In general, if we know the average speed of an object, then we can use the following relation to find how far it moves in a time t:

$$d = \bar{v}t$$

The bar over the *v* means that we are using the average speed for a trip. In effect, this is one definition of what we mean by average speed.

We can rewrite the previous numerical example to show explicitly how the distance for that trip is related to the average speed for the trip. We start as we did before, then rearrange terms slightly by writing 0.5 hours as (1 hr/2):

$$(30\,\mathrm{km/hr})(0.5\,\mathrm{hr}) + (90\,\mathrm{km/hr})(0.5\,\mathrm{hr}) = \left(\frac{30\,\mathrm{km/hr} + 90\,\mathrm{km/hr}}{2}\right)(1\,\mathrm{hr}) = 60\,\mathrm{km}$$

The term in the big parentheses is by definition the average speed for the trip in our example. Multiplying by the total time for the trip gives us the distance traveled during the trip.

This example was simple because there were only two relevant speeds during the trip. However, even if the speed constantly changes we can use this notion of average speed to deduce how far an object moves in a given time. If the object accelerates at a constant rate, then we can find a general expression for the distance it travels, as we now show.

First, imagine an object that starts from rest and accelerates at a constant rate. We do not have to start from rest, but it makes the argument a little simpler to follow; adding an initial speed would shift the distance up or down, but it would not change the outline of the argument we present.

For the sake of argument, imagine that the object accelerates at 1 m/s/s, for 10 seconds. How fast will this object be moving at the end of the 10 seconds? Well, after one second it is moving 1 m/s, after 2 seconds it is moving 2 m/s, and so on. After 10 seconds it will be moving 10 m/s. If we plot its speed against time, it will look like the graph below.

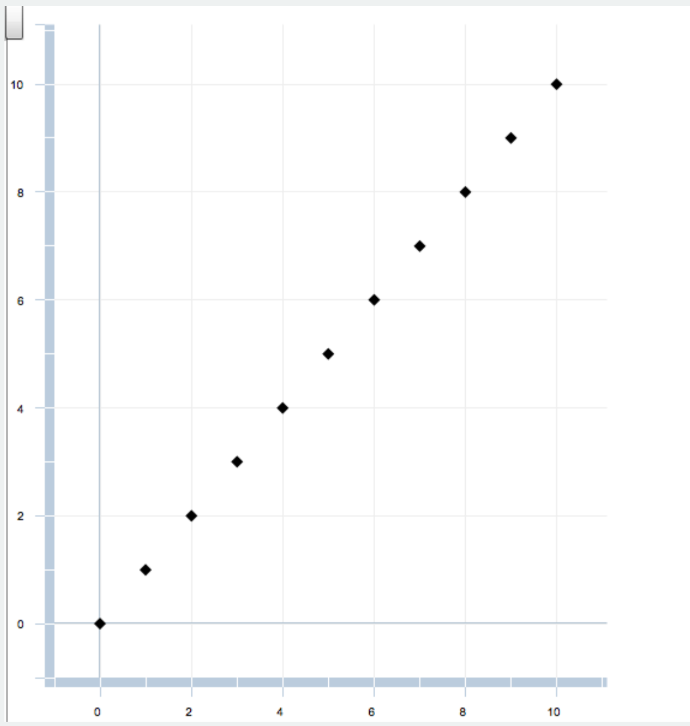


Figure 10.2.6: Graph shows velocity (vertical) vs. time (horizontal) for an object with constant acceleration.

The average speed for the trip is right in the middle: 5 m/s. Notice how the object spends as much time above its average speed as below it. This is because the speed of the particle is symmetric about its average. Under these circumstances the average can be computed simply from the mean of the initial and final speeds. Check that this formula works for the example above.

$$\bar{v} = \frac{v_0 + v_f}{2}$$

But the final speed (v_f) is just the initial speed (v_0), plus whatever change in speed is due to the object's acceleration, $v_f = v_0 + at$. So we can rewrite the last expression as follows.

$$\bar{v} = \frac{v_0 + (v_0 + at)}{2}$$

or

$$\bar{v} = \frac{v_0 + v_0 + at}{2}$$

We can simplify this expression slightly.

https://phys.libretexts.org/@go/page/31348

$$\bar{v} = v_0 + \frac{1}{2}at$$

Now we can substitute this into our earlier expression for the distance in terms of the average speed.

$$d = \left(v_0 + \frac{1}{2}at\right)t$$

This leads to the following expression for the distance traveled by an object that accelerates at a constant rate.

$$d = v_0 t + \frac{1}{2}at^2$$

We can use this expression to find how far an object moves when it starts from some initial speed and accelerates at a constant rate. An example would be an object falling under the influence of gravity near Earth's surface. Recall that the gravitational acceleration is essentially constant if we remain near the surface.

In the preceding examples we saw how the path of light was dramatically different for observers in an accelerated frame than for those in a frame moving at constant velocity. According to the equivalence principle, there should not be a difference between what is seen in an accelerated frame and what is seen in the presence of a gravitational acceleration. So does that mean that if we view light moving where there is a gravitational acceleration we should see it bend? According to the equivalence principle, the answer is "yes." This is the sort of thought experiment that Einstein used to come to the realization that space must be curved. To oversimplify for the moment, light, by definition, follows a "straight path," but in the presence of gravity, that "straight path" must follow the curvature created by the gravitational acceleration. But we do not observe light curving due to Earth's gravitational acceleration. Is there a contradiction here between the equivalence principle and reality? The next activity addresses this apparent contradiction.

Light Deflection

Consider the ship at sea one more time. Now replace the gliding seagull with a speeding photon, as in Figure 10.2.7.

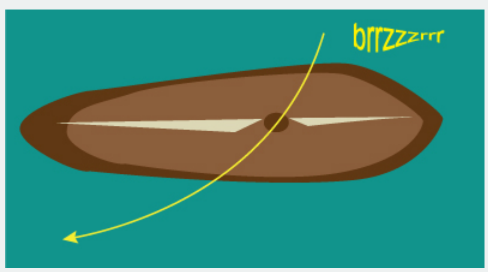


Figure 10.2.7: A photon passes over a ship that is accelerating, and is seen to be deflected by 1 mm. Drawing is not to scale. Credit: NASA/SSU/Aurore Simonnet

1.

2.

3.

The idea of curved space seems strange, but if you take a trip from New York to Tokyo, how do you go? The straightest path between the two cities must follow the curvature of Earth's surface. It is not a "straight line" in the usual sense. Even if we do not think about it very much, we are already somewhat familiar with curved spaces because we happen to live on one: Earth's surface! And just as in general relativity, the curvature is only apparent when we travel long distances. On small scales the surface of Earth seems flat. We might create an "equivalence principle" for Earth's surface that says the world is flat on small scales. Only on larger scales does its curvature become apparent. We will explore these ideas in more detail below, but first we will look at how gravity affects time.

10.2.2: Gravity Curves Time

To understand the effect of gravity on the time part of spacetime, we can again consider the motion of light in a rocket ship, as in our previous example. For this example, however, we imagine what happens if we shine the light beam parallel to the ship's motion. For instance, imagine that we shine light forward from a source at the back of the ship. If an observer at the back measures the frequency of the light to be f_0, what frequency will a different observer, placed at the front of the ship, measure? Will it be different from, or the same as, the frequency measured by the observer at the back? There are three possible outcomes: either the frequency at the front is larger, smaller, or unchanged. We can reason our way to the answer.

In the case of no acceleration, the answer is straightforward: we know that the velocity of the source and observer are the same at all times. Since they are both attached to the same rocket ship their relative velocity is zero. The critical aspect of this case is that we know the speed of both sources remains unchanged from the moment the light is emitted to the moment it is detected. Therefore, the frequency of the light is the same at both the source and the detector. It does not matter if the rocket is "moving" or "stationary" according to an outside observer: the observers on the ship are not moving relative to one another.

Now imagine what happens if we do the same experiment on a ship that accelerates in the forward direction (Figure $10.2.8$). At the instant the light leaves the source, the ship has a certain velocity, call it v_0. However, in the time required for the light to travel from the source to the detector the ship speeds up. Its final velocity, v, is larger than the velocity it had when the light was emitted. Note the difference between this case and the previous one. If the rocket does not accelerate, then there is a single reference frame in which both source and detector are at rest at all times. If the rocket accelerates, then there is no single reference frame in which both are always at rest. At the instant the light is emitted there is a reference frame in which both are at rest. Then, when the light is detected there is a different reference frame in which they are at rest. The two reference frames are different because the rocket has increased its speed in the short time required by the light to travel from source to detector. It is as if the source and detector have different velocities, and this is because of the time delay caused by the finite speed of light.

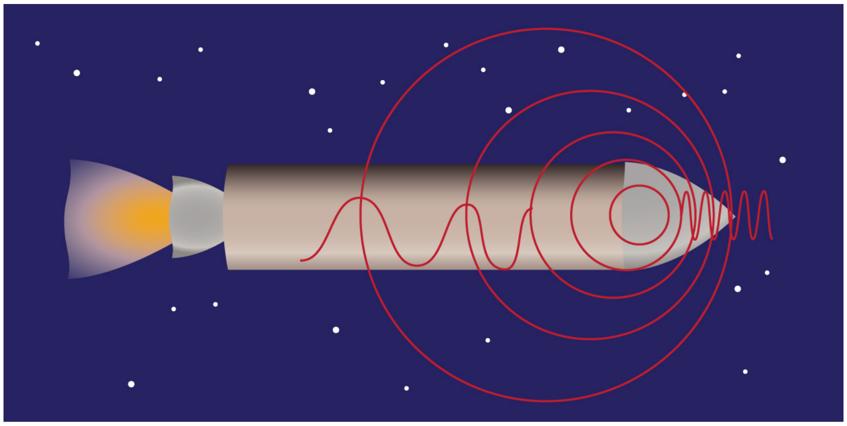


Figure 10.2.8: On a ship that is accelerating, the velocity of the ship at the time the light is sent is less than the velocity of the ship at the time the light is received. This causes a shift in the frequency of the light. Credit: NASA/SSU/Aurore Simonnet

From the Doppler shift formula we know that if a source of light and a detector of light have different velocities, then there will be a shift in the frequency of the light as measured at the source and detector. In this case, since the detector is accelerating away from the point where the light was emitted, the frequency will be lower (there will be a redshift) than it was at the source— which is also accelerating away from the point of emission, of course. This is the critical difference between the accelerating rocket and the constant velocity rocket. The following example shows the size of the shift when the acceleration is small.

In terms of frequency, the Doppler relation is described by the following equation.

$$f = f_0\left(1 - \frac{v}{c}\right)$$

The frequency f is that measured by an observer moving at a speed v relative to the source of the light, and f_0 is the frequency of the light measured in the frame of the source. For positive velocities (when we are moving away from the source, inducing a redshift) the frequency decreases. For negative velocities (when we move toward the source, inducing a blueshift) the frequency increases, just as we expect.

The difference in velocity of the rocket between the time the light is emitted at the source and when it is detected is due only to the acceleration of the rocket over the time the light is in transit. If we call the light travel time T, and the acceleration a, then the velocity of the rocket at the time the light is detected can be written as follows.

$$v = aT$$

We are assuming that the acceleration is low, and we further assume that the length of the rocket is not very big. Under these assumptions the rocket does not move very far while the light travels,. Then the time required for the light to get from the back of the ship to the front is the rocket's length divided by the speed of light. If the length/height of the rocket is \(*H*\) we have:

$$T = \frac{H}{c}$$

Combining this with the previous two relations we get:

$$f = f_0\left(1 - \frac{aH}{c^2}\right)$$

This is the shift in frequency we expect for light if it is on an accelerating rocket ship. (For more information, see Going Further 10.2: Another Look at Gravitational Redshift.) But remember, according to the equivalence principle we should see the same shift wherever there is a gravitational acceleration. That is, if we shine a beam of light directly upward against the acceleration of gravity, just as in the last example where the light traveled against the acceleration, there should be a shift in its wavelength as it travels upward.

As an example, on Earth's surface where the acceleration is 1g ($9.8 m/s^2$), there should be a small shift in the frequency of a beam of light that is directed upward as shown in Figure 10.2.9.

Figure 10.2.9: Because of the acceleration due to gravity, there is a small shift in the frequency if we shine a light upward from Earth’s surface (blue dot). The effect is exaggerated in the illustration. Credit: NASA/SSU/Aurore Simonnet

The Effects of a Gravitational Acceleration on Light

Imagine that you are shooting a laser beam up away from Earth.

1.

2.

Now imagine that you are on an orbiting spaceship and were shooting the laser beam down toward the Earth.

3.

4.

https://phys.libretexts.org/@go/page/31348

The next activity illustrates why we generally do not notice these shifts.

Gravitational Redshift and Blueshift

Worked Example:

1. Imagine that we have a tower 100 meters high, and that we shine a green laser pointer from the bottom to the top of the tower. If the wavelength of the laser light is 420 nm at the bottom of the tower, what will its wavelength be upon reaching the top?

- Given: λ = 420e-9 m, a = 9.8 m/s^2, H = 100 m
- Find: The shifted frequency , f
- Concepts: $f_0 = c/\lambda$, $f = f_0(1 - aH/c^2)$
- Solution: First we need to find the frequency associated with the emitted wavelength: $f_0 = c/\lambda$ = (3E8 m/s) / (420E-9 m) = 7E14 Hz
- Now we can find the new frequency:

$$f = (7E14\ Hz)\ [(1 - (9.8 m/s^2)\ (100\ m)/(3E8\ m/s)^{2]}$$

$$= (7E14\ Hz)\ (\ 1 - 1.1E\text{-}14\ )$$

$$= (7E14\ Hz)$$

The shift in frequency is far too small to be noticeable by an observer in the weak gravity of Earth. It is only about a part in 100 trillion! However, with a stronger gravitational acceleration the effect does become noticeable. We will consider those cases later on.

Questions

In the worked example we considered the shift of frequency for a green laser pointer that is fired from the ground to the top of a 100 meter tall tower. Now imagine that we have a red laser pointer with wavelength 650 nm at the top of the tower. We fire straight down toward the ground.

1.

2.

3.

4.

5.

Going Further 10.2: Another Look at Gravitational Redshift

In the examples so far we have assumed that the equivalence principle holds, and so we have computed the shift in wavelength of light in a gravitational acceleration based upon what we expect on an accelerating rocket ship. We will now find what the shift will be for a photon on Earth if we fire it from the bottom of a tower to the top. We will again use the equivalence principle to accomplish this.

One way to state the equivalence principle is that an accelerated frame is indistinguishable from a frame where there is a gravitational acceleration, at least for small scales. Another way to state it is as follows: On small scales, freely falling frames are equivalent to the inertial frames of special relativity. So if we are in a frame that is freely falling in a region with a gravitational acceleration, then we expect all our measurements to be the same as we would have in an inertial frame. We will now follow this reasoning through.

Imagine we have a colleague who directs a laser beam upward from the bottom of a tower to the top, where we are located. Further imagine that our colleague measures some wavelength for the beam, call it λ_0 — it might be 420 nm, like in the worked example from the activity. What wavelength will we observe at the top of the tower? To answer this question, we make a rash decision and step off the tower platform at the instant the beam is fired upward.

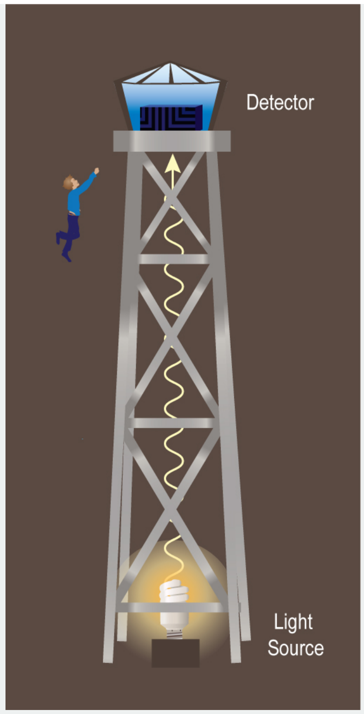


Figure 10.2.10: Light is emitted at the base of a tower and travels upward as the observer steps off the top of the tower and freely falls. Credit: NASA/SSU/Aurore Simonnet

According to the equivalence principle, once we leave the support of the top of the tower and begin to fall, we are in an inertial frame because we are only under the influence of gravity. The light is also freely falling, influenced only by gravity. We neglect the effects of the air, as usual.

For a moment imagine another observer, next to us but traveling upward at a speed v. From the point of view of such an observer, the wavelength of the light would be increased by an amount given by the observer's speed multiplied by the time required for one wavelength to go by the observer. We will call this time T; it is the period of the wave. To understand why this happens, you must realize that the observer would move slightly, in the same direction as the light wave, as the wave moves rapidly upward.

So for the tiny shift in wavelength, we may write:

$$\Delta\lambda = vT$$

We can rewrite the time period, T, in terms of the frequency, f, of the light using the relation $T = 1/f$, and then substitute for T.

$$\Delta\lambda = \frac{v}{f}$$

We can also rewrite the shift in wavelength using the relation $\Delta\lambda = \lambda - \lambda_0$, where λ_0 is the unshifted wavelength of the light (the wavelength in the frame of the source) and λ is the wavelength we measure. Our expressions is now the following.

$$\lambda - \lambda_0 = \frac{v}{f}$$

And now, since we know that the product of wavelength and frequency is the speed of the wave, or (λ = c/f), we can combine.

$$\frac{c}{f} - \frac{c}{f_0} = \frac{v}{f}$$

Rearranging this leads to the expression below.

$$\frac{f}{f} - \frac{f}{f_o} = \frac{v}{c}$$

This equation simplifies as follows.

$$1-\frac{f}{f_o}=\frac{v}{c}$$

Now we can solve this for the frequency, f, measured by the observer.

$$\frac{f}{f_o}=1-\frac{v}{c}$$

Or...

$$f=f_0\left(1-\frac{v}{c}\right)$$

So an observer moving upward at the instant we step off the tower would measure a shift of this size for the light's wavelength. Of course we are not moving upward; we have just stepped off the tower and are beginning to fall to its base. However, if a friend of ours remains on the tower, as we begin to fall, it looks to us as though our friend is moving upward. Since we are in free fall, and therefore in the inertial frame of the also freely falling light, it must also look to the light as if our friend is moving upward relative to its own frame of reference. And just as in the case of the rocket, the velocity we attain downward is very small in the short time needed for the light to reach the top of the tower: this means that the frame of the top of the tower is only moving slowly relative to the free-falling frame of the light.

We can calculate our speed when both the acceleration due to gravity (g) and the height of the tower (H) are "small," and thus determine the speed of our friend's reference frame. We realize that the time we have been falling is just the height of the tower divided by the speed of light: t = H/c. Our speed is therefore v = gt = gH/c. This gives us the following expression for this new frequency.

$$f=f_0\left(1-\frac{gH}{c^2}\right)$$

We have found exactly the expression we got before for the accelerating rocket. In this example we explicitly used the symbol g for Earth's gravity. Of course it is valid not only on Earth, but on any other objects for which our assumption of low gravity and small *H* are valid.

The redshift relationship we have just found for frequency is known as *gravitational redshift*. It has an important implication: it does not only relate to light; it is true for any periodic phenomenon. For instance, imagine we make a clock that keeps time by measuring the oscillation of a specific frequency of light, ticking once for each wave that goes by. An observer at the top of a tower will see the hands on a clock sitting on the ground moving more slowly than an identical local clock that uses a local light source. This is because light on the ground oscillates more slowly from the observer's vantage point, or equivalently, *time runs more slowly where gravity is stronger*. From this result we can deduce that an observer in Earth orbit will notice that clocks on the ground and on the tower move more slowly than his local clock. The effect is called gravitational time dilation. It is one of the important predictions/tests of general relativity. Written in terms of time, the relationship is the following.

$$t=\frac{t_0}{\left(1-\frac{gH}{c^2}\right)}$$

Our derivation is only appropriate for very weak gravitational accelerations, like that of Earth. And we assumed that the tower was not very big. In that case the shift in frequency/wavelength of the light is not very big. A more general derivation would give a result that could also be used in strong fields, like those around black holes and neutron stars. Such a derivation is beyond the mathematical techniques appropriate for these modules. It requires the full mathematical apparatus of general relativity. However, derivation given here gives us a flavor for the physical reality underlying some of the more bizarre-seeming effects of general relativity. We will use the results of the full calculation in the rest of the modules.

As an aside, it is good to realize that popular devices like GPS navigation tools make use of effects from general relativity. GPS relies on signals from satellites in Earth orbit. Without taking account of general (and special) relativistic effects, the precision attainable by GPS is quite noticeably degraded. In the next activity, we will explore the time dilation effects that must be taken into account for a GPS.

Time Dilation and GPS

Worked Examples:

1. GPS satellites are in orbit at about 36,000 km above the Earth's surface. Due to general relativistic effects, clocks at a distance above the Earth run faster than those closer to the Earth. The fractional change in the time is given by $(t – t_0)/t$. For a clock on a GPS satellite, what is the fractional change in time?

- Given: $g = 9.8\ m/s^2$, $H = 3.6E4\ km = 3.6E7m$
- Find: The fractional change in time, $(t – t_0)/t$
- Concepts: $t = t_0/(1 – gH/c^2)$
- Solution: First let's rewrite the expression for time dilation in terms of the fractional time:
  $t = t_0/(1 – gH/c^2) \rightarrow t_0/t = (1 – gH/c^2) \rightarrow 1- t_0/t = gH/c^2 \rightarrow (t - t_0)/t = gH/c^2$
  Plugging in the numbers for the GPS satellite:
  $(t - t_0)/t = (9.8m/s^2)\ (3.6E7\ m) / (3E8\ m/s)^2 = 3.9E-9$

2. By how many seconds will the clock on a GPS satellite lead an identical clock on the ground after one day in orbit?

The fractional change affects each interval of time, so we multiply the fractional change by the duration, in this case, a day. To get the answer in seconds, we also need to convert by the number of seconds in a day. The final answer is:

(3.9E-9)(1 ~~day~~)(24 ~~hours~~/~~day~~)(3600 seconds/~~hour~~) = 3.4E-4 seconds

To see a comparison of special and general relativistic effects, see Math Exploration 10.2.

3. Due to special relativistic effects, the fractional change in the time $(t – t_0)/ t_0$ for a GPS satellite is about 1 part in ten billion (1×10^{-10}). How much smaller is this than the general relativistic effect?

- To compare, take a ratio:
- GR time dilation/SR time dilation = 3.9e-9/1E-10 = 39
- The effect due to special relativity is nearly 40 times smaller, or the effect of general relativity is about 40 bigger, for a GPS at this distance in Earth's gravitational acceleration.

Questions

1.

2.

This page titled 10.2: Gravity and Curvature is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.3: What is Curvature?

What Do You Think: Flatness and Curvature

10.3.1: Curvature in 2D Space

We have seen that light should follow a curved path when there is a gravitational acceleration. We have also seen that the rate time passes depends upon the strength of gravity. You should be getting the idea that gravity does some unexpected things to both space and time. These effects are somewhat reminiscent of the ways that space and time are distorted in special relativity. However, once the relative motions are chosen in special relativity, the distortions are the same everywhere, and they never change. The distortion of spacetime is more complicated in general relativity. We say that spacetime is flat and static in special relativity. In contrast, spacetime is curved and dynamic in general relativity.

What is meant by flat and curved? In this section you will explore the meaning of these words in some detail in order to help you better understand what Einstein's view of gravity is really all about.

You are accustomed to "flat" and "curved" objects in the world. You might think of a soccer field as being flat, whereas a soccer ball is curved. While the terms "flat" and "curved" have a specific scientific and mathematical meaning, the everyday view of curvature is useful in helping to understand curvature of spacetime. In fact, the description of curvature in higher-dimensional spaces (even those in which one of the dimensions is time!) is based entirely upon our (mathematical) description of curvature in the normal space in which we live.

Think for a moment about a soccer field and a soccer ball. What do we mean when we say that one of them is flat and that the other is curved? There are many technical aspects involved in the answer to this question, but fundamentally, the best way to think about curvature is that it has to do with the way we measure the distance between two points in a space. So, for instance, on the soccer field we know that the distance between two points will be given by the Pythagorean theorem.

$$d^2 = \Delta x^2 + \Delta y^2$$

The letter d represents the straight-line distance between the points, and Δx and Δy are the x and y separations between them on a coordinate grid. (Figure $10.3.1$). We will illustrate what we mean here with some examples in the next activity.

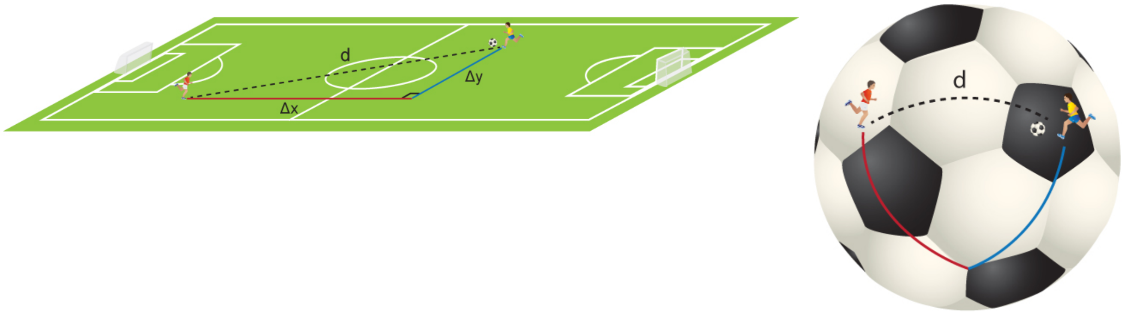


Figure 10.3.1: Left panel: On the surface of a soccer field, the distances between two players can be computed using the Pythagorean theorem. However, on a curved surface like that of a soccer ball (right panel), the distance between two points does not follow the same relation. Although the soccer ball looks like it is made of flat hexagonal and pentagonal pieces, in reality, they each bulge out a bit due to the pressure inside the ball and its stuffing. Credit: NASA/SSU/Aurore Simonnet

DISTANCE ON A SOCCER FIELD

Imagine that two soccer players are on a field, as depicted in Figure $10.3.2$ We can use a grid of points and the Pythagorean theorem to calculate the distance between them.

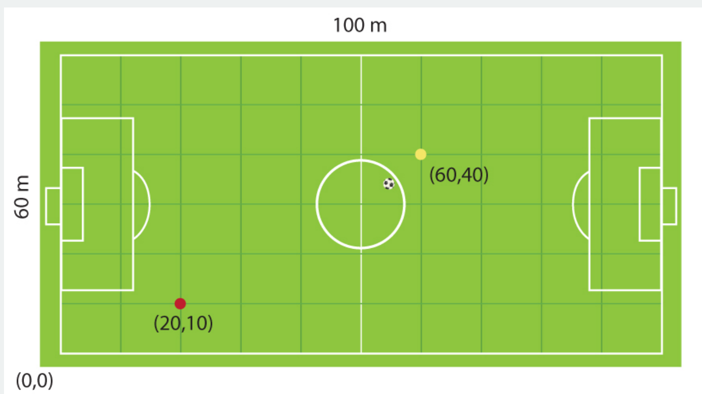


Figure 10.3.2: This soccer field has the origin at the lower left corner, and the coordinates indicate the positions of two players. Credit: NASA/SSU/Aurore Simonnet

Worked Example:

1. If we take the origin of our coordinate system to be at the lower left corner of the field, then the red player is at (x = 20 m, y = 10 m). The yellow player is at (x = 60 m, y = 40 m). What is the distance between them?

- Given: Red: x = 20 m, y = 10 m; Yellow x = 60 m, y = 40 m
- Find: the distance between the two soccer players: d
- Concept: $d^2 = (\Delta x)^2 + (\Delta y)^2$
- Solution: First compute Δx and Δy:
- Since the red player is at x = 20 m and the yellow player at x = 60 m, we have: Δx = 60 m - 20 m = 40 m.
- Similarly, Δy = 40 m - 10 m = 30 m.
- Now, from the Pythagorean relation we get: $d^2 = (\Delta x)^2 + (\Delta y)^2 = (40\text{ m})^2 + (30\text{ m})^2 = 1600\text{ m}^2 + 900\text{ m}^2 = 2500\text{ m}^2$
- Taking the square root of both sides: d = 50 m

Questions

In the worked example we saw that the distance between the two soccer players was 50 meters. We assumed that the origin of our coordinate system was at the lower left of the soccer field. Now, assume that the origin is at the upper left to answer the following questions.

1.

2.

https://phys.libretexts.org/@go/page/31349

3.

It should be clear from the activity that the choice of coordinates will have no effect on the outcome. We can shift the coordinate origin around to any point we wish. We could even use a rotated coordinate system, one whose axes did not line up with the edges of the soccer field: the distances between the players would still be the same. It would have to be! This will also be true in curved spaces. However, in a curved space we will not be able to use the familiar Pythagorean relation. In curved spaces, distances are not related that way.

The surface of a soccer ball is an example of one kind of curved space: it is a sphere. If we know the positions of two points on a sphere, can we still compute the distance between them? We can, but the procedure is more complicated than it is for a flat soccer field.

Consider a larger sphere than a soccer ball, for example, Earth. Now think about the distances between points on Earth's surface, perhaps the distances between cities. We can lay out an x–y grid on Earth's surface just as we did on the soccer field. We might put the origin at the intersection of the Greenwich Meridian and the equator. Then we could measure the position of every point on Earth's surface using that grid.

As a first try, you might lay out a lot of soccer fields, each touching the adjacent ones at its edges. You could build up a huge grid this way by matching the grid lines from one field to each adjacent field and simply attaching the coordinates of one field to the next. In this way, you could try to cover Earth with a grid of patchwork fields, like a big quilt. But would it work? Unfortunately, no. There is no way to cover Earth's surface with flat, square soccer fields. If you try, you would find that some of the fields overlap, or else there are gaps between them. This is true no matter how small you make the fields. Even tiny ones will have some overlap or leave some gaps that they do not cover.

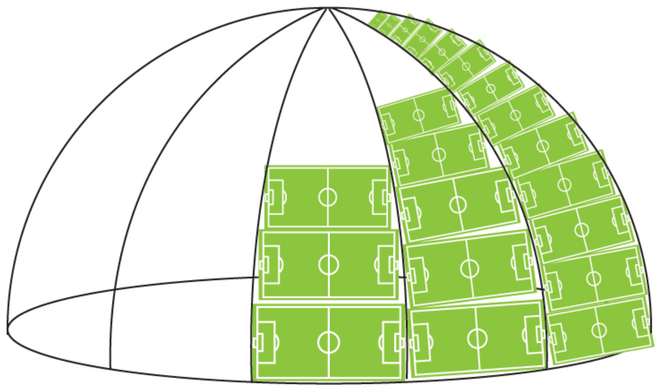


Figure 10.3.3: We cannot cover Earth's surface with flat soccer fields. We have to deform the fields by stretching them in some places, compressing them in others. This is because soccer fields are flat, whereas Earth's surface is curved. Credit: NASA/SSU/Aurore Simonnet

To understand why soccer fields cannot cover Earth's surface, just think about the grid system that we actually do use on Earth's surface: The grid is called longitude (perhaps like our x–coordinate on the soccer field) and latitude (like our y–coordinate). There is something a little peculiar about these coordinates, as we can see if we consider two pairs of points. The first pair has x = 10, y = 10 and x = 20, y = 10. The second pair of points has x = 10, y = 70, and x = 20, y = 70. Figure 10.3.4 shows how these points are laid out.

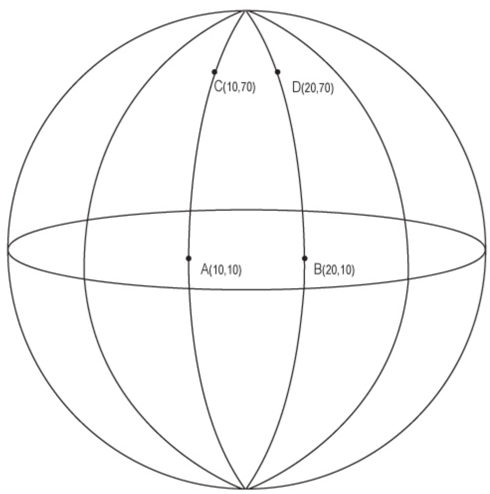


Figure 10.3.4: Two points separated by the same longitude near the equator (A and B) are farther from each other than two points with the same longitudes that are near the poles (C and D). Credit: NASA/SSU/Aurore Simonnet

If this were flat space, then we know that the distance between the first two points would be the same as the distance between the second two. For both sets of points the y coordinate is the same, so the only difference lies in the x coordinate. In a flat space we would have $d = \Delta x = 10$. But, that is not so on a sphere. The points near the pole are clearly closer together than the points near the equator. If you have a globe handy, have a look at it and you will see that this is true. And for an extreme example, if we had set the y value of the points to 90 degrees, then they would coincide; there would be no distance between them at all!

This property of coordinates in spherical geometries is distinctly different from the properties of coordinates in a flat coordinate system. In a flat system, lines that start out parallel remain so forever. In a spherical system, the lines of longitude that start out parallel at the equator do not remain so. In fact, they actually cross at the poles. The extreme example is two lines of longitude separated by 90 degrees: they start out parallel (i.e., they are both perpendicular to the same line, the equator), but when they reach the pole they are perpendicular and cross at a right angle.

If we insist on holding strictly onto ideas appropriate for flat spaces, the attributes of spherical ones seems a little bit absurd. Nonetheless, spherical geometries are real, and they do have these characteristics. They are simply different from the flat geometries with which most of us are more familiar. Probably the best thing we can do when learning about them is to relax our grip on some of the ideas instilled when we studied flat (Euclidean) geometry in school. And keep in mind that the geometry of general relativity, like that of a sphere, is definitely not Euclidean.

So how can we measure the distances between two points in a curved space? It must have *something* do to with the coordinates of the points. If it did not, there would not be any reason to use coordinates to describe the position of points. Have a look at the map of Earth's surface in Figure 10.3.5. The map is a kind of projection that we often see on road maps or wall maps or books. To create this projection you can imagine a cylinder wrapped around Earth such that it just touches Earth's surface at the equator, as in Figure 10.3.6. The points on the globe are projected out onto the cylinder to make the map. There are many ways to do this, and they have various advantages and disadvantages. However, none of them faithfully reproduces the distances between points over their entirety. There are always distortions introduced because the curved space of the sphere cannot be represented in the flat space of the cylinder without introducing warps or stretches. (Cylinders have flat geometry. We will explain this somewhat non-intuitive statement below.)

Figure 10.3.5: This kind of map is quite common. It projects the curved surface of Earth onto a flat map for ease of viewing. In doing so it distorts distances, with those near the pole appearing to be much larger than those near the equator. Credit: Shutterstock

In the projection we are using, called a **Mercator projection**, the distortions are noticeable because the parts of the map near the poles are represented as being much larger than they actually are relative to the parts near the equator. Put another way, distances near the poles appear to be much greater than they are on Earth's surface. For instance, Greenland seems to be as big or bigger than South America. In fact it is roughly the size of Argentina.

To help the users of these maps take account of their distortions, map makers usually place a small scale somewhere on the map. The scale indicates how to read distances at various places on the map. So for example, the scale might indicate that near the equator one inch is equal to 100 kilometers, but at 70 degrees north of the equator this same one inch could correspond to 50 kilometers. Near the poles it takes more inches to represent a given distance, and objects look bigger because more inches are needed to represent them. It can be very misleading. This is an example of map coordinates not being related to map distances in a simple way, as they would be in a flat space. It is the essence of what we mean when we say a space is curved.

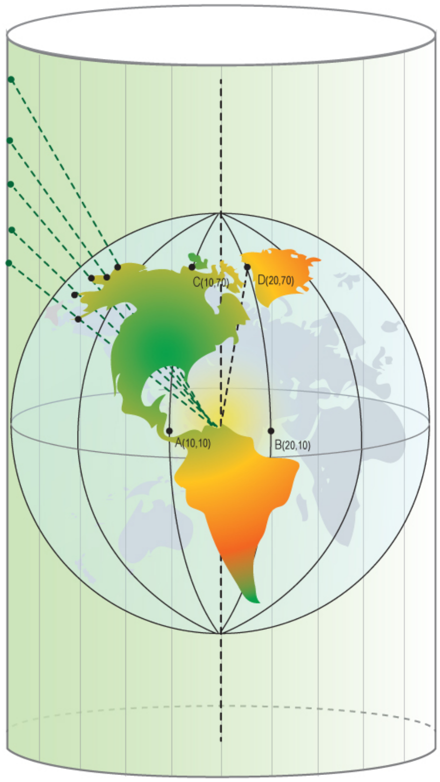


Figure 10.3.6: How a Mercator projection map is made, using a cylinder wrapped around a globe. Credit: NASA/SSU/Aurore Simonnet

For a sphere, writing down a scale for the coordinates is actually pretty simple, though the math needed to derive it is more than we want to tackle. We will just state that the correct way to find the distance d between two points on the surface of a sphere of radius R, assuming small changes in latitude, θ, and longitude, α is written as follows.

$$d^2 = (R\Delta\theta)^2 + \cos^2\theta(R\Delta\alpha)^2$$

This expression does not differ so much from the Pythagorean relation we used in flat space. It still has the square of the distance related to the squares of the difference in coordinates, in this case $\Delta\theta$ for the change in latitude and $\Delta\alpha$ for the change in longitude. These must both be expressed in radians, not degrees, for the expression to be valid. (Remember that there are π radians in 180 degrees.) In addition, there is the radius of the sphere, R. We need that to give us a size scale. After all, we travel farther going half way around Earth than we do going half way around a soccer ball, even though the change in coordinates amounts to 180 degrees in either case. But it is not the R that tells us the space is curved. The giveaway that we are dealing with a curved space is the $\cos\theta$. It multiplies the longitude part of the distance, and it changes for different values of the latitude. But what does it do, exactly?

Remember that the cosine function varies between 1 and -1. It is 1 when its argument is zero (on the equator in this case) and it is zero when its argument is +90 degrees or -90 degrees, at the North or South Pole. Figure $10.3.7$ shows a plot of the cosine for one complete trip around a circle, from zero to 360 degrees. If you go more than 360 degrees the cosine just repeats; it is just like you started counting again from zero every time you go once around the circle.

The cosine in the second term causes the longitude to make no contribution when for angles (latitudes) on either pole (where the cosine is zero), and it causes the change in longitude to have an equal effect with the change in latitude when the points are on the equator. In between the two the longitude contributes some, but not as much as the change in latitude.

So in effect, the cosine of the latitude scales the contribution of the change in longitude to the distance between two points. If you consider how longitude and latitude work on Earth's surface, you will see that this is exactly what we expect. Lines of longitude get closer to each other as they leave the equator and approach the poles. Thus, we know that a difference in longitude will make a smaller contribution to the total distance between two points as the points approach the poles.

https://phys.libretexts.org/@go/page/31349

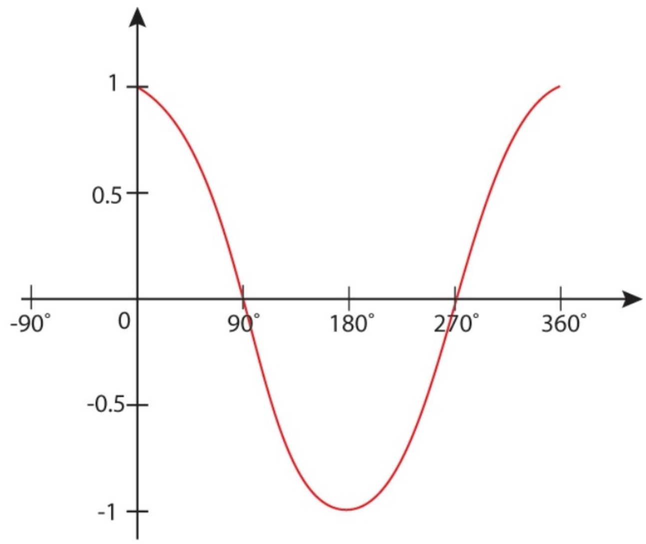


Figure 10.3.7: The cosine is a function of an angle. It varies between -1 and 1, and it repeats every 360 degrees. Credit: NASA/SSU/Aurore Simonnet

The following activity gives you a chance to use this relationship to find the distance between two points on the surface of a ball. Do not forget to convert the angles from degrees to radians (remember that π radians are 180 degrees).

Distance on a Soccer Ball

Worked Example:

1. Imagine there are two ants on the surface of a soccer ball. The ants are close to the “equator” of the soccer ball, at “latitude” +10 degrees. One ant is at “longitude” 10 degrees and the other is at 15 degrees, as measured in the soccer ball’s coordinate system. If the soccer ball is 22 cm in diameter, how far apart are the ants?

- Given:
  The radius of the soccer ball, R = 22 cm
  The ants are at coordinates (lat,long) = (10,10) and (10,15). These values are expressed in degrees, and we must convert them to radians so that we can use our distance relation:
  10 degrees = (10 degrees) (π radians / 180 degrees) = 0.1745 radians
  15 degrees = (15 degrees)(π radians / 180 degrees) = 0.2618 radians
- Find: The distance between the two coordinates, d
- Concept: We will use the expression for distance:
  d2=R2[(Δθ)2+cos2θ(Δα)2]
- Solution:
  First we need to find the cosine of the latitude, or cos(10) = 0.9848. Note that we do not have to convert the 10 degrees to radians when we take its cosine. The cosine of an angle is just a number. However, we must be careful to use the correct mode on our calculator when we take the cosine. Many scientific calculators can switch between using angles expressed in degrees or in radians. You can use either one, but make sure you know which your calculator is set to use.

Now we are ready to plug our numbers into the equation:

$d^2 = R^2 (\Delta\theta^2 + \cos^2\theta\ \Delta\alpha^2) = R^2 (0.1745 - 0.1745)^2 + (0.9848)^2 (0.2618 - 0.1745)^2 = 0.0075 R^2$

Taking the square root we get d = 0.0866 R. This is the separation of the two points in radians.

To find the distance between them in centimeters we have to multiply by the value for R, the radius of the ball:
d = (22 cm)(0.0866) = 1.9 cm

Questions

1.

2. W

If you wanted to be more precise, your simple cosine relation would not be good enough. In reality, the surface of Earth is not smooth; it is bumpy. It has mountains and valleys. Distances between points are affected by these, as you will have noticed if you ever had to hike over a mountain or down into a canyon to get to where you were going. It is not possible to write down a general relation for the distances between two points if we have to take into account these sorts of surface features. But that is okay, we know we could do it if we needed to, perhaps in some small area like that shown in Figure $10.3.8$. We would just have to hire a surveyor to precisely map out the region of interest. Usually we do not have to know the geometry to that kind of precision, but in principle we could measure it.

Figure $10.3.8$: If we wanted to measure the distance between two points on a bumpy surface we would have to know the curvature locally in that region. Generally we do not know this, although we could measure it if we wanted to. This figure of the Grand Canyon indicates a region where the Earth's surface is bumpy. Credit: Shutterstock

Before we discuss curvature in spacetime, we want to illuminate one point that can cause confusion for people when they first learn these ideas about geometry. We said that we can project curved spaces like the surfaces of spheres onto cylinders to get a flat-space representation of them. But cylinders look curved, so how do we get a flat map by using a cylinder? The confusing part is that cylinders only *look curved*. In fact, cylinders are flat, at least geometrically speaking. From our definition of curvature— the way we measure distances in the space depends upon where we are— we can show that cylinders are flat.

The easiest way to understand how this works is to consider a flat piece of paper… We know that a piece of paper is flat because we can write down a coordinate system in which the distances between all the points are given by the familiar Pythagorean theorem. There are no extra terms anywhere— like the cosine term for spherical geometry, for instance. That is a flat space.

Now bend the piece of paper over so that opposite edges meet; we have made a cylinder. In doing this we did not have to distort the sheet of paper at all. We did not stretch it or crinkle it or tear it, we merely had to bend it so that the opposite sides coincided. This means that the distances between all the points are still related by the same Pythagorean relation, and the surface must still be flat. In general, a 2D space that can be “unfolded” or “unwrapped” and laid directly onto a flat space like a table without any distortions, like tearing, stretching or crinkling, must itself be a flat space.

Another important point relates to the use of the radius *R* as a scaling factor. A cylinder does have a radius, and a larger cylinder has a larger radius, as shown in Figure $10.3.9$. We know that the bigger the cylinder, the larger the distance between any two points with coordinates (x,y). So the radius scales the distances to coincide with the size of the cylinder. However, the radius is constant over the entire surface we are considering. It is merely a constant multiplying, or scaling, factor that does not vary from point to point.

A sphere has a radius that scales distances too. This works the same way as it does for a cylinder: on a sphere the radius multiplies the longitude and the latitude parts of the distance relation and does not vary over the surface being scaled. However, on a sphere we also have the cosine term. It varies around the sphere and only multiplies the latitude term in the distance relation. Contrast that with the way the radius scales the size for both the cylinder and the sphere.

The critical difference between a curved space like the surface of a sphere and a flat space like the surface of a cylinder is that the former has extra terms that vary from place to place and that are required to compute distances between points in the space. This is what tells us that a sphere is curved, whereas a cylinder, lacking such terms, is flat.

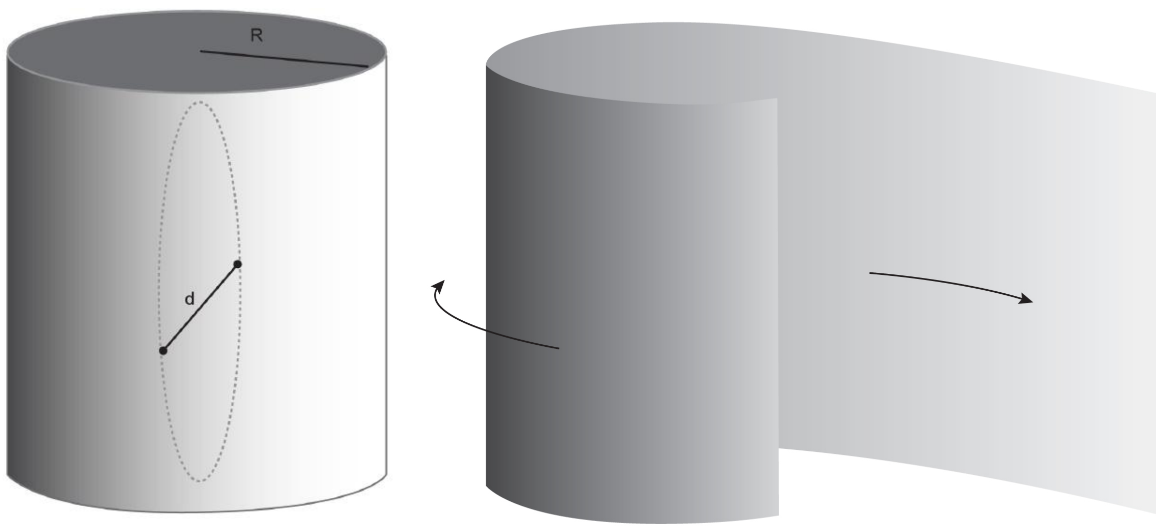


Figure $10.3.9$: (Left) If we wanted to measure the distance between two points on a cylindrical surface we would have to know the size of the cylinder in addition to the (x,y) coordinates of the individual points. (Right) Despite having a radius and looking curved, by our definition the cylinder has a flat geometry because it can be unrolled onto a flat surface Credit: NASA/SSU/Aurore Simonnet

We want to emphasize that when we talk about a space with a flat geometry, what we mean is that the distance between any two points is given by the Pythagorean theorem.

$$d^2 = \Delta x^2 + \Delta y^2$$

When we talk about a space with a curved geometry, we mean that the distance relation will have some other terms, *A* and *B*, where *A* and *B* in general would depend on the values of *x* and *y*.

$$d^2 = A(x,y)\Delta x^2 + B(x,y)\Delta y^2$$

In our example of the sphere, *A* was constant everywhere and equal to 1, and *B* was $cos^2\,\theta$. Here we are ignoring the scaling constant *R* because it just tells us about the size of the ball— Earth or soccer ball, say, just as in our description of the cylinder; it has nothing to do with the geometry or curvature.

Notice that we have not required our 2D space to exist in a 3D space in order to define the curvature. We can envision it that way if we like, and it seems a natural and convenient way to think about the surface of Earth. But it is not necessary.

What about the radius? Doesn't it imply some sort of curved surface? No, not necessarily. R has come in as a global scaling factor. We identify it as the radius of a ball, either Earth or a soccer ball in the examples above, but that is not essential. In general, R is just a number that tells us how large the distances in a space are. It is better to think of curvature in this way; once we go to curved spaces with more than two dimensions, it is not possible to visualize any higher dimensional spaces in which they might exist, so attempting to visualize them that way only leads to confusion.

Is a Circle Curved?

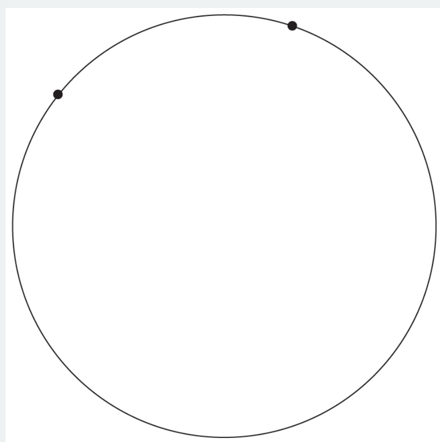


Figure 10.3.10: We generally think that the arc describing a circle is curved. Is this borne out using our more precise idea of what we mean by “curvature?” Credit: NASA/SSU/Aurore Simonnet

10.3.2: Distances and Curvature Extended to three Dimensions

Everything we have said about 2-dimensional spaces can be generalized to three dimensions. The distance between two points in a flat 3-dimensional space is given by the 3D version of the Pythagorean theorem:

$$d^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

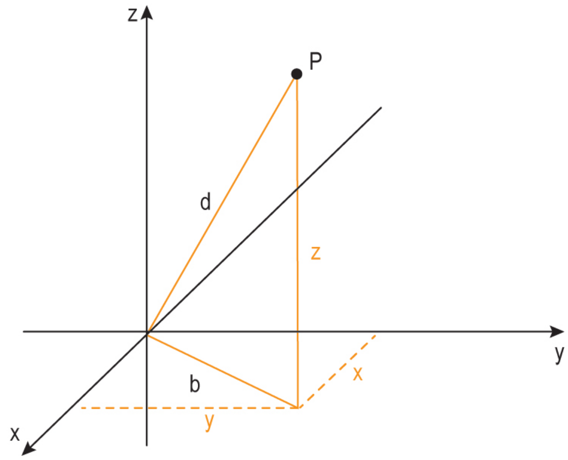


Figure 10.3.11: In three dimensions the Pythagorean theorem still applies. Notice how we can get a 3D version of the relation by successive applications in two-dimensional planes: $b^2 + z^2 = d^2$, *and* $x^2 + y^2 = b^2$, *so* $x^2 + y^2 + z^2 = d^2$. Credit: NASA/SSU/Aurore Simonnet

If the space is curved, then there could be extra multiplicative factors that depend on the coordinates, just as in 2 dimensions. We cannot easily draw an example of how that space might appear, but we could imagine and describe its behavior by adding appropriate extra terms to the distance relation. As an example:

$$d^2 = A(x,y,z)(\Delta x)^2 + B(x,y,z)(\Delta y)^2 + C(x,y,z)(\Delta z)^2$$

In this relation *A*, *B*, and *C* depend on position (x,y,z), and they are different for different spaces; if *A*, *B*, and *C* are constant everywhere then the space is not curved. We can use expressions like this to compute distances between points in any space for which we know how *A*, *B*, and *C* vary with position. Vitally, we do not have to be able to actually visualize the space to do this.

We will worry about the detailed properties of curved 3-dimensional spaces. We mentioned it just as a bridge to describing the four curved dimensions of spacetime in general relativity. Curvature in 4-dimensional spaces can be described the same that it is in three dimensions.

10.3.3: Curvature in Spacetime

There are four spacetime dimensions: three of space, one of time. If we choose the x-direction to be in the direction of relative motion between different reference frames, then the other two dimensions are not affected by the motion, so we ignore them. Under this assumption, distances in spacetime are given by a relation reminiscent of the Pythagorean theorem, but with a negative sign on the time term.

$$s^2 = (\Delta x)^2 - (c\Delta t)^2$$

Here we label the distance in spacetime, or the spacetime interval, with the symbol *s*. The name helps us remember that it is different from a distance *d* in normal space, which has no time part.

For the case of special relativistic flat spacetime, it is easy to add in the other space dimensions if we wish. We do just what we did in the last section for normal 3D space:

$$s^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (c\Delta t)2$$

The expression is a little bit more complicated than the previous one, and ignoring the y and z directions in special relativity allowed us to simplify our analysis somewhat without giving up a complete description of a system's behavior. Unfortunately, we cannot do that in general relativity. Though it is always possible to take the relative velocity to define the x–direction in special relativity, the curvature of spacetime that is gravity is fundamentally bound up into the geometry of all four dimensions of spacetime. Just as the curvature on Earth's surface depends on both latitude and longitude, we cannot describe it properly with just a single dimension. We must keep all four dimensions of spacetime to describe curvature in general relativity.

What are the extra terms that multiply the different parts of the expression to express the curvature in general relativity? That depends on the distribution of mass and energy, which are the sources of gravity. If we consider the gravity created by a spherical object like a star or planet. It will have a simple general form.

$$s^2 = A[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2] - B(c\Delta t)^2$$

In this case, we have already found the curvature term for the time part of the spacetime interval: In Section 10.2 we looked at how gravity curves time. That gave us an expression for the gravitational redshift, and that is exactly the distortion of the time dimension that we seek. Substituting in $g = GM/r^2$, we can write an expression for B.

$$B = \left(1 - \frac{2GM}{rc^2}\right)$$

We have expressed B using polar coordinates rather than Cartesian coordinates, for reasons that will be clear below. The spacetime interval, including the temporal curvature, can then be written as below.

$$s^2 = A\left[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\right] - \left(1 - \frac{2GM}{rc^2}\right)(c\Delta t)^2$$

What about the spatial part of the interval? Does that also have to be modified? It turns out that it does, but there is no simple intuitive way to derive it. We would have to use the full mathematics of general relativity, and that is beyond the level of this unit. Instead we will simply state it, for later use, and note that the term seems reasonable given that it is the same as the temporal curvature term. Notice that for the spatial part we must divide rather than multiply by the curvature term.

$$s^2 = \left(1 - \frac{2GM}{rc^2}\right)^{-1}\left[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\right] - \left(1 - \frac{2GM}{rc^2}\right)(c\Delta t)^2$$

Using the fact that $(\Delta r)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$, we can express this interval in polar coordinates if we like.

$$s^2 = \left(1 - \frac{2GM}{rc^2}\right)^{-1}(\Delta r)^2 - \left(1 - \frac{2GM}{rc^2}\right)(c\Delta t)^2$$

This is the spacetime interval for a spherically symmetric object. The spherical symmetry means it is the same in all directions, x, y and z; they are all multiplied by the same curvature term, and that term depends only on the distance from the source, *r*, not on the direction. That is why polar coordinates are convenient. Also, note that the speed of light causes the curvature term to be very close to 1 in the Solar System. There is no object within the Solar System for which the ratio *M/r* is large enough to offset c^2, and it explains why we are generally not aware of curvature's effects.

We will look at this interval and its consequences in much more detail in the next section on tests of general relativity.

The Gravitational Radius

If you take a look at the spacetime interval for a spherically symmetric object that we have just introduced, you notice that the curvature terms behave strangely at a certain point, namely, where $r = 2GM/c^2$. We call GM/c^2 the gravitational radius.

Notice that the gravitational radius depends on the mass of the object creating the gravity. The mass is a type of scaling factor; it plays a role similar to that the radius, *R*, did when measuring distances on the surface of a sphere or cylinder in the previous section.

Now let's calculate the gravitational radius for some actual objects.

1. F

2.

3.

This page titled 10.3: What is Curvature? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.4: Tests of General Relativity

What Do You Think: Global Positioning System

In this section we will discuss some of the observational tests of the General Theory of Relativity. The theory makes a number of predictions about the world that are distinct from those of Newtonian gravity. We have already looked at several. Here we will consider a few more, and in somewhat greater detail. A number of these topics will be covered in later chapters, and this section serves as an introduction for deeper investigations to come.

10.4.1: The Orbit of Mercury

In 1915 when Einstein was completing his General Theory of Relativity, the orbit of Mercury had long been a puzzle to astronomers. Unlike the other planets, Mercury displays a slight variation in its orbit from the predictions of Newtonian gravitational theory. The orbit is not a closed ellipse as Newtonian theory predicts. It has a small amount of overshoot. This means that the entire orbit is slowly turning as the planet orbits the Sun. What we mean by this is that the point of closest approach to the Sun, called the perihelion, slowly rotates around the Sun, as does the point of greatest distance from the Sun, the aphelion (Figure 10.4.1).

Top View

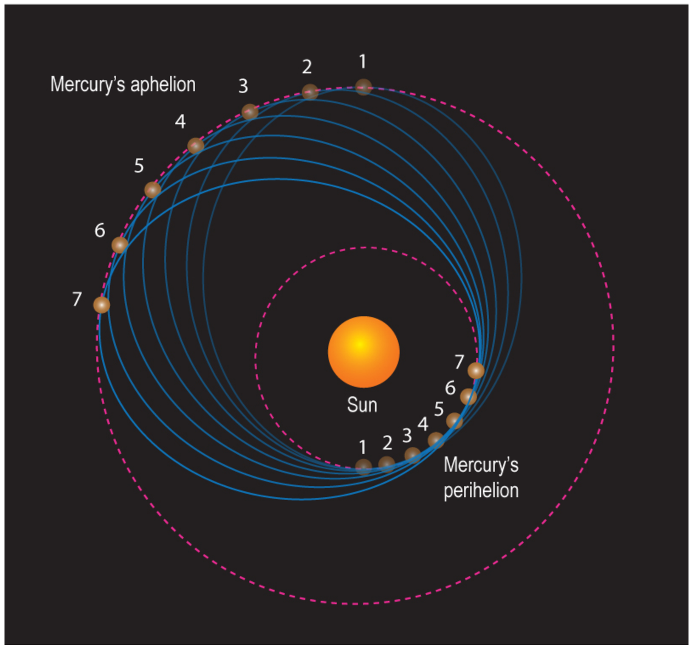


Figure 10.4.1: As Mercury orbits the Sun, its orbit slowly precesses around the Sun. This means that the position of the perihelion or aphelion slowly circles the Sun. Newtonian gravity cannot account for all of this motion, but general relativity does. The eccentricity of the orbit has been greatly exaggerated in this figure. Credit: NASA/SSU/Aurore Simonnet

The amount of this orbital shift is very small, and most of it is caused by gravitational tugs exerted by the other planets (primarily Jupiter), on Mercury as it goes around the Sun. However, even when the gravitational effects of the other planets are accounted for, Mercury still travels an extra 0.1036 arcseconds of rotation (beyond 360 degrees) for each orbit. This is an extremely small amount, too small to be seen, but it accumulates over time. After a century, Mercury will have rotated an extra 43 arcseconds beyond the predictions of Newtonian gravity. Even this angle is small, but it is measurable. The discrepancy indicates that something is not quite right with the Newtonian understanding of the Solar System.

The first attempts to account for Mercury's orbital anomaly proposed the existence of a planet orbiting nearer to the Sun than Mercury. The planet, tentatively named Vulcan, was never seen, though many astronomers searched for it for many years. The idea of an unseen planet affecting the motion of a visible one was not unprecedented. The planet Neptune had been discovered in exactly this way. Astronomers had noticed anomalies in the position of Uranus. They were able to predict where an additional planet would have to be in order to cause those anomalies, and in 1846 Neptune was discovered in the area that had been predicted. A similar solution for Mercury's strange motions was a reasonable suggestion, but no planet was ever found. The orbital anomalies remained an unsolved puzzle.

Albert Einstein was aware of Mercury's anomalous orbit. When his new theory seemed complete enough to be applied to the orbit he did the calculation. There were no free parameters in Einstein's theory that could be used to adjust it to the correct value. The only relevant inputs were the mass of the Sun and the size and eccentricity of Mercury's orbit. When Einstein made the calculation, his theory predicted that Mercury should precess around the Sun at a rate of 43 arcseconds per century, precisely the amount of the observed anomaly.

This was one of the great moments in scientific discovery. Einstein's theory, built from his imagination and tempered by his understanding of physics, could have been right or it could have been wrong. Only comparison with observations of nature would determine which. When Einstein made his calculation and his theory gave the correct answer, he knew that he had gained a profound insight into the workings of the world— he knew his theory was right. This was the culmination of 10 years of hard work, and on that evening in 1915, Einstein was the only person in the history of the world to have that insight. He would not remain so for long.

10.4.2: Light Bends Around the Sun

We have already discussed how, according to general relativity, gravity should cause light to follow a curved path. (See the discussion about the equivalence principle in Section 10.2.1.) However, we also showed that the expected deflection of a photon's path on Earth is so small that it is completely unmeasurable. Earth simply does not produce enough spacetime curvature. Its mass is too small. If we could find an object with a larger mass, perhaps we could measure the deflection of a beam of light.

Fortunately, we have such an object quite close to us. The Sun is 300,000 times Earth's mass, and there are many stars behind it that conveniently emit photons in all directions, including ours. Photons from these stars provide an excellent test of general relativity's light-bending prediction. But is the Sun, massive as it is, able to create gravity strong enough to deflect starlight by a detectable amount?

We can make an estimate of the amount by which a beam of light will be deflected by the Sun. To do this, we will use the equivalence principle, as we did when we found the gravitational redshift in Section 10.2.2. We imagine that a beam of light begins its journey far away from the Sun, and that it passes close to the Sun as it travels. When it is close enough, it will begin to be affected by the Sun's gravity. The beam will be deflected by some amount, just as we computed for a beam of light on Earth's surface. The situation is shown schematically in Figure 10.18.

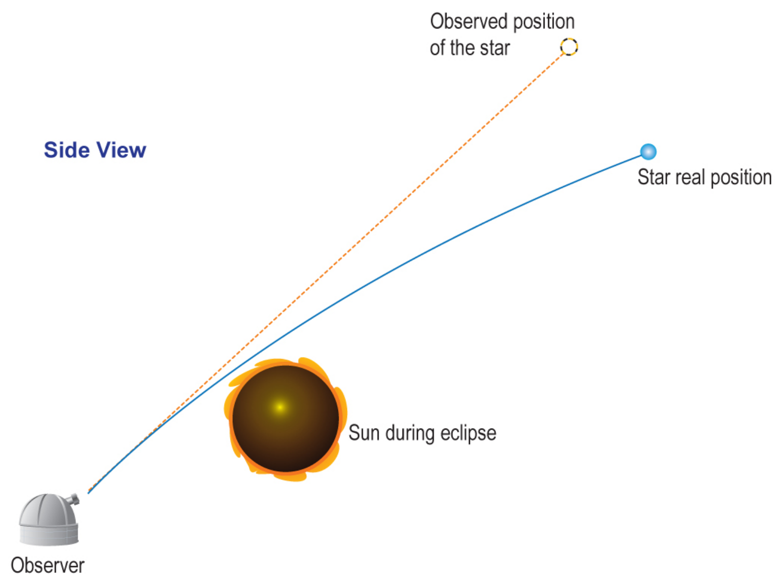


Figure 10.4.2: As a beam of light passes by a massive object like the Sun, it will be deflected from its original direction by an amount that depends on the mass of the gravitating object and the distance of closest approach. The trajectory of an undeflected photon is shown as the dotted path. Credit: NASA/SSU/Aurore Simonnet

If the photon were not deflected, then it would travel along the straight line indicated by the dotted path in Figure 10.18. However, the spacetime curvature bends that straight line into the curved path shown by the solid line. From the point of view of Newtonian gravity, we would say that the photon falls in the gravitational acceleration for some amount of time, and that its falling is what causes its direction to change.

Using the weak equivalence principle, we derive the angle of deflection in Going Further 10.3: Angle of Light Deflection. The total amount of deflection the photon experiences is given below.

$$\theta = \frac{2GM}{bc^2}$$

The angle of deflection is represented by θ, M is the mass of the Sun and b is the distance of closest approach. In the next activity, you will calculate the angle of deflection for the case of light passing close by the Sun.

Calculating Newtonian Deflection of Light

Newtonian arguments indicate that the angle θ by which the light path should change as it travels near a massive body is given by $\theta = 2GM/bc^2$, where M is the mass of the Sun (2×10^{30} kg), and b is the distance of closest approach. Remember that θ is given in radians.

Worked Example:

1. Calculate the angle θ (in radians) when b = the distance from the Sun to Earth = 1.5×10^8 km.

- Given: b = 1.5E8 km = 1.5E11 m, M = 2E30, c = 3E8 m/s, G = 6.67E-11 Nm^2/kg^2
- Find: θ
- Concept: $\theta = 2GM/bc^2$
- Solution: $\theta = 2(6.67E{-}11\ Nm^2/kg^2)(2E30\ kg)/[(1.5E11\ m)(3E8\ m/s)^2] = 1.98E{-}8$ radians

Questions

1.

2.

The relationship used in this activity relies only the equivalence principle and our understanding of Newtonian gravity. It is remarkably close to the correct answer. In fact, what we have found is the amount by which Newtonian gravity alone would deflect light traveling close to the Sun. In doing so we have employed the curved-time part of the spacetime interval Recall that we were able to deduce time curvature using only the equivalence principle in Section 10.2.2.

However, in general relativity there is another curvature term that must be taken into account, that of space. The contribution of space curvature is the same size as the time curvature contribution, so it increases the deflection by a factor of two. The angle of deflection is very small, but it is large enough to be easily measurable, even with small telescopes.

The trouble, of course, is that we cannot usually view stars near the Sun. They are (obviously) up during the daytime, when the sky is bright enough to make the stars invisible. The situation changes during a total solar eclipse. For the brief few minutes that the Moon covers the Sun, it is possible to see stars near the solar limb.

In order to take advantage of rare solar eclipses, scientific expeditions were undertaken in 1919 to test this “light-bending” prediction of general relativity. Expeditions, under the direction of the British astronomer Arthur Eddington, traveled to the island of Principe, off the West Coast of Africa, and to Sobral, in Brazil, to view the eclipse. They photographed the sky around the Sun while the Moon covered it, and then compared the images to others taken at night, when the Sun was in another part of the sky. The results, though of poor quality, confirmed the predictions of general relativity. It was for this result that Albert Einstein became a household name and an icon. News of the eclipse results made huge headlines in newspapers around the world. Since the first measurements in 1919, the results have been confirmed several times with ever-improving precision. Modern measurements are generally done using radio telescopes and background galaxies.

The Sun is not the only object in the sky that deflects light as it travels to us. In fact, so-called gravitational lenses have become an important tool used by astronomers to study the Universe.

GOING FURTHER 10.4: NEWTONIAN GRAVITY AND CURVATURE

10.4.3: Gravitational Redshift

We have seen how the equivalence principle leads to a prediction of gravitational redshift — the shifting of light to lower frequency as it climbs out of a region of relatively stronger gravity to a region where gravity is weaker. The effect is too weak to be measured on Earth’s surface, as we calculated in Section 10.2.2. Even on the surface of the Sun, where gravity is more than 10 times greater than on Earth’s surface, the effect is too small to detect. However, nature has again provided us with a region where we can put the predictions of general relativity to the test.

When a star like the Sun dies, it leaves behind an object composed of the ash remaining from its lifetime of nuclear fusion. Such an object is called a white dwarf. The typical white dwarf is comparable in size to Earth, but has a mass comparable to the Sun's.

Surface Gravity of a White Dwarf

You should have found that the surface gravity of a white dwarf star is much larger than the gravity on Earth’s surface. If you recall, we deduced the gravitational redshift relation in Section 10.2.2 under the assumption that the gravity is weak. We also assumed that it was nearly constant. Neither of these assumptions is valid when we view the spectrum emitted at the surface of a white dwarf. Under these conditions we have to use the full expression for the gravitational redshift, valid for strong gravity as well as weak. That relationship is written below.

$$\lambda=\frac{\lambda_0}{\sqrt{1-\frac{2GM}{Rc^2}}}$$

The measured wavelength, λ, of a line is measured by an observer far away from the white dwarf, λ_0 is the unshifted wavelength, R is the radius of the white dwarf, and M is its mass. G and c are the gravitational constant and speed of light, respectively. In the next activity we will employ this relationship to find the expected shift in the wavelength of a line in the white dwarf spectrum.

Gravitational Redshift on a White Dwarf

We can use the general relation for gravitational redshift, along with the value of the surface gravity of a white dwarf that you just found, to calculate the shift in wavelength of the sodium-D lines in the white dwarf spectrum. The sodium-D lines are two closely spaced lines that have rest wavelengths near 589 nm.

The measurement of the gravitational redshift of a white dwarf was first attempted in the 1920s for the companion of the star Sirius. At first the results seemed to agree with the predictions of general relativity. Later it was realized that the result was spurious, one reason being that most of the light being measured was from Sirius itself, not its faint companion star (see Figure $10.4.1$). In addition, the white dwarf radius used to compute the expected shift turned out to be wrong. More modern measurements using more powerful telescopes and detectors, as well as a better value for the radius of the white dwarf, have confirmed the predictions of general relativity. Ironically, the best constraints on the gravitational redshift turn out to have been done on Earth's surface after all. These measurements used extremely sensitive equipment in a laboratory; the Pound-Rebka experiment was performed at Harvard University in 1959.

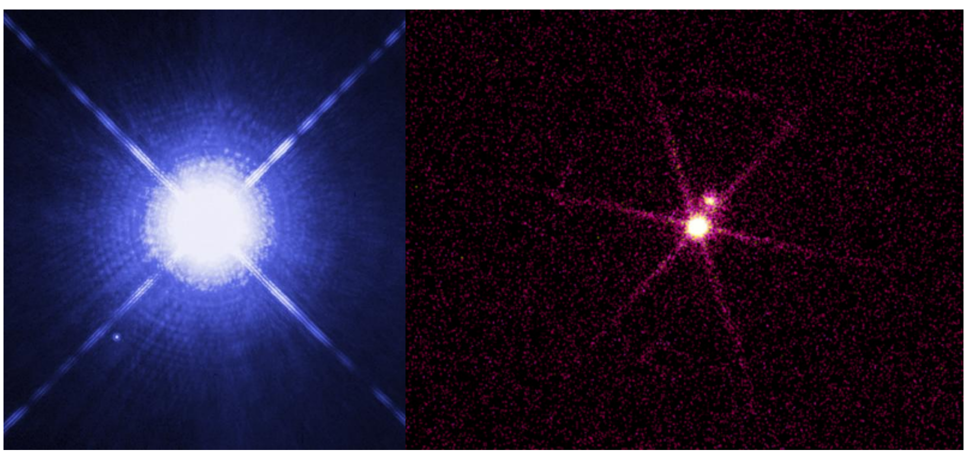


Figure 10.4.1: The figure on the left shows the Hubble Space Telescope image of Sirius A, the brightest star in our nighttime sky, along with its faint, tiny stellar companion, Sirius B. Astronomers overexposed the image of Sirius A [at center] so that the dim Sirius B [tiny dot at lower left] could be seen. The cross-shaped diffraction spikes and concentric rings around Sirius A, and the small ring around Sirius B, are artifacts produced within the telescope's imaging system. Credit: NASA, H.E. Bond and E. Nelan (Space Telescope Science Institute, Baltimore, Md.); M. Barstow and M. Burleigh (University of Leicester, U.K.); and J.B. Holberg (University of Arizona) The figure on the right shows the Chandra X-ray Observatory image of the same two stars. In this case, however, the brighter star is Sirius B, which has a surface temperature of about 25,000 K, producing very low energy X-rays. The spike-like pattern is due to the support structure for Chandra's transmission grating. Credit: NASA/SAO/CXC

10.4.4: Dragging of Reference Frames by Moving Objects

General relativity makes many predictions that seem strange to us. One of the strangest is called frame-dragging. Frame-dragging is also called the Lense-Thirring effect, after the two Austrian scientists, Josef Lense and Hans Thirring, who first studied it in 1918. Frame-dragging is the tendency of massive objects to pull space along with them as they move, sort of the way a boat pulls water along with it as it moves through the sea. It is clearly a tiny effect, since we are not aware of it at all from our normal experiences.

As a result of frame-dragging, spinning objects tend to wrap spacetime around them as they spin. Again, the effect tends to be so small as to be negligible in most cases. However, for a large object spinning rapidly, it is possible to measure the effects of frame-dragging with a sensitive enough experiment. Such an experiment, called Gravity Probe B (GPB), was conducted in Earth orbit during 2004 and 2005. GPB used sensitive gyroscopes to measure the effects of frame-dragging and confirmed the predictions of general relativity to its experimental precision, about 20%. For details about the GPB experiment, see Going Further 10.5: Gravity Probe B.

Note

The ideal gas law is easy to remember and apply in solving problems, as long as you get the **proper values a**

GOING FURTHER 10.5: GRAVITY PROBE B

Going Further 10.5: Gravity Probe B

The Gravity Probe B experiment was able to measure the extremely small frame-dragging effects from the spinning Earth. However, frame-dragging is not always a small effect. If an object is massive enough, and if it spins rapidly enough, then the effects of frame-dragging can be extreme. In these cases it can even be impossible to "sit still" in space because space itself is not sitting still: an observer would be forced along with the rapidly spinning space. Environments like that only exist close to rapidly spinning black holes, and we will consider those objects in the next chapter.

10.4.5: Gravitational Radiation

The final prediction of general relativity was only recently confirmed, in 2015, by direct measurement. To understand this prediction it is best to think first about Newton's idea of gravity, described by his universal law of gravitation:

$$F_g = \frac{GM_1M_2}{r^2}$$

There is nothing in this law that describes how the effects of gravity are transmitted from one place to another. In fact, according to this law, if we move one of the masses to a different point in space, then the other mass "knows" this instantly, and it reacts accordingly. The ability of gravity to act instantly - according to this law - across any distance bothered many of Newton's contemporaries. It continued to bother scientists who came after. And any scientists who lived after 1905, the year that Albert Einstein published his first paper on special relativity, knew that Newton's law could not be correct. According to special relativity, it is not possible for any signal or information to cross space faster than light can travel.

In general relativity, gravity is not a force as envisioned by Newton. Instead, gravity is the result of spacetime being distorted, and the distortions of spacetime are caused by the distribution of mass. Changing the mass distribution will generally change the spacetime curvature, and thus also the gravitational effects on objects in that spacetime. However, when a mass distribution is changed, the effects of that change are not carried instantaneously to all points in space. The information is carried at the speed of light by ripples in the spacetime fabric. These ripples are called gravitational waves.

An analogy might help to understand how this works. We can imagine an observer living in a houseboat at the edge of the sea. If there is a storm far out to sea, the observer might not be aware of that fact. The storm could be over the horizon and out of view. However, storms generally have strong winds that create large waves. These waves travel outward from the storm in all directions, moving at speeds of a few meters per second. Eventually the waves will reach the shore where the observer lives, and the large swells will cause the houseboat to rock back and forth. In this way the observer might become aware of the storm's existence even if the storm never approaches the shore at all— surfers are aware of this and use satellites to track storms so that they can predict when the waves at a certain beach might be good for surfing.

Replace the storm in our analogy with a binary star system and the surface of the sea with spacetime. You can see how we might be able to detect the stars, at least in principle, by measuring the rippling waves of spacetime caused by the stars' motion. As the ripples pass us, the compressing and stretching of spacetime allows us to infer the existence of the stars even if we cannot see them by other means.

In practice, it is extremely difficult to detect the presence of gravitational waves. The stretching of spacetime is minuscule, and it tends to be swamped by non-gravitational effects. Most of these effects are local and have nothing to do with distant orbiting stars.

There are currently several experiments underway that are designed to measure gravitational waves directly. The first of these to have success is the Laser Interferometer Gravitational-wave Observatory, or LIGO. It detected a gravitational wave signal from a merging black hole system on September 14, 2015. To learn more about the LIGO experiment, see Going Further 10.6: Gravitational Wave Observatories.

> **Note**
>
> The ideal gas law is easy to remember and apply in solving problems, as long as you get the **proper values a**

GOING FURTHER 10.6: GRAVITATIONAL WAVE OBSERVATORIES

LIGO is the first experiment to detect gravitational waves directly, but even before this detection, scientists were sure the waves were there. The use of indirect means demonstrated the existence of the waves over the past several decades. This was somewhat analogous to the way a biologist might infer the presence of a bobcat if she saw its tracks or scat on a trail. Of course, gravitational waves don't leave paw prints or other traces of their passing, but they do leave traces of themselves on the systems that emit them.

Let us return for a moment to the example of a binary star system. We said earlier that such a system would emit gravitational waves as the stars move in their orbits. These waves will carry energy. Since the waves are being produced by the movement of stars in space, the stars must be giving some of their own energy to the waves. There is no place else for the energy of the waves to come from.

If the stars lose energy to gravitational waves, then their motion will be affected. Different orbits have different amounts of energy, and the lower the energy of the orbit the smaller it is. So if binary stars create gravitational waves, then they must slowly spiral in toward each other as the waves carry the system's energy away. Their orbital decay should be observable if we wait long enough.

In general, the in-spiraling of binary stars is difficult to measure. For instance, in the Solar System, the amount of energy loss from gravitational radiation is so small as to be utterly unimportant over the lifetime of the Sun, a period around 10 billion years. However, in some astrophysical systems the energy loss is quite a bit larger.

It turns out that the amount of energy radiated in gravitational waves is proportional to the square of the orbital frequency, or put another way, it depends on the size of the system. Larger systems have lower orbital frequencies, and so they produce fewer gravitational waves. This is one reason that the Solar System loses so little energy this way: the periods of the planets in the Solar System are months or years, making for tiny orbital frequencies. For certain systems though, the orbital frequency can be quite high.

> **Gravitational Luminosity**
>
> In this activity, we will calculate the amount of energy that could potentially be carried by gravitational radiation. We will compare our value with more familiar sources of radiation.
>
> The energy emitted every second in gravitational waves by an object of mass m orbiting another of mass M is predicted by general relativity to be given by the expression below.
>
> $$P_{gw} = \frac{2}{5}\left(\frac{GM}{Rc^2}\right)^5 \left(\frac{m}{M}\right)^2 \left(\frac{c^5}{G}\right)$$
>
> If we are considering a planet orbiting the Sun, then M is the mass of the Sun, m is the mass of the planet, and R is the size of the planet's orbit. As usual, c and G are the speed of light and the gravitational constant, respectively. This expression is somewhat simplified if we rewrite it in terms of the gravitational radius.
>
> $$P_{gw} = \frac{2}{5}\left(\frac{R_g}{R}\right)^5 \left(\frac{m}{M}\right)^2 \left(\frac{c^5}{G}\right)$$

1.

2.

The previous activity demonstrated that gravitational radiation can in principle carry large amounts of energy. However, this is only important for certain kinds of systems. Such a system was discovered in the early 1970s by two scientists working at the Arecibo radio telescope in Puerto Rico, then the most sensitive radio telescope in the world. Sadly, that telescope collapsed in December of 2020 after suffering damage during Hurricane Maria in 2017.

The telescope was being employed in the 70s by a graduate student named Russell Hulse, who wished to find pulsars. These are rapidly rotating, highly magnetic neutron stars, the first examples of which had been discovered only a few years before by another graduate student in England, Jocelyn Bell. Pulars show flashes of emission analogous to those from a light house. They generally have very stable pulse separations that vary from a few seconds to tiny fractions of a second, depending on the object being observed. In fact, pulsar periods are so stable that they constitute the most accurate “clocks” known.

However, one of the pulsars discovered by Hulse, called PSR 1913+16, showed an unusual periodic variation in its pulse separation. Pulses would arrive slightly early on successive pulses, then switch and arrive slightly late. The early-late variation pattern repeated roughly every eight hours.

Upon discussing this pulsar with his advisor, Joseph Taylor, it became clear that what they were observing was a pulsar in orbit around another star. However, given the orbital period of only eight hours, the companion had to be extremely small. It was also dim, since no companion was visible. Further careful observations suggested that the companion itself was also a neutron star, though not a pulsar since no pulses were seen from it.

This so-called **Hulse-Taylor pulsar** is interesting in several respects. First, it allows the most accurate measurements of the mass of neutron stars of any astrophysical system. The pulsar is measured to have a mass of 1.44, and its companion 1.38, in solar mass units. More interesting is that the stars are in highly elliptical orbits, so the effect of orbital precession is easily measured in the strong gravity of the pulsars: it is four degrees per year, gargantuan compared to the tiny precession of Mercury’s orbit around the Sun. In addition, the effects of gravitational redshift can be measured as the pulsar climbs in and out of the gravitational potential of its companion. The pulse delays are caused by the combined effects of Doppler shifts due to orbital motion and the gravitational redshift due to the changing gravitational potential around the orbit.

However, by far the most interesting observed property of this system is that the orbital period is slowly decreasing. The two neutron stars are gradually spiraling into one another. The rate of energy loss implied by the in-spiraling can be compared to that expected from general relativistic gravitational radiation (see previous activity), and the two match very closely, within the uncertainties associated with the various measured properties for the system.

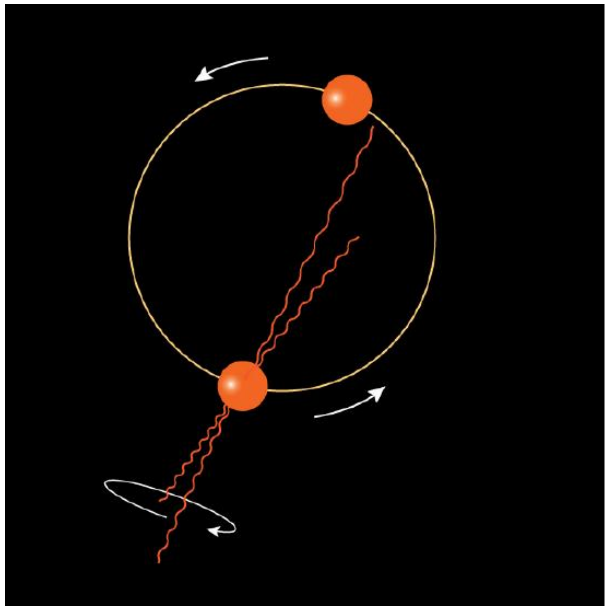


Figure 10.20 In the binary pulsar system discovered by Hulse and Taylor, a pulsar and a non-pulsing neutron star orbit each other, similar to the system drawn here. We have simplified things by making the orbits of each object circular. In the actual binary pulsar system, the orbits are elliptical. Credit: NASA/SSU/A. Simonnet

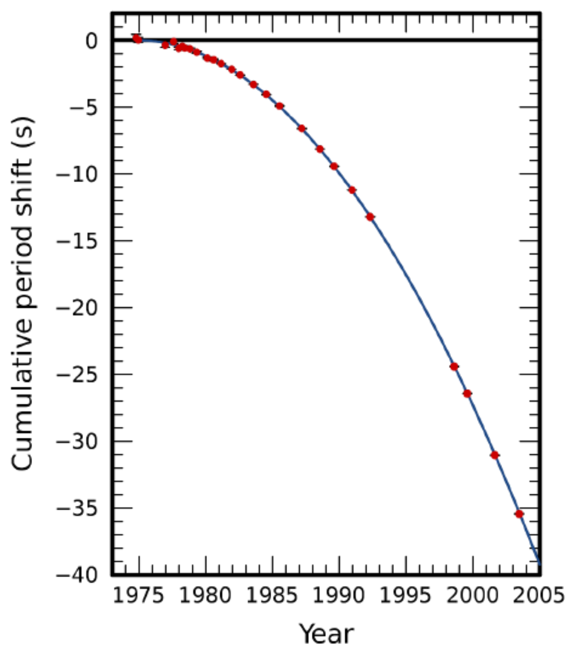


Figure 10.21: This figure shows almost 30 years of data illustrating the changing size of the orbit of the Hulse-Taylor binary pulsar. The shrinking orbit leads to a change in the orbital period that is measured each year. The measured points (red squares) are plotted along with the theoretical prediction (blue line). Credit: J. M. Weisberg and J. H. Taylor (from Wikimedia).

We can use the relation for emission of energy by gravitational waves to estimate the luminosity, and thus the lifetime, of this system, as we will see in the next activity.

In the Hulse-Taylor pulsar system, the stars will spiral into each other for about 100 million years before the neutron stars finally merge. Their current luminosity is still much lower than it will be in the final seconds before the merger occurs, and much lower than the gravitational luminosity you found in the previous activity.

While not a direct detection of gravitational waves, the orbital decay of PSR 1913+16 is convincing evidence that such waves exist in the form that general relativity predicts for them. The evidence was convincing enough that Hulse and Taylor received the 1993

https://phys.libretexts.org/@go/page/31350

Nobel Prize in physics for their work on this binary pulsar system. Since their original discovery, other similar systems have been found that exhibit similar orbital decay.

In the late summer of 2015 scientists finally got a direct detection of gravitational waves. As a bonus, they also got their first direct detection of black holes. The LIGO gravitational wave observatory measured the signal from a merger event between two black holes with masses 29 and 36 solar masses, respectively. The event, which took place in a galaxy 1.4 billion lightyears from Earth, was called GW150914 because it was detected on September 14, 2015. The detection happened before LIGO had even begun its scientific program; the machine was in its engineering shake-down phase, having been turned on not even a week before. This first detection was only one of many to come, and not all of them have been black hole - black hole mergers.

About two years after GW150914, a merger of a different sort was seen. On August 17, 2017, NASA's Fermi Gamma-ray Space Telescope detected a short burst of gamma-rays. The event was dubbed GRB170817A because it was the first gamma-ray burst seen on that date. The timing of this burst, which lasted only about 2 seconds, coincided with the detection of gravitational waves from the source GW170817, seen by both LIGO detectors and a similar instrument in Italy called VIRGO that had been brought online only a few weeks before. The directions of the two sources, rough though they were, were consistent with being from the same object.

Extensive follow-up by many research teams using various instruments spanning the entire electromagnetic spectrum confirmed that the source of the gamma-rays and gravitational waves was the same system: a pair of neutron stars in the galaxy NGC 4993, located at a distance of around 140 million light years from Earth in the constellation Hydra. This was the first time that a source of gravitational waves could be tied to a corresponding source of electromagnetic waves. This event marked the direct detection of gravitational waves from a binary neutron star system as the stars merged to form a black hole - in this event we are seeing the eventual fate of the Hulse-Taylor system discussed above. In addition, the event answered a long-standing question in astrophysics related to the origin of certain elements beyond iron in the periodic table: copious amounts of these elements were ejected during the merger.

Because of this event (and another like it), we now know with fair certainty that the gold in your wedding ring or earrings was ejected into space during a merger between two orbiting neutron stars in our own galaxy. These stars were formed, lived, died and then merged to form a black hole, all long before the Sun and Earth ever existed.

In Math Exploration 10.3 you can explore the gravitational radiation we expect to be emitted by planets in the Solar System. You will probably not be surprised to find out that the amount of energy is very small. Only for much more massive systems, like merging black holes or the neutron star systems discussed above, can we detect the gravitational waves being produced.

Math Exploration 10.3

Of the handful of classical tests of general relativity, all have now been confirmed.

This page titled 10.4: Tests of General Relativity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.5: The Source of Gravity

❓ What Do You Think: Where Does Gravity Come From?

In the descriptions of gravity we have looked at so far, we have left off one important part: What is the source of spacetime curvature? We have described what curvature is, at least in a basic sense, and we have described how curvature affects the motion of particles and the perceptions of space and time that different observers have. But where does the curvature actually come from?

This is like asking where the force of gravity comes from in Newtonian theory. Newton had a ready answer: from masses. Newton's law of gravitation describes precisely what the gravitational force between two point masses will be, given their separation. In Newton's view, gravity comes from the interaction between the two masses (and not from either of the masses by themselves). What is the corresponding law in general relativity?

It turns out that it is possible to write down a precise mathematical relationship in general relativity, akin to Newton's law of gravitation, but the meaning of the relation is not as easy to convey. We will write it down, and you will see why: The so-called Einstein equation can be written as:

$$\mathbf{G} = \frac{8\pi G}{c^4}\mathbf{T}$$

This expression is not very revealing, so we will break it down a bit. The capital G on the right is the familiar gravitational constant from Newton's law, c^4 is the speed of light (raised to the fourth power). The 8 and π are numbers, of course. It is hard to understand how this simple expression leads to spacetime curvature, but that is because we are hiding a lot.

The **G** and **T** terms actually stand for fairly complicated mathematical entities known as tensors. That is why we have written them in boldface; we want them to stand out to convey that they are not just normal numbers. **G** is an object that describes the four-dimensional spacetime curvature. **T** describes the four-dimensional mass–energy distribution. This deceptively simple expression is actually 10 independent equations relating the curvature to the mass–energy distribution. In a similar way, Newton's gravitational law is actually three equations that describe the gravitational forces in the independent x, y, and z directions, but we generally write it as a single equation involving the 3-dimensional force and position vectors. The Einstein equation is something like that, but it describes a more complicated view of gravitation than can be conveyed in just a single (or three!) equation.

Solving the 10 equations contained in the Einstein equation produces the spacetime interval for that particular distribution of mass–energy. For a spherically symmetric, time-independent mass–energy distribution we get the spacetime interval we saw earlier in the chapter. We have already discussed how that describes gravity around Earth and the Sun. It also describes the spacetime around black holes. For other mass–energy distributions we get different curvature described by different spacetime intervals. (See Going Further 10.7: The Einstein Equation for more details.)

Because the mathematics required to solve the Einstein equation are advanced, we have not shown you an example of how this is done. Even the simplest example requires math beyond the level assumed for this unit. However, we did not want to leave our discussion of general relativity without at least giving you a brief description of the connection between mass–energy, the source of gravity, and spacetime curvature, the way gravity gets expressed.

As in Newtonian gravity, the source of gravitational effects in general relativity is mass. However, in general relativity, other forms of energy produce a gravitational effect too, even the energy in gravity itself. That is why we have been careful to say "mass–energy" when discussing the source part of the Einstein equation. This should not be surprising if you recall the mass–energy equivalence from special relativity.

$$E=mc^2$$

General relativity carries over this equivalence from special relativity, as it does many other ideas. But in general relativity we are not limited to inertial frames, where any motion must be unchanging. In general relativity, which was initially intended to be a mere expansion into accelerated motion, what we have found is a radical new way of describing the gravitational interaction that underlies the dynamics of the Universe on large scales. In later chapters we will see that the general relativistic view of gravity leads to truly astounding notions regarding the origin and evolution of the cosmos.

Going Further 10.7: the Einstein Equation

We have just been introduced to two terms: the four-dimensional spacetime curvature, $\mathbf{G_{\mu\nu}}$, and the four-dimensional mass–energy distribution, $\mathbf{T_{\mu\nu}}$. How can we visualize these terms to try to gain a deeper understanding of what they represent?

We introduced the concept of a vector when we discussed velocity, and we went into more detail when we discussed forces. A vector is a physical quantity that has both strength and direction. We can specify its components in the three spatial dimensions: x, y, and z. By specifying the sizes of these components, we also specify the direction that the force vector is pointing.

For example:

- F_{grav} = (10,0,0) indicates a force with strength of 10 N in the x-direction
- F_{grav} = (10,10,0) indicates a force with strength = $(10^2 + 10^2)^{1/2}$ = 14.14 N at an angle 45 degrees between the x and y axes.
- F_{grav} = (10,10,10) indicates a force with strength = $(10^2 + 10^2 + 10^2)^{1/2}$ = 17.32 N at an angle 45 degrees between the x and y axes and also 45 degrees between the x–y plane and the z-axis.

These force vectors are illustrated in Figure B.8.7.

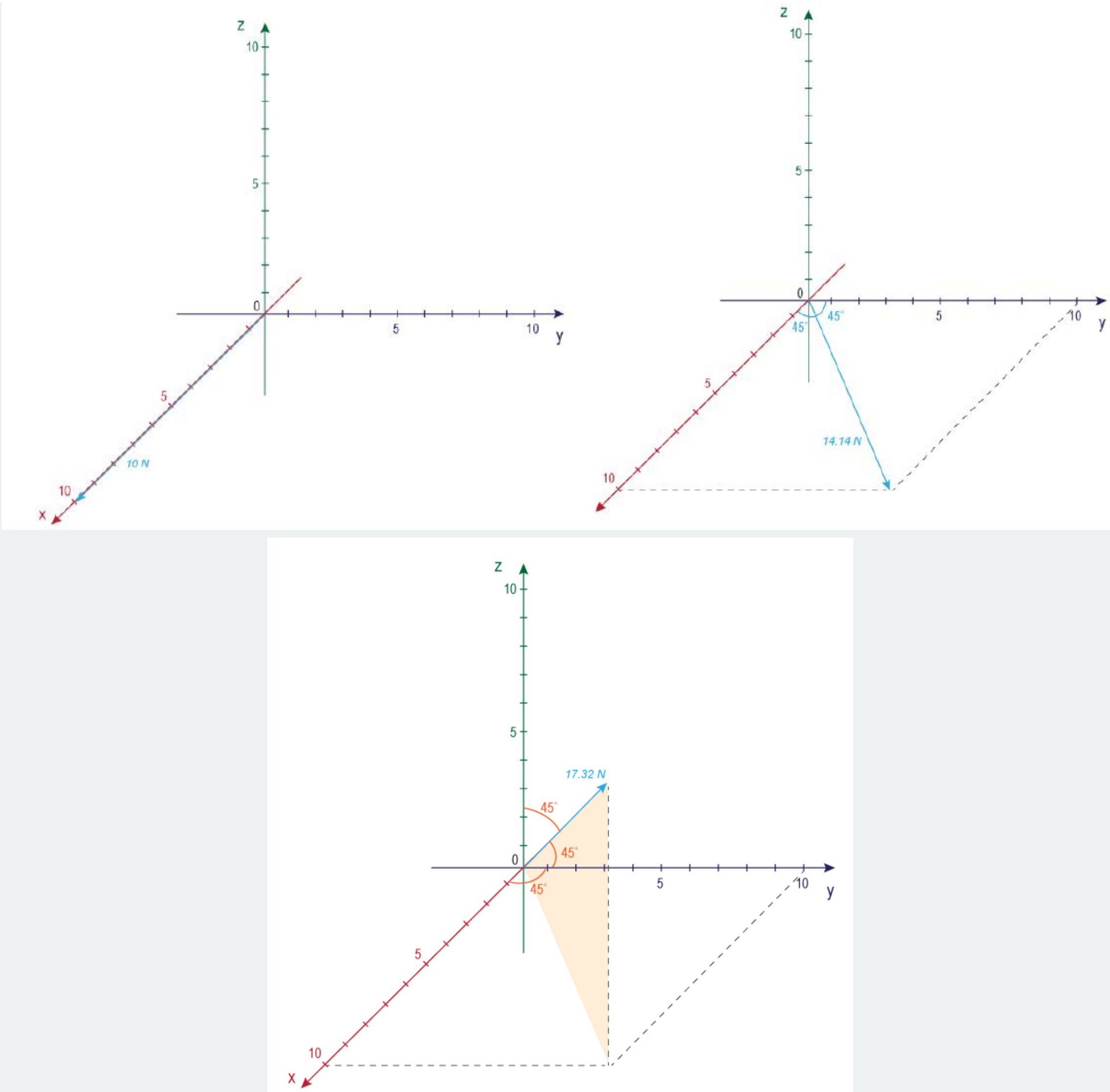


Figure B.8.7: (a) A force vector with magnitude 10 N in the x-direction; (b) a force vector with magnitude 14.14 N at an angle 45 degrees between the x- and y-axes; and (c) a force vector with magnitude 17.3 N at an angle 45 degrees between the x- and y-axes and also 45 degrees between the x-y plane and the z-axis. Credit: NASA/Sonoma State/A. Simonnet

Vectors are useful mathematical tools. They can express forces that point in one particular direction. However, when we use vectors to represent (for example) the gravitational force, we are also assuming that the objects that are being affected can be treated as points. That is to say, we assume they are very, very small, so that the force can act on the objects as though all their mass was concentrated at one tiny point. Furthermore, we can describe the location of any point by its coordinates in three-dimensional space.

But what if we have really large objects? For example, Earth is rather large compared to satellites that are in orbit around it. So approximating the satellites as tiny points compared to Earth works pretty well. (However, as we have seen in our discussion of GPS, there are errors that arise that need to be corrected, which are about 1 part in one billion.) But Earth is not that much larger than the Moon, and both have bulges and craters. We might therefore expect that neither one of these objects looks like a tiny point to the other. Instead, we need to consider the entire Earth–Moon system when we do an accurate calculation of how they affect one another.

In these types of situations, we cannot describe the gravitational force as acting in a specific, linear three-dimensional direction. Instead, we need to specify the components of the gravitational acceleration acting in different directions in four-dimensional spacetime. Then we can account for the curvature correctly. The four-dimensional spacetime curvature **G**, which has components, $G_{\mu\nu}$, is the mathematical way that we express this more complicated situation. Rather than being a series of three numbers (like a vector), it can be represented mathematically by a special type of 4×4 matrix called a tensor.

This is how the components of $G_{\mu\nu}$ are defined:

$$G_{\mu\nu}=\begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix}$$

In this case, the diagonal components of the matrix appear in the expression for curvature that we defined previously:

$$s^2=g_{11}(\Delta x)^2+g_{22}(\Delta y)^2+g_{33}(\Delta z)^2-g_{44}(c\Delta t)^2$$

For special relativity in flat space, the components (g_{11}, g_{22}, g_{33}, g_{44}) = (1,1,1,–1) yielding the familiar expression for the spacetime interval:

$$s^2=(\Delta x)^2+(\Delta y)^2+(\Delta z)^2-(c\Delta t)^2$$

The matter-energy tensor, **T**, which has components $T_{\mu\nu,}$ is a similar 4 × 4 matrix that describes the field associated with the distribution of the matter and energy in spacetime.

$$T_{\mu\nu}=\begin{pmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ T_{41} & T_{42} & T_{43} & T_{44} \end{pmatrix}$$

To be clear, the individual matrix elements in **G** and **T** each describe some aspect of local curvature, in the case of **G**, or the local mass-energy content, for **T**. Putting both of these tensors together into Einstein's equation leads to a series of 10 independent equations - there would be more, but certain symmetries reduce the total number to 10. These equations must be solved simultaneously in order to correctly calculate the gravitational effect (spacetime curvature) created by the corresponding mass-energy distribution. In a few simple cases, these equations can be solved with pencil and paper; however, in all other cases, the solution requires numerical calculations by supercomputers.

This page titled 10.5: The Source of Gravity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.6: Wrapping It Up 10 - Curved Spacetime Around Astronomical Objects

Imagine you are studying a binary system in which a pulsing neutron star is orbiting a supermassive black hole and you are viewing the system from the direction of the black hole. Over time you have detected that the time interval between the pulses from the pulsar is changing. Intrigued, you set about to discover what is happening by studying a graph of the pulsar's periods.

The *Binary Pulsar Simulation* contains four pieces:

- The Options box allows you to select up to two different effects: the Doppler Shift and Gravitational Redshift .
- The Observer View section shows you what the observer is seeing from the binary pulsar system.
- The Simulation View section shows the orbit of the pulsar around its black hole companion and has a button for you to start the simulation.
- The last section is the Graph, which displays the light intensity of the pulsar over time. After running the simulation, you can use your mouse to hover over points on the graph to read off intensities and times for individual data points.

The big arrow near the point marked "A" points in the direction of you, the observer.

When you first start up the activity, the default setting has no options selected.

Play Activity

10.6.1: Part I: The Period of the Pulsar

In this part of the activity, you will measure the period between pulses emitted by the pulsar with no options selected. This will give you the "rest" or "laboratory" period of the pulsar.

Click the Start button to run the simulation for one orbit and examine the graph of the pulses. The letters A, B, C, and D describe different pulsar locations as it orbits around the black hole.

1.

2.

3.

4.

5.

10.6.2: Part II. Doppler Shift

Now select the Doppler Shift box and run the simulation again.

1.

2.

3.

4.

5.

https://phys.libretexts.org/@go/page/31352

Now convert your results from period to frequency:

6.

7.

10.6.3: Part III: Gravitational Redshift

Next, select the gravitational redshift box (and de-select the Doppler shift box) and run the simulation again.

1.

2.

3.

4.

10.6.4: Part IV: Doppler Shift and the Gravitational Redshift

So far in our simulations, we have been looking at the two effects on a pulsar's period individually. But in the real world, both the Doppler shift and gravitational redshift occur simultaneously. In this part of the activity, we will see what happens when we include both effects.

1.

2.

https://phys.libretexts.org/@go/page/31352

This page titled 10.6: Wrapping It Up 10 - Curved Spacetime Around Astronomical Objects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.7: Mission Report 10: Curved Spacetime Around Astronomical Objects

A.

B.

C.

D. Questions to be graded for accuracy:

1.

2.

3.

4.

This page titled 10.7: Mission Report 10: Curved Spacetime Around Astronomical Objects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

10.8: Formulae, Constants, and Conversion Factors for Chapter 10

Formulae	**Relationship**
$g=\frac{GM}{R^2}$	Surface The universal force of attraction between all matter. gravity:
$d=v_0t+\frac{1}{2}at^2$	The length of space between two objects. " id="term-2350855" tabindex="0">Distance and acceleration:
$v=at$	How fast an object moves in a given direction, i.e., the **speed** of an object in a given direction. Velocity differs from **speed** because **speed** is how fast something moves without regard to direction. " id="term-2350855" tabindex="0">Velocity and acceleration:
$t=\frac{t_0}{\left(1-\frac{gH}{c^2}\right)}$	the distortion of time between **inertial frames**. Observers in different inertial frames will measure time to pass more quickly in their own frame than in any inertial frame moving relative to their own. The slowing of clocks that are in motion relative to an observer when compared to clocks at rest with respect to the observer. Predicted by Einstein's **Special Theory of Relativity.** " id="term-2350855" tabindex="0">Time dilation (weak field approximation):
$f=f_0\left(1-\frac{gH}{c^2}\right)$	Gravitational This is the name given to the apparent change in the wavelength of light due to the **Doppler effect**. Scientists know what the regular spectrum of a galaxy should look like (based on the spectrum of light emitted from known elements). If the light waves from a galaxy appear to have shifted towards higher frequency (blue), it is moving towards us, and if they have shifted toward a lower frequency (red), it means the object is moving away. " id="term-2350855" tabindex="0">redshift (weak field approximation, A property of a wave that describes how many wave patterns, or cycles, pass by a point in a given time. Frequency is often measured in hertz (Hz), where one hertz is one cycle per second. " id="term-2350855" tabindex="0">frequency, A quantum (particle) of light or electromagnetic energy. Photons are have zero rest-mass and no electric charge. " id="term-2350855" tabindex="0">photon traveling upward):
$\lambda=\frac{\lambda_0}{\left(1-\frac{gH}{c^2}\right)}$	Gravitational redshift (weak field approximation, The distance between adjacent peaks in a series of periodic waves. Also see **electromagnetic spectrum**. " id="term-2350855" tabindex="0">wavelength):
$f=f_0\sqrt{1-\frac{2GM}{Rc^2}}$	Gravitational redshift (full expression, frequency):
$\lambda=\frac{\lambda_0}{\sqrt{1-\frac{2GM}{Rc^2}}}$	Gravitational redshift (full expression, wavelength):
$d^2=(\Delta x)^2+(\Delta y)^2$	in any right triangle with sides a and b and hypotenuse c, $a^2+b^2=c^2$ " id="term-2350855" tabindex="0">Pythagorean theorem:
$d^2=(R\Delta\theta)^2+\cos^2\theta(R\Delta\alpha)^2$	Distance on a sphere:
$d^2=(\Delta x)^2+(\Delta y)^2+(\Delta z)^2$	Pythagorean Theorem in 3-D:

Formulae	**Relationship**
$s^2=(\Delta x)^2+(\Delta y)^2+(\Delta z)^2-(c\Delta t)^2$	the four-dimensional system employed in special relativity that merges three spatial dimensions with one dimension of time and describes all events as points in that system: *(t, x, y, z)*. Three coordinates, *x*, *y*, and *z* describe the position of an event in space, and one, *t*, describes its position in time. " id="term-2350855" tabindex="0">Spacetime interval in space in which the **Pythagorean Theorem** holds. " id="term-2350855" tabindex="0">flat space:
$s^2=\left(1-\frac{2GM}{rc^2}\right)^{-1}\left[(\Delta x)^2+(\Delta y)^2+(\Delta z)^2\right]-\left(1-\frac{2GM}{rc^2}\right)(c\Delta t)^2$	the four-dimensional system employed in special relativity that merges three spatial dimensions with one dimension of time and describes all events as points in that system: *(t, x, y, z)*. Three coordinates, *x*, *y*, and *z* describe the position of an event in space, and one, *t*, describes its position in time. " id="term-2350855" tabindex="0">Spacetime interval in spherically curved space:
$\theta=\frac{2GM}{bc^2}$	Angle of deflection of light:
$P_{gw}=\frac{2}{5}\left(\frac{GM}{Rc^2}\right)^5\left(\frac{m}{M}\right)^2\left(\frac{c^5}{G}\right)$	Power emitted by ripples in spacetime caused by the changing of mass distributions. " id="term-2350855" tabindex="0">gravitational waves:
$\mathbf{G}=8\pi G\,\mathbf{T}/c^4$	Einstein equation:

This page titled 10.8: Formulae, Constants, and Conversion Factors for Chapter 10 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

11: Black Holes

Chapter 11 centers on black holes, including their properties and how we know they exist. Motivated by groundbreaking observations of our own Milky Way's supermassive black hole, active galaxies, as well as solar mass black holes, you will examine the physical characteristics that result from the strong gravitational fields of these not-so-uncommon exotic objects. You will also engage in calculations of black hole sizes, temperatures, densities, and lifetimes.

The supermassive black hole at the core of supergiant elliptical galaxy Messier 87, with a mass about 7 billion times that of the Sun, as depicted in the first false-colour image in radio waves released by the Event Horizon Telescope (10 April 2019). Visible are the crescent-shaped emission ring and central shadow,[19] which are gravitationally magnified views of the black hole's photon ring and the photon capture zone of its event horizon. The crescent shape arises from the black hole's rotation and relativistic beaming; the shadow is about 2.6 times the diameter of the event horizon. (CC BY 4.0; Event Horizon Telescope)

This page titled 11: Black Holes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

11.0: Black Holes Introduction

Video Transcript

ESOcast

[Narrator] In an unprecedented 16 year long study using several of ESO's flagship telescopes, astronomers have produced the most detailed view ever of the surroundings of the monster lurking at our galaxy's heart: a supermassive black hole. The research has unraveled the hidden secrets of this tumultuous region by mapping the orbits of almost 30 stars.

[Narrator] This is the ESOcast, cutting edge science and life behind the scenes of ESO, the European Southern Observatory, exploring the far reaches of the universe with our host Dr. J., AKA Dr. Joe Liske.

[Dr. J.] Hello and welcome to the second episode of the ESOcast. Today we have a very cool piece of science for you. A team of German astronomers, with characteristic precision and patience, have spent 16 years mapping out the motions of 28 stars orbiting the very center of our Milky Way galaxy. Now, astronomers have believed for quite a while that the center of our galaxy is the site of a supermassive black hole. Black holes are a consequence of general relativity. They are objects that are so dense and whose gravity is so strong that not even light can escape them. These observations that we're going to how you today are the best evidence yet that black holes are not just theoretical constructs but actually do exist in reality. This is truly a milestone result.

[Narrator] Observers under dark skies far from the bright city lights can marvel at the splendor of the Milky Way arching in an imposing band across the sky. Zooming in towards the center of our galaxy, about 25,000 light-year away, you can see that it is composed f myriads of stars. This is a pretty impressive sight, but much is hidden from view by interstellar dust and astronomers need to look using a different wavelength, the infrared, that can penetrate the dust clouds. With large telescopes, astronomers can then see in detail the swarm of stars circling the supermassive black hole in the same way that the Earth orbits the Sun. The galactic center harbors the closest supermassive black hole known and the one that is also the largest in terms of its angular diameter on the sky making it the best choice for a detailed study of black holes.

[Dr. J.] So what this team did was that, at various points of the past 16 years, they kept taking images of the very central region of the Milky Way. Now, from these images they were able to map out the motions of a total of 28 stars, and what these motions showed was that these stars aren't just moving about randomly, but that they are clearly orbiting a very massive central object - and the point is that this central object is completely unseen. Now, from the motions it's also possible to deduce the mass of the central object. It came out to be a little over 4 million times the mass of the Sun. Now, what's more, that enormous mass has to fit into a tiny little volume, and so one cannot escape the conclusion that the central object really is a black hole.

[Narrator] The observing campaign started with observations made in 1992 with a SHARP camera attached to ESO's 3.5 meter New Technology Telescope, NTT, housed at the La Silla observatory in Chile. More observations have subsequently been made

in the last few years using two instruments mounted on ESO's 8.2 meter Very Large Telescope, VLT. Over the 16 years of this study, ESO's telescopes have stared at this one region for 50 full nights.

[Dr. J.] This new research marks the first time that so many of these central stars have had their orbits determined so precisely. The data also revealed a lot about the characteristics of these stars and how they must have formed. For one of the stars, the astronomers were even able to follow it for a complete orbit. The star approached the central black hole to within just one light-day – that's just five times the distance between Neptune and the Sun. Professor Reinhard Genzel from the Max Planck Institute for Extraterrestrial Physics in Germany is the leader of the team that made the discovery. Reinhard, why is it so important to study the center of the Milky Way?

[Prof. Genzel] Well you see, the Milky Way center is one of the most important laboratories we have to study, in very great detail, what's happening in the centers of galaxies - in much more detail than we can ever hope to do in all other galaxies. Yet, here we are: we can study whether there's a central black hole, what happens around it, and so forth; all very general issues which we would like to explore and which we cannot really study that much in detail in other galactic nuclei.
[Dr. J.] Doctor Stefan Gillessen is the first author of the paper reporting the study. So Stefan, tell us: what's the most important result you obtained?

[Dr. Gillessen] The most important result of our research really is that we now have empirical evidence for the existence of a massive black hole in the center of our Milky Way. The mass of this black hole is around 4 million solar masses and we know the mass at the percent level.

[Dr. J.] This is, of course, and amazing result. But the team doesn't plan to stop here. Now, in the past they've used the novel technique of adaptive optics to remove the blurring effects of the atmosphere. In the future, they plan to do even better, and to get even higher resolution images by using another new technique called interferometry. This is where you combine the light from two or more of the VLT's Unit Telescopes together. So Reinhard, what's the next step?

[Prof. Genzel] Well, you see, at this point we really are fairly sure that there is a massive black hole at the center of our Milky Way, and the next thing we want to actually play with it, play with it in the sense that we want to use it as a tool to test whether general relativity, the theory of Einstein, is actually wrong or right.

[Dr. J.] Wow, playing with a black hole to test relativity – that's pretty cool stuff. I'm Dr. J., signing off for the ESOcast. Join me again next time for another cosmic adventure.

Black holes have captivated the public imagination like no other astronomical object. Most of us have seen them depicted in movies, on television and in books. Unfortunately, the popular treatments given to black holes do not usually do them justice. What's more, they tend to leave people completely misinformed as to their nature. In this chapter we hope to give you a more accurate picture of what black holes are and how they fit into the Cosmos.

This page titled 11.0: Black Holes Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

11.1: What Are Black Holes?

Learning Objectives

- Explain the following laws within the Ideal Gas Law

Black Holes in Galaxies

The modern notion of a black hole has its origins in general relativity. In fact, the term "black hole" was itself coined by the American physicist John Archibald Wheeler (1911–2006), one of that theory's greatest proponents. Wheeler and his students are largely responsible for the great advances in general relativity that took place after World War II. However, in its simplest form, the idea of black holes predates these advances by almost two centuries.

Toward the end of the 18th century, two scientists had independently considered the possibility of an object whose gravity was so strong that light would be unable to escape it—the essence of what we mean by "black hole." One of these scientists was John Michell (1725–1793), in England. Michell was a colleague of the much more widely known Henry Cavendish (1731–1810), famous for the "Cavendish experiment" to measure the gravitational constant , G. In fact, Cavendish used an experimental apparatus devised by Michell for his classic experiment. In 1784 Michell sent a letter to Cavendish in which he speculated about an object whose escape velocity might exceed the speed of light:

> *"...if the semi-diameter of a sphere of the same density with the Sun were to exceed that of the Sun in the proportion 500 to 1, a body falling from an infinite height towards it, would have acquired at its surface a greater velocity than that of light, and consequently, supposing light to be attracted by the same force in proportion to its vis inertiae, with other bodies, all light emitted from such a body would be made to return towards it, by its own proper gravity."*

This is the earliest known discussion of such an object. At around the same time, the French scientist Pierre-Simon LaPlace was entertaining similar ideas. He published them in a book, which he called *Exposition du Systeme du Monde,* in 1796.

Both Michell and LaPlace were well versed in the Newtonian theory of gravity, and they knew about the concept of escape velocity. The escape velocity from an object of mass M and radius R is given below.

$$v_{esc} = \sqrt{\frac{2GM}{R}}$$

If we set the escape velocity to the speed of light (c) and rearrange the expression, we find that the size (R) of such an object - one with $v_{esc} = c$ - given that it has a mass M, is the following.

$$R = \frac{2GM}{c^2}$$

This is the maximum size R an object of mass M must have if its escape velocity is to equal the speed of light. Of course, if the object is smaller, then light still cannot escape. Also, the size of these objects scales linearly with mass, so if we double the mass of an object we double the minimum size needed for it to trap light. Any object whose size is less than or equal to that given by the previous expression will be a so-called "dark star," as Michell referred to them. No light will be able to leave its surface, so it will be completely dark. The notion of a "dark star" (as opposed to black hole) is purely Newtonian. We have not made any reference to general relativity at all yet. However, in section 11.2, we will see that the characteristic size of a black hole, as determined by general relativity, is described by this same expression.

Size of Black Holes

Worked Example:

1. What is the size of a black hole with the same mass as the Sun, 2×10^{30} kg?

- Given: $M = 2 \times 10^{30}\ \mathrm{kg}$
- Find: R, the radius in meters
- Concept(s): R = 2GM/c^2 where the gravitational constant, \(G = 6.67\times 10^-11}\, {\rm Nm^2/kg^2,}\) and the speed of light, $c = 3 \times 10^8\ \mathrm{m/s}$.
- Solution: $R = (2)(6.67 \times 10^{-11}\ \mathrm{Nm^2/kg^2})(2 \times 10^{30}\ \mathrm{kg})/(3 \times 10^8\ \mathrm{m/s})^2 = 2.95 \times 10^3\ \mathrm{m}$

For an object with the Sun's mass to be a black hole, or in the Newtonian view a "dark star," such that light is unable to escape it, it would have to have to be about 6 km across - or less.

Questions:

1.

2.

What if the Sun Were Replaced With a Black Hole?

Our discussion thus far has only considered Newtonian gravitation, and so we should expect that these objects behave gravitationally just as other objects do in Newtonian gravity.

1.

2.

3.

At the time of Michell and LaPlace, most scientists thought of light as being composed of waves, not particles—though Michell was thinking in terms of particles. No one could understand how gravity would be able to affect the motion of waves, and there was no good understanding of the equivalence principle. As a result, nobody really took the ideas of “dark stars” seriously. It turns out the same was true for black holes after general relativity was published, including with Einstein himself. Well, at least at first.

We have deduced the size of black holes (“dark stars” in the Newtonian view) by setting an object’s escape velocity to be the speed of light. The Newtonian intuitions gained this way can be helpful to a certain point; however, the Newtonian form of gravity is not

https://phys.libretexts.org/@go/page/31356

the proper one to use when thinking about black holes. It leaves out much of what we know to be true about gravity. To get a full picture of what black holes are like, we should use the ideas of general relativity. That is what we do in the next section.

This page titled 11.1: What Are Black Holes? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

11.2: Spacetime Near Black Holes

Learning Objectives

- You will understand that in general relativity black holes affect space and time.
- You will understand that outside observers never actually see anything cross the event horizon because of time dilation.
- You will understand that outside observers see a clock get redder, dimmer, and tick slower as it approaches a black hole.
- You will be able to describe "spaghettification" and compression that occurs when something falls into a black hole.
- You will understand that observers falling into a black hole do cross the event horizon and that time runs differently for them compared to outside observers.
- You will understand that material that falls into a black hole cannot get back out.
- You will understand that an object can orbit a black hole in a variety of ways without inevitably falling into the black hole.
- You will understand that it is only close to black holes that the GR effects become important; far from the event horizon Newtonian gravity is a good description.
- You will understand that while black holes are primarily described by their mass, including spin and charge provides a complete description of them.

What Do You Think: The Shape of Black Holes

11.2.1: The Geometry Near Black Holes

In general relativity, the spacetime around stars or planets is described by a spherically symmetric spacetime interval. The spacetime interval tells us what the separation is between points in spacetime; it is analogous to finding the distance in space using the Pythagorean theorem in flat space, but including a time part, and accounting for the curvature of spacetime.

We can make the spherical symmetry more explicit by separating out the angular and radial parts of the spacetime interval, where we are using the fact that any point in space can be described by its distance from the origin in a particular direction. In Figure 11.1, we compare different types of coordinate systems (Cartesian: (x,y,z,t) vs. spherical: (d,θ,A,t)).

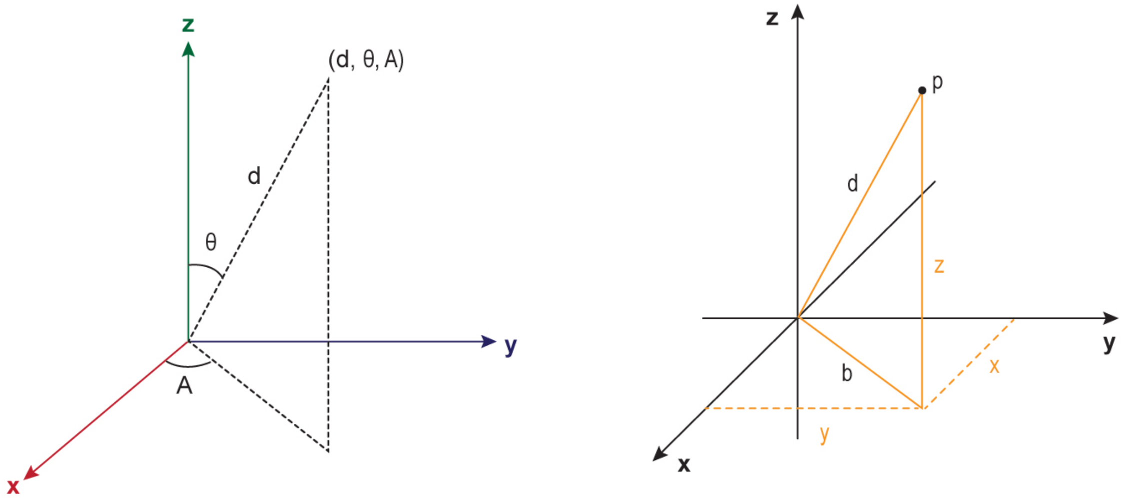


Figure 11.1: Points in space can be described by an xyz position, called Cartesian coordinates (after the French mathematician René Descartes), or by an dθA set of coordinates, as shown. The angles θ and A are like longitude and latitude, respectively. However, since we are not restricting ourselves to the surface of a globe, as on Earth, we must also specify how far from the origin the point is located: that is what the coordinate d does. Credit: NASA/SSU/Aurore Simonnet

In spherical coordinates, the angle θ is like latitude on Earth and the angle A is like longitude. For a spherically symmetric space (one that is the same in every direction) the orientation of the poles is arbitrary. Only the distance from the origin is important. Under this circumstance we can ignore the angular parts of the coordinates in spacetime because they don't tell us anything meaningful about the geometry of the space.

This means for a spherically symmetric spacetime (one in which the spatial part forms a spherically symmetric 3-space) we do not need to worry about any terms in the spacetime interval with θ and A. They don't change when developing the spacetime interval from the 3-space interval. We can focus solely on the terms with the radial distance (d) and time (t), since those are the only things that affect the geometry. With these considerations, the spacetime interval (s^2) around a spherically symmetric object is that stated below.

$$s^2 = \left(1 - \frac{2GM}{dc^2}\right)^{-1} (\Delta d)^2 - \left(1 - \frac{2GM}{dc^2}\right)(c\Delta t)^2 + r^2 \Delta\Omega^2$$

Here G is the universal gravitational constant, c is the speed of light, M is the mass of the object, d is the distance from the origin, and t is the time. The angular part is given by the last term, with which we do not have to concern ourselves; it is the same as for flat space and does not contribute to spacetime curvature.

The spacetime interval may look complicated, but the curvature depends only on the distance from the origin, d, and time, t. Those are the only terms where we find a difference from the flat space interval, the one used for special relativity. The curvature has no dependence on the direction considered. This is clear from the lack of modifications to the angular term in the expression for the interval. That is what is meant by spherically symmetric.

The interval describes the Schwarzschild geometry, discovered by the German astronomer Karl Schwarzschild (1873–1916) in 1915. It applies to any spherically symmetric gravitating system, including non-spinning black holes—we will have a look at spinning black holes later.

Figure 11.2: Karl Schwarzschild was a German astrophysicist. He was the first person to solve Einstein's equations for a spherically symmetric system. Credit: Bildarchiv Preussischer, Kulturbesitz, Berlin

An interesting historical aspect of Schwarzschild's discovery is that it was made during his service in the German army during World War I. Schwarzschild had volunteered for the military at the outbreak of the war in 1914. He was stationed on the Russian front as an artillery officer when he wrote his two papers on general relativity in the following year—the very year Einstein published the theory. The papers presented the solution we have presented, that for a spherical, nonrotating body.

Because Schwarzschild was occupied at the front, he was unable to present his papers at scientific meetings in Berlin. Instead, he sent copies of them to Einstein, who presented them in his absence. The papers were well received, and Einstein wrote to Schwarzschild that he had been surprised to see that the solution could be so simple.

While we now use Schwarzschild's solutions to describe the spacetime around black holes, Schwarzschild himself stated that he thought their description of objects whose escape speed exceeded that of light was mere mathematics, with no basis in reality—an opinion which Einstein also held. Schwarzschild never had the chance to explore the idea further. While on the Russian front, he contracted pemphigus, an autoimmune disease in which the immune system attacks the skin. He died of the condition in the spring of 1916, shortly after his relativity papers had been presented by Einstein.

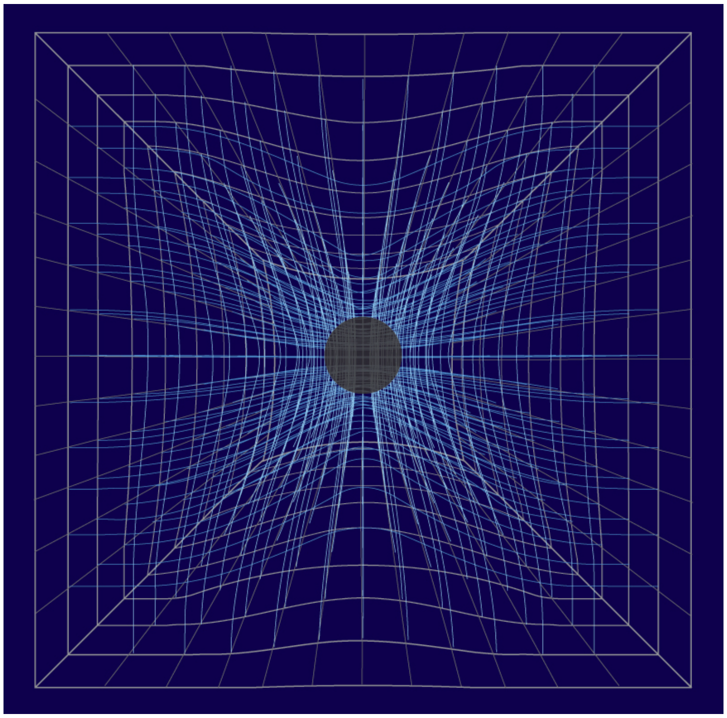


Figure 11.3: We can visualize the spatial part of the stretching of spacetime as shown in this diagram. The stretching occurs in all three spatial dimensions and is spherically symmetric. The time part of the distortion cannot be shown in this diagram, of course. Credit: NASA/SSU/Aurore Simonnet

The Schwarzschild spacetime interval applies to any spherically symmetric gravitating object. We can use it to study the orbits of planets around the Sun, and even the bending of light as it passes through the Sun's gravity. However, there is something a little peculiar about the Schwarzschild geometry. Notice the curvature terms in the large parentheses. They contain the ratio:

$$R_S = \frac{2GM}{c^2}$$

You might have noticed that this expression is the same one we derived in Section 11.1 for the size of a black hole. In the Newtonian derivation, it is the point where the escape speed becomes equal to the speed of light. This quantity has been given the name **Schwarzschild radius**, R_S, and it is quite important, as we will see. We can rewrite the Schwarzschild interval in terms of the Schwarzschild radius, and its importance will become more apparent.

$$s^2 = \left(1 - \frac{R_S}{d}\right)^{-1} (\Delta d)^2 - \left(1 - \frac{R_S}{d}\right) (c\Delta t)^2 + r^2 \Delta\Omega^2$$

Look at what happens to the curvature terms as the radial coordinate (d) approaches R_S: The time curvature term goes to zero, but the spatial (radial) curvature does not. If $R_S = d$, then $R_S/d = 1$ and $1 - R_S/d = 1 - 1 = 0$. In fact, since the radial curvature term is the reciprocal of the time curvature term, it must grow without bound as r approaches R_S, because $1/0$ is infinite. What does this mean? Clearly something interesting is happening as we approach R_S.

11.2.2: Time Dilation

In the next activity, we will examine what happens as a clock moves closer to a black hole. The effects become particularly pronounced as it approaches the Schwarzschild radius.

? Time Dilation

A. Predictions

First, predict what you think will happen:

1.

2.

https://phys.libretexts.org/@go/page/31357

3.

https://phys.libretexts.org/@go/page/31357

4.

B. Observations

Now test your predictions by using the slider at the bottom to bring the clock closer and farther away from the Schwarzschild radii of a black hole.

Play Activity

1.

2.

3.

4.

In the last activity you saw what happens to an object falling into a black hole as seen from an observer far from the black hole. Several different effects were evident. First, the object was gravitationally redshifted as it fell deeper into the gravitational potential well, where the gravity is stronger. This made it shift in color to appear redder. The clock also became dimmer and time seemed to slow down. All of these effects are caused by the time dilation of an object's clocks as seen by an observer outside the region of strong gravity.

The slowing of clocks is very dramatic. So much so that as the black hole event horizon is reached, the clocks are seen to cease running altogether; according to an outside distant observer, time ceases to flow at the event horizon of a black hole. So, according to an outside observer, nothing ever quite crosses the event horizon. Instead, an object falling in slows and seems to freeze just as it reaches the event horizon. It also dims to the point of invisibility at the same time. So does anything ever actually fall into a black hole? Keep that question in mind as we undertake the journey from the perspective of the object that is falling.

11.2.3: Falling Into a Black Hole

? Falling Into a Black Hole

In the previous activity, you discovered the effects of time dilation on an object falling into a black hole as seen from the point of view of an observer outside of the region of strong gravitational acceleration of the black hole.

In this activity, we will consider the viewpoint of an object falling in toward the black hole, in this case a circular ring of balls.

1.

2.

Now play the activity to observe what happens as a circular ring of balls falls into a black hole.

Play Activity

3.

If you are an intrepid space traveler who decides to take a trip into a black hole, the last activity should have convinced you that you do indeed fall all the way to the center in a finite time. You do not freeze above the event horizon, with clocks completely stopped. That is the viewpoint of someone far outside the black hole, but remember, they experience space and time quite differently from you. That is one of the fundamental results of relativity.

On the way to your doom, you experience some unpleasant side effects of the intense gravity near the black hole. One of these stems from the increasingly strong tidal effects you encounter. Tides, like ocean tides on Earth, are caused by the difference in the strength of gravity across an object: the side of Earth facing the Moon feels stronger gravity than the opposite side, and this difference, when added up over the entire surface of Earth, causes a bulge in the height of sea water that is nearest to (and farthest from) the Moon. The same thing happens to an object falling into a black hole, but on a much more dramatic scale.

Imagine yourself falling feet-first into the black hole. At first you will not register any difference in acceleration across your body. However, if a ring of free-falling balls surrounds you, you will see them begin to separate, as shown in Figure 11.4. This is because the balls nearest to the black hole experience a somewhat larger gravitational effect than the balls farthest from the black hole. As a result, they fall faster than the balls farther away. This differential gravitational effect (gravitational tide) is tiny if you are quite far from the black hole, but it becomes larger as you fall closer to the black hole. By the way, this is not any different for non-collapsed objects like Earth, the Moon, Sun and so on. That is why even the tiny little Moon can exert tidal effects on Earth.

There will eventually come a point where you begin to feel that your feet are being attracted much more strongly than your head toward the center of the black hole, tugging on you to separate your head from your toes. Farther in, the tides would grow large enough that you would begin to feel uncomfortable. As you fell deeper and deeper, the effect would grow to the point that it would begin to rip you apart, stretching you as if you were undergoing some sort of medieval torture. In fact, you would eventually be stretched to the point of breaking. Your feet would be falling so much faster than your head that your body could not overcome the tension; it would break. This would happen over and over again, on smaller and smaller scales. The remaining bits of you would be stretched and rendered into tinier bits, and those bits would in turn be ripped apart as the gravity on their side nearest to the black hole increasingly exceeded the gravity on their far side.

Eventually, even the molecules of which you are made would be ripped asunder, as would the atoms in those molecules. This process would continue to arbitrarily small scales as you approached the center of the black hole, where all matter would be destroyed by the intense gravitational tides.

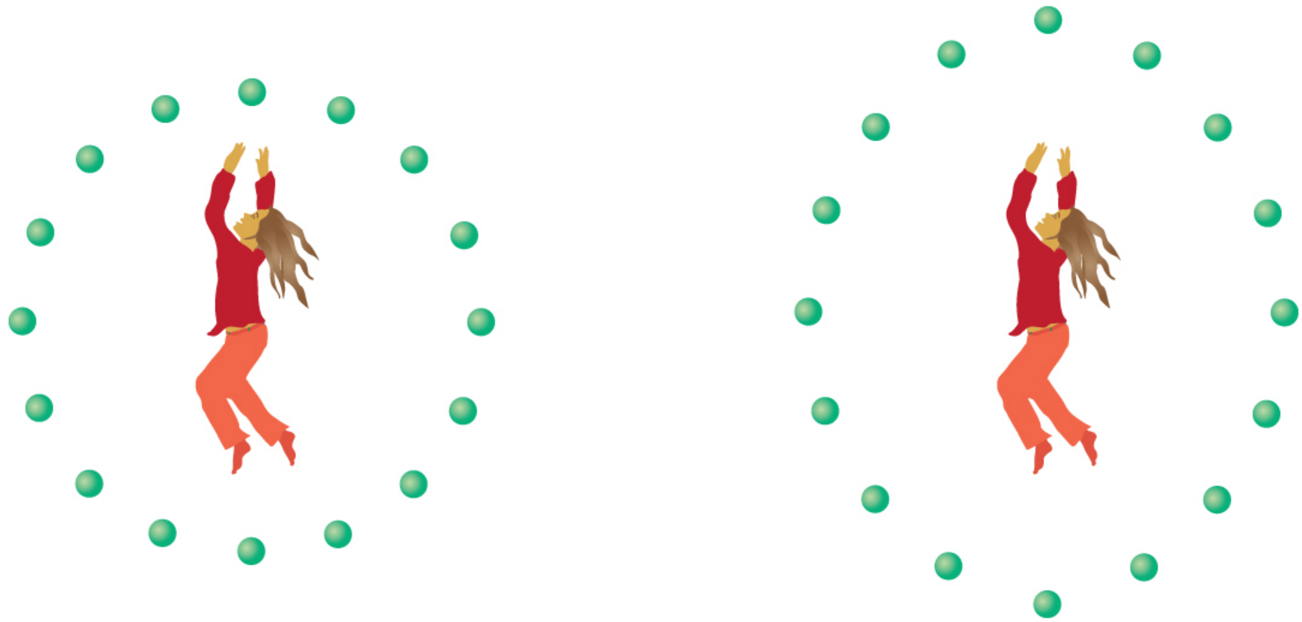


Figure 11.4: If you fell freely into a black hole while surrounded by a ring of freely falling balls, you would eventually begin to see the balls deviate from their initial circular arrangement. This is because of the growing tidal forces of the black hole. Eventually the tides would be strong enough to begin to rearrange you and your atoms in a manner similar to the rearrangement of the balls. Credit: NASA/SSU/Aurore Simonnet

Tidal stretching is not the only unpleasantness in store for anyone venturing into a black hole. Not only does the gravitational acceleration increase the closer you get to the center, but it changes direction as well. Gravity always points directly toward the center of the black hole, but that is different in different places. The gravitational acceleration on your left side is not exactly in the same direction as the acceleration on your right side. When you are far from the center of attraction, the two directions are nearly the same—consider the direction to Earth's center from your left and right shoulders as an example. However, as you approach the center the difference in direction begins to make itself felt. The strong gravity begins to pull your shoulders inward toward your center. You are squeezed from the sides. This effect, too, just as the tides, grows stronger as you get closer to the center. Eventually you are crushed from the sides by the converging attraction of gravity.

Figure 11.5: If you were to fall feet-first into a black hole, the stronger gravitational force at your feet and weaker gravitational force at your head would stretch you out like a thin strand of spaghetti. Credit: NASA/SSU/Aurore Simonnet

This process of being crushed from the sides while being stretched from head to toe as you approach the center of a black hole has been given the fanciful name **spaghettification**, since you would tend to be drawn out into a long, thin noodle-like shape -

spaghetti - if you fell into a black hole. It is not a pleasant notion to contemplate.

11.2.4: Orbiting a Black Hole

A person need not suffer the dramatic (and fatal) effects encountered upon falling into a black hole to see the effects of general relativity. Even in the relatively much weaker gravity near the Sun, the orbit of Mercury is slightly different from what Newtonian gravity predicts it should be. Objects orbiting a black hole also have modified orbits, and much more profoundly so. We will explore some of these effects in the next activity.

Orbits Around a Black Hole

In this activity, you will use a simulation created by John Walker and presented by the HubbleSite team (NASA/STScI/eduweb) to explore the different possible types of orbits that are possible around a black hole.

Your challenge is to adjust the position and velocity of an object to create four different orbits. Describe what combinations of position and velocity are needed to achieve each type of orbit. For example, a close distance and a small velocity might cause an object to fall into a black hole.

NOTE: This activity does not currently work properly. We are aware and correcting the issue.

Play Activity

1.

2.

3.

4.

In the last activity, you studied how objects are able to orbit a black hole. There were some orbits that did not allow the object to continue orbiting forever; eventually it would cross the event horizon of the black hole. Once inside the horizon, no object can escape. On the other hand, you should have found that some orbits are stable. An object in those orbits can persist outside the black hole indefinitely, and it will not fall in. So black holes do not suck objects in. As you have seen with other gravitating bodies, depending on the combination of distance and velocity, an object can have an array of different orbits. Newtonian gravity and general relativity agree on this.

However, Newtonian gravity and general relativity make very different predictions for orbits when gravity becomes strong enough. In Newtonian gravity, all orbits are stable. Near a black hole, general relativity predicts that some of these orbits are not stable. While an object need not fall into a black hole, general relativity predicts that it will if it ventures too close, even if it does not initially cross the event horizon. But if it remains far enough away, and object can pass by the black hole. It can have an elliptical orbit, or it can have a more complicated orbital pattern.

Such possibilities are different from that predicted by Newtonian gravity, which cannot explain the complicated “spirograph” orbits, and which also allows any object to escape no matter how close it approaches the center of attraction. This is not at odds with our discussion of “dark stars” at the beginning of the chapter. In that discussion, we asked how small an object should be for its escape velocity to be the speed of light, we did not consider an object falling in from infinity. Such an object will escape from a classically conceived “dark star” because it will be able to travel faster than light. However, we know that relativity does not allow for that case: no object can travel faster than the speed of light, and so no object can emerge from the event horizon of a black hole.

Earlier we considered the possibility of replacing the Sun with a black hole and asked what the effect would be on Earth’s orbit. Are there any additional effects we must take into account from general relativity that will cause us to change our conclusion? In the case of weak gravity, as is the case far from the Sun in the Solar System, general relativity and Newtonian gravity must give the same results. After all, Newtonian physics works perfectly well in the outer part of the Solar System. If general relativity is correct then it must duplicate the results of Newtonian gravity in the areas where Newtonian physics works. As a result, our conclusion that Earth’s orbit would not change if the Sun were replaced with a black hole remains valid. In fact, for the limiting case in which gravity becomes weak, general relativity simply reduces to Newtonian gravity. Both theories give exactly the same results in that limit.

11.2.5: Black Hole Properties

Thus far, we have only considered the effect of mass on the spacetime near a black hole. That was the first solution found for the Einstein equations, as we have already discussed. But might there be other properties of a black hole that can affect the spacetime close to it? It turns out that there are only two.

The first is charge. You could imagine that if a black hole attracted a lot of electrons, for example, then it would eventually build up a large amount of negative charge, and this would create a strong electric field similar to the field created by rubbing a balloon on your head to create “static electricity.” However, the very fact that such a black hole would create a strong electric field means that charged black holes are not likely to exist. In our example, the electric field from all the negatively charged electrons would attract positively charged protons, and they would then fall into the black hole, neutralizing its charge. The same thing would happen if a lot of positively charged particles fell into the black hole: they would attract negative charges that would cancel the total charge. For this reason, we do not expect to observe any black holes in which electric charge plays an important part.

The only other property of a black hole that can affect the nearby spacetime is spin. Like a normal star, a black hole can rotate. In fact, we expect black holes to be rotating. We will discuss why this is the case in Section 11.3. But how does rotation affect the spacetime curvature? The answer, of course, is contained in the spacetime interval for a rotating black hole, first derived by Roy Kerr in 1963. This Kerr spacetime interval differs somewhat from the Schwarzschild interval we have discussed. First of all, it is not spherically symmetric. Rotation picks out a preferred direction, namely that of the spin axis. Consider how the spinning Earth looks different from different directions (see Animated Figure 11.6): from above the poles you see one particular view, and that is completely different from the view you have above the equator or point in between.

However, if you change your position *around* the axis, the view does not change. We call this type of symmetry, where our view of space depends on our latitude but not on our longitude, azimuthal symmetry. Sometimes it is called cylindrical symmetry or longitudinal symmetry, but these usages are less common. Cylinders and disks both have azimuthal symmetry. So do spinning objects like Earth and spinning (Kerr) black holes, and the spacetime around them reflects this symmetry.

Play Animation

Figure 11.2.1: Animated Figure 11.6: Earth looks different depending on the point from which you view it. If you are directly above one of the poles, you see only one hemisphere, and it appears to spin around its center. If you view it from above the equator you will see the entire surface of Earth pass by, but you can only see one half of it at a time. If you are placed intermediate between the equator and one of the poles you will see a hybrid view: Some of the surface will always be visible, some never visible, and some will be visible part of the time, when it is facing your direction. So a spinning Earth is not spherically symmetric because it can look different to you depending on your viewing location. Credit: Animation - NASA/SSU/Kevin John

If we view a black hole that is spinning, several differences from a nonspinning black hole become immediately apparent. These differences are all described by the Kerr spacetime interval. However, with the addition of even something as simple as rotation, the interval becomes more complicated than we would like to deal with in these modules (books on introductory general relativity discuss the Kerr metric for those interested). Instead we will investigate some of the qualitative changes that occur when we add spin to a black hole.

One difference has to do with frame-dragging. For Earth, the effect of its rotation on orbiting satellites was described in Going Further 10.5: Gravity Probe B. A similar effect happens near a spinning black hole, except that it is much stronger than for Earth. The effect of the dragging of frames near a spinning black hole becomes so strong that at some distance it becomes impossible for any object to remain at rest: it must rotate with the black hole. The region where this is true is called the ergosphere of the black hole. In Figure 11.7, you can see the ergosphere depicted; it extends out along the equator of the black hole and is smaller along the spin axis. The shape of the ergosphere is that of an oblate spheroid, the shape you would get if you compressed an inflated spherical ball from its sides. All spinning objects tend to have this shape. Earth and the Sun both do, for example. For slowly spinning things, the departure from pure sphericity is quite small. For spinning black holes the departure can be very large.

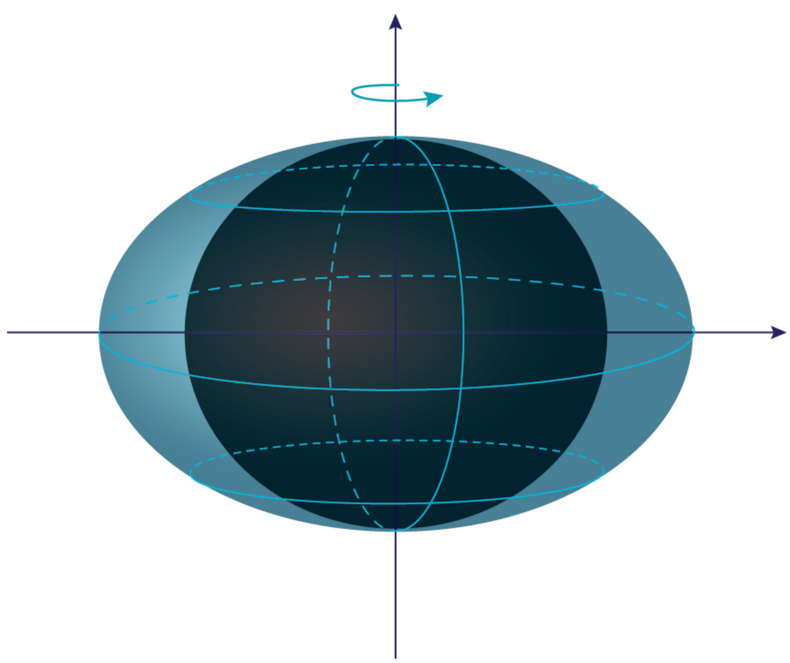


Figure 11.7: A spinning black hole has a different appearance from one that does not spin. It contains an ergosphere (depicted in blue), where spacetime is dragged around by the motion of the hole. Inside that region is a horizon, similar to the one for a nonspinning black hole, but with a slightly oblate shape. Credit: NASA/SSU/Aurore Simonnet

The black hole spin has a strong effect on the surrounding spacetime, and that in turn has a strong effect on the orbits of particles around the black hole. If we confine ourselves to orbits in the equatorial plane of the black hole, then orbits in which the particle moves with the spin of the hole are markedly different than orbits for which the particle moves opposite the spin.

For orbits that are not confined to the equatorial plane, the particle motions become complicated. Such orbits do not have to conserve angular momentum , for example. So a particle that approaches the black hole from some angle can be dragged by the spin of the black hole into an orbit with a completely different orientation. In fact, the effects of the spinning spacetime can cause an orbiting particle to circle the spin axis many times, while it makes only a single circuit from above the equator to below and then back again. To an outside observer such orbits look like a helix, as shown in Figure 11.8. The details of these orbits are not easily computed by simple computer exercises like the ones we are using in these modules.

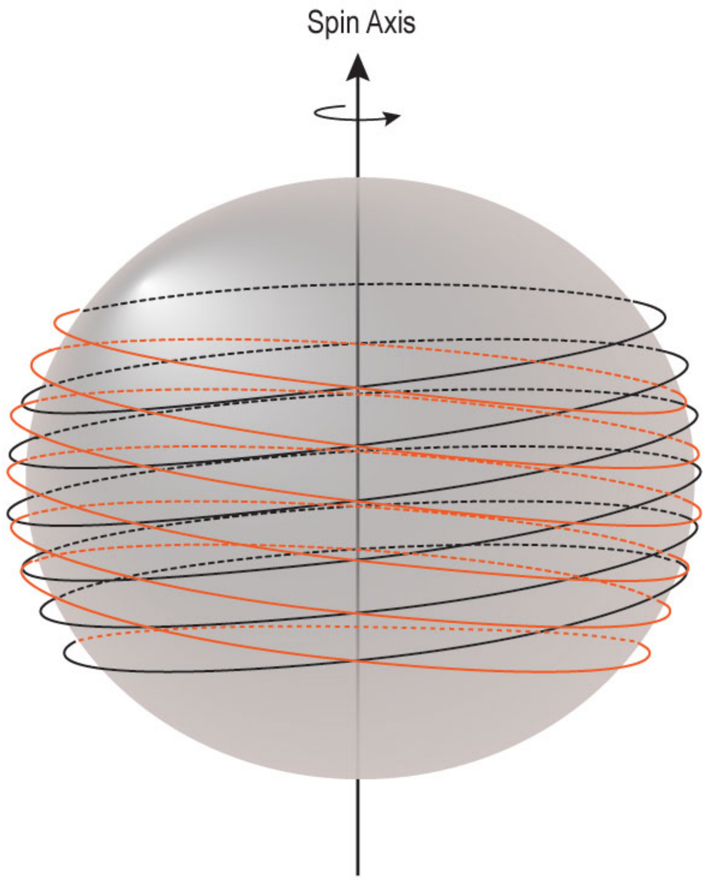


Figure 11.8: For orbits around a spinning black hole that are not confined to the equatorial plane, the orbits can appear quite complicated. The rotating spacetime near the hole can drag the particle around the spin axis much faster that its own motion would carry it. As a result, the orbit can appear to be a helix that circles the black hole as the particle moves up and down parallel to the spin axis. Credit: NASA/SSU/Aurore Simonnet

This page titled 11.2: Spacetime Near Black Holes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

11.3: Quantum Effects Near Black Holes

Learning Objectives

- You will be able to calculate the temperatures of black holes.
- You will understand that black holes can evaporate, and that their rate of evaporation depends strongly on their mass.
- You will understand why particle accelerators are not going to devour the world by producing mini black holes.
- You will understand that general relativity does not completely describe what ultimately happens to material that falls into a black hole at the singularity because it is not a quantum theory.
- You will understand that a mathematical singularity is a term that goes to infinity, but that the singularity of a black hole would not be infinitely dense in real life.
- You will understand that a quantum theory of gravity has not yet been successfully developed, but that we have an idea of the length and time scales where it would be applicable.

An important aspect of black holes that lies outside the scope of general relativity is their quantum nature. General relativity is not a quantum theory of gravity, and so our understanding of how black holes will behave on extremely small scales is still far from complete.

11.3.1: The Temperature of Black Holes

One of the first attempts to understand the quantum aspects of black holes was undertaken by Stephen Hawking (b. January 8, 1942, d. March 14, 2018) in the early 1970s. Hawking was studying the thermodynamics of black holes. He found that black holes have a temperature given by:

$$T_{bh} = \frac{hc^3}{16\pi^2 GMk_B} \tag{11.3.1}$$

where h is Planck's constant (h = 6.626 × 10^{-34} J s), and k_B is known as Boltzmann's constant (k_B = 1.38 × 10^{-23} J/K). The other symbols have their usual meaning: c is the speed of light, G is the gravitational constant, and M is the mass of the black hole. If SI units are used for these quantities, then the temperature is in kelvin (K). If we combine all of the constants into a single number, Equation 11.3.1 becomes:

$$T_{bh} = \frac{1.23 \times 10^{23}}{M}$$

where the temperature of the black hole (T_{bh}) is in kelvin, its mass (M) is in kg, and the constant in the numerator has untis kg K. From the expression, we see that the temperature of a black hole is inversely proportional to its mass: the greater the mass, the lower the temperature of the black hole. In the next activity, we will explore this relationship for black holes with different masses.

Black Hole Mass and Temperature

Worked Example:

Find the temperature of a one solar mass black hole.

- Given: A black hole with mass, $M = 2 \times 10^{30}$ kg (one solar mass)
- Find: T, the temperature of this black hole in Kelvin (K)
- Concept(s):

$$T_{bh} = \frac{1.23 \times 10^{23}\,\mathrm{kg\,K}}{M}$$

- Solution:

$$T = \frac{1.23 \times 10^{23}\ \cancel{\mathrm{kg}}\,\mathrm{K}}{2 \times 10^{30}\ \cancel{\mathrm{kg}}}$$
$$= 6.1 \times 10^{-8}\,\mathrm{K}$$

This is a tiny temperature. The average temperature in the Universe is about 3 K, so a one solar mass black hole has a temperature even colder, by a factor of about 50 million, than the average temperature of the Universe.

Questions:

1.

2.

We discussed earlier that a black hole is an object where the spacetime curvature is so extreme that not even light can escape. Yet, Hawking found that black holes emit radiation in a spectrum identical to a Planck emitter of the same temperature. So a black hole like the one whose mass you found in the previous activity (with a temperature of 6000 K) would emit light with a spectrum identical to that of the Sun; recall that a Planck emitter (sometimes referred to as a blackbody) has a spectrum that depends only on its temperature. How can this be if black holes by definition do not allow light to escape? The answer to that question requires us to consider the quantum aspects of the problem.

11.3.2: Hawking Radiation and the Evaporation of Black Holes

Special relativity allows for the conversion of mass to energy and vice versa. So if we imagine an electron, say, and its antiparticle, a positron, coming together, we know that they will annihilate and form two photons with equal energy and opposite direction of travel. Each photon will have an energy γmc^2, where *m* is the mass of the electron or positron. The two particles are identical except for their electric charge, which is negative for an electron and positive for a positron. The γ factor is the relativistic boosting factor that depends on a particle's speed. It is 1 if the particles are not moving, but larger than 1 if they are moving. We will assume for simplicity that the particles approach each other with equal speeds: we can always consider the collision from a frame of reference where this is true. Our frame of reference cannot affect the outcome of the collision.

In this interaction there are several quantities that must be conserved, or in other words, there are several quantities that must be the same before the collision and after it. These include energy, momentum and electric charge. There are some others, too, but these three are the important ones to consider at the moment. Momentum is described in Going Further 7.1: Newton's Third Law. If you have not read it, now would be a good time to do so.

Because each photon has the energy of one of the initial particles, we know that energy will be conserved. Similarly, because photons have no electric charge, we can see that total charge will also be conserved. Before the collision, the total charge (we call the charge of an electron "*e*") was: *e - e = 0*. Afterward it is still zero. The fact that the two photons move in opposite directions with equal energy means that momentum is also conserved. Recall that momentum is a vector, so it has strength and direction. The strength of the momentum of a photon is *E/c*, its energy divided by the speed of light. The direction of each photon is opposite the other, both before and after the collision. Beforehand they approach one another. Afterward they move away. In both cases the particle momenta add to zero.

Of course, we could also imagine the opposite process, where two photons of sufficient energy collide and are converted into a positron and electron pair. That process would be just like the first if we imagined filming it and running the film backwards. The reverse process would also satisfy all known laws of physics. Furthermore, it is observed to happen.

Quantum mechanics adds an interesting twist to this scenario by relaxing the requirement that energy be strictly conserved. It does not completely do away with conservation of energy: clearly, it is not possible to create something from nothing, at least not as we observe in the laboratory and our everyday existence. But it is possible to cheat a bit, and create something out of nothing for a very short time. This absurd sounding statement is one possible phrasing of the Heisenberg uncertainty principle, first discovered by the German physicist Werner Heisenberg (1901–1976). It can be stated more succinctly with mathematics.

$$\Delta E \Delta t \geq \frac{h}{4\pi}$$

In this expression, h is Planck's constant. Planck's constant is always present when quantum considerations are important. The ΔE on the left side of the equation is the uncertainty in the energy for a particular process, and Δt is the uncertainty in the time required for the process to complete. There are always uncertainties associated with any measurement, but the uncertainty principle is more fundamental than that. It states that no matter how carefully you build your measuring device, no matter how great its precision, there will always be an inherent uncertainty imposed by nature itself. That uncertainty is quantified by the uncertainty principle.

Notice that the quantum uncertainty is very small. Planck's constant is only about 10^{-34} in SI units. While this is not zero, it is too small to be noticed in day-to-day life. Notice further that it is the *product* of time and energy uncertainty that is constrained. So the uncertainty in one can be made arbitrarily small, but only at the cost of making the uncertainty in the other arbitrarily large. That is the important consideration for the radiation emitted from black holes.

✔ The Uncertainty Principle

Imagine a volume of empty space. The volume contains no light, no particles, nothing. According to the uncertainty principle, if we watch that empty volume of space, we should from time to time see a positron–electron pair (for example) spontaneously pop into being, briefly exist, and then annihilate and disappear. This happens in empty vacuum and is allowed as long as the pair does not exist for longer than the limit set by the uncertainty principle. So just how long is that? We can answer that using the uncertainty principle.

Worked Example:

1. How long can an electron–positron pair spontaneously pop into being and exist in a vacuum? The mass of each is 9×10^{-31} kg.

- Given: $m_e = 9 \times 10^{-31}\ \text{kg}$.
- Find: The time limit set by the uncertainty principle in seconds, Δt
- Concept(s):

If we wish the uncertainty in the energy to be large enough to allow for the creation of an electron and positron, then the uncertainty in the time must be small enough such that the uncertainty principle is not violated. So we can use the uncertainty principle, with the equals sign as a limit on the range.

For the energy, we can use the familiar relationship between matter and energy, $E = mc^2$.

The uncertainty principle is: $(\Delta E)(\Delta t) = h/4\pi$, where $h = 6.626 \times 10^{-34}\ \text{J s}$.

The energy of a positron–electron pair is: $\Delta E = 2m_e c^2$ where $c = 3 \times 10^8\ \text{m/s}$

- Solution:

$$\begin{aligned}(\Delta t) &= \frac{h}{4\pi(\Delta E)} \\ &= \frac{h}{4\pi(2m_e c^2)} \\ &= \frac{h}{8\pi m_e c^2} \\ &= \frac{6.626 \times 10^{-34}\ Js}{(8)(\pi)(9.1 \times 10^{-31}\ kg)(3 \times 10^8 m/s)^2} \\ &= 3.2 \times 10^{-22}\ s\end{aligned}$$

Questions

1.

https://phys.libretexts.org/@go/page/31358

2.

In this activity we have found an *upper limit* on the time that particles that have been spontaneously created out of empty space can exist: if the energy is large enough to create an electron and positron, for example, then the uncertainty in the time must be large enough that the product $\Delta E \Delta t$ remains no smaller than $h/4\pi$. That means if the particles exist for less time than this uncertainty, there is no violation of the uncertainty principle. An electron and positron of typical mass must annihilate within about 10^{-22} seconds to satisfy the uncertainty principle. You should have found that more massive particles, such as a proton—anti-proton pair would exist for less time. Such particles are not *real* in the sense that they cannot exist indefinitely, but only for a time so brief that they cannot even be measured individually. To distinguish them from real (long lasting) particles, physicists call these sorts of particles **virtual particles**.

Of course, the uncertainty principle is a theoretical prediction, and this has all been a bit of mathematics. How do we know that the spontaneous creation/annihilation process really happens, and that virtual particles actually exist at all? See Going Further 11.1: Testing the Uncertainty Principle, to learn why physicists think that the uncertainty principle gives a correct description of virtual particles, and that even they, in some sense, are "real."

Going Further 11.1: Testing the Uncertainty Principle

The Heisenberg uncertainty principle is one of the more bizarre aspects of quantum mechanics, a fairly non-intuitive description of the world to begin with. Why should anyone believe any of the strange predictions of the theory? Well, nobody should believe them; scientists are forced to accept the strangeness of the theory because its predictions are borne out by experiment. The uncertainty principle is basic to quantum mechanics and affects all aspects of the theory, but it has two consequences that are particularly useful in demonstrating that the theory has merit.

One consequence of the uncertainty principle is the Lamb shift, a tiny shift in the energy level of the electron in a hydrogen atom. The shift is caused by the sort of quantum vacuum fluctuations associated with the production of virtual particle–antiparticle pairs as discussed for electrons and positrons. Even though any particle–antiparticle pair exists for a time that is too short to measure directly, new pairs are popping up (and then disappearing) all the time. These fluctuations of the vacuum perturb the electrons in an atom, gently shifting its energy—in effect, the hydrogen atom constantly absorbs and then re-emits virtual photons.

Because the effect is averaged over many virtual particle pairs constantly appearing and disappearing, the energy shift looks like a slight broadening of the energy of an atomic orbital. Using the quantum theory of the electromagnetic interaction, called **quantum electrodynamics** (QED), it is possible to calculate the size of this broadening. In a hydrogen atom, the shift slightly separates the orbitals known as $^2S_{1/2}$ and $^2P_{1/2}$, which would otherwise have the same energy. The splitting was actually measured in 1947 before the theory of QED was fully developed. However, once the theory was available, its prediction for the splitting matched the observed amount to one part in a trillion. Or in other words, the theory matched the measurement to the eleventh decimal place, which is the limit of experimental precision. Such spectacular agreement between theory and experiment suggests that the notion of virtual particles is a valid description of nature, no matter the philosophical discomfort it might cause.

An additional verification of the quantum view of nature is the Casimir effect. In this experiment, two grounded conducting plates are placed very close together, in effect creating a region of space where only certain energy states of particles, including virtual particles, can exist. The reason is that quantum mechanics predicts that particles should have a wave-like nature in addition to their particle nature. The wavelength of a particle depends on its energy. Within an enclosed space, an even number of half-wavelengths must be able to fit. This is a bit like the vibrating string on a violin or guitar, which is fixed at both ends, as shown in Figure B.11.1. The restriction on wavelengths is a restriction on energy states of any particles within the space between the plates. Since any particle energies are possible outside the two plates, the exclusion of certain states between them causes a pressure difference between the inside and outside of the plates. As a result they are pushed together. This has been shown to happen in accord with the predictions of quantum electrodynamics, just as the Lamb shift does.

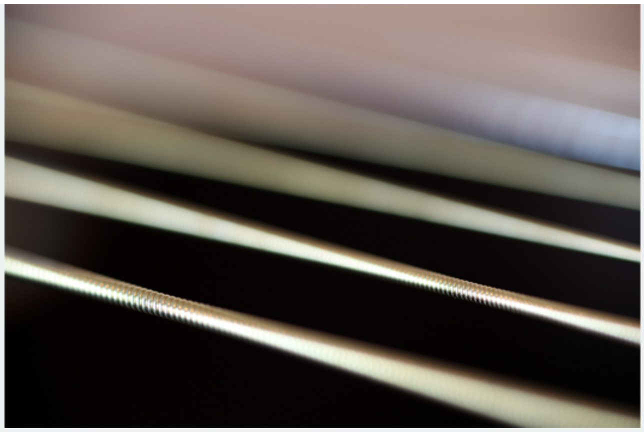


Figure B.11.1. The wave-like nature of particles is analogous to waves on a stringed instrument. Just as the strings on the guitar pictured here can only vibrate with certain frequencies because the strings are tied down at two ends, only certain particle energies are possible in the Casimir effect, where the particles are constrained between two plates. Credit: Shutterstock

11.3.2.1: Heisenberg Uncertainty and Hawking Radiation

Stephen Hawking used this strange quantum uncertainty effect to explore the radiation emitted by black holes. He imagined what would happen if a particle–antiparticle pair was created near the event horizon of a black hole , as in Figure 11.9. Under these circumstances it would be possible for one of the particles to cross the horizon, while the other did not. In this case, the two could not recombine to annihilate. The one particle would have been lost inside the black hole. It's partner would have to become a real particle. But that would require the energy to create it, which would have to come from somewhere. The only source of that energy is the mass–energy of the black hole itself. So these particle pairs provide a way for black holes to evaporate. The process is quite slow for a large black hole.

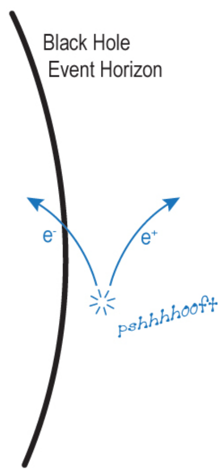


Figure 11.9: Hawking radiation. A virtual particle pair is formed near the event horizon of a black hole. One of the particles crosses the horizon and cannot get back out to annihilate with its partner. The partner must therefore become a real particle, stealing the energy required to do so from the only source available, the black hole itself. The figure shows an electron–positron pair, but any pair of antiparticles will work. Credit: NASA/SSU/Aurore SImonnet

To estimate the time required for a black hole to evaporate, we can use the temperature expression for a black hole and several pieces of information that you have learned previously:

- The relationship between the temperature (T) and flux (F) of a an object emitting energy with a Planck spectrum

$$F = \sigma T^4$$

- The relationship between flux (F), luminosity (L), and distance (r)

$$L = (F)(4\pi r^2)$$

- The relationship between mass (m) and energy (E)

https://phys.libretexts.org/@go/page/31358

$$E = mc^2$$

- The relationship between energy (E), luminosity (L), and lifetime (t)

$$t = \frac{E}{L}$$

To combine these ideas, we use the mass of the black hole (M), the temperature of the black hole (T_{bh}), and the Schwarzschild radius (R_s) as the distance. We find that the thermal luminosity of a black hole is:

$$L_{th} = \frac{3.56 \times 10^{32}}{M^2}$$

where the thermal luminosity L has units of W, the mass M has units of kg, and the constant has units of W kg^2.

The lifetime of a black hole is given by:

$$t \approx 2.5 \times 10^{-16} M^3$$

where the lifetime t has units of seconds, the mass M has units of kg, and the constant has units of s/kg^3.

The thermal luminosity of a black hole is inversely proportional to mass squared. The time for a black hole to evaporate is proportional to mass cubed. This means that the *larger* the mass of a black hole, the *lower* its thermal luminosity and the *greater* its lifetime.

If you are interested in a mathematical derivation of these relationships where we have included each of the steps, see Going Further 11.2: Deriving the Thermal Luminosity and Lifetime of a Black Hole.

You might object that we cannot simply divide the rest energy of the black hole by the thermal luminosity because both change as the black hole evaporates and loses mass. You would be right to make this objection, however, a more careful mathematical treatment that takes these changes into account shows that we have overestimated the lifetime by only a factor of 3. We will ignore this factor for our purposes of illustration, but you can keep it in mind if you like. The next activity will show why we are not going to concern ourselves with it.

Going Further 11.2: Deriving the Thermal Luminosity and Lifetime of a Black Hole

Recall that a black hole is a Planck emitter (or blackbody). The amount of energy lost by a Planck emitter per unit area per unit time (the energy flux, F) is given by the Stefan-Boltzmann equation:

$$F = \sigma T^4$$

where the constant $\sigma = 5.67 \times 10^{-8}\ \mathrm{W/m^2\ K^4}$ and is needed to account for the system of units we use.

The relationship between flux and luminosity is: $F = L/(4\pi r^2)$. This means that the total amount of energy emitted by the black hole, its luminosity, L, is this energy flux multiplied by the total area of the event horizon, essentially the area of the black hole, A_{BH}. Since a black hole is spherical, we can use the area of a sphere: $A = 4\pi R^2$.

$$L = A_{BH} \sigma T^4 = 4\pi R_{Sch}^2 \sigma T^4$$

We can now substitute into this expression for the Schwarzschild radius, the appropriate radius for the event horizon of a black hole, and the temperature of the black hole to find the black hole thermal luminosity:

$$L_{th} = 4\pi\sigma \left(\frac{2GM}{c^2}\right)^2 \left(\frac{hc^3}{16\pi^2 GMk_B}\right)^4$$

Or simplifying:

$$L_{th} = \frac{\sigma h^4 c^8}{2^{12}\pi^7 G^2 k_B^4 M^2}$$

This expression looks complicated, but it is mostly just a bunch of physical constants, or in other words, a product of a bunch of numbers. That means it will simplify to a single number divided by the square of the black hole mass. In SI units, it becomes:

$$L_{th} = \frac{3.56 \times 10^{32}}{M^2}$$

where the thermal luminosity L is in W, the mass M is in kg, and the constant is in W kg2.

We have found the thermal luminosity of a black hole, the amount of energy lost to the black hole per unit time due to Hawking radiation. Now we would like to determine how long it can radiate its energy. That depends on how much energy it has to radiate. The total amount of energy in the black hole is contained in its rest mass: $E = Mc^2$. We can estimate its lifetime, t, if we divide the energy in the black hole by its thermal luminosity. This is analogous to the way we estimate the lifetime of a star.

$$t \approx \frac{E}{L} = \frac{Mc^2}{\left(\frac{\sigma h^4 c^8}{2^{12}\pi^7 G^2 k_B^4 M^2}\right)}$$

Simplifying, we have:

$$t \approx \frac{2^{12}\pi^7 G^2 k_B^4 M^3}{\sigma h^4 c^6}$$

Again this looks complicated, but we can simplify further by multiplying out all the constants as we did previously:

$$t \approx 2.5 \times 10^{-16} M^3$$

where the lifetime t is in seconds, the mass M is in kg, and the constant is in $\mathrm{s/kg^3}$.

Evaporating Black Holes

Worked Example:

1. Determine the luminosity for an evaporating one solar mass black hole.

- Given: mass $M = M_{Sun} = 2 \times 10^{30}$ kg
- Find: luminosity, L
- Concept(s): $L \sim 3.56 \times 10^{32}/M^2$
- Solution:

$$\begin{aligned} L &\sim \frac{3.56 \times 10^{32}\, W\ \cancel{\mathrm{kg}^2}}{(2 \times 10^{30}\ \cancel{\mathrm{kg}})^2} \\ &= 8.9 \times 10^{-29}\, W \end{aligned}$$

2. Determine the evaporation time in years for a one solar mass black hole.

- Given: mass $M = M_{Sun} = 2 \times 10^{30}$ kg
- Find: t, the estimated evaporation time in units of seconds
- Concept(s): $t \sim 2.5 \times 10^{-16} M^3$
- Solution:

$$t \sim (2.5 \times 10^{-16}\ \mathrm{s}\ \cancel{\mathrm{kg}^{-3}})(2 \times 10^{30}\ \cancel{\mathrm{kg}})^3 = 2 \times 10^{75}\ \mathrm{s}$$

Converting to years

$$(2 \times 10^{75}\ \cancel{\mathrm{s}})(1\mathrm{yr}\ /\ 3.15 \times 10^{7}\ \cancel{\mathrm{s}}) = 6.3 \times 10^{67}\mathrm{years}$$

Astronomers have measured the current age of the Universe to be about 14 billion years. Clearly the time for a one solar mass black hole to evaporate is enormously long compared to the age of the Universe. The ignored factor of three plays no role in this conclusion.

Questions

1.

2.

3.

In the previous activity, we saw that it takes an extremely long time for solar mass black holes to evaporate. So we should not expect to see any of them evaporating anytime soon. However, the time required for a black hole to evaporate depends on its mass cubed, so as the mass of a black hole approaches zero, the time required for it to evaporate by Hawking radiation also approaches zero, and very quickly. Furthermore, if you look at the luminosity we derived, we expect that the black hole gets brighter and brighter as it gets smaller. In fact, the luminosity grows without bound as the mass of the black hole shrinks to zero. So just before a black hole disappears, it emits a very brief, but very bright, flash of light. Its temperature also grows without bound as its mass shrinks, so we would expect this light to be mostly X-rays and gamma-rays.

But do we expect to see any such objects? Given that the Universe is about 14 billion years old, it is possible that small enough black holes might just be evaporating now, assuming they were formed as part of whatever process formed the Universe itself. Astronomers refer to these hypothetical objects as primordial black holes because they would have been formed from processes that formed the Universe. In the next activity you will compute the size of such black holes.

Primordial Black Holes

Use the expression for the evaporation time of black holes to find the mass of primordial black holes, given the age of the Universe. Remember to convert to SI units (kg, m, s) where necessary.

Questions

1.

2.

What about even smaller black holes, such as proton-mass black holes? There were stories in the news when the LHC at CERN was turned on (and also several decades ago when the Tevatron at Fermilab came online) about the possibility of Earth being swallowed by mini-black holes that might be produced in the particle collisions. The black holes, it was argued by some, would start out with masses around the mass of a proton, but would then grow in size until they eventually consumed the entire Earth. What a scary thought—but is it reasonable?

Some of those making these arguments went so far as to file lawsuits asking that the accelerators be barred from starting. How reasonable were their claims? We can answer that question with a calculation like the ones we have just done.

MINI BLACK HOLES

Questions:

1.

You should have discovered that it takes a very short time for a proton-mass black hole to evaporate. It is as close to instantaneous as we are likely to encounter. So long before anything can fall into such a tiny black hole it will evaporate.

Given the short lifetime of tiny black holes, why were some people concerned about the possibility that mini-black holes in the Tevatron and LHC could grow large enough to consume the whole planet? If you are unable to understand why they were concerned, then you are in agreement with physicists. You are also in agreement with the judges for the lawsuits, which were all thrown out.

11.3.3: Singularities and Quantum Gravity

As we have already explored, close to the event horizon of a black hole we expect quantum effects to become important. If we imagine journeying beyond the horizon and down into a black hole, we can ask what happens as we approach the center. We have seen that before we even arrive there, we are stretched radially and compressed from the sides by spaghettification. And we have discussed how this process of destruction will continue on to smaller and smaller scales, growing more extreme as we approach r = 0. This happens because the spacetime curvature increases without bound as we approach r = 0, and there is energy in spacetime curvature. In fact, if we look at the Schwarzschild interval, we see that both the time and space curvature terms become infinite at r = 0.

This point is called a **singularity**, a mathematical term used to describe places where mathematical expressions become undefined. If an expression becomes infinite at some point, that certainly satisfies the conditions for being singular there. One often hears that all the mass that falls into a black hole gets squeezed into a point of infinite density at the singularity. Certainly infinite density, or infinite anything, would constitute a singularity. But is a singularity really a meaningful concept when applied to the actual Universe?

At some point we expect that the energy of the curved spacetime will begin to interact with the fluctuating virtual particles, or quantum fluctuations, in the spacetime, converting gravitational energy into particles. Of course, those particles cannot get across the horizon; they are imprisoned inside the black hole near its center. They fall back to the center and are converted back to curvature; the situation could be extremely chaotic.

In order to understand the conditions near a singularity, we require a **quantum theory of gravity** —such a theory would encompass both gravitational and quantum phenomena. Physicists have not yet created such a theory. Still, there are some rudimentary considerations that we can make to give us an idea of what a quantum theory of gravity would encompass and where it would be applicable.

The first such consideration is to think about the size scale on which we expect quantum fluctuations to begin to interact strongly with gravity. We can use the relevant physical constants to determine this scale. Any theory of quantum gravity must contain combinations of h, c, and G: the h because the theory encompasses quantum phenomena, the c because it must include relativity, and the G because it must include gravity. These quantities are strictly determined by various properties we measure for the Universe. They are independent of any specific model we might devise, and so we expect them to be in some sense unbiased and to provide fundamental information that any successful future theory must incorporate. But how can we combine these constants in such a way to give us a size scale, or in other words, a length? We can assume that the constants combine in the following way.

$$[l]=[h^{\alpha}\,c^{\beta}\,G^{\gamma}]$$

We use the square brackets to mean that we are comparing only the dimensions of the quantities (essentially their units), not their numerical values. That reduces our task to finding values of the exponents α, β, and γ (alpha, beta and gamma) that give the right-hand side of the equation the value "length to the first power." Note that we could have used a, b and c instead, but we are already using c to mean the speed of light, and we want to avoid confusion. Speaking of which, don't confuse γ with the relativistic γ factor we learned about in Chapter 9. Here we are using it to mean a simple parameter for this exercise.

If we use SI units, then the quantities must combine to give meters. It turns out that the appropriate values of α, β and γ are 1/2, -3/2 and 1/2, respectively. See Going Further 11.3: Planck Units and Dimensional Analysis for the details of how we determine this.

We can check this result by inserting units for each of the constants, still ignoring their numerical values:

- [h]= J s
- [c] = m/s
- [G] = N m^2 / kg^2

Now we can use these values to check that our exponents are correct.

$(J\ s)^{1/2}\ (m/s)^{-3/2}\ (N\ m^2 / kg^2)^{1/2} = ((kg\ m^2/s^2)\ (s))^{1/2}\ (m/s)^{-3/2}\ (\ (kg\ m/s^2)\ (m^2/kg^2)\ )^{1/2} = m$

We expressed joules and newtons in terms of meters, kilograms, and seconds in the second expression. Combining terms we find that they cancel except for one factor of meters, just as we had desired.

Now that we know this set of exponents result in a length, we can determine what the length scale is by inserting the numerical values for the constants. This is the length scale on which gravity and quantum mechanics interact strongly. It is referred to as the

Planck length because it was Max Planck who first determined it:

$$l_{planck} = \left(\frac{hG}{c^3}\right)^{1/2}$$

Numerically it is equal to 4 × 10^{-35} m. On size scales smaller than this, we expect that quantum effects will become important for gravity. In a sense, our idea of spacetime as smooth and continuous breaks down on this scale; space becomes grainy. On these scales, general relativity should no longer be adequate to explain gravity, and so any singularities contained in the theory should not be important. Notice that the Planck length is tiny. A proton is about 10^{20} Planck lengths across, so we do not expect to see any quantum gravity effects in atoms or even in currently known subatomic particles. But l_{planck} is still much larger than zero, and therefore long before we get to any singularity, general relativistic predictions should cease to be a valid description of nature. We should not have to worry about things like points of infinite density.

From the Planck length one can construct a Planck time, the time required for light to travel one planck length:

$$t_{planck} = \frac{l_{planck}}{c} = \left(\frac{hG}{c^5}\right)^{1/2}$$

Numerically the Planck time is 10^{-43} s. Again, this is a tiny quantity. We expect that quantum gravitational effects will be important for processes with durations similar to this timescale. We also expect that time itself will be grainy on this scale, not the smooth continuous stream we generally think about.

We have not made any predictions about what happens at the center of a black hole. All we have done is make an argument that we do not expect our current understanding of gravity, as expressed in the theory of general relativity, to be an adequate description there. On these scales we might see the structure of spacetime change abruptly, forming tiny black holes or creating other abrupt changes in the local curvature, but these changes would only last for a time of order t_{planck} and affect length scales of order l_{planck}. We will likely have to wait for physicists to create a quantum theory of gravity before a full theoretical description of black holes emerges, and before we know if our assumptions about the quantum nature of spacetime are in any sense correct.

For an artist's depiction of what space and time might be like on scales where quantum gravity is important, see Figure 11.10. Both space and time are grainy, resembling foam, and are not continuous as in our everyday experience. The concept of quantum gravity not only has important implications for the nature of the interiors of black holes, but also for the earliest moments in the history of the Universe.

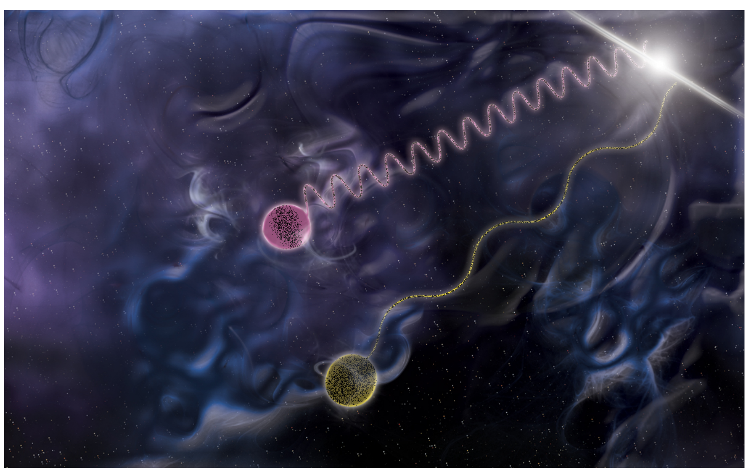


Figure 11.10: On tiny length scales and short timescales it is possible that spacetime has a quantum nature, giving it the structure of a foam. Random bubbles form and disappear. The bubbles have a size comparable to the Planck length and exist for times comparable to the Planck time. Thus spacetime might look something like the foam that forms in a rapidly boiling pot of sauce, but on minuscule scales in space and time. Credit: NASA/SSU/Aurore Simonnet

Going Further 11.3: Planck Units and Dimensional Analysis

To find the proper combination of physical constants to find the Planck length, we can use a procedure called dimensional analysis. In this procedure we use the dimensions (or units) of physical quantities to find how they must be combined in order to give us another physical quantity of interest to us. This is a technique used by scientists when they do not have a theory to describe some phenomenon.

As an aside, the technique can (and should) be used by students of science to check the answers to computations they make, because if the units of their answer come out wrong, the answer itself must be wrong. Dimensional analysis is a quick and easy self-check, although it does not guarantee an answer is correct just because the units are.

Generally in dimensional analysis square brackets surrounding a quantity indicate that we should consider only its dimensions (or units), not its numerical value. The numerical value will be used at the end, but first we consider just the units. So, in determining the Planck length, we have a relation that says:

$[l] = [h^{\alpha} c \beta G^{\gamma}] = [h]^{\alpha} [c]^{\beta} [G]^{\gamma}$

We will use the SI units for each dimension to simplify the analysis. Of course, this procedure will work with other units as well; it is only the dimension, whether length, time, mass, etc., that matters. We pick SI for convenience and specificity.

The units for the three quantities we have are:

- [h] = J s
- [c] = m/s
- [G] = N m^2 / kg^2

Substituting these units into the above expression, we get:

$\mathrm{m} = (\mathrm{J\ s})^{\alpha} \cdot (\mathrm{m/s})^{\beta} \cdot (\mathrm{Nm^2/kg^2})^{\gamma}$

We have to break down joules and newtons into their equivalent combinations of meters, kilograms, and seconds to proceed. These are

J = kg m^2/s^2
N = kg m/s^2

And now we substitute these in.

$\mathrm{m} = (\,(\mathrm{kg\,m^2/s^2})(\mathrm{s})\,)^{\alpha} \cdot (\mathrm{m/s})^{\beta} \cdot (\,(\mathrm{kg\,m/s^2})(\mathrm{m^2/kg^2})\,)^{\gamma}$

We have to combine like terms. This gives us

$\mathrm{m} = (\mathrm{kg\,m^2/s})^{\alpha} \cdot (\mathrm{m/s})^{\beta} \cdot (\mathrm{m^3/kg\,s^2})^{\gamma}$

and finally

$\mathrm{m} = \mathrm{kg}^{(\alpha-\gamma)} \cdot \mathrm{m}^{(2\alpha+\beta+3\gamma)} \cdot \mathrm{s}^{(-\alpha-\beta-2\gamma)}$

We have derived three expressions, and these can be used to relate the exponents to each other. By comparing the units on either side of the equation, we find that

$$\alpha - \gamma = 0$$

since there are no kilograms to the left of the equal sign.

$$-\alpha - \beta - 2\gamma = 0$$

since there are no seconds to the left of the equal sign.

$$2\alpha + \beta + 3\gamma = 1$$

since there is only one power of meters to the left of the equal sign.

This gives us three equations in three unknowns, and we can solve them in the normal way of substitution. We find the values α = γ = 1/2, β = –3/2.

This leads us to the Planck length:

$$l_{planck} = h^{1/2} G^{1/2} c^{-3/2} = \left(\frac{hG}{c^3}\right)^{1/2}$$

If instead of a length scale we had wanted to find a mass scale, we could have used the same procedure, except that the left-hand side of our equation would have had one power of kilograms instead of one power of meters. Everything else would have been the same. In that case we would have found the Planck mass:

$$m_{planck} = h^{1/2} c^{1/2} G^{-3/2} = \left(\frac{hc}{G^3}\right)^{1/2}$$

If you are interested, see if you can figure out how this expression is determined. The procedure is identical to the previous one, but for the minor difference we have already pointed out. The Planck mass will be important in our consideration of the earliest moments of the Universe, when its age was of order the Planck time. If you like you can also try to work out the Planck time this way and compare it to the value mentioned in the main text—they should be the same.

The German physicist Max Planck first undertook dimensional analysis of this sort in 1899. He thought of these units, l_{planck}, t_{planck}, m_{planck}, etc., as natural units, in that they are independent of any particular theory or particle. The basic units of length, mass, and time can be combined in various ways to derive expressions for energy, force, pressure, etc., so we can also compute a Planck energy, a Planck force, and so on.

In addition to the three mechanical constants we have used here, Planck also used the Coulomb constant (like the gravitational constant G, but for the electrical force) and the Boltzmann constant, which is important in thermodynamics and statistical physics. These are used when we wish to venture beyond mechanics and include electromagnetic and thermal phenomena.

Planck units are just one set of many other natural units that are possible, but in some sense they are the most fundamental because, as mentioned already, they do not depend on the properties of any particular particle or theory. They depend only on physical constants related to space, time, and the interactions between particles.

This page titled 11.3: Quantum Effects Near Black Holes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

11.4: Astrophysical Black Holes

Learning Objectives

- Explain the following laws within the Ideal Gas Law

What Do You Think: Detecting Black Holes

So far we have discussed black holes from a theoretical perspective. But how do we know that any of this has a basis in reality? How can we even begin to study objects that emit no light and are more massive than the Sun? For a long time the answer was, "we can't." That began to change several decades ago when advances in astronomical instruments opened new vistas on the Universe. Black holes popped out, sometimes glowing brightly in the darkness, other times remaining hidden except for their influence on visible neighbors. In this section we look at how black holes are detected and how they affect the Universe and its contents.

11.4.1: Stellar Mass Black Holes

The first observational evidence that black holes are more than mere theoretical oddities was obtained using instruments borne aloft by sounding rockets in the 1960s. The rocket payloads were simple and small, and the flights were short, lasting just long enough to carry their loads above the atmosphere before falling back to Earth. Nonetheless, these flights were a critical demonstration that astrophysical x-ray sources exist.

Without going above the atmosphere it is impossible to know if x-ray sources are in the sky. The atmosphere is completely opaque to x-rays and higher energy photons. Sounding rocket flights proved that x-ray sources are present, but they gave only short views of those sources. The situation improved with the launch of the first orbiting x-ray telescope, Uhuru, in late 1970.

Uhuru takes its name from the Swahili for "freedom," a reference to the location of its launch platform in Kenya. It was the first of a series of three similar satellites launched by NASA in the 1970s. Uhuru was a survey instrument, and it provided a view of the sky very different from sounding rockets. The rocket flights were short and their instruments primitive. Because their detectors were only able to view the sky for the few minutes that the rocket payload was above the atmosphere, only a small number of very bright sources were detected. Most astronomers of the time did not expect an x-ray telescope to see much more than this. They were wrong.

Over its three-year lifetime, Uhuru cataloged more than 300 x-ray sources spread over the entire sky. That does not sound like a lot when compared to the thousands of stars visible to even the naked eye, but it was only a start. Subsequent x-ray missions have cataloged more than 100,000 sources. We now know that many of these sources are black holes, but the evidence for that was decades building.

One of the most interesting sources observed with Uhuru was a very bright source called Cygnus X-1 (see Figure 11.11), so designated because it is the brightest x-ray source in the constellation Cygnus. The source had been discovered nearly a decade earlier and tentatively identified with a blue supergiant star. The improved x-ray observations possible with Uhuru showed that the source varied extremely rapidly, up to several times each second. The variations set limits on the size of the emitting region because any source can only vary as quickly as light can travel across it. In this case it could be no bigger than about a quarter of the Earth–Moon distance. See Going Further 11.4: Light Travel Time and Variability.

The identification of the blue supergiant with the x-ray source was strengthened as additional observations of the source were made in various wave bands. But blue supergiants do not emit x-rays in the amounts detectable from Earth. That suggested the supergiant had a companion, invisible in optical light, which was responsible for the x-rays. Careful study of the supergiant's spectrum revealed that it was indeed a member of a binary system; the star's spectral lines were seen to shift back and forth with a period of about five and a half days. These motions placed constraints on the mass of the unseen companion, requiring it to be at least 10 solar masses. A massive invisible companion that is a powerful x-ray emitter suggests the presence of a black hole, and the evidence for a black hole has slowly increased over the subsequent four decades.

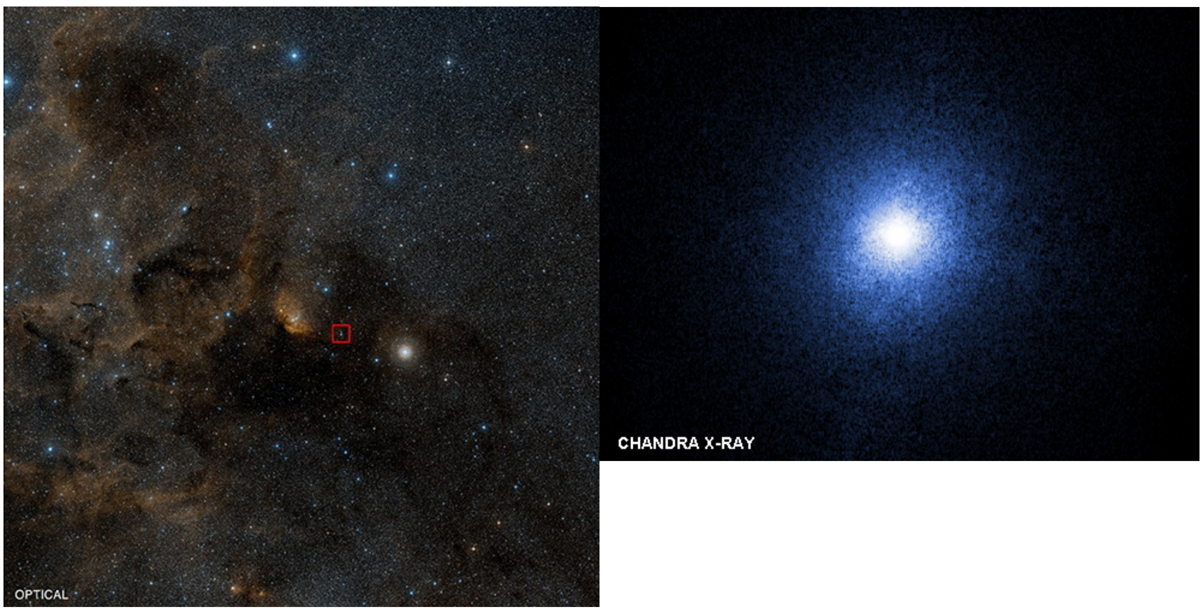


Figure 11.11: The optical image on the left shows the region of the galaxy containing Cygnus X-1 (red box). The x-ray image on the right was taken by the Chandra X-ray Observatory. Credit: Optical (Digitized Sky Survey); x-ray (NASA/CXC)

GOING FURTHER 11.4: LIGHT TRAVEL TIME AND VARIABILITY

Cygnus X-1 belongs to a class of objects called high-mass x-ray binaries. In these systems, a massive normal star orbits a compact object, usually a neutron star or black hole. Rarely, the compact object is a white dwarf. White dwarfs, neutron stars, and (stellar mass) black holes are the remnants left over when stars have died. The x-rays are produced when the compact companion captures some of the wind produced by the normal star. Very massive stars like blue supergiants have powerful stellar winds, amounting to the loss of perhaps 10 solar masses every million years. This rate is about a billion times larger than the mass lost by the Sun due to the solar wind.

The material captured from the stellar wind forms an accretion disk around the compact object and spirals down into it. The gravitational energy lost during the accretion process is enough to heat the gas in the accretion disk to millions or even tens of millions of kelvin. At those temperatures, much hotter than the surface of even the hottest normal star, the disk glows in x-rays. There can also be secondary x-ray emission from the surface of the star facing the disk. As x-rays from the disk heat the stellar photosphere, they cause it to also glow in x-rays.

The x-ray emission from high-mass x-ray binaries looks essentially the same whether the accretion is onto a neutron star or a black hole. It is usually necessary to make a careful study of the motion of the normal star in order to distinguish between the two. Black holes are more massive than neutron stars, and so they affect the motion of the companion to a larger degree than a neutron star.

Once the motion of the star is known, and the mass of the compact object has been determined, scientists can deduce whether the invisible component is a black hole or a neutron star: the upper limit for the mass of a neutron star is about three times the mass of the Sun. Thus, if the accreting object is larger than about three solar masses, it is likely to be a black hole. In the case of Cygnus X-1, the black hole mass is determined to be about 15 solar masses.

In addition to high-mass x-ray binaries, there are low-mass x-ray binaries. In these systems a normal star, typically with a mass similar to that of the Sun, orbits a massive compact object, either a neutron star or a black hole. Occasionally a white dwarf is present in these systems instead of a normal star. Material is transferred from the star to its compact companion. Essentially, material on the surface of the normal star will find itself equally attracted by the gravity of the normal star and the compact object. It can thus fall off the normal star onto its companion. This will only happen if the separation between the two is quite small.

As the material is transferred it forms an accretion disk, allowing it to shed angular momentum and thus fall into the compact object. Just as in high-mass x-ray binaries, the accretion disk is heated to millions of kelvin by the release of gravitational energy, and thus it glows in x-rays. Figure 11.12 shows several artistic renderings of x-ray binary systems. X-ray binaries are also sometimes sources of gamma-rays and radio waves.

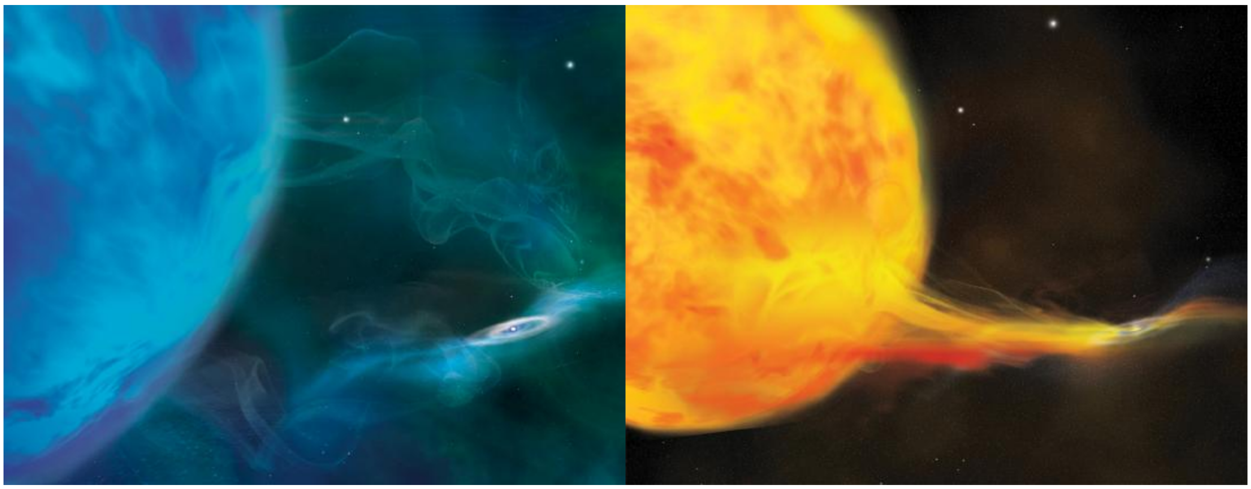


Figure 11.12: X-ray binary systems are systems in which a normal star orbits a neutron star or black hole. These artist's renderings depict two examples of such systems. In the first, at left, a massive star orbits a black hole. In the other system a low-mass star orbits a black hole. If the star and black hole are close enough to each other, then some of the material of the normal star can fall onto the black hole, forming an accretion disk that is hot enough to glow brightly in x-rays, giving the systems their name. Credit: NASA/SSU/Aurore Simonnet

We have suggested that the energy needed to power the x-ray emission from x-ray binary systems comes from the release of gravitational energy. The energy is released as material falls onto the compact object from its companion. But is this a reasonable model for the emission? We have learned enough gravitational physics in the past several chapters that we can answer that question.

Recall that the gravitational potential energy of an object of mass m in the potential of another object of mass M a distance r away is given by

$$PE = -\frac{GMm}{r}$$

This says that the energy is proportional to the product of the masses and inversely proportional to the distance between them. The power produced, or luminosity is the energy emitted per unit time. We can compare this to the observed luminosity of x-ray binary systems. The luminosity that might be produced from gravitational potential energy is therefore:

$$L = \frac{\Delta PE}{\Delta t}$$

where Δt is the time interval and ΔPE is the energy emitted over the time interval. If this power is produced by mass accretion then we can write the following:

$$L = \frac{\Delta PE}{\Delta t} = -\left(\frac{GM}{r}\right)\left(\frac{\Delta m}{\Delta t}\right)$$

We are assuming that the mass (M) of the accreting object, the black hole in this case, is not changed appreciably by the mass Δm falling onto it over the time interval Δt. Under this assumption, the only thing changing on the right side of the equation is the mass being lost by the companion star and attracted by the black hole; this is the accretion rate, $\Delta m/\Delta t$. Recall that the minus sign in the equation for potential energy is just a convention in the definition.

Accretion Rate and Luminosity

The luminosity of x-ray binaries ranges from about 10^{26} W to 10^{31} W, or from about one solar luminosity to 100,000 solar luminosities. In this activity we see how to use these numbers to test the feasibility of the mass accretion model for x-ray binaries.

Worked Examples:

1. Find the accretion rate for a black hole emitting 10^{26} W of luminosity.

Given:

- luminosity $L = 10^{26}\ \mathrm{W}$
- gravitational constant: $G = 6.67 \times 10^{-11}\ \mathrm{Nm}^2\,/\,\mathrm{kg}^2$
- mass of black hole: $M = 15 M_{Sun} = 3 \times 10^{31}\ \mathrm{kg}$
- speed of light: $c = 3 \times 10^8\ \mathrm{m/s}$

Find: Δm/Δt, the accretion rate in units of kg/s

- Concept(s):

  $$L = -\left(\frac{GM}{r}\right)\left(\frac{\Delta m}{\Delta t}\right)$$

  and r is the Schwarzschild radius,

  $$R_{Sch} = \frac{2GM}{c^2}$$

Solution:

- We can rearrange the equation in terms of $\frac{\Delta m}{\Delta t}$, substitute the Schwarzschild radius for r, and then plug in the appropriate numbers.
  First solve for $\frac{\Delta m}{\Delta t}$ by multiplying by $-\frac{r}{GM}$.

  $$\frac{\Delta m}{\Delta t} = -\frac{Lr}{GM}$$

  Now substitute R_{Sch} for the radius, r:

  $$\frac{\Delta m}{\Delta t} = -\frac{L}{G\cancel{M}} \cdot \frac{2\,G\cancel{M}}{c^2} = -\frac{2L}{c^2}$$

Notice that the mass of the black hole cancels. Only the power emitted matters.

$$\frac{\Delta m}{\Delta t} = -\frac{2L}{c^2} = -2\left[\frac{10^{26}\ \mathrm{W}}{(3 \times 10^8\ \mathrm{m/s})^2}\right] = -2.29 \times 10^9\ \mathrm{kg/s}$$

2. Express the accretion rate in terms of solar masses.

Converting units,

$$(-2.29 \times 10^9\ \cancel{\mathrm{kg}}/\ \cancel{\mathrm{s}})(1 M_{Sun}\,/\,2 \times 10^{30}\ \cancel{\mathrm{kg}})(3.15 \times 10^7\ \cancel{\mathrm{s}}\,/\,1\ \mathrm{yr}) = -3.5 \times 10^{-14}\ M_{Sun}/\mathrm{yr}$$

This is a very small amount of mass to transfer. For the higher luminosity systems, the accretion rate would be about a hundred thousand times larger, but it is still very small and not a challenge for these systems. The mass model for X-ray binaries seems quite reasonable, at least from these simple energy arguments.

As an aside, the final value for $\Delta m/\Delta t$ is negative because the star that the black hole is accreting matter from is losing mass. When this negative quantity is multiplied by the minus sign in the original equation for luminosity, we see that luminosity is a positive quantity, as it should be.

Questions

1.

2.

Notice from the last activity that the mass of the black hole does not affect the accretion rate. If we drop a given amount of mass from a great distance onto any black hole, we will get the same luminosity out. Can you think of a reason why? We will look at this process in more detail in the next section.

11.4.2: Supermassive Black Holes

Active galactic nuclei (AGN) are another class of bright x-ray sources. They are found in active galaxies, naturally. Only a few percent of galaxies are active. AGN are bright sources of radio and gamma-ray emission as well as x-rays. Optically, AGN are usually not very impressive, being well below the threshold of detection by the unaided eye. In fact, even the brightest optical AGN are barely visible with the aid of a small telescope. This is due to the fact that all AGN are very far away.

Various types of active galaxies are shown in Figure 11.13.

https://phys.libretexts.org/@go/page/31359

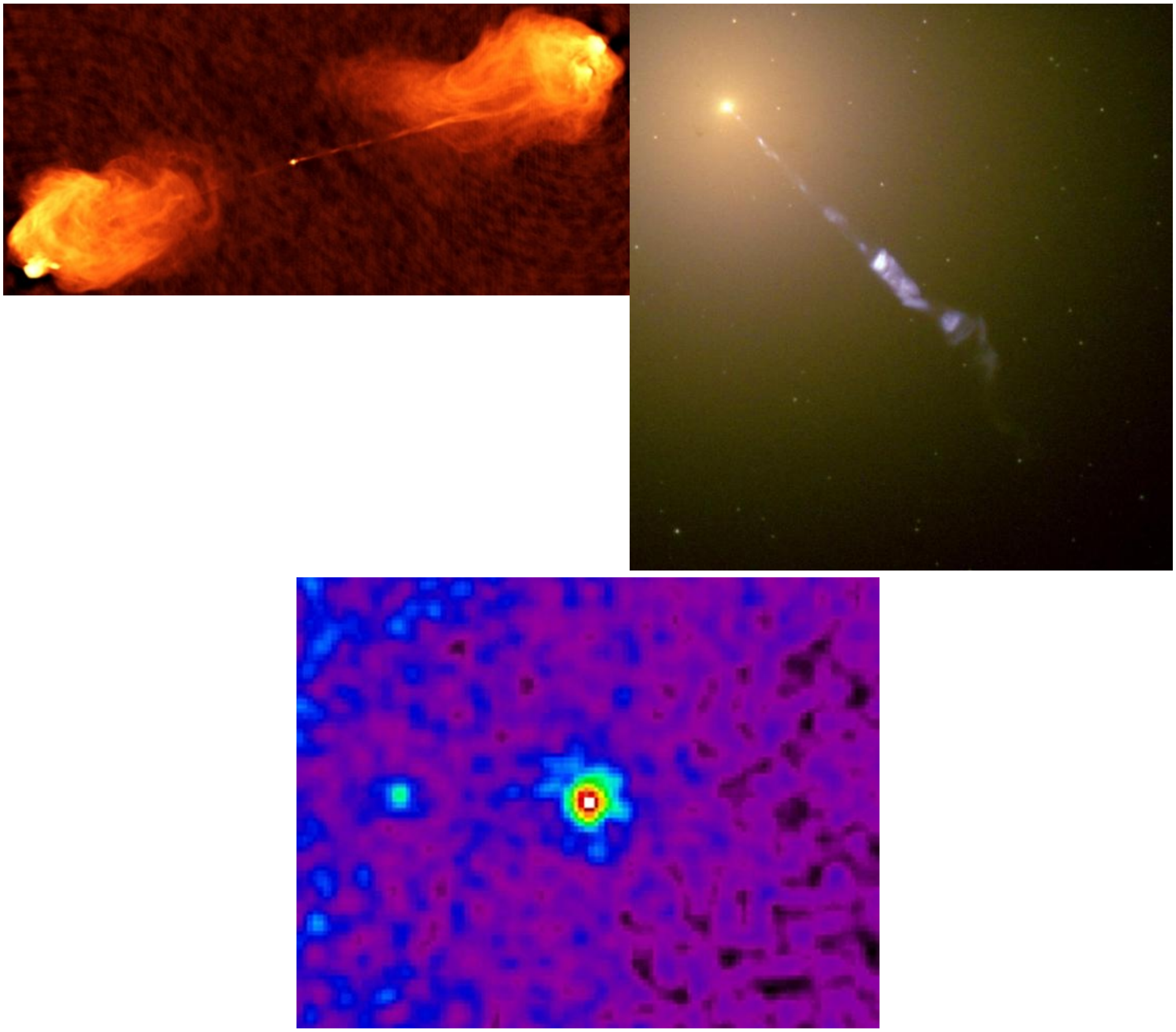


Figure 11.13: AGN are seen in various wave bands: (a) the radio galaxy Cygnus A, (b) the galaxy M87 as seen in optical light, and (c) the Seyfert galaxy NGC 1275 as seen in gamma-rays. AGN are seen at various distances and with varying energy outputs. Credit: (a) Image courtesy of NRAO/AUI, (b) NASA and The Hubble Heritage Team (STScI/AURA), (c) NASA/DOE/Fermi LAT Collaboration

It turns out that most AGN lie at immense distances, hundreds of millions or even billions of light-years from the Milky Way. From that perspective it is clear that their luminosities must be enormous. In fact, the brightest AGN in the sky put out the energy of billions, or sometimes trillions, of Suns, comparable to or even exceeding the brightness of a large galaxy like the Milky Way. Even more impressive is that these luminosities are seen across the entire spectrum, from radio up through infrared and optical, all the way to x-rays and gamma-rays: comparable amounts of energy are seen in each of these wave bands, see Figure 11.14. Something interesting is going on in AGN.

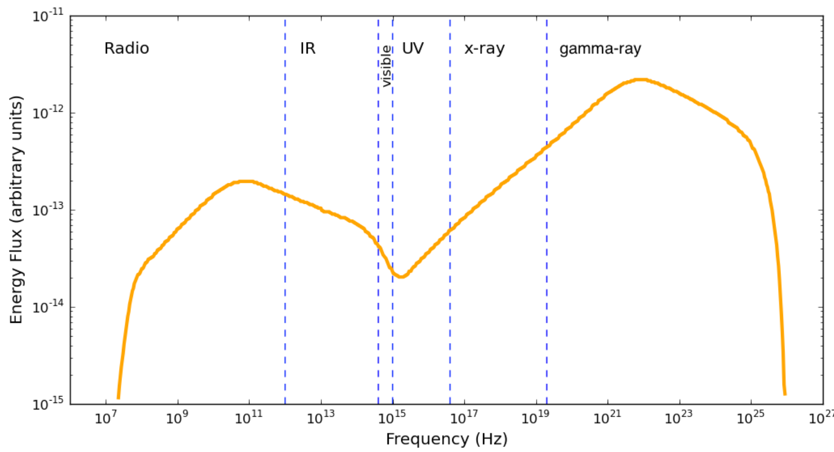


Figure 11.14: AGN emit tremendous energy all across the electromagnetic spectrum. The two-humped shape is common to all AGN, but notice that the energy being emitted extends across ~20 orders of magnitude. Credit: NASA/SSU/Carolyn Peruta

AGN are not just immensely bright, their brightness can vary on extremely short timescales (see Figure 11.15). Fluctuations range from seconds, minutes, or hours to months or years. Typically the longer the time for the variation the larger the AGN is. Some AGN change their brightness by a factor of a hundred over roughly year-long timescales, but the shorter fluctuations are the interesting ones for our purposes. They set upper limits to the size of the emitting region. These fluctuations suggest that the emitting region of AGN is no more than a few light-hours across, a size comparable to the region of our Solar System occupied by the gas giant planets. The shortest timescales for changes in AGN brightness are thought to be caused by emission processes in powerful relativistic jets ejected from their cores. These do not place any constraints on the size of the total emission region.

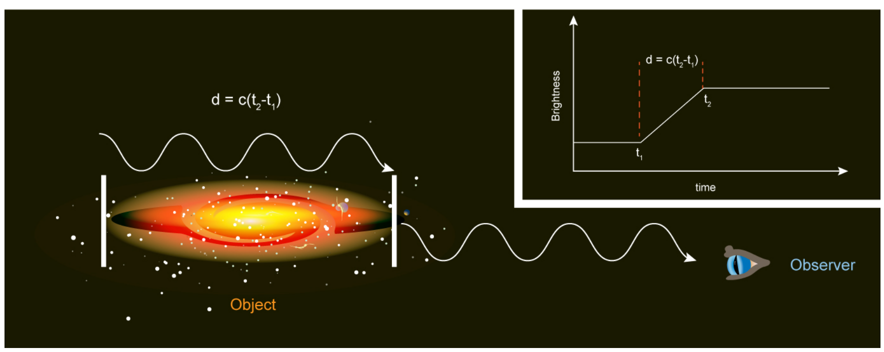


Figure 11.15: The brightness of AGN varies with time. The variations are seen on timescales as short as milliseconds and as long as years. Light travel time allows us to estimate the size of the emitting region. Credit: NASA/SSU/Aurore Simonnet

In the 1960s, when AGN were first beginning to be understood, astronomers were faced with a seemingly unsolvable puzzle. How was it possible to explain the energetics of a region comparable in size to the solar system that exceeded the energy output of an entire galaxy? These regions were always seen in the very centers of large galaxies. Sometimes they glowed brightly across the entire electromagnetic spectrum, but sometimes they would shine only in radio, or just in optical. As x-ray astronomy developed in the 1970s and afterward, it became clear many of the brightest x-ray sources were associated with AGN. How to explain these weird objects?

By a process of elimination, the idea of accretion onto a black hole was fairly quickly deemed the most likely explanation. Thus, AGN became some of the earliest evidence for the reality of black holes, just around the time when x-ray binaries did as well. The following activity, similar to the previous example on the energetics of x-ray binaries, shows why this accretion model was deemed feasible.

Accretion Onto a Supermassive Black Hole

Worked Example:

1. Assume that AGN are powered by accretion onto a black hole. The black holes at the centers of galaxies are likely to be larger than the ones found in stellar binary systems, but as we showed earlier, the mass of the black hole does not matter.

Estimate the accretion rate required to produce a luminosity of 10^{12} solar luminosities. First find the rate in units of kg/s and then convert to M_{Sun}/yr.

- Given: A luminosity 10^{12} times larger than the Sun which has a luminosity of $4 \times 10^{26}\ \mathrm{W} \rightarrow L = 4 \times 10^{38}\ \mathrm{W}$
- Find: Δm/Δt, the accretion rate in units of kg/s and M_{Sun}/yr
- Concept(s): $\frac{\Delta m}{\Delta t} = -\frac{2L}{c^2}$
  where the speed of light , $c = 3 \times 10^8\ \mathrm{m/s}$
- Solution:
  $$\frac{\Delta m}{\Delta t} = -2\left(\frac{4 \times 10^{38}\ \mathrm{W}}{(3 \times 10^8\ \mathrm{m/s})^2}\right) = -8.9 \times 10^{21}\ \mathrm{kg/s}$$

Converting to solar masses per year:

$$(-8.9 \times 10^{21}\ \cancel{\mathrm{kg}} / \cancel{\mathrm{s}})(1\ M_{Sun} / 2 \times 10^{30}\ \cancel{\mathrm{kg}})(3.15 \times 10^7\ \cancel{\mathrm{s}} / \mathrm{yr}) = -0.16\ M_{Sun} / \mathrm{yr}$$

You should have found a much larger accretion rate is required to produce the typical bright AGN luminosities, a trillion times the previous one. At this rate, the black hole must devour about one solar mass every six years. That is quite a lot. It is too much for a small black hole to consume, and in any event , any black hole that feeds at such a prodigious rate cannot stay small for long.

Questions

1.

The unifed model of AGN, depicted in Figure 11.16, explains how AGN are powered— by an accretion disk surrounding a supermassive black hole. As with accretion disks around forming stars, the accretion disks around supermassive black holes also have jets of material beaming out. The next activity will lead you through how the very different looking observations of active galaxies can all be explained by this single model.

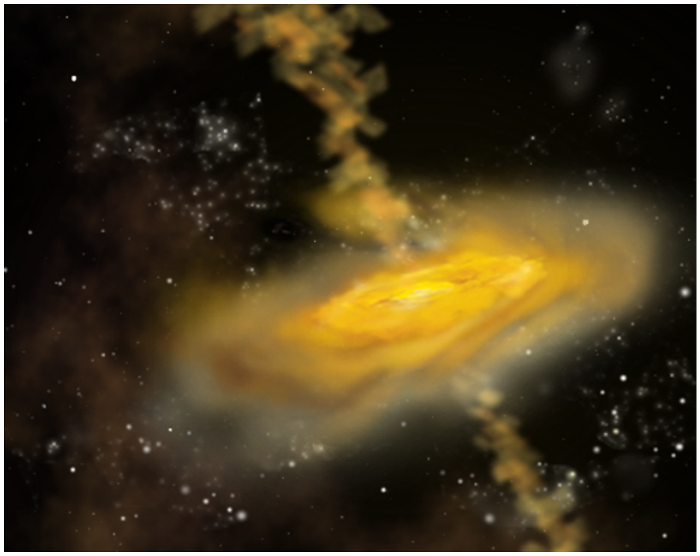
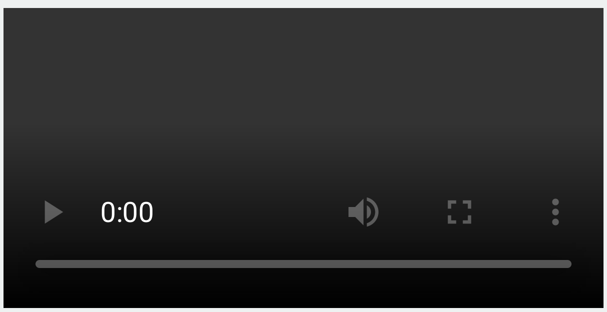


Figure 11.16: This artist’s conception illustrates the unified model for AGN. Astronomers have found that all large galaxies contain supermassive black holes in their centers, with masses of millions or billions of solar masses. If material is falling into these black holes, then the nucleus is said to be active. It glows brightly from radio to gamma-ray energies, in some cases dominating the emission from the galaxy by a factor of one hundred. The detailed workings of AGN are still not understood. Credit: NASA/SSU/Aurore Simonnet

THE UNIFIED AGN MODEL

Watch the movie and then answer the questions.



VIEW TRANSCRIPT

1.

2.

3.

4.

5.

Figure 11.17 shows the AGN Centaurus A in several different wave bands. Each wave band reveals different aspects of the galaxy’s structure.

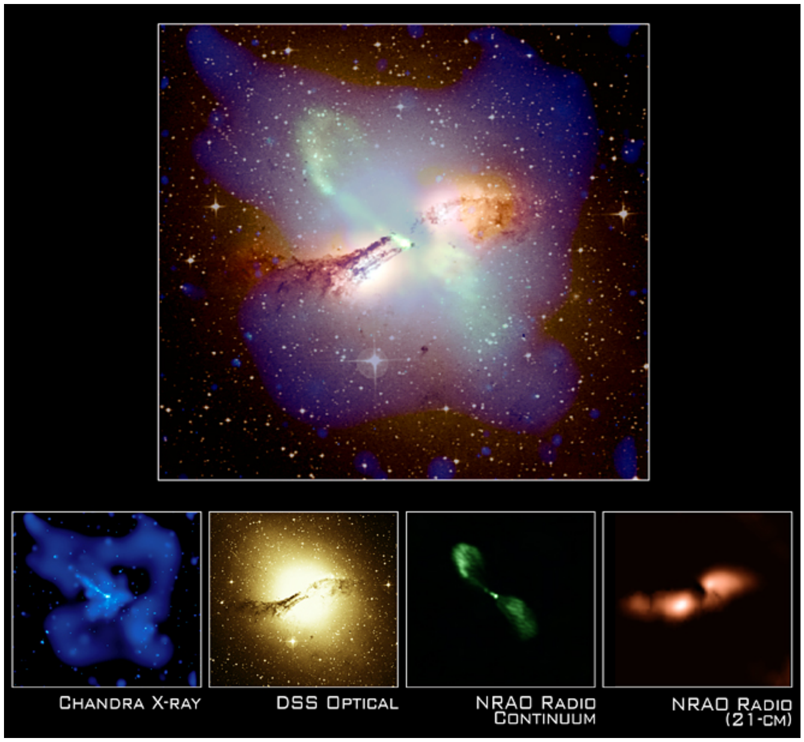


Figure 11.17: This image gives an example of the active galaxy Centaurus A as seen in x-ray, optical, and radio emission. Each wave band reveals different aspects of the galaxy's structure. About a percent of galaxies seem to be active at any given time, but there is evidence that all galaxies go through an active phase at some point (or points) in their evolution. Credit: x-ray (NASA/CXC/M. Karovska et al. http://chandra.harvard.edu/photo/2002/0157/); optical (Digitized Sky Survey U.K. Schmidt Image/STScI); radio continuum image (NRAO/VLA/J. Condon et al.); radio 21-cm image (NRAO/VLA/J.Van Gorkom/Schminovich et al.).

Under the assumption that AGN are powered by accretion onto a black hole, we have found that higher accretion rates are needed for a higher luminosity. But how do we know that the highest accretion rates correspond to the most massive black holes? We might intuitively expect this to be the case since it is easier to collect more material through a large event horizon than through a small one. The surface area of a black hole event horizon depends on its radius squared, so that means it depends on the square of the black hole mass as well. Since more massive black holes have much larger event horizons, we expect that they could support higher accretion rates. While this argument is sensible, it is not very quantitative. How much mass should we expect to be able to fall into a black hole of a given size?

The answer to this question was first explored by Arthur Eddington in a very different context. Eddington was the same British astrophysicist who measured the bending of starlight around the Sun during the eclipse of 1919. Eddington's approach was slightly different from what we have outlined above. He did not consider the size of the black hole event horizon—in fact, he was not thinking about black holes at all. He was thinking about the stability of the most massive stars, and he compared the inward pull of gravity to the outward force of radiation pressure felt by the material in the stars' atmospheres. Eddington reasoned that radiation pressure must act on the gas , and the higher a star's luminosity, the higher the outward pressure on the gas. Eventually, if the luminosity were large enough, the outgoing radiation would push on the material in the atmosphere hard enough to overcome gravity, blasting it off the star. In essence, he said that at the point where the luminosity is so great that radiation pressure cancels gravity, the star would become unbound; radiation will blow it apart.

The same argument can be applied to accreting black holes. As the accretion rate increases, so does the luminosity. As the luminosity increases, the radiation pressure it exerts on the in-falling material also rises. If the luminosity becomes large enough—as the result of a large accretion rate—then it will be able to halt the accretion. There should therefore be an upper limit to the amount of material that can fall onto a black hole, a limit imposed by the very radiation produced by the gravitational energy released through accretion.

This limiting value for the luminosity is now called the **Eddington limit** or Eddington luminosity. It has the value

$$L_{edd} = 6.3\, M_{BH}$$

where L_{edd} is the luminosity in W, M_{BH} is mass of the black hole in kg, and the constant 6.31 has units of W/kg. The expression gives the highest accretion rate possible for a spherically accreting object. The Eddington luminosity is proportional to the mass, which means that the higher the mass, the higher the maximum luminosity. See Going Further 11.5: Derivation of the Eddington Luminosity to see how we arrive at this expression.

Eddington Luminosity

Use the *Eddington Luminosity Tool* to find the maximum luminosity for several black hole masses.

1.

2.

3.

4.

We do not expect spherical accretion onto a black hole, we expect it to be via an accretion disc. As a result, the Eddington luminosity is really just a guide. We can use it to give us an idea of how the accretion onto a black hole is related to its mass, but there could be (and probably are) large variations among different black hole accretion systems.

GOING FURTHER 11.5: DERIVATION OF THE EDDINGTON LUMINOSITY

Energy arguments suggest strongly that the gravity of black holes provides the power for AGN. However, these arguments are not completely convincing. After all, it could be that some other undiscovered object or process is powering AGN. In the last decade any such doubts have been mostly set aside. Careful studies of the motions of visible material near AGN have shown that a black hole lies at their core. One of the strongest cases for the AGN-black hole connection is the nearby weak AGN, NGC 4258, also called M106 (Figure 11.18). The motion of gas in the galactic nucleus indicates the presence of an unseen 40 million solar mass object within 0.1 parsecs of the center— the typical distance between stars in the vicinity of the Sun is roughly one parsec. So we have about one solar mass of material per cubic parsec in our region of the Galaxy. This stands in contrast to 40 million solar masses in 0.001 cubic parsecs in the nucleus of NGC 4258.

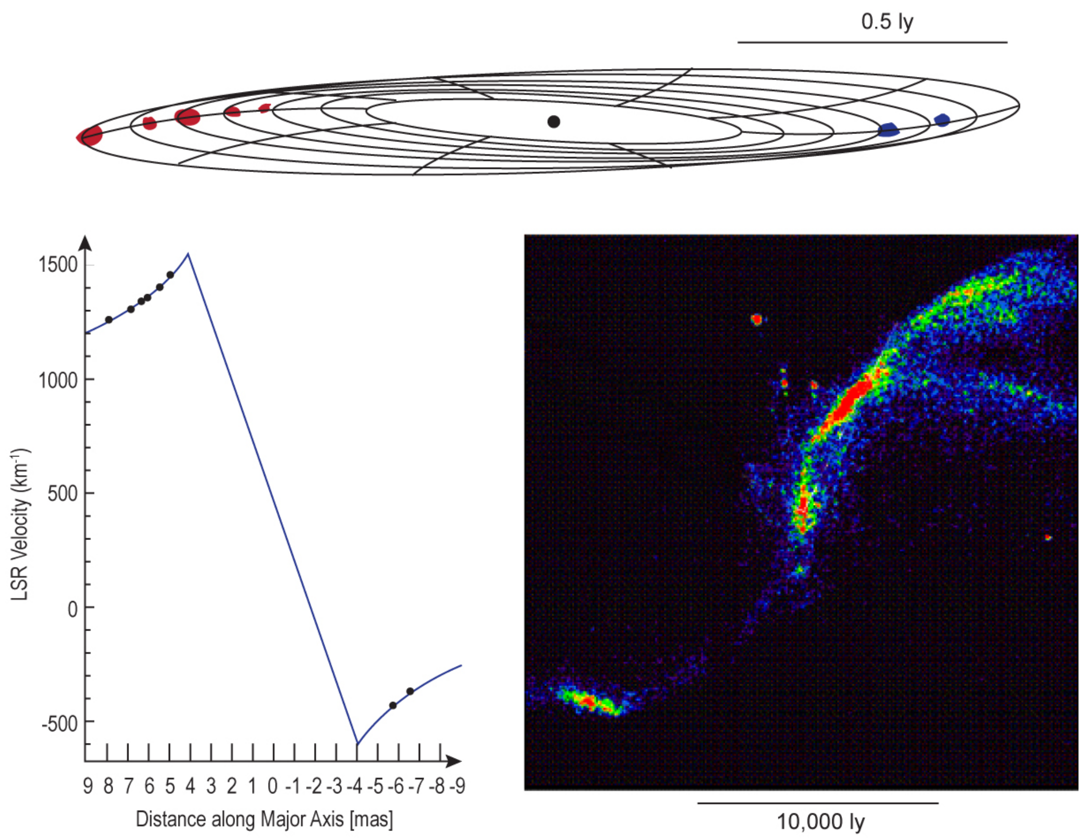


Figure 11.18: Analysis of the gas motions in the center of the nearby AGN NGC 4258 indicates the presence of a central object with the mass of 40 million Suns. Top: the area within 0.5 light-years of the black hole is a warped disk. Bottom left: the rotation curve of material near the black hole. Bottom right: An expanded view of NGC 4258. Credit: NASA/SSU/Aurore Simonnet, based on an image by L. Greenhill courtesy of NRAO/AUI.

In addition to detecting the presence of black holes in galaxies, astronomers have also started to determine properties other than mass for some of them. One of these properties is spin. The spin can be deduced by examining emission spectra produced in the accretion disk. For instance, in the active galaxy NGC 1365, astronomers used two x-ray telescopes to measure the spectrum of the material in the accretion disk (Figure 11.19). The XMM-Newton telescope measured the spectrum between 3 and 10 keV. At these energies interpretation of the spectrum is ambiguous because its shape could be the result of either the black hole spin or the presence of obscuring clouds in our line of sight. However, the NuSTAR telescope was able to measure the spectrum out to 79 keV. With this higher energy part of the spectrum, the ambiguity drops away, and only a spinning black hole remains consistent with all the observations, including both of these x-ray observations and other observations at infrared and optical energies.

The black hole at the center of NGC 1365 is observed to be spinning almost as fast as general relativity predicts a black hole should be able to spin. This has important implications for the formation of supermassive black holes, and by extension, for galaxies themselves.

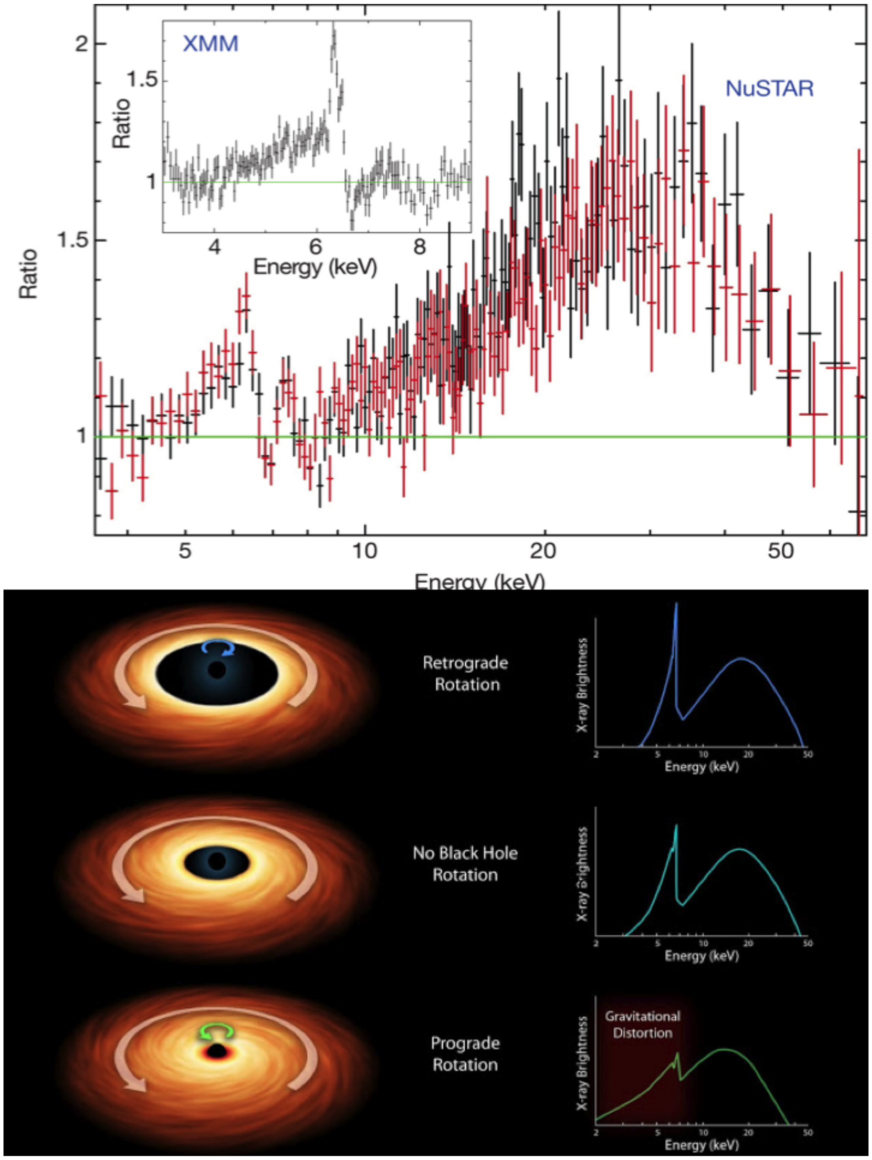


Figure 11.19: The plot at the top shows the x-ray spectra of NGC 1365 from two telescopes, XMM-Newton and NuSTAR. Combining the spectra allows astronomers to test models for the formation of the emission spectrum, one based upon the presence of absorbing clouds in our line of sight, the other based upon the spin of the supermassive black hole in the center of the galaxy. Only the spinning black hole model is consistent with the observations. The bottom image shows the effects of black hole rotation on theoretical energy distributions. Credit: Risaliti, G., et al. (2013), *Nature*, 494, 449; NASA/JPL-Caltech

11.4.3: Detecting the Milky Way Black Hole

Currently, the strongest observational case for the presence of a supermassive black hole is not in AGN at all. It is in the center of our own Galaxy, the Milky Way. Research carried out by teams at UCLA and at the Max Planck Institute for Astrophysics, near Munich, Germany, have shown that the center of our Galaxy contains a black hole with the mass of four million Suns. This conclusion is the result of watching stars move in the central region. They are seen to be orbiting a massive, invisible object.

Figure 11.20 shows a series of observations of our Galactic center by a team at UCLA led by Andrea Ghez (Figure 11.21). The UCLA team used the Keck Telescope, one of the largest telescopes in the world, located atop the Mauna Kea volcano on the Big Island of Hawaii. The blobs of light in Figure 11.20 are images of stars near the center of our Galaxy, and the dots show the measured positions of several of those stars during almost two decade of observations, from 1995 to 2012. You can see that these stars are all orbiting the same unseen object. That is the location of the black hole at the center of the Milky Way. Figure 11.22 shows an animated version of the data.

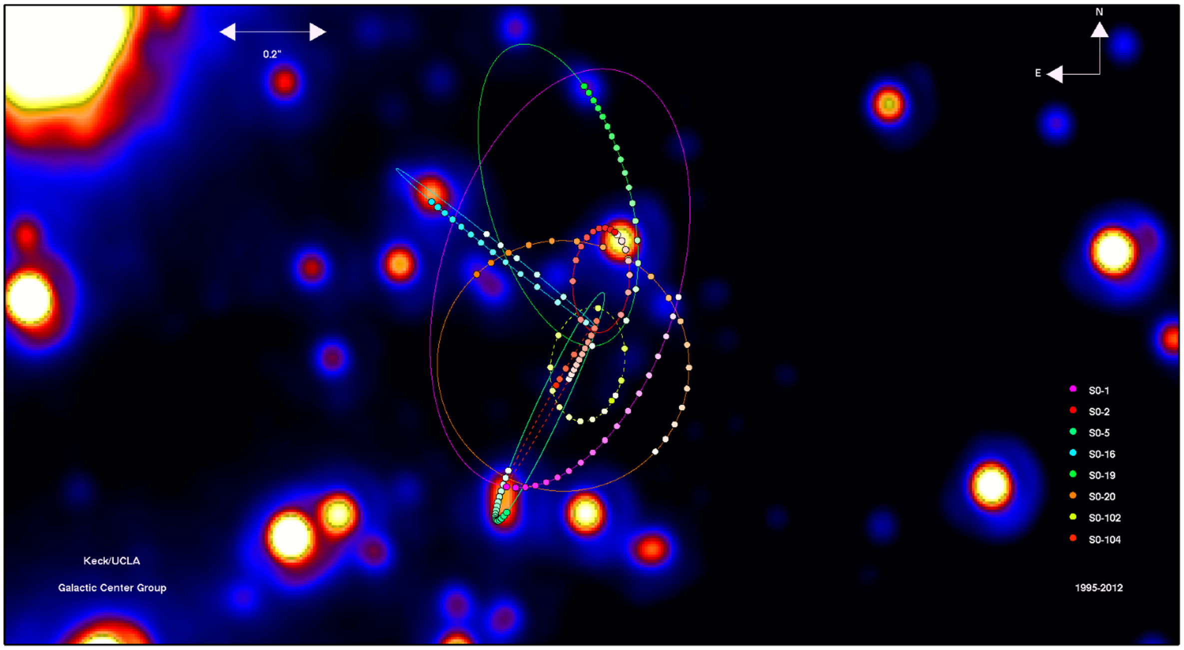


Figure 11.20: The orbits of stars within the central 1.0 × 1.0 arcseconds of our Galaxy. While every star in this image has been seen to move over the past 17 years, estimates of orbital parameters are only possible for the eight stars that are closest to the black hole. The annual average positions for these eight stars are plotted as colored dots, which have increasing color saturation with time. Also plotted are the best fitting simultaneous orbital paths. These orbits provide the best evidence yet for a supermassive black hole, which has a mass of 4 million times the mass of the Sun. Credit: Keck/UCLA Galactic Center Group

Figure 11.21: The UCLA Galactic Center Group, led by Andrea Ghez (lower right). For more information, visit the group's website: http://www.astro.ucla.edu/~ghezgroup/gc/. Credit: Mary Watkins

Figure 11.22: Images taken by Andrea Ghez and her group using the Keck telescope are used to track specific stars orbiting the black hole at the center of the Milky Way. These orbits, and a simple application of Kepler's laws, provide the best evidence yet for a supermassive black hole, which has a mass of 4 million times the mass of the Sun. Especially important are the stars S0-2, which has an orbital period of only 15.78 years, and S0-16, which passes a mere 90 AU from the black hole. Credit: Keck/UCLA Galactic Center Group

Orbits are the key to measuring the masses of astronomical objects. If we want to measure the mass of something, we need to find one or more objects that orbit it and measure the properties of those orbits. Once we know the orbital period and average orbital distance of an orbiting object, we can use that information to measure the force of gravity needed to hold that object in orbit. Then, from the force of gravity, we can determine the total amount of mass in the system, because gravity is directly related to mass.

Consequences of all three of Kepler's laws of orbital motion are evident in the image and animated image. The orbits of stars around the black hole are ellipses (Kepler's first law). The stars move faster when they are close to the black hole and slower when they are farther away (Kepler's second law). Finally, stars orbiting at greater distances have longer orbital periods (Kepler's third law).

Newton's version of Kepler's third law relates the orbital distance, *a*, and orbital period, *p*, to the total amount of mass in the system.

$$M = \frac{a^3}{p^2}$$

Here *M* is in units of solar masses, *a* is in AU, and *p* is in years. It is easy to see how this formula works for the Earth–Sun system. Plugging in a = 1 AU and *p* = 1 year gives a total mass for the Earth–Sun system of 1 M_{Sun}, and you would get the same mass by plugging *a* and *p* for any of the other planets. If we apply the formula to the stars in the center of the galaxy, using AU for their orbital sizes and years for their periods, then they determine the mass of the central black hole in terms of solar masses.

By determining the orbital distances and orbital periods of the stars near the Milky Way's central black hole in this way, the UCLA group reports a mass for the black hole of $4.1 \pm 0.4 \times 10^6$ M_{Sun}.

In addition to these stellar motions, evidence for our Galaxy's supermassive black hole comes from radio emission and occasional x-ray flares. Presumably these happen when gas falls into the black hole.

In near infrared wavelengths, ESO's Very Large Telescope (VLT) has observed a gas cloud being tidally stretched (spaghettifiied) by the Milky Way's black hole. Figure 11.23 is a series of images taken by the VLT from 2004 to 2013. The horizontal axis shows the stretching of the head vs. the tail of the cloud over time and the vertical axis shows how the spread in the velocity of the material has increased over time. The closest approach of the cloud to the black hole is about 25 billion km (about 170 AU).

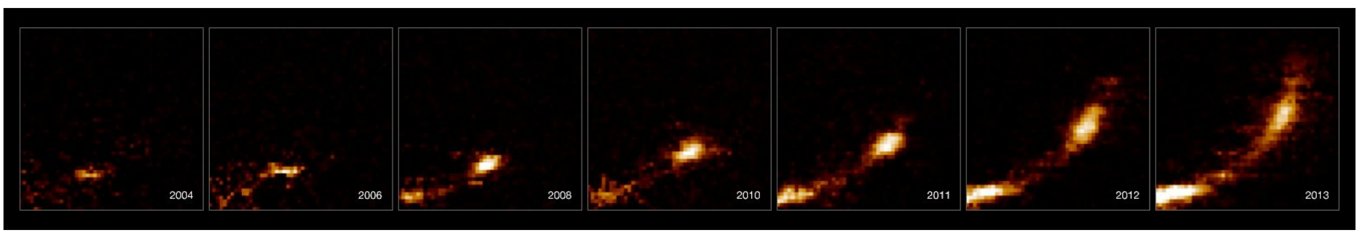


Figure 11.23: Images taken by the VLT from the years 2004 through 2013 of a cloud near the galactic center. The cloud is being ripped apart by the Milky Way's central black hole. The horizontal axis in the figure shows the size of the cloud; it has stretched more than 160 billion km. The vertical axis shows velocity; the head of the cloud is moving millions of km/h faster than the tail. The cloud is moving at nearly 1% of the speed of light. Credit: ESO/S. Gillessen

Our galaxy definitely does not host an AGN, though it does host a gigantic black hole in its center, just as do AGN. Why the difference? The answer seems to be the environment of the black hole. In galaxies with large amounts of gas in their centers, the black holes induce AGN behavior as the gas falls into them, releasing vast amounts of gravitational potential energy. In most galaxies, including ours, the galactic nuclei are mainly devoid of gas. Since there is no material falling into the black hole there is no energy released and therefore no AGN.

The fraction of galaxies that exhibit AGN behavior is only around a percent or so at the present time. The proportion is observed to have been larger in the past, and it seems that the AGN phenomenon is related to galactic formation and evolution.

11.4.4: Intermediate Mass Black Holes

So far in this section we have described black holes with masses comparable to stars. These are called stellar mass black holes. We have also discussed the supermassive black holes found in the centers of galaxies. Supermassive black holes are immense, with masses measured in the millions or billions of solar masses, while stellar mass black holes are no larger than about a dozen, or maybe a few dozen, times the mass of the Sun. But what about black holes with masses in the middle? Have astrophysicists found any so-called intermediate mass black holes? The answer, somewhat surprisingly at first, seems to be “maybe.” There is some evidence for these objects, but the evidence is not conclusive.

If we consider how black holes form and grow, we might set aside our initial surprise at the absence of any with intermediate mass. The masses of these black holes would range from several hundred to perhaps tens of thousands of solar masses. What objects would form a black hole of that size?

Stars do not seem able to produce black holes of intermediate sizes. The largest stars are around 200 solar masses, at least in the present-day Universe. But such stars have enormous winds. Their winds allow them to lose most of their mass over their lifetimes. When they eventually do undergo collapse to a black hole, the remnants they leave are in the stellar mass range, measured in tens of solar masses.

We know that most stars are formed in groups, and perhaps these groups could form many massive stars. These would in turn create many stellar mass black holes. Then perhaps these stellar mass black holes could somehow combine to form a single black hole with a mass of a few hundred solar masses, or even a few thousand.

While this scenario seems reasonable, it must not be very common, simply because astronomers do not see very many intermediate mass black holes. In fact, there is debate about whether any have been detected at all. A few candidate objects have been detected via their x-ray emission. These ultra-luminous x-ray sources are brighter than even the brightest x-ray binaries by a factor of ten or so. Some astronomers think that the extra x-ray luminosity is the result of accretion onto a large black hole. Several models have been developed along these lines. None have been completely successful.

Additional evidence for intermediate mass black holes could come from their gravitational effects on stars and gas near to them. That is how the presence of the black holes in the center of the Milky Way and other galaxies has been uncovered. However, intermediate mass black holes have weaker gravity than the supermassive black holes at the centers of galaxies. So they would have a much smaller effect on objects close to them. Detecting any motions they might induce in nearby objects stretches the capabilities of the largest ground-based telescopes, like those used to study the Milky Way center. Even the Hubble Space Telescope is not quite up to the task. Perhaps later generations of telescopes and detectors will be powerful enough to look for the tell-tale signs of distant black holes too small to provide the gravitational wallop of their supermassive cousins.

There is another possibility. Perhaps the merger of stellar mass black holes to form intermediate mass black holes is one step on the way to creating the supermassive black holes we detect. After all, we have seen that only a short time is required (on relevant timescales) to grow a black hole to an extremely large size if material is present to accrete onto it. If this is the correct description, then the reason we do not see many intermediate mass black holes anymore is that they have all been merged into the monsters in galactic centers. This could be part of the normal processes in the early Universe that formed the galaxies or their building blocks.

However, if intermediate mass black holes are still being created in star forming regions, then they should still be around. There has not been time for them to sink into galactic centers, nor is there a lot of material present to greatly increase their masses. This means we should be able to see them if they are around. Whatever the truth, the means of formation for black holes more massive than stars is still an open question.

This page titled 11.4: Astrophysical Black Holes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

11.5: Wrapping It Up 11 - Black Hole Densities

Learning Objectives

- Explain the following laws within the Ideal Gas Law

Once material crosses the Schwarzschild radius of a black hole there is no escape; not even light is moving fast enough to get back out. Imagine a sphere around the black hole singularity reaching out to the Schwarzschild radius. This sphere represents the effective volume of the black hole. Yes, it's mass is seemingly concentrated in a point, but the sphere encompassing the region of no return is much larger. More massive black holes have larger Schwarzschild radii, so how does their average density compare to smaller examples?

11.5.1: **Part I: Black Hole With an Average Density Like Water**

The average density of any object is its mass divided by the volume it occupies. For a spherical object with radius R, the volume is $\frac{4}{3}\pi R^3$. Liquid water has an average density of 1000 kg/m^3. Using the Schwarzschild radius to estimate the volume, how massive must a black hole be to have the average density of water? How does this mass compare to the sun?

1.

2.

3.

4. If the average density is equal to that of water (1000 kg/m^3), what is the mass of the black hole in kg?

5.

11.5.2: **Part II: Black Hole at the Center of the Galaxy**

1. The supermassive black hole at the center of the Milky Way has a mass of $4 \times 10^6\ \mathrm{M_{Sun}}$. What is the average density of this black hole?

2.

3.

4.

This page titled 11.5: Wrapping It Up 11 - Black Hole Densities is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

11.6: Mission Report 11 - Black Hole Densities

A.

B.

C.

D. Questions to be graded for accuracy:

1.

2.

3.

4.

This page titled 11.6: Mission Report 11 - Black Hole Densities is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

12: Gravitational Lenses

Chapter 12 delves into the phenomenon of gravitational lensing. Light from distant sources is deflected by the curved spacetime around massive objects, providing astrophysicists with the ability to map out mass, both seen and unseen. In the first part of the chapter you will examine the geometric properties of simple gravitational lenses. In the second part, you will use astrophysical gravitational lenses as instruments for studying dark matter, weighing galaxy clusters, and finding extra-solar planets.

This page titled 12: Gravitational Lenses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

12.0: Gravitational Lensing Introduction

Video Transcript

Just as a wanderer in a desert sees a mirage when light from remote objects is bent by the warm air hovering just above the sand, we may also see mirages in the universe. The mirages we see with a modern telescope such as the Hubble Space Telescope do not arise from warm air, but instead from remote clusters of galaxies; huge concentrations of matter.

Long ago, some people thought the Earth was flat. This is not surprising, since, in our everyday lives, we're not aware of the curvature of our planet. Space is actually curved, although we can't see that for ourselves on a starry night. The curvature of space does, however, produce phenomena that astronomers can see.
One of Albert Einstein's predictions is that gravity warps space and therefore distorts rays of light in the same way that ripples on a pond create a warped honeycomb pattern of light on the sandy bottom. Light from distant galaxies is distorted and magnified by the gravitational field of massive galaxy clusters on its path to Earth. The effect is like looking through a giant magnifying glass and the result is called gravitational lensing.

The weird patterns that rays of light create when they encounter a weighty object depend on the nature of the lensing body; thus the background object can appear in several guises. Einstein rings, where the whole image is boosted and squeezed in a circle of light; multiple images, ghostly clones of the original distant galaxies; or distorted into banana-like arcs, then arclets.

Though Einstein realized in 1915 that this effect would happen in space, he thought that it could never be observed from the Earth. However, in 1919 his calculations were indeed proved to be correct. During a solar eclipse expedition to Principa Island near the west coast of Africa led by the renowned British astronomer Arthur Eddington, the positions of stars near the obscured solar disc were observed. It was found that the stars had moved a small but measurable distance outwards on the sky compared with when the sun was not in the vicinity.

Nowadays, faint gravitational images of objects in the distant universe are observed with the best telescopes on Earth and, of course, with the sharp-sighted Hubble.
Hubble was the first telescope to resolve details within the multiple arcs, revealing the form and internal structure of the lensed background objects directly.
In 2003, astronomers deduced that a mysterious arc of light on one of Hubble's images was the biggest, brightest, and hottest star-forming region ever seen in space.
I takes fairly massive objects, for example, clusters of galaxies, to make space curve so much that the effect becomes visible in deep images of the distant universe – even with Hubble's astonishing resolution. And so far, gravitational lenses have been observed mainly around clusters of galaxies, which are collections of hundreds or thousands of galaxies and are thought to be

the largest gravitationally bound structures in the universe.

Astronomers know that the matter we see in the universe is just a tiny percentage of the mass that must be there, for matter exerts a gravitational force and the visible stuff is simply not enough to hold galaxies and clusters of galaxies together. Since the amount of warping of the banana-shaped images depends on the total mass of the lens, gravitational lensing can be used to weigh clusters and to understand the distribution of the hidden dark matter.
On clear images from Hubble, one can usually associate the different arcs coming from the same background galaxy by eye. This process allow astronomer to study the details of galaxies in the young Universe and too far away to be seen normally with the present technology and telescopes.
A gravitational lens can even act as a kind of natural telescope. In 2004 Hubble was able to detect the most distant galaxy in the known Universe, using the magnification form just such a gravitational lens in space.

The measurement of the bending of starlight around the Sun during the total solar eclipse of 1919 is what made Albert Einstein a world celebrity. In this chapter we will explore gravitational lenses and the phenomenon of the bending of light by gravity. Some lenses are point-like masses such as stars. Others are extended mass distributions such as galaxies and galaxy clusters. We will discover that the lens phenomenon is an important tool for astrophysics. Beyond being an interesting consequence of general relativity in its own right, the presence of lenses can provide information about mass that is otherwise unseen, but that still affects the path of light through space.

This page titled 12.0: Gravitational Lensing Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

12.1: What Are Gravitational Lenses?

Learning Objectives

- You will understand that gravity can bend the path of light.
- You will understand that this bending of light depends on the mass of the source of the gravity.
- You will understand that objects other than the Sun can act as lenses and that many instances of gravitational lensing have been seen.

What Do You Think: What is a Lens?

You are probably familiar with the glass lenses used to correct vision and construct instruments like cameras, microscopes and telescopes. These lenses use the fact that light travels more slowly through glass than through air, thereby causing the light to alter its direction of travel when passing from one substance to the other. By carefully shaping the glass/air interface, it is possible to bring all the light passing through the lens to a single point called the focus. To learn more about how optical lenses bend and focus light, see Going Further 12.1: Optical Lenses.

Gravity can also alter the direction that light travels. The effect is somewhat analogous to the way glass lenses work. Gravity, as a result, can sometimes act as a lens, focusing the light from a source. An optician would not consider the gravitational lenses we will study in this chapter to have very high quality; they have many imperfections that cause distortions of various kinds. Nonetheless, gravitational lenses provide astrophysicists with a powerful tool to probe the Universe.

GOING FURTHER 12.1: OPTICAL LENSES

Gravity bends space and time. Light rays follow this curvature and thus have their directions distorted by gravity. We can calculate the angle of deflection *(α)* of a light ray passing by a mass *(M)*. If we include both the space and time curvature pieces, the deflection angle is:

$$\alpha = \frac{4GM}{bc^2} = \left(\frac{4G}{c^2}\right)\left(\frac{M}{b}\right)$$

G and *c* have their usual values of 6.67×10^{-11} Nm^2/kg^2 and 3×10^8 m/s, respectively. The parameter b is called the impact parameter. It is the closest approach of the light to the massive object.

The expression has been written in two ways to make it more clear which parts are constant (the left hand bracket containing only constant terms) and which can vary. Only the mass and impact parameter can vary from one lens to another. The other terms are all constant.

The geometry of the deflection is shown in Figure 12.1.

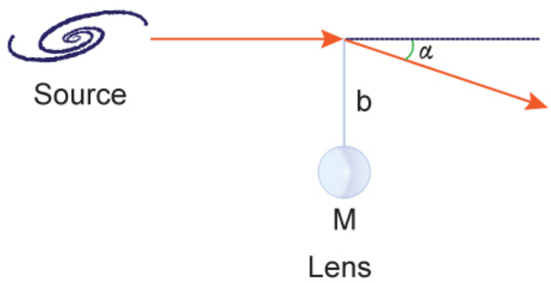


Figure 12.1 Light rays from a source are bent by the gravity of mass *M* (the gravitational lens). The angle of deflection is *α* and the distance of closest approach is *b*. Here we have made the simplification that a light ray bends only at the midpoint of the lens, whereas it actually follows the continuous (spherical) curvature of spacetime around the lens. Credit: NASA/SSU/Aurore Simonnet

For starlight passing by the Sun, the bending was only appreciable for light rays passing very close to the surface of the Sun. For light passing farther away the deflection became too small to measure. We can see that the expression for the deflection angle depends on the distance from the lens, and that the deflection becomes smaller as the impact parameter (b) becomes larger. But what if the Sun had a larger mass? Would the bending of starlight become noticeable at larger distances for a more massive object?

Again, looking at the expression for the deflection angle, if the mass of the lens is bigger, the deflection angle will be bigger, too. So the effect will be more noticeable if the lens has a bigger mass, regardless of how close the light passes to the object. Whether or not the deflection is noticeable just depends on the size of α, which itself depends only on the ratio of the mass to the impact parameter - the constant term is the same for any lens. As long as the mass and impact parameter combine such that the deflection angle is big enough, the lensing effect will be visible.

Gravitational Lensing by Objects Other Than the Sun

In this activity, we will compute the angle of deflection for light passing by objects other than the Sun. If we include both the space and time components of curvature, then light going just past the Sun is bent at an angle of about 2 arcseconds.

Worked Examples:

1. Calculate the angle α in radians for light skimming just past the outer edge of a white dwarf. Use a mass of 1.4 solar masses and a radius of 6.4 × 10^6 m for the white dwarf.

- Given: M = 1.4 solar masses = $1.4 \times 2 \times 10^{20}\ \mathrm{kg} = 2.8 \times 10^{30}\ \mathrm{kg}\ b = 6.4 \times 10^6\ \mathrm{m}$
- Find: deflection angle α
- Concept: $\alpha = \dfrac{4GM}{bc^2}$
- Solution:

$$\alpha = \frac{(4)(6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}})(2.8 \times 10^{30}\ \mathrm{kg})}{(6.4 \times 10^6\ \mathrm{m})(3 \times 10^8\ \mathrm{m\,s^{-1}})^2} = 1.3 \times 10^{-3}\ \mathrm{radians}$$

2. Now convert this angle to arcseconds.

Use the conversion factor $1\ \mathrm{radian} = 2.06 \times 10^5\ \mathrm{arcsec}$.

$$1.3 \times 10^{-3}\ \cancel{\mathrm{radian}} \times (2.06 \times 10^5\ \mathrm{arcsec}\ /\ \cancel{\mathrm{radian}}) = 267\ \mathrm{arcsec}$$

3. Now compare to the deflection of light by the Sun.

The deflection by the Sun is about 2 arcsec, so this angle is 267/2 = 133 times greater.

Questions:

1.

2.

3.

This page titled 12.1: What Are Gravitational Lenses? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

12.2: Lensing by Point Masses

Learning Objectives

- You will understand how the lensing effect depends on the relative positions of the source, lens and observer.
- You will understand that the mass of the lens can be determined using gravitational lensing.
- You will understand that objects such as dim stars, brown dwarfs, and stellar remnants (white dwarfs, neutron stars, black holes) can act as gravitational lenses. Collectively they are called MACHOs.
- You will understand that microlensing observations imply that a large percentage of the dark matter in our Galactic halo is not regular matter (MACHOs) but rather a yet-to-be-determined exotic form of matter.

What Do You Think: Microlensing

12.2.1: Modeling Gravitational Lenses

Gravitational lenses are easiest to understand if we employ a simple model, similar to the one used to understand conventional lenses in optical systems. We assume that the light travels in a straight line until it reaches the lens, and then its direction is altered by some angle. After that the light travels in a straight line to the observer. The lens itself is assumed to be infinitely thin and to act on the light only as it crosses the plane of the lens. This model captures the essential behavior of the lens system: the lens causes the source to appear to be displaced from the point of view of the observer. At the same time, it avoids the complicated mathematical treatment needed to find the actual path of the light through the region of strong gravity that creates the lens in the first place.

The assumptions used for all lenses, including gravitational ones, are the following:

1. Light travels in a straight line from the source to the infinitely thin lens.
2. The light is bent at the lens plane (and only at the lens plane).
3. The light then travels in a straight line to the observer.

Students who are familiar with the geometrical approach to optics or who have read Going Further 12.1: Optical Lenses will see some similarities between these assumptions and those used in the treatment of glass lenses. The geometry of a lens system is shown in Figure 12.2. A source located at point S is emitting light that is deflected by a lens located between the source and some observer, located at point O. The lens could be anything with mass: a star or planet, or a galaxy or cluster of galaxies. If the lens did not deflect the light, then the observer would see the source at its true position. However, due to the deflection of the light caused by the lens's gravity, the image of the source is observed to be at position I.

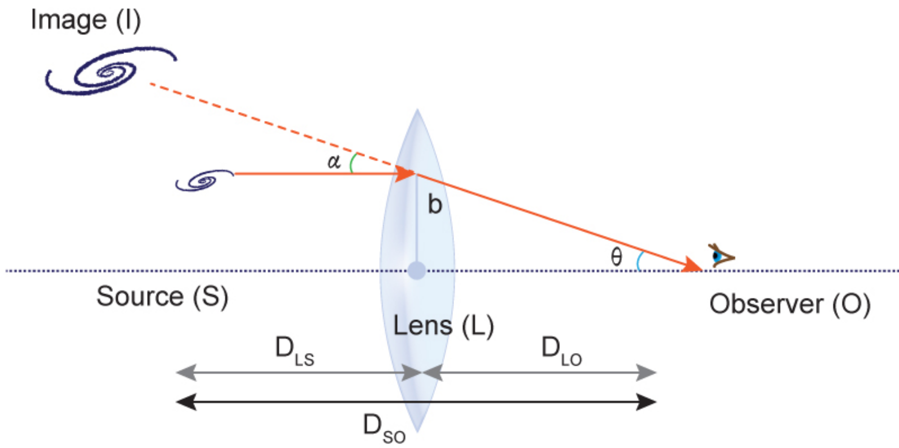


Figure 12.2 The figure shows a side view of the basic elements of a gravitational lens system. The observer is at point O, the source is at S. These two are separated by a distance D_{SO} . The angle α is equal to the deflection angle. The deflection of the light makes the image appear to be at position I as seen by the observer. The image has thus been offset by a small amount compared to the source, i.e. the case if no lens were present. The angle θ is the angular separation of the image from the observer. The lens is taken to be somewhere along the path of the light a distance D_{LO} from the observer and D_{LS} from the source. These two distances do not have to be the same. Credit: NASA/SSU /Aurore Simonnet

In Figure 12.2, the line connecting the observer to the lens is called the **optical axis**. We can imagine the optical axis extending behind the lens indefinitely. The source can be located on the optical axis, but in general it does not have to be. That is why we have located the point S slightly off the optical axis. Since gravity always bends a light ray toward the optical axis, the position of the image, point I, will always be located farther from the optical axis than the actual source of the light for the configuration shown. There is an additional image formed below the optical axis for which this is not true, but we will consider that image later. The distance between the source and the observer is D_{SO}, the distance between the lens and the observer is D_{LO}, and the distance between the lens and the source is D_{LS}. The angle θ in Figure 12.2 is the angular separation between the observer, the image, and the optical axis.

In Going Further 12.2: The Lens Equation, we examine the geometry of gravitational lenses in more detail, and derive a general equation relating the angles and distances for several cases.

It turns out that if the source, lens, and observer are all lined up along the optical axis, the image of the lensed object makes a ring around the lens, as shown in Figure 12.3. The ring-like image is called an Einstein ring.

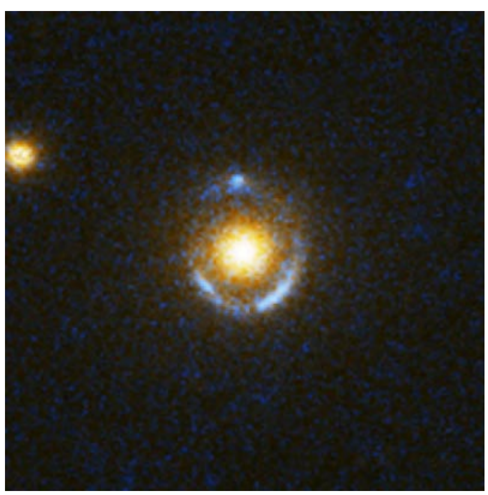


Figure 12.3 This object (known as SDSS J073728.45+321618.5) is one of several Einstein rings discovered as part of the Sloan Lens ACS Survey (SLACS). The width of this image is about 8 arcsec. For more information, see the press release. Credit: NASA, ESA, A. Bolton (Harvard-Smithsonian CfA) and the SLACS Team

The angular radius, θ_E, of an Einstein ring is given below.

$$\theta_E = \sqrt{\left(\frac{4GM}{c^2}\right)\left(\frac{D_{LS}}{D_{LO}D_{SO}}\right)}$$

This is called the **Einstein radius of the lens** system. We distinguish it by adding the subscript E to the angle θ from Figure 12.2. As usual, G and c are the gravitational constant and the speed of light. The size of the Einstein ring varies only with the geometry of the lens, as expressed inside the parentheses on the right-hand side, and on the total projected mass within the impact parameter, $M(b)$. For example, if the mass within the impact parameter is larger, then θ_E will be larger. If the distance between the source and observer (D_{SO}) is bigger, then θ_E will be smaller.

The following activities will give you an idea of the size of the Einstein radius for objects with masses relevant to astrophysics.

The Einstein Radius

In this activity, you will be able to adjust the parameters in the lens equation in order to determine their effect on the size of the Einstein ring.

There are three sliders, one for each of the following:

- The mass of the lens (M)
- The distance to the source from the observer (D_{SO})
- The relative positioning of lens, source, and observer (the ratio D_{LS}/D_{LO})

Play Activity

A. Mass of the Lens

1.

2.

https://phys.libretexts.org/@go/page/31365

3.

4.

B. Distance to the Source from the Observer

1.

2. I

3.

4.

C. Relative Position of the Lens

1.

https://phys.libretexts.org/@go/page/31365

2.

Size of the Einstein Radius

In this activity, we will compute the sizes of Einstein radii for solar mass lenses in our Galaxy as well as for extragalactic lenses like the one pictured in Figure 12.4.

Worked Examples:

1. Compute the Einstein radius for a distant galaxy being lensed by a foreground trillion solar mass galaxy. Assume that all of the mass of the lens lies within the impact parameter, b. Assume the distance to the gravitational lens is 1000 Mpc and the

distance between the lens and source is 2000 Mpc.

- Given: Mass M = 1e12 × 2e30 kg = 2e42 kg, D_{LS} = 2000 Mpc, D_{LO} = 1000 Mpc
- Given: Mass $M = 10^{12} \times 2 \times 10^{30}\ \text{kg} = 2 \times 10^{42}\ \text{kg}$, D_{LS} = 2000 Mpc, D_{LO} = 1000 Mpc
- Find: θ_E, the Einstein radius in radians and arcseconds
- Concept(s):

$$\theta_E = \sqrt{\left(\frac{4GM}{c^2}\right)\left(\frac{D_{LS}}{D_{LO}D_{SO}}\right)}$$

- where $c = 3 \times 10^8\ \text{m/s}$ and $1\ \text{Mpc} = 3.09 \times 19^{22}\ \text{m}$
  and D_{SO} = D_{LS} + D_{LO}
- Solution:

First, get the distances in SI units

(meters):
D_{LS} = 2000 ~~Mpc~~ x (3.09E22 m / 1 ~~Mpc~~) = 6.18E25 m
D_{LO} = 1000 ~~Mpc~~ x (3.09E22 m / 1 ~~Mpc~~) = 3.09E25 m
D_{SO} = D_{LS} + D_{LO} = 3000 ~~Mpc~~ x (3.09E22 m / 1 ~~Mpc~~) = 9.27E25 m

Now we can plug the numbers in to get θ_E:

$$\theta_E = \sqrt{\left(\frac{(4)(6.67\times 19^{-11}\ \text{N m}^2\ \text{kg}^{-2})(2\times 10^{42}\ \text{kg})}{(3\times 10^8\ \text{m/s})^2}\right)\left(\frac{6.18\times 10^{25}\ \text{m}}{(3.09\times 10^{25}\ \text{m})(9.27\times 10^{25}\ \text{m})}\right)}$$

$$= 1.13 \times 10^{-5}\ \text{radian}$$

- Now express this angle in arcseconds. The conversion factor for radians to arcseconds is: 1 radian = 2.06E5 arcseconds.
- θ_E = (1.13E-5 ~~radians~~)(2.06E5 arcseconds / ~~radian~~) = 2.3 arcseconds.

To see how to compute the Einstein radius for a solar mass object in our galaxy, see Math Exploration 12.1.

Math Exploration 12.1

Questions:

1.

Determining the Mass of a Lens

In this activity, you will analyze gravitationally lensed images to determine the mass of the lens.

Worked Example:

1. The image in Figure A.12.1 is an Einstein ring from the SLACS survey.

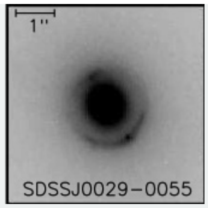


Figure A.12.1 This object, SDSS J0029-0055, is one of the Einstein rings discovered as part of the Sloan Lens ACS Survey (SLACS). Credit: Adapted from Bolton et al. (2008), *Astrophysical Journal*, 682, 964

In their paper, the SLACS team has determined that the radius of the Einstein ring is 1.11 arcsec, the distance

to the source is 9.67×10^{25} m, and the distance to the lens is 2.84×10^{25} m. Based on this information, what is the mass of the lens?

- Given: θ_E = 1.11 arcsec, D_{SO} = 9.67E25 m, D_{LO} = 2.84E25 m
- Find: M, the mass of the lens
- Concept(s):

$$\theta_E = \sqrt{\left(\frac{4GM}{c^2}\right)\left(\frac{D_{LS}}{D_{LO}D_{SO}}\right)}$$

where $c = 3 \times 10^8 \text{ m/s}$ and $1 \text{ Mpc} = 3.09 \times 10^{22} \text{ m}$ and $D_{SO} = D_{LS} + D_{LO}$

- Solution:

First, get the distance between the lens and the source: $D_{LS} = D_{SO} - D_{LO}$ = 9.67E25 m - 2.84E25 m = 6.83E25 m

We will also need the Einstein radius in radians: θ_E = 1.11 ~~arcsec~~ × (1 radian / 2.05E5 ~~arcsec~~) = 5.39E-6 radians

Now rearrange the equation to solve for M.

$$M = \left(\frac{(\theta_E c)^2}{4G}\right)\left(\frac{D_{LO}D_{SO}}{D_{LS}}\right)$$

Plugging in numbers, we get the solution.

$$M = \left(\frac{[(5.39\times10^{-6})(3\times10^{8}\text{ m/s})]^2}{(4)(6.67\times10^{-11}\text{ N m}^2\text{ kg}^{-2})}\right)\left(\frac{(2.84\times10^{25}\text{ m})(9.67\times10^{25}\text{ m})}{6.83\times10^{25}\text{ m}}\right) = 3.94\times10^{41}\text{ kg}$$

We can express this in solar masses if we divide by 2E30 kg.

$M = 1.97 \times 10^{11}$ solar masses

Questions

1.

2.

In the previous activities you saw how changing the geometry or mass of the lens changed the size of the Einstein ring. However, in all cases you might have noticed that the ring is extremely small. For typical Galactic lenses the Einstein radius is only a few milli-arcseconds, too small to be seen. For extragalactic lenses the size is larger, though still usually less than an arcsecond.

GOING FURTHER 12.2: THE LENS EQUATION

12.2.2: Microlensing

The simplest mass distribution for a gravitational lens is that of a point mass, which creates two images, one on each side of the optical axis. When the lens, source, and observer line up we have the case of the last section, an Einstein ring. The images are distorted into arcs that completely encircle the lens.

However, if the source is moved off the optical axis, then the images separate. One, which we will call θ_-, moves inward toward the optical axis. This is the image that we mentioned, but ignored in Figure 12.2 earlier; it “passes under” the optical axis. The other image, which we will call θ_+, the one shown in Figure 12.2, moves outward. See Figure 12.4 for an illustration.

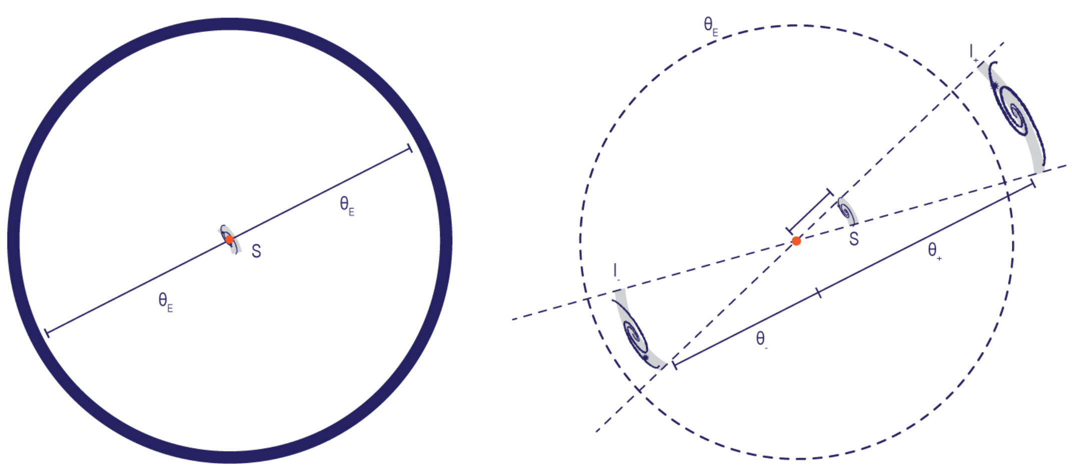


Figure 12.4 A point-mass gravitational lens will create two images of a source. In these diagrams, the gravitational lens is depicted as a small red dot and the source is shown as a blue spiral galaxy . The diagrams are drawn from the viewpoint of an observer looking at the sky. (a) If the source lies directly behind the lens, then the images overlap on the Einstein ring, as shown at left. (b) If the source is offset slightly, as on the right, then the images separate, one lying inside the Einstein ring (θ_-), the other lying outside it (θ_+). Credit: NASA/SSU /Aurore Simonnet

The following activity allows you to explore how gravitational lensing will modify the appearance of an image as the angle between the source and lens is adjusted.

> **Lensing by a Point Mass**
>
> In this activity, you will use an interactive diagram that provides a simulated view through a telescope of a source object that has been lensed by a point mass. When the source object lies directly behind a massive lens, it is possible to have it lensed such that the image forms a ring running entirely around the lens, which might or might not itself be visible. The simulation in the activity lets you change the alignment of the lens and source and then see how the images formed by the lens change.
>
> **Play Activity**

The simulation begins with a perfect alignment between the source, lens, and observer. The lens is a point mass, represented by a white dot, the source is a larger blue disk in the center of the screen. Because of the perfect alignment, the image is a ring, shown in blue.

Use the slider to move the source slightly off the center of the optical axis. The image will change as you do this.

Move the source farther to the left and right. You should notice that the image changes more dramatically as you move the source farther from the center.

1.

2.

3.

4.

5.

6.

7.

8.

In Section 12.2.1, we saw an HST picture of an extragalactic Einstein ring, but calculated that the sizes of Einstein rings due to lensing events within our Galaxy would be too small too see. Because of the tiny size of the images produced, this kind of lensing is called **microlensing**. However, the small size of microlensing does not mean that its effects are not measurable.

In addition to being displaced, you may have noticed in the previous activity that the images are also magnified in size compared to the source. For a spherically symmetric lens, the image on the outside of the Einstein ring (θ_+) is larger than the one on the inside of the Einstein ring (θ_-). The greater the distance off of the center of the optical axis the source is, the smaller both images will be. If the source is at an angular separation from the lens that is greater than the Einstein radius, the θ_- will be smaller than the source; the farther away, the smaller, until eventually no lensing effect is seen.

These changes in size turn out to be extremely useful. Because gravitational lenses preserve the surface brightness of the source in the images it produces, if the image is larger than the source it also becomes brighter. On the other hand, if the image becomes smaller than the source, then the image is dimmer because it has a smaller area.

Even if we cannot see the individual images, we can detect the changes in brightness caused by the lens; a gravitational microlensing event will cause the brightness of the source to appear to increase over what it would be without lensing. Figure 12.5 illustrates this effect.

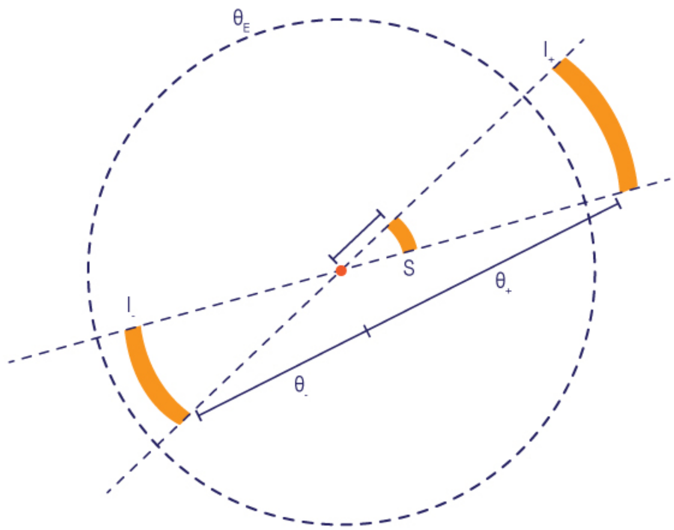


Figure 12.5 A point-mass lens changes the size of the images relative to the source. Because the lensing effect preserves surface brightness, a larger image will be brighter than the source. Credit: NASA/SSU/Aurore Simonnet

In Figure 12.5, we see schematically how the images produced in a lens will increase the source's apparent brightness when the images are too close to each other to be resolved. In that case they look like a single image. For an observer, the brightness of the individual images combine to create an apparent brightness that is the sum of the two. This magnification is due only to the fact that the images each have the same surface brightness as the source, while being larger in area. The magnification for a point-mass lens (like a star, for example) is given by the expression that follows.

$$m=\frac{1}{1-\left(\frac{\theta_E}{\theta}\right)^4}$$

When θ is smaller than the Einstein radius (the θ_- case), the magnification is negative, meaning that the image of the source is inverted (upside down). That is not important for a point source, but it is important when we consider lensing of extended sources like galaxies. When the angle θ is larger than the Einstein radius (the θ_+ case), the magnification is positive, so the image is right-side up relative to the source. In the next exercise, you will plot the magnification as a function of θ and use your plot to draw some conclusions about the brightness of a lensed source.

Magnification by a Point Source

1.

2.

3.

4.

5.

Of course, the brightness of an image in a lens with static geometry does not change, and whether it is dimmer or brighter than the source alone cannot be determined. However, if the geometry is not static, in other words, if the angular separation between the lens and source changes in time, then the source will vary in brightness with time. In real life, we cannot control the geometry of astrophysical lenses the way we did in the previous interactive activity. However, sometimes naturally occurring motions will change the geometry for us. This effect was first suggested as a means to look for non-luminous objects in our Galaxy by the Polish astrophysicist Bohdan Paczyński (1940–2007) in the mid-1980s.

Paczyński proposed to monitor millions of stars in the Large Magellanic Cloud (LMC), a small satellite galaxy of the Milky Way, and wait for a massive object (a lens) to pass in front of one of them as it orbited through the halo of the Milky Way (see Figure 12.6). The possible lensing objects in our Galactic halo could be: dim stars, large planets, brown dwarfs, black holes, neutron stars, or white dwarfs. Together these objects are known as MACHOs, for Massive Astrophysical Compact Halo Objects.

The resulting change in brightness of the background LMC star as it was lensed would reveal the presence of the otherwise invisible lens object, the MACHO, in our halo. A chance alignment of a background star with a foreground lens is extremely rare. However, by increasing the number of background stars - by observing the relatively dense stellar field of the LMC - the likelihood grows. That is exactly the advantage gained by monitoring an object like the LMC.

The LMC lies outside the disk of our Galaxy at a distance slightly over 160,000 light-years. Its distance makes it a promising target for microlensing surveys, first, because its stellar density on the sky is increased by its distance and, second, because a sightline to the LMC passes through a large part of our Galaxy's halo. In addition, the LMC is close enough to cover a large fraction of the sky, at least compared to other galaxies. Because of all these favorable properties, Paczyński reasoned that the LMC would make a good candidate for finding microlensing events. In practice, astronomers realized that the Small Magellanic Cloud (SMC) and even the Galactic bulge would also be suitable targets for microlensing surveys.

Figure 12.6 (a) Light from a star in the Large Magellanic Cloud (LMC) is lensed by an unseen object in the halo of our Galaxy. (Note: LMC and lens not to scale in this drawing.) (b) The LMC is the largest known of several dozen small galaxies that orbit the Milky Way. The LMC and another satellite, the Small Magellanic Cloud, are easily visible from the southern hemisphere, though they are too far south to be seen from most of the northern hemisphere. Credit: (a) NASA/SSU /Aurore Simonnet , (b) upload.wikimedia.org/wikipedi ...arp.750pix.jpg

One complication with Pacyński's strategy is that microlensing is not the only process that can cause a star's brightness to vary. Many stars are inherently variable. We have already discussed the Cepheid variable stars in Chapter 4, but many other types of variable stars exist. Fortunately, they can all be distinguished from microlensing events. For instance, variable stars often show periodic variations (Cepheids being just one example), whereas microlensing of a star will be extremely unlikely to repeat. For other kinds of variable stars, the color of the star changes with its brightness due to a changing surface temperature— novae and supernovae exhibit this kind of behavior. Microlensing, like all gravitational lensing, behaves the same in all wavelengths of light. So in a microlensing event the color of the source will not appear to change over the course of the brightening. In addition, the brightening and dimming phases in a microlensing event should by symmetric in time. That is not the case with other transient astrophysical sources. A model light curve for a microlensing event is shown in Figure 12.7.

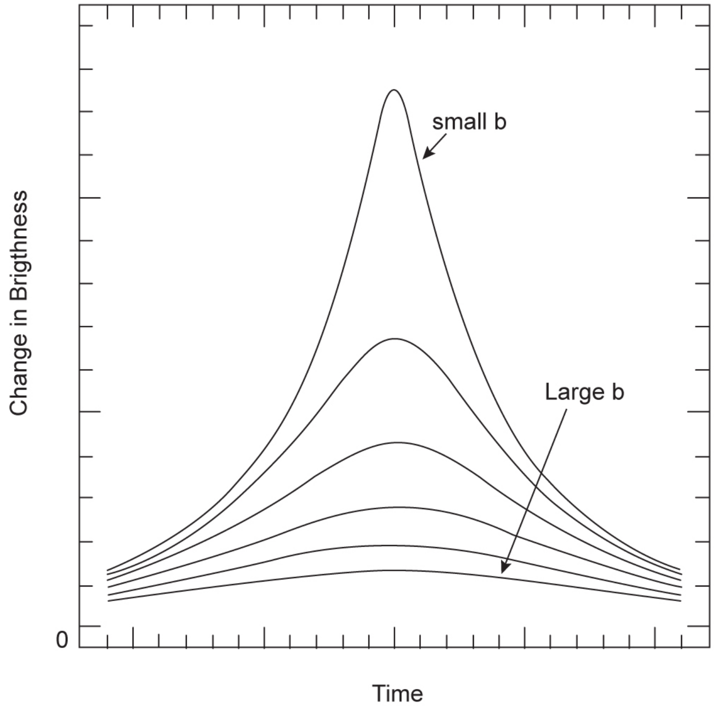


Figure 12.7 Gravitational microlensing events are distinct from other sorts of variable sources because they are perfectly symmetric in time, rising and falling in precisely the same manner. In addition, the brightening and dimming is the same at all wavelengths, which is not the case for most other variable astrophysical sources. Finally, the brightening will not repeat; it is a one-time event. Credit: NASA/SSU/Aurore Simonnet

Several microlensing surveys were undertaken through the 1990s, two of them of fairly large scale. The first of these was the Polish OGLE project, led by Pacyński and Andrezj Udalski of the University of Warsaw. OGLE (Optical Gravitational Lensing Experiment) has been running since 1992 and, as of this writing (2021), is still in operation. Observational data are collected with a dedicated 1-meter telescope at Las Campanas Observatory in Northern Chile. It has used wide-field imaging to monitor many hundreds of millions of stars in the Galactic bulge and the LMC and SMC. Another large project, started about a year after OGLE, was called the MACHO Project, which ran until 1999. The MACHO Project was a collaboration between astronomers at Mt. Stromlo and Siding Springs Observatories in Australia, and teams from several campuses of the University of California and the Lawrence Livermore National Laboratory in the United States. Like OGLE, the MACHO Project surveyed stars in the Galactic bulge and the LMC and SMC. Other projects were also undertaken over the past two decades to monitor the LMC, SMC, and Galactic bulge for gravitational microlensing events. Light curves from several microlensing events are shown in Figure 12.8.

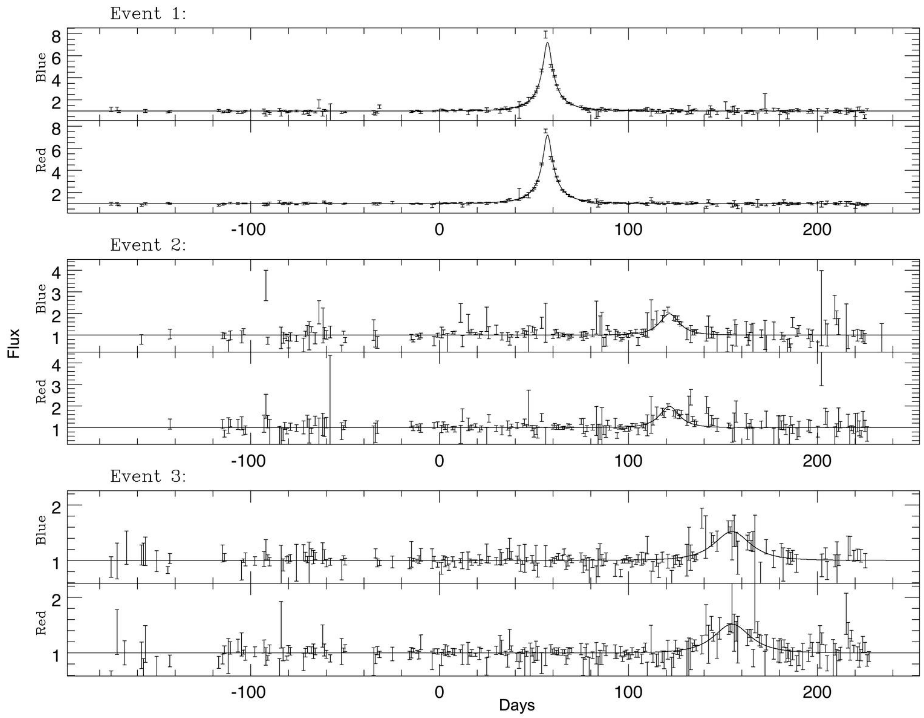


Figure 12.8 Light curves (flux vs. time) for three microlensing events. The MACHO team observed these events by monitoring 8.6 million stars in the LMC for a little over a year. Each event shows the same signature in red and blue light, as expected from microlensing, but not other potential sources of variability. Credit: NASA/SSU/Aurore Simonnet. Adapted from Alcock et al. (1995) *Physical Review Letters*, 74, 286

The upshot of all these projects is that massive objects like white dwarfs, neutron stars, black holes, and brown dwarfs—the most likely kinds of objects to be MACHOs—cannot make up more than half of (much less all of) the dark matter of our Galaxy's halo. There are simply too few microlensing events. Still, MACHOs do seem to account for a significant fraction of the dark matter in the Galaxy's halo, most likely around 20%. This means that at least half of the dark matter must be made of some sort of exotic matter, and the most likely amount is about 80%. From the MACHO surveys we know that most of the dark matter in galaxies is not composed of neutrons, protons, and electrons, the stuff of familiar matter. It is made of some form of matter still to be discovered.

The following activity allows you to explore how a MACHO's mass, velocity, and impact parameter influence a gravitational microlensing light curve.

MACHOs

In this activity, you will manipulate the mass, impact parameter, and velocity of a MACHO to determine how each of these variables affects the light curve observed when the MACHO lenses a distant star.

A peak in the light curve is created when the lensing MACHO passes in front of the star, causing the starlight to be magnified.

Play Activity

1.

2.

3.

4.

5.

6.

7.

This page titled 12.2: Lensing by Point Masses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

12.3: Lensing by Extended Mass Distributions

Learning Objectives

- You will understand that galaxies and clusters of galaxies can act as gravitational lenses.
- You will understand that many astrophysical lenses are complex in nature. You will understand that lensing suggests that most of the mass of galaxies and galaxy clusters is dark.
- You will understand that lenses provide an independent prediction of dark matter, distinct from that based upon Newtonian dynamics and the motions of stars, gas and galaxies.

What Do You Think: Gravitational Lensing With Galaxies

12.3.1: Extended Lenses Provide More Complex Patterns of Images

In the previous sections we considered gravitational lenses that are point masses. These would be objects like stars or black holes, objects whose mass is relatively small and compact. Other kinds of gravitational lenses exist: those whose mass is distributed over large volumes of space. Objects like galaxies and clusters of galaxies are examples of extended lenses. In this section, we will see how these extended lens systems differ from point lenses, and we will explore what they can tell us about the Universe.

In essence, extended masses are collections of point masses; galaxies are collections of stars, with some amount of gas mixed in. However, the simplest treatment of these systems assumes that they are smooth distributions of mass. It is generally not necessary to consider the stars and gas clouds individually if we want to understand the behavior of a galaxy as a whole. Instead, a distribution of mass is assumed, the expected image pattern for that assumption is computed, and then that prediction is compared to the lenses we see in the sky. One of the simplest assumptions to make is that the stellar velocities (or velocity dispersions in an elliptical galaxy) of the galaxy are constant. Indeed, this is a property generally seen in galaxies anyway, though if we prefer, other velocity structures can be used instead. In any event, the velocity of material in a galaxy is related to the mass distribution of that galaxy.

The first gravitationally lensed object observed was the so-called Double Quasar, detected in 1979. The system shows what appears to be two quasars separated by a mere six arcseconds in the sky. The quasars are peculiar in that they have nearly identical spectra. The similarity is so great that it caused astronomers Dennis Walsh and Robert Carswell of Great Britain, and American Ray Weymann, to suggest that the pair were actually two images of the same quasar. They surmised that the quasar was being lensed by an intervening and unseen galaxy. Their interpretation was confirmed over the next several years by several other teams who used radio observations to make detailed images of the quasar system. The intervening lens was also found and studied. It turned out to be a galaxy cluster, not a single galaxy. Figure 12.9 shows an image of the Double Quasar.

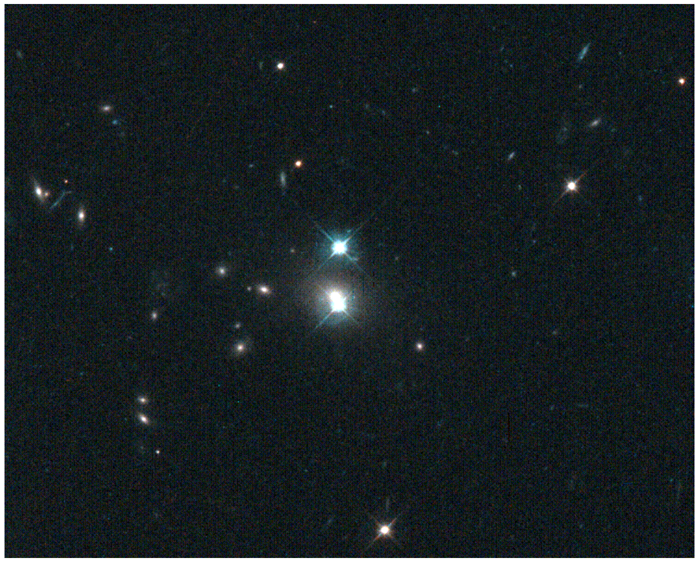


Figure 12.9 This image shows two apparently identical quasars separated by just a few arcseconds in the sky. In 1979, a team of astronomers suggested that we are actually seeing a single quasar that is being lensed by a galaxy along the line of sight. The system is known as the Double Quasar, or as QSO 0957+561 A and B. A bright galaxy, part of the cluster responsible for the lensing, is seen almost on top of the B image of the quasar. Credit: NASA/ESA/HST/WFPC2/George Rhee

Since 1979, many more examples of gravitational lenses have been found. They are especially conspicuous in the deep, high-resolution images taken with the Hubble Space Telescope and the larger ground-based telescopes using adaptive optics to correct atmospheric distortions. Most of these lenses are far more complex than the Double Quasar.

As we have discussed, the mass distribution in a lens affects the distribution of images seen in that system. For the Double Quasar, the mass distribution in the lens is relatively simple, but in other systems that is not the case. Careful study of the image distribution around a lens can provide important information about the total mass of the lens as well as its distribution. The lens geometry—the relative positions of the source, lens, and observer—will also have an effect on the distribution of the images seen. Sometimes it can be difficult to disentangle the mass distribution and geometry effects from each other, but rigorous analysis of the images often removes any such confusion, at least when the lens effect is strong enough to produce many images over a large region.

As we have seen, point sources give two images aligned with the lens itself. In an extended source there can be more than two images. Figure 12.10 shows a quasar that is lensed by a galaxy, producing four images. Images are only co-linear with the lens if the mass distribution of the lens has circular symmetry in the plane of the sky. That is not generally the case for galaxies or galaxy clusters. Figures 12.11 and 12.12 show examples of gravitational lensing by galaxy clusters. The type of image pattern seen in Figures 12.9–12.12 is called strong lensing because the lensing distortions (multiple arced images) are pronounced and easy to identify.

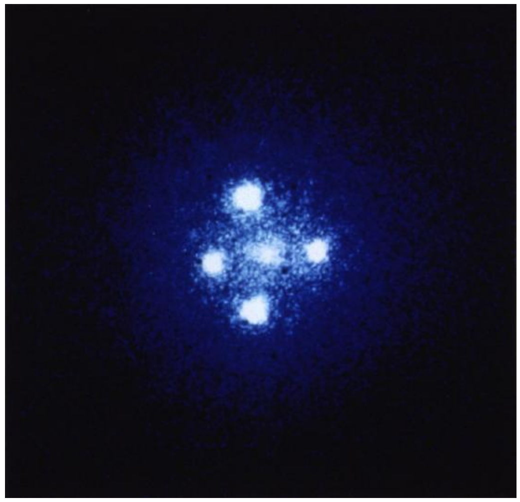


Figure 12.10 For the right mass distribution and geometry it is possible to observe four quasar images rather than two. Such an image is known as an Einstein Cross. Credit: NASA/Hubble Space Telescope

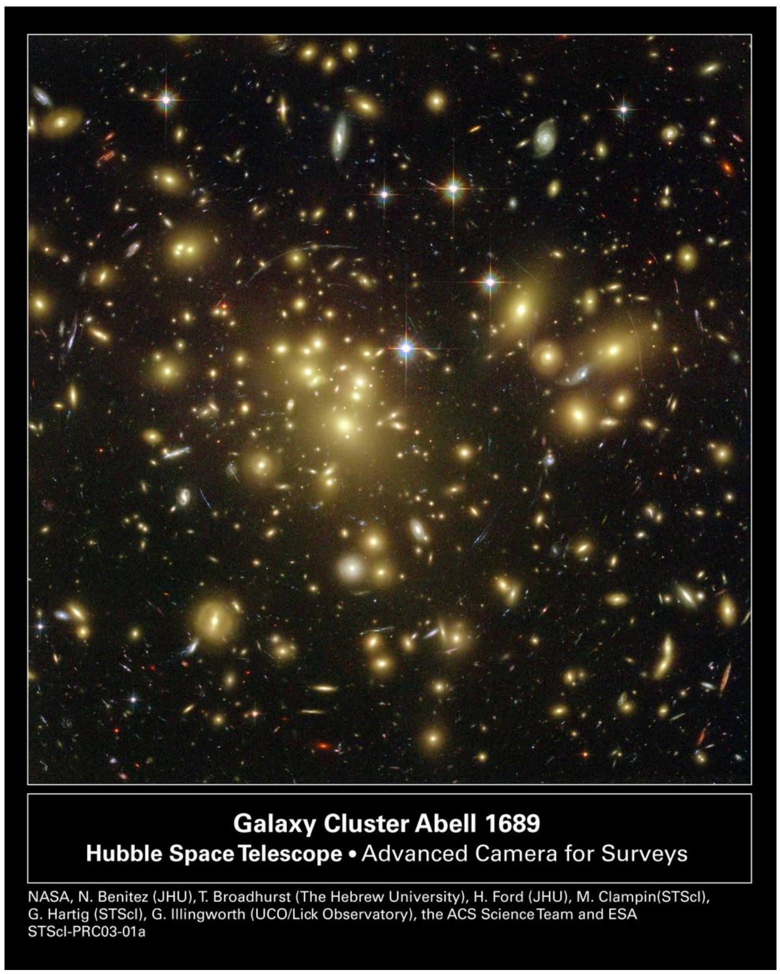


Figure 12.11 Many arcs are seen in and around this galaxy cluster, called Abell 1689. These arcs are background galaxies magnified and distorted by the mass within the cluster. The fact that the arcs all seem to be concentric indicates that the mass of this cluster is fairly symmetrically distributed around the central galaxy. Credit: NASA/Hubble Space Telescope

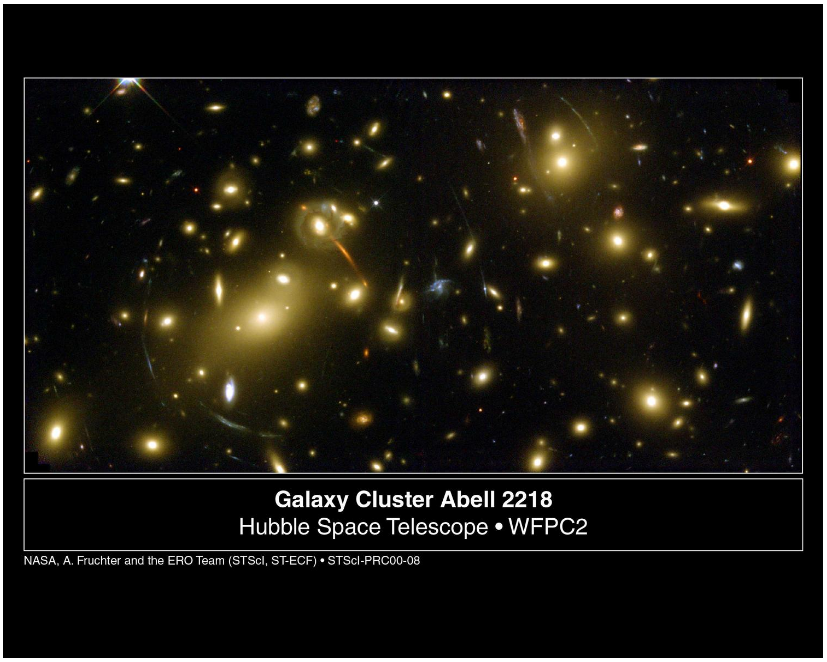


Figure 12.12 Unlike in the previous image, in this image the mass distribution is not symmetric about a single central galaxy. This is clear by the different arcs that are centered on different regions. In this cluster, called Abell 2218, the mass is much clumpier than it is in Abell 1689. Credit: NASA/Hubble Space Telescope

In addition to the magnification seen in point-source lenses, extended lenses create another effect called shear. The shear causes images to be stretched or distorted: it is a kind of astigmatism, similar to that found in poorly shaped glass lenses—or in the eye lenses of some people. The image of a background galaxy will not only appear larger than the galaxy itself, it will also be distorted. The effects of both magnification and shear are easily seen in Figures 12.11 and 12.12. Figure 12.13 gives a schematic rendering of how magnification and shear work when a gravitational lens is caused by an extended mass distribution. Both effects will usually be present in a gravitationally lensed image. Of course, the exact amount of magnification and shear depend upon the particular mass distribution causing the lens effect, as well as the geometry of the source-lens-observer system.

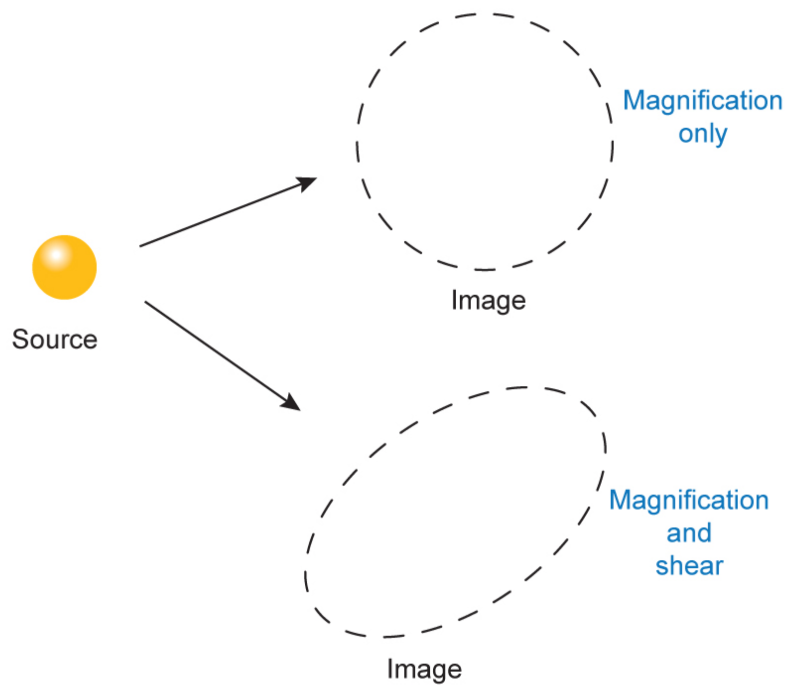


Figure 12.13 A circular source, as shown on the left, will be modified in two distinct ways by a gravitational lens with an extended mass distribution. At top right simple magnification is shown. The object does not change shape, but is enlarged by the lens. At the lower right the effect of shear is shown. The object is distorted such that its shape is modified. In general, gravitational lenses will produce both kinds of effects. Credit: NASA/SSU/Aurore Simonnet

To get a more realistic idea of how background sources are distorted by gravitational lenses, it is necessary to use a computer to simulate the bending of light rays. This is how astronomers study gravitational lens systems. Using computer models, the scientists are able to vary the amount of mass and its distribution, as well as the geometry of the lens. They can then compare what they see in the sky to the output of their simulation and draw conclusions about the nature of the lens.

The following activity provides an opportunity for you to examine a variety of images and to determine the type of lens that produced the gravitational image distortions.

Different Types of Lenses

In this activity you will be presented with several gravitational lens images. Your job is to identify the type of lens that produced each image: spherical, elliptical, or clumpy. The source object in all cases is a galaxy.

- The images are laid out in a grid. Click on the tile for an image that you would like to classify in order to see a larger version of the image and decide what type of lens produced it.
- If you are correct, the tile will pop back into place and turn blue.
- If you are incorrect, you will be given a hint.
- Once the lens types for all of the images are properly identified, your task is complete.

Play Activity

Image credits for this activity: NASA/ESA/Hubble Space Telescope

There are certain sorts of lens geometry for which the magnification of the background image is particularly strong. The images tend to be larger, though generally more distorted, for these kinds of geometry. This happens when the source of light lies behind a caustic. This is a term from optics and describes places where parallel light rays are brought to a focus. For a well-behaved spherically symmetric lens, a caustic is simply the focus of the lens, but even when lenses are not perfectly formed (as with water waves or gravitational lenses), there are places where parallel rays of light will be collected. Figure 12.14 shows examples of caustics and how they are produced.

Figure 12.14 (Top) Caustics are places where parallel rays are directed by a lens. (Bottom) Caustics in a lens system composed of a glass of water or water in a pool. They appear as the bright areas next to the glass of water or at the bottom of the pool.Light passing through these points will be parallel after passing through the lens, so objects behind a caustic receive particularly strong magnification and distortion. Credit: NASA/SSU/Aurore Simonnet and Kevin McLin

12.3.2: Masses of Gravitational Lenses

When astronomers compare their models to observed lenses, they are able to determine many properties of the lens. In particular, they can deduce the amount of mass in the lens and its distribution. Scientists understood quite early that these analyses might provide a way to measure the mass of galaxies. And gravitational lenses provide a mass measurement that is totally independent of dynamical methods such as the rotation curves of galaxies or the motions of galaxies in a cluster. Because the two methods are distinct, the the lens masses are an excellent check on the dynamical method. If both give consistent values for the masses of systems, we will have added confidence that the masses obtained are correct.

However, the lens masses do have limitations, just as dynamically determined masses do. For instance, the mass distribution of the various lens components must be assumed. In complex lens systems there can be confusion between the mass distribution in individual galaxies and the effects of nearby galaxies, and it is not always possible to disentangle these using the image distribution alone. Furthermore, the lens analysis is only sensitive to mass within the Einstein radius of the lens, and that does not extend far enough out to probe the entire mass structure of a galaxy. The lens effect can also be subject to additional gravitational effects of unseen matter along the line of sight—in effect, this additional mass can mimic extra mass in the assumed lens itself.

In any event, when mass determinations are made using gravitational lenses they typically show the presence of much more mass than is seen directly via emission of electromagnetic radiation (light in any wavelength), just as the dynamical analyses do. In fact, lensing analysis often requires two or three times as much mass as is inferred using dynamics alone. The discrepancy is likely caused in large part by the uncertainties mentioned above.

This page titled 12.3: Lensing by Extended Mass Distributions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

12.4: Weak Lensing

Learning Objectives

- You will understand the difference between strong and weak lensing.
- You will understand that weak lensing reveals the presence of dark matter throughout the Universe.

What Do You Think: What Kind of Matter Can Be a Lens?

All of the lensing we have discussed so far has been strong lensing. That occurs when the source of light is within or close to the Einstein radius of the lens. In strong lensing, obvious arcs and distortions are present. So are multiple images of the background lensed objects. In microlensing these distortions are blurred together so that a strong magnification is noticed, even if the individual images are too close together to be distinguished.

However, even when a source is far away from the Einstein radius, there can be small distortions. They can be so small that they are not noticeable in any individual source. Despite that, if many sources are present in the background, it is still possible to detect the effects of lenses by looking for correlated distortions in background objects. This technique, called **weak lensing**, has proven to be a very useful tool for cosmologists. They can use weak lensing to trace the distribution of totally unseen matter in regions of space far from galaxies. The effect is illustrated in Figures 12.15 and 12.16.

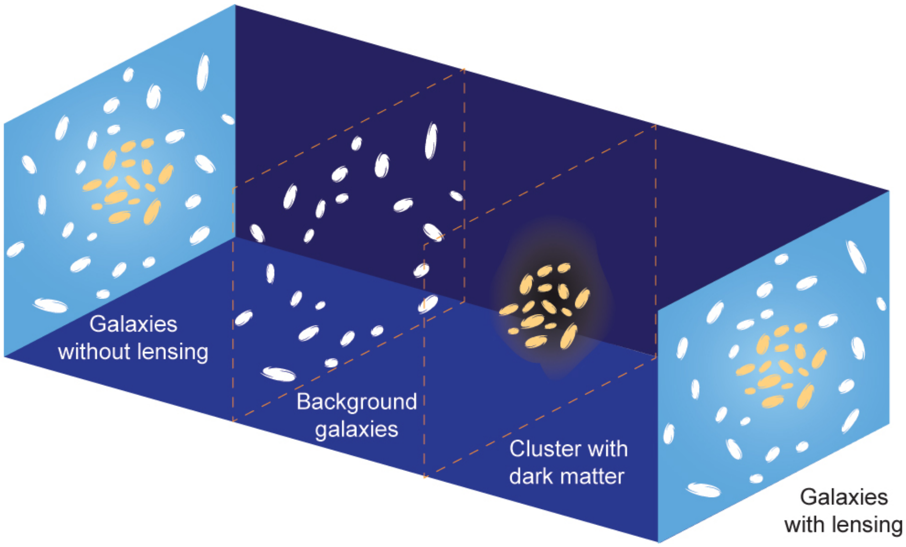


Figure 12.15 In weak gravitational lensing, a small distortion is introduced into the appearance of background objects as their light travels past foreground objects. The effect is generally too small to be seen in individual objects. However, the distortions of galaxies around a mass concentration are such that they will tend to align in circular arcs around the mass. The presence of correlated small distortions among many background objects can thus be used to detect otherwise unseen material between an observer and a source. Credit: NASA/SSU/Aurore Simonnet

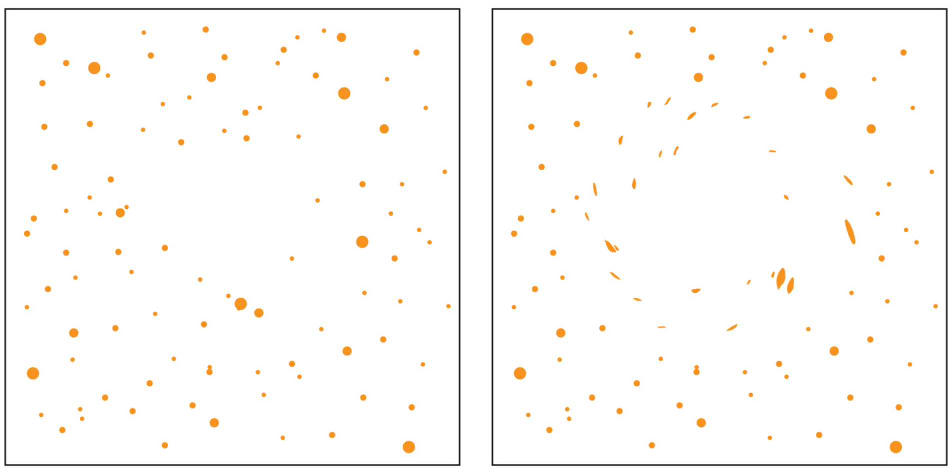


Figure 12.16 The panel on the left is an undistorted field. The panel on the right shows what is expected to happen if mass is concentrated along the line of sight: distortions are seen in the background field of galaxies. Here the distortions have been hugely exaggerated to make them more obvious. In general they are quite small and are not visible, but instead are revealed only with statistical image analysis. Credit: NASA/SSU/Aurore Simonnet

The Deep Lens Survey has been surveying the sky and has "reconstructed" the mass distribution (both seen and unseen) in their fields based on the weak lensing signal. One such reconstruction, of a 4 square degree field, is shown in Figure 12.17. Brightly colored areas in the figure represent regions with a higher lensing signal. They are therefore likely to have a greater density of matter. Notice the filamentary and clumpy structure. The team followed up with deeper imaging and found optical counterparts (e.g. visible galaxy clusters) for many of the peaks in this map.

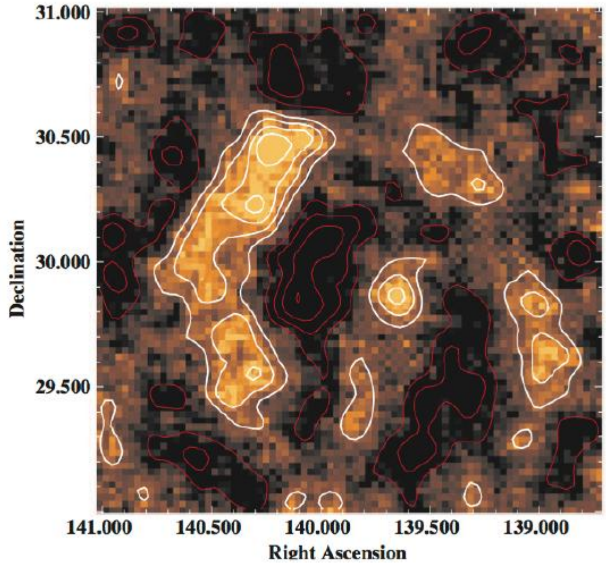


Figure 12.17 This map of a 4 square degree region of the sky was reconstructed from the weak lensing signal in the Deep Lens Survey. The brightly colored areas and white contours represent areas with a greater lensing signal, where there is more likely to be matter, whether dark matter or visible matter. The dark areas and red contours represent areas where there is less likely to be matter. Credit: NASA/SSU/Aurore Simonnet. Adapted from Kubo et al. (2009) *Astrophysical Journal,* 702, 980. For the team's first such reconstruction, see Wittman et al. (2006) *Astrophysical Journal,* 643, 128.

Figure 12.18 shows the results of a weak lensing analysis of Hubble Space Telescope (HST) data. In this study, the distortions caused by weak lensing reveal the presence of dark matter throughout the space along the sightline. Furthermore, the data show that the dark matter is more clumped together the closer it is to us. Recall the concept of lookback time; light that we are seeing now left closer objects more recently and farther objects longer ago. Thus these HST data suggest that the dark matter has become more concentrated over time. This observation is important for our understanding of the Universe and its origin. We will learn why in later chapters.

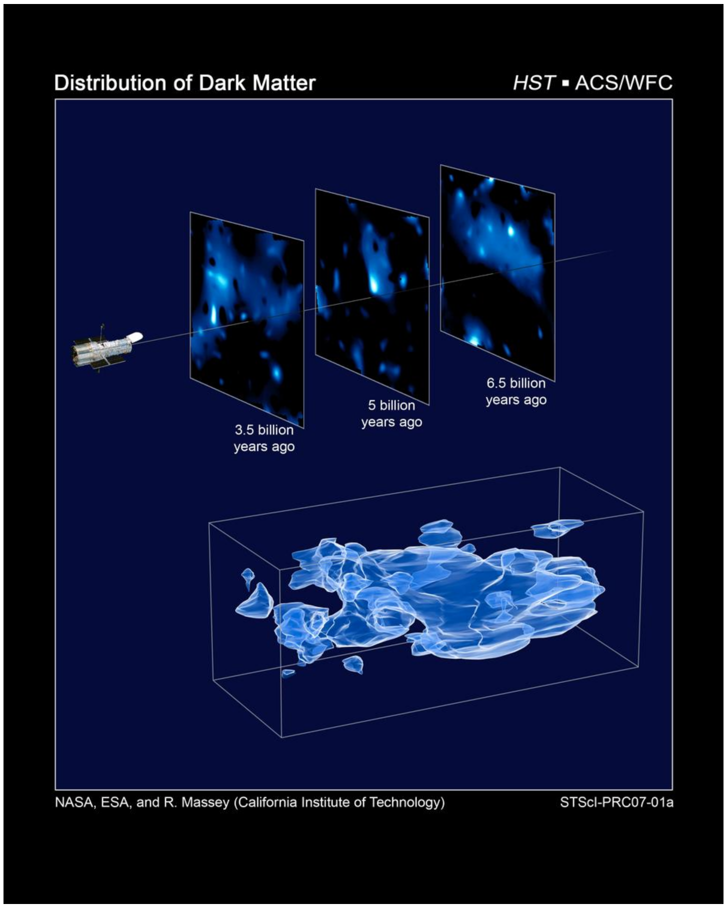


Figure 12.18 By studying the weak-lensing signature along this line of sight, astronomers have been able to determine that the matter causing the lensing, mostly unseen, is more concentrated for the closer galaxies than for the more distant ones. This implies that the dark matter is becoming denser in certain places with the passage of time. The top panel of this figure shows three slices of the dark matter distribution at three different epochs in the history of the Universe. The bottom panel shows the entire distribution vs. time. More recent times are on the left. For more information, see the press release. Credit: NASA/ESA/CalTech/Richard Massey

Since weak lensing is a statistical method, it requires the analysis of a large number of background galaxies. As a result, it has become an important method to use with large surveys of galaxies. The surveys use specially designed telescopes with very wide fields of view, allowing them to image many galaxies at once. By studying galaxies at different distances and noting their weak lensing distortions, astronomers can discern the distribution of dark matter in space. They these methods we are able to measure material that would otherwise be undetectable.

Several such surveys are planned for the coming decade. The Dark Energy Survey and Large Synoptic Survey Telescopes are two examples that will be studying weak lensing, among other things. A current survey called Pan-STARRS will study weak gravitational lensing as one of its science goals. It will be an important tool for astronomers in their investigations of galaxies, clusters, and larger structures.

This page titled 12.4: Weak Lensing is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

12.5: Wrapping It Up 12 - Measuring Gravitational Lenses

Learning Objectives

- You will be able to put the above concepts together to determine the mass and the level of complexity of the geometry of the lens for various systems.

In this activity, you will be working with images and data from the CASTLES survey, a collaboration between the Harvard Center for Astrophysics and the University of Arizona. The team used HST to obtain high-resolution optical and infrared images of 100 gravitationally lensed systems.

12.5.1: Part I: Determining the Masses **of Lenses**

You will be analyzing three image: a double quasar, an Einstein cross, and an image with several arcs. For each of these images, you will need to

- determine the ratio of the distance between the lens and observer to the distance between the lens and source.
- determine the Einstein radius from the image.
- use this information to compute the mass of the lens.
- determine what type of object the lens might be, from the mass of the lens.

Play Activity

12.5.1.1: A. The Double Quasar

The first image for you to analyze is a double quasar known as SBS0909+523.

1.

2.

3. Using the mass calculator, enter your values for the Einstein radius (θ_E) in radians, the distance from the observer to the source (D_{SO}), and the ratio (D_{LO}/D_{LS}) to determine the mass of the lens.

Play Activity

4.

https://phys.libretexts.org/@go/page/31368

5.

12.5.1.2: B. The Einstein Cross

The next image for you to analyze is an Einstein cross known as Q2237+030.

1.

2.

3.

4.

5.

6.

12.5.1.3: C. Several Arcs

The final image for you to analyze is known as SDSS1004+4112.

1.

2.

3.

4.

5.

6.

12.5.2: Part II: Determining the Complexity of Lenses

Rank the following CASTLES images based on the complexity of the lens. Use the pull-down menus to order them, 1 being the least complex lens and 4 being the most complex lens.

Play Activity

This page titled 12.5: Wrapping It Up 12 - Measuring Gravitational Lenses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

12.6: Mission Report 12 - Measuring Gravitational Lenses

A.

B.

C.

D. Questions to be graded for accuracy:

1.

2.

3.

4.

This page titled 12.6: Mission Report 12 - Measuring Gravitational Lenses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

13: The Expansion of the Universe

Chapter 13 presents the observational evidence for the expansion and age of the Universe and the basis of the Big Bang theory. The chapter begins by presenting some historical background that sets the context for understanding Edwin Hubble's observation that the Universe is expanding. You will determine the expansion rate, and then, using Hubble's Law, learn how to estimate the age of the Universe. The last section of the chapter introduces Big Bang theory; exploring properties of the expansion, including changes in density and temperature.

This page titled 13: The Expansion of the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.0: The Expansion of the Universe Introduction

You might have heard that the Universe is expanding, but what does that mean? Does the Universe expand into pre-existing space, or is the expansion caused by the creation of new space? What does the expansion look like? Is expansion even the best phrase to use to describe what the Universe is doing? The opening video is a simulation of the uniform expansion of space. Galaxies are shown receding from each other as the space between them is stretched. The blue grid is provided to help you achieve a better sense of the three dimensional motions.

This video contains no audio

Notice how every galaxy in the visualization is moving away from every other galaxy. If you were located in any of these galaxies, you might not naturally conclude that you were moving at all. Instead you would see every other galaxy receding from you. From the video's point of view, however, we can see that there is no special place in the expansion; each place is just like every other. There is no sense of a center to the expansion, nor of a boundary or an edge. Also, notice that the galaxies themselves are not expanding, only the space between them is.

Every scientific visualization has been made to illustrate only some simplified aspects of the real Universe. At the same time, there are quite often liberties taken with other aspects. In the expansion visualization we are able to see the consequences of a uniform stretching of space, but it also contains several simplifications that we should be aware of.

First, the Universe is stretching very slowly. In the animation, distances are doubling in a few seconds. In the real Universe, several billion years are required for distances to double. Additionally, in the real Universe it is not individual galaxies that are being moved away from one another. Instead, every gravitationally bound structure rides the stretching, somewhat like a coin glued to a stretching piece of a rubber band. The largest bound structures are clusters of galaxies, so we observe clusters separating, not individual galaxies. Within the clusters, individual galaxies might actually be approaching each other, not receding. Another simplification is related to the evolution of galaxies. In the animation we do not see the galaxies changing (evolving or developing) with time. Over billions of years, the timescale depicted by the animation, galaxies change significantly. Finally, in the clip we see all of the galaxies at the same time. The real Universe is so large that it takes light a long time - perhaps hundreds of millions or billions of years - to travel across intergalactic distances. As a result of these enormous distances (and the finite speed of light), we observe distant objects as they were (and where they were) when the light from them was emitted, we do not see them where and as they are now.

In this chapter, we will begin to investigate the expansion of the Universe. In particular, we will look at the evidence for the cosmic expansion. We will see how an expanding Universe implies that at some time in the past everything we see around us must have been extremely close together, and we will learn how to calculate when that was. We should also not that the stretching history we see is consistent with the predictions of general relativity, our key physical theory for understanding everything we observe on large scales.

This page titled 13.0: The Expansion of the Universe Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.1: Some History

Learning Objectives

- You will understand some context for our place in the Universe according to Western traditions.

Center of the Universe

Cosmology is the study of the large-scale structure and evolution of the Universe. People in early and isolated societies had a wide range of ideas concerning the formation of the Universe. The Western tradition of cosmology began with the ancient Greeks, and below we give a very condensed version of the development of cosmological ideas in Europe and the Mediterranean region, starting with Ancient Greece.

The Greeks applied logic and geometry to the study of astronomy. Starting around 600 BCE, the Greeks began developing a geocentric model of the Universe. In their model, Earth was at the center of a series of shell-like spherical domains carrying the Sun, Moon, planets, and stars around Earth. They built upon this idea for many centuries, culminating in a model made by Claudius Ptolemy (100–170). The Ptolemaic model worked well at predicting astronomical phenomena and was in use for hundreds of years. In addition to their model of the heavens, the Ancient Greeks also knew that Earth was round, and its size was measured by the Greek astronomer Erastosthenes (276–196 BCE).

When the Roman Empire superseded the Greek civilization that came before it, there was little advancement of cosmological ideas. The Roman interest in science tended more to the practical, like the building of roads and water transport systems, metallurgy and other pursuits with more immediate application. The Romans adopted the Greek ideas and, through their vast Empire, were able to spread them all around the lands they controlled.

Classical civilization in Europe and the Western Mediterranean came to an end with the fall of the Western Roman Empire in the 5th century AD. After the destruction of the Library of Alexandria in 415, which housed millions of volumes of ancient books, much of classical learning was lost. However, some of the ancient knowledge was preserved by Islamic scholars working in Baghdad and other Islamic cities. While Western Europe was experiencing the Dark Ages, the Middle East was a center of enlightenment.

Islamic scholars of many religions and cultures worked together. They created algebra and made detailed observations and maps of the sky, and this knowledge was shared with learned people in the Eastern Mediterranean. After the fall of Constantinople in 1453, which put an end to the Eastern Roman Empire (The Byzantine Empire), many scholars moved to Europe. The influx of trade and knowledge spurred the European Renaissance.

Among the pieces of knowledge preserved was the spherical shape of Earth. It is a common myth that this information was not known at the time of Columbus and other European explorers, but that is not true. Scholars certainly knew that Earth is spherical, although the average person, illiterate and uneducated at that time, probably did not.

Over the centuries, using the old Ptolemaic ideas to predict celestial motions became increasingly difficult. The discrepancies between observations and predictions became too big to ignore. Nor could they be explained away by simple means. Eventually, it became implausible to explain the observed motions of the planets while insisting that they were traveling around Earth. A better idea was needed.

Nicolaus Copernicus (1473–1543) provided a simpler explanation for planetary behavior. He postulated that the planets (including Earth) orbit the Sun. Copernicus imagined the planets to move on giant, concentric circular paths with the Sun at their common center. Thus, he displaced humans from the center of the Universe. Such a Sun-centered model is known as heliocentric. The Greek astronomer Aristarchus (310–230 BCE) had proposed a heliocentric model centuries before. He was not taken seriously at the time, and the details of his model were lost when Classical civilization collapsed.

Despite being simpler conceptually, the Copernican heliocentric model had its challenges. For example, it did not work any better than the Earth-centered model it replaced. Furthermore, it was not very intuitive. If Earth is moving, why do we not see any of the stars appear to move as a result of that motion? These were questions that made people doubt the Copernican model when it was first proposed.

The answer to their questions is obvious to us now, but it was not at the time; the stars are so far away that their apparent displacements due to Earth’s orbital motion are tiny, well beyond the technical capabilities of any ancient or medieval observers to measure. Eventually, however, the invention and development of telescopes allowed measurements of stellar parallax, and from those measurements it was possible to calculate stellar distances. In addition, Johannes Kepler (1571–1630), who lived in the century after Copernicus, showed that planetary orbits are actually ellipses, not circles. Kepler's elliptical orbits provided an excellent fit to the careful observations made by Tycho Brahe (1546–1601) - and they still are an excellent fit to any observations of planets or asteroids made now, no matter how good the precision. This is because the Keplerian model of elliptical orbits with the Sun occupying one of the foci of the ellipse is the correct description of orbits.

If Copernicus and Kepler displaced people from the center of the Universe, subsequent observations that explored space beyond the solar system soon demoted the Sun as well. Slowly the picture emerged of the Sun as an average star among a huge population of similar bodies. Each star is so far away from its neighboring stars that even light requirs years to travel between them. Beyond the nearby stars, those few with distances we can measure geometrically, lie many others: the great band of the Milky Way was resolved by telescopes into millions more stars. The band was recognized by Thomas Wright (1711–1786) and Immanuel Kant (1724–1804), among others, as the effect of our being located within a huge disk of stars—a galaxy. By the early years of the last century, our Galaxy was the largest structure known, but some scientists were proposing that many fuzzy blobs (then known as nebulae) might be other galaxies. It is at this point that our current ideas about the cosmos begin to take root.

Edwin Hubble’s first major contribution to cosmology was to identify a few special stars—Cepheid variables—in a number of these "nebulae." The luminosity of Cepheids is correlated with their period of variability, a fact discovered by Henrietta Leavitt of Harvard Observatory. The more luminous the Cepheid, the longer its period of variation. Hubble was able to monitor the variability of several Cepheids and thus determine their luminosities. From the ratio of their apparent brightness to their calculated luminosity, he could deduce their distances and show that in many cases they were far beyond our own Galaxy. They must, therefore, be located within galaxies of their own. Figure 13.1 shows the first Cepheid star Hubble recognized in another galaxy. The star was in the Andromeda galaxy, our nearest large neighbor. It has a size similar to the Milky Way.

With this discovery, Hubble settled the debate about the structure and size of the Universe, at least for the time. He showed that the Milky Way is but one galaxy among many others filling the cosmos. Just how many galaxies exist, and out to what distances, remained unknown.

Figure 13.2 shows Edwin Hubble peering into the eyepiece of the 100-inch telescope at Mt. Wilson in California. This is something he did only when posing for photos like this one. Photographic film is a much more sensitive light detector than the human eye, and so astronomical observations were being done using film at that time. Essentially this turned the telescope into a giant telephoto lens for the camera mounted onto it.

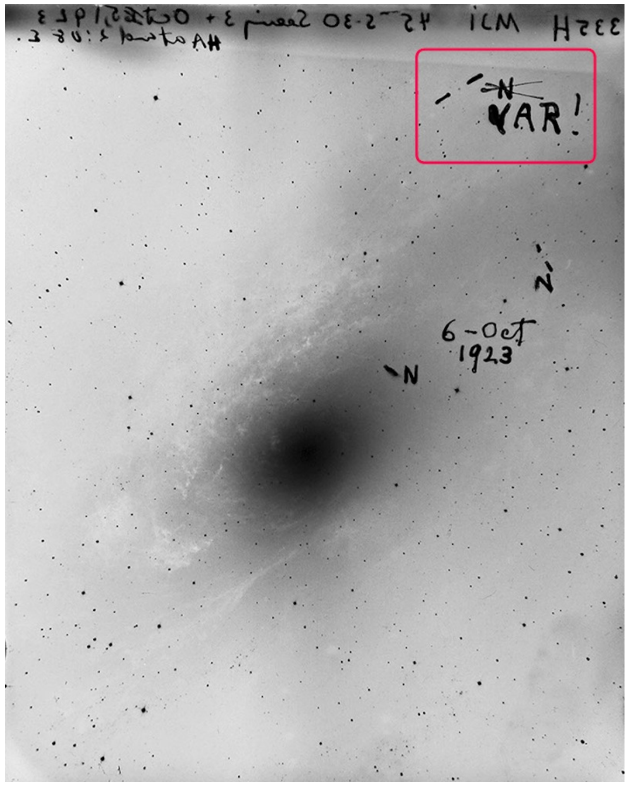


Figure 13.1: Edwin Hubble took this photographic image of a Cepheid variable star in the Andromeda galaxy in 1923 using the 100-inch telescope at Mount Wilson Observatory. The results were published in 1925 and revolutionized the cosmic distance scale. Credit: Mount Wilson Observatory Historical Archives

Figure 13.2: Edwin Hubble looks through the 100-inch telescope at Mount Wilson Observatory in 1937. Hubble used this telescope throughout the 1920's and 1930's and made several important discoveries. Credit: Margaret Bourke-White, Time & Life Pictures/Getty Images

Over a period of three thousand years the best human models for the cosmos evolved from being local, to geocentric, to heliocentric, to stellar, to galactic, to extragalactic. The scale of the models has grown from hundreds of kilometers to, as we will soon see, billions of light-years. Each light-year is approximately 10 trillion kilometers, so there has been a vast expansion of in the scale of the known Universe. Similarly, as we have grasped the great ages of stars by understanding their nuclear mechanisms and history, so has our estimate of the age of the Universe, determined by entirely independent means, grown from a few human generations to billions of years.

Humans, central to the early models they proposed, have become less and less conspicuous in each succeeding model. However, we still carry a special significance as the only part of the Universe we know of that engages in cosmology—we are the part of the

Universe that can think about the Universe. In the remainder of this chapter we will take the first steps on our journey to comprehend how the current view of the cosmos has come into being. We will start out by understanding the primary evidence underpinning the Big Bang theory: cosmic expansion.

This page titled 13.1: Some History is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.2: The Hubble Law

Learning Objectives

- You will be able to construct and interpret a Hubble diagram from the distances and velocities of galaxies.
- You will be able to perform calculations and understand the Hubble law conceptually.

What Do You Think: the Motions of Galaxies

Once Hubble had made measurements of galactic distances, the next step was to look for relationships between galaxies' distances and other properties. One property Hubble considered was the shift of lines in the spectrum of the galaxy, caused by its velocity toward or away from us. What Hubble found was surprising. Recall that spectral features such as emission or absorption lines shift to longer wavelengths (redshift) if the source moves away from the observer. On the other hand, if a source moves toward an observer, then the perceived wavelengths are shifted to shorter values - they are blueshifted. The faster the relative motion, the larger the observed shift. This interpretation of redshifts and blueshifts is called the Doppler effect. It is analogous to the Doppler effect for sound, in which the pitch of a passing source of sound, for example a train whistle, gets higher as the source (train) moves toward the observer, and it gets lower as the source (train) moves away.

In the activity that follows you will measure distances and redshifts of galaxies in order to learn about Hubble's most surprising result. You will use a standard ruler method to estimate galactic distances. There are many alternative methods for determining distances. Each method is good over some range of distances, and the ranges overlap. This overlap allows us to fit each technique into a consistent system—the cosmological distance ladder—for measuring cosmic distances.

Making a Hubble Diagram

Play Activity

In this activity you will determine the relationship between a galaxy's distance and its redshift.

- In part A, you will find the distances to galaxies from their images using a standard ruler approach: if we know the inherent sizes of the galaxies, we can calculate their distance because the farther away they are, the smaller they will appear.
- In part B, you will find the redshifts of the galaxies from their spectral lines.
- In part C, you will combine your data in a diagram.

Credit: This activity is based on a lab developed at the University of Washington by Ana Larson. Galaxy spectra are from Kennicutt, R. C., Jr. (1992) *Astrophysical Journal Supplement Series*, 79, 255. Images of the galaxies are from the Palomar Observatory Sky Survey, which was digitized under NASA contract by the Space Telescope Science Institute, operated by AURA. Data were retrieved from the NED database.

A. Finding the Distances to the Galaxies

For this part of the activity, we will be using the images of galaxies. The images are negatives. That is why bright objects, like stars and galaxies, appear dark. Negatives make it easier to see the faint outer edges of the galaxies.

There may be more than one galaxy in the image; the galaxy of interest is always the one closest to the center.

https://phys.libretexts.org/@go/page/31428

B. Finding the Velocities of the Galaxies

Previously, you have learned how to measure the Doppler shift of galaxies using emission lines (bright line spectra) and then used those measurements to calculate the velocities of the galaxies. Here we will be doing something similar.

A small part of each galaxy's spectrum is provided, showing several spectral lines; in this case they will be absorption lines, so dark lines. The two most prominent absorption lines are the Ca K and Ca H (calcium) lines. You will be using the Ca K line to make your measurements. The rest wavelength for Ca K is 3934 Angstroms (Å). Recall, 1 Å = 10^{-10} m.

https://phys.libretexts.org/@go/page/31428

C. Making a Hubble Plot

From the activity, you should have found that if we take spectra of many galaxies, we find that most of the galaxies are redshifted, as opposed to finding some redshifted and others blueshifted. In fact, out of the millions of galaxies whose redshifts have been measured by astronomical surveys, only a handful of the most nearby galaxies are blueshifted. Furthermore, there is a simple *linear* relationship between the distance to a galaxy and the shift in its spectrum: the amount of redshift increases linearly with increasing distance to the galaxy. This is a key observational feature that any model of the Universe must explain. These results, known since the late 1920s, have since been confirmed for a vast sample of galaxies, but they certainly came as a huge surprise when Hubble discovered them nearly a century ago. Hubble's original measurements are discussed in **Going Further 13.1: Edwin Hubble's Data**. Current measurements are shown in Figure 13.3. The slope of line that best fits the data is known as the Hubble constant, H_0.

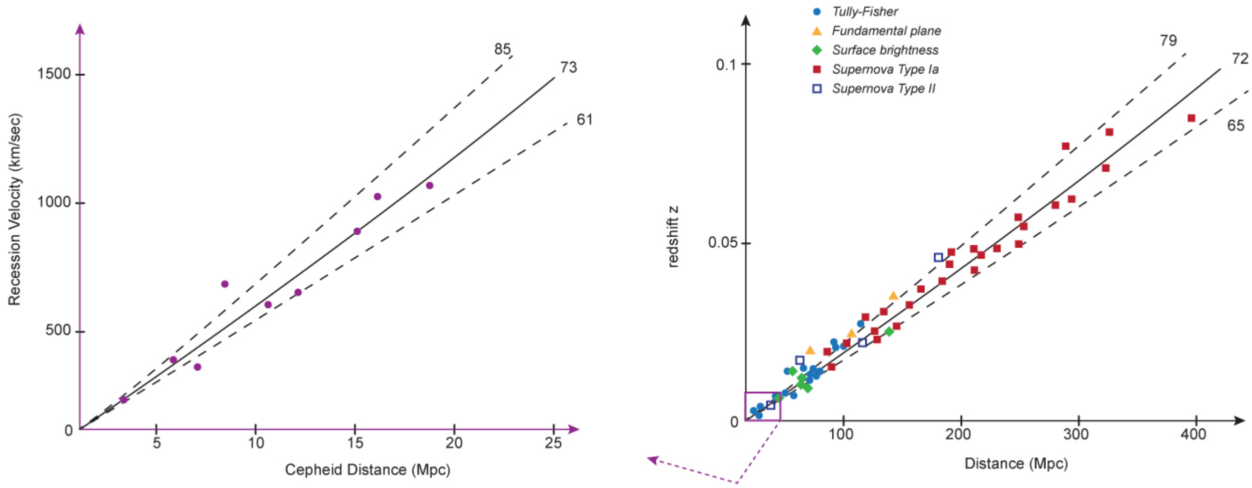


Figure 13.3: Hubble Diagram. The relationship expressed by the Hubble law: the farther away galaxies are, the greater their velocity (or redshift). It is illustrated on the left by points determined relatively recently by the HST Key Project. These data were acquired by using Cepheid variable stars as standard candles to obtain distances to their host galaxies. Lines are fits to the data, with the corresponding values for the Hubble constant labeled. Plots on the right include non-Cepheid data in distance determinations, which provide more points out to larger distances. Credit: NASA/SSU/Aurore Simonnet based on data from Freedman et al. (2001) *Astrophysical Journal*, 553, 47

The relationship between galaxies' velocities and their distances is now referred to as the **Hubble Law**. It can be expressed in words by saying that the velocity of a galaxy is directly proportional to the distance to that galaxy. Mathematically we express the same idea more succinctly as an equation.

$$v = H_0 d$$

Here, d is the distance to the galaxy and v is its velocity. H_0 is the constant of proportionality, or the slope of the best-fit line to the data. So again, from both the equation and the graph, we see that galaxies are moving away from us, and the farther away a galaxy is, the faster it is moving away. Hubble originally used a *K* to represent the proportionality constant in his expansion law, but it has since been changed to H_0, called the **Hubble constant,** to honor him. In the next activity, you will measure the Hubble constant using your data.

Measuring the Hubble Constant, H_0

4. Since the Hubble constant is the slope of the line, the units of the Hubble constant will be the units of the vertical axis divided by the units of the horizontal axis, or:

$$\mathrm{km/s/Mpc}$$

Sometimes this is written as $\mathrm{km\,s^{-1}\;Mpc^{-1}}$.

Using your own data, you have now determined the value of one of the most famous numbers in astronomy!

The Hubble constant is one of the most important numbers in astronomy. An enormous amount of effort has been devoted to determining H_0 with increasingly better precision. For most of the second half of the 20th century the value of H_0 was estimated by different investigators to be between 50 and 90 km/s/Mpc—not very high precision at all. The situation improved tremendously in the two decades following the launch of the Hubble Space Telescope (HST) in 1990. One of the primary missions of HST was to determine the value of the Hubble constant more precisely. The Hubble Key Project to Measure Cosmological Distances, led by Wendy Freedman of Carnegie Observatories, was an effort of 28 astronomers (some shown in Figure 13.4, below) to use observations from HST to establish the value of the Hubble constant to 10% precision. The project made the most precise optical determination of the constant in May 2001 of 72 ± 8 km/s/Mpc. The consistency of values derived from many different independent techniques is now impressive and reassuring.

Figure 13.4: Wendy Freedman of Carnegie Observatories was one of the team leaders for the HST Key Project.

Going Further 13.1: Edwin Hubble’s Data

Edwin Hubble’s early published data are shown in Figure B.13.1. The first data (top), from 1929, were the basis for the initial announcement of the Hubble law, and correspondingly, the expansion of the Universe. In the figure, the solid circles and the blue line are for the galaxies treated individually, while the open circles and the red line combine the galaxies into groups. There were few galaxies in the sample and a wide scatter in the data, due largely to measurement errors. Hubble also mislabeled the y-axis ‘km’ rather than ‘km/s’. The 1931 data (bottom), collected with his assistant Milton Humason, demonstrate the Hubble Law far more convincingly, mainly by extending measurements to much more distant galaxies; most of the 1929 data are contained within the small box at the extreme lower left of the 1931 plot.

The values for the Hubble constant derived from the slopes of the lines in these plots were wrong by a large factor (about 10) due to several unrecognized systematic errors. Unknown at the time, there are different classes of Cepheid variables, and Hubble applied the period-luminosity law of one class to objects of the other. Hubble did not know about the cosmic distance ladder, so he had to improvise. For instance, in distant galaxies Cepheids were too faint for him to detect, forcing him to find (or invent) alternative ways of measuring the distances to those galaxies. Even now, when we look at the most distant galaxies visible in our most powerful telescopes, we cannot see the Cepheids in them because they are too faint. Hubble had no chance of seeing them in the 1920s.

This error was not corrected for many years and resulted in serious overestimates of the speed of expansion of the Universe—and correspondingly short age estimates for the Universe, as we will see.

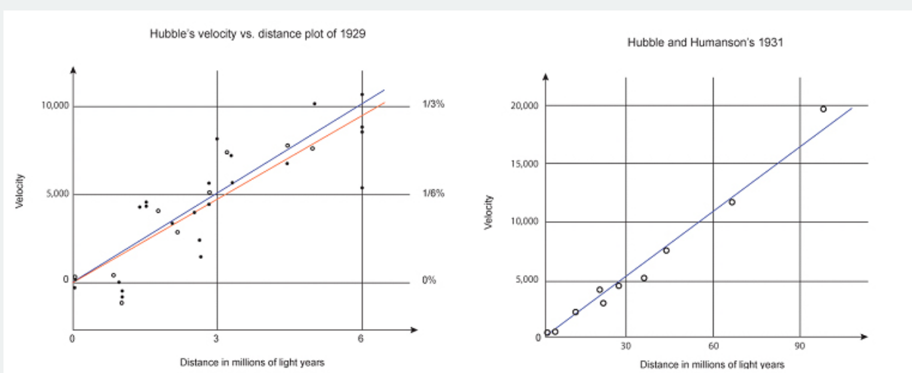


Figure B.13.1: Hubble’s original data from 1929 (top) and Hubble and Humason’s data from 1931 (bottom). Solid circles and blue lines are data and fits for the galaxies treated individually, while the open circles and the red line in the top figure aggregate the data. Credit: NASA/SSU/Aurore Simonnet based on Hubble and Humason’s data.

One possible interpretation of a relation like Hubble’s, and the most obvious, is that it is due to the motions of galaxies *through space*. If we interpret the redshifts of galaxies as Doppler shifts, then we can convert redshifts into velocities through space. While the Doppler interpretation is obvious and easy to understand, it turns out not to be the correct interpretation. It is not consistent with other observations, which we will come to in a moment.

This page titled 13.2: The Hubble Law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.3: The Universe is Expanding

Learning Objectives

- You will know that the Hubble law is evidence that the Universe is expanding.
- You will know that the Universe is expanding, but individual galaxies are not.
- You will know that there is no center or edge of the Universe.

Expansion or Explosion?

13.3.1: Interpretation of the Hubble Law

The observation of galactic redshifts from the 1920s and thereafter was the main motivation for the Big Bang theory. It is very difficult to explain these observations in terms of independent motions of galaxies through space. After all, the galaxies are billions of light-years away from each other. How could they all conspire to move in the same way? Many people, when they hear the term ‘Big Bang,’ imagine some sort of explosion, like a giant firecracker, but that is problematic because it is not consistent with a number of observed properties of the Universe. The actual Big Bang theory is something quite different. From the perspective of general relativity it is much more natural to explain the redshift-distance relation in terms of galaxies that are embedded in a space that is uniformly expanding.

If we abandon the scenario in which galaxies are moving through space, we are left with one where space itself expands and carries galaxies along as it does so. General relativity provides a context in which the notion of expanding space makes sense. In fact, if general relativity is an accurate theory of the behavior of space and time, it is very difficult to contrive a situation in which the Universe could be static. Given the success of general relativity in explaining many individual phenomena (gravitational lenses, precession of orbits, etc.), it is natural to think of space overall as either expanding or contracting.

Even in a Newtonian analog the idea of a static Universe is difficult to make work: think how surprised you would be to see a ball simply hovering some distance above Earth’s surface. If you looked at a photograph showing a ball in that position, you would assume that it must be either rising or falling. An analogous scenario is true for the Universe on grander scales: when we look at any given galaxy we are essentially seeing a snapshot of it. General relativity provides a framework for understanding how that snapshot relates to the ongoing evolution of the Universe.

It is worth elaborating on how an expanding space differs from a static space with objects moving through it. Without general relativity, we are not used to thinking of space as having properties of its own, so we will use several analogies.

One way to imagine space is like a stretchy band with galaxies stuck onto it (Animated Figure 13.5). As the band stretches, the galaxies all move away from each other, but they are moving due to the stretching of the space, they are not moving through the space. Furthermore, we see that the galaxies themselves are not stretching, they are not growing, only the space between them is. That is because the internal (gravitational) forces within galaxies overwhelm the expansion. So our Universe is expanding, but our Galaxy and Solar System are not. On average, all galaxies are moving away from each other, but within a galaxy, the stars are not moving away from each other because gravity keeps them bound together.

Animated Figure 13.5: An analogy for the expansion of space: a stretchy band. The galaxies move away from each other as space stretches. They never actually move along the band; they are just along for the ride. The galaxies themselves do not expand, only the space between them. Credit: NASA/SSU

Continuing with this analogy, we can explain the Hubble Law: galaxies are getting farther apart from each other more quickly the farther apart they are from each other. Figure 13.6 illustrates this principle in one dimension. We can imagine that the galaxies (A, B, C, D) are on the band in a line and are a certain distance apart (say 1 cm) at a given moment. The band then expands such that all of the galaxies will be 1 cm farther apart from adjacent galaxies at some time (say 1 second) later.

In this scenario, not only will the galaxies be farther apart than before, but the farther apart they started from each other, the more they will have been carried along with the expansion. From the perspective of galaxy A, galaxy B started off 1 cm away and is 2 cm away 1 second later. So from A's perspective, B moved away at 1 cm/s. Similarly, galaxy C started 2 cm away from A and is 4 cm away 1 second later. It moved the distance between A and B plus the distance between B and C, or a total distance of 2 cm. That means C moved away from A at 2 cm/s. Galaxy D started off 3 cm away from A and moved away at 3 cm/s. This follows a linear relationship. If we plotted the distances of the points A, B, C and D vs. their motions, we would get a line, just like in a Hubble diagram.

It does not matter which point you choose as a reference in this exercise. You will always get the same linear expansion law. From the perspective of Galaxy B, for example, galaxy A was 1 cm away and moved away at 1 cm/s. Galaxy C was also 1 cm away from B and moved away at 1 cm/s due to the expansion of the band. Galaxy D was 2 cm away and moved away from B at 2 cm/s.

Again, an important thing to realize is that none of these galaxies exerted any effort to move. They were carried along with the expansion, or stretching, of the fabric of space. For simplicity we have done this example in one dimension (a line), but this sort of stretching happens in all three spatial dimensions simultaneously as the Universe expands.

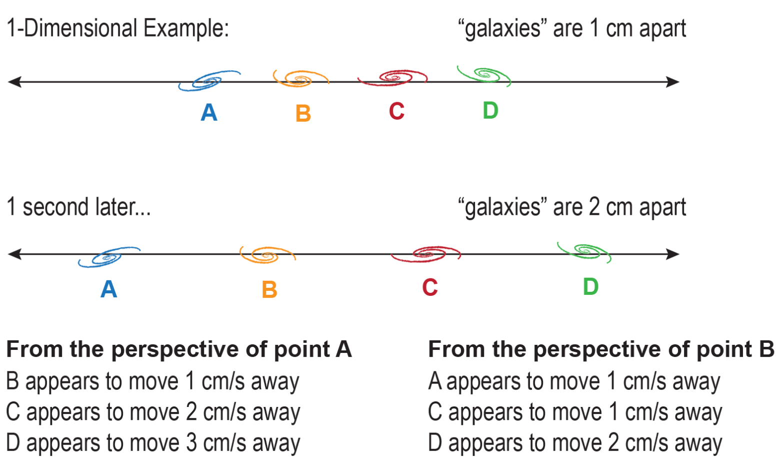


Figure 13.6: An analogy for the expansion of space: a stretchy band. The galaxies move away from each other as space stretches. The expansion follows the Hubble Law: the farther away a galaxy is from another, the faster its velocity. This is true from the perspective of any of the galaxies. Only one dimension is shown here but the fabric of space stretches in all three spatial dimensions as the Universe expands. Credit: NASA/SSU/Aurore Simonnet

In addition to being carried along with the expansion of space on the largest scales, galaxies are also moving in ways not due to the expansion of the Universe. These motions, known as peculiar velocities, are caused by local conditions near the galaxies, for example the tug of gravity from galactic neighbors. An analogy for this would be to imagine ants crawling around on an expanding picnic blanket. The crawling around motions of the ants would be analogous to the peculiar velocities of galaxies. On average, all of the ants would be getting farther away from all of the other ants because of the expansion of the blanket, just as all galaxies on average are getting farther away from all other galaxies due to the expansion of the Universe. Any particular ant would see its closest neighbors moving in many different directions, even after the blanket has begun to stretch.

For nearby ants, the peculiar motions are greater than the very slow stretching separation between them. In contrast, each ant sees the most distant ants moving rapidly away from itself due to the accumulated stretching of the blanket that is occurring between them. The peculiar motions within a distant group become comparatively unimportant and are even undetectable between groups with large separations. This illustrates what is meant by saying that space is expanding and carrying objects along with it, in contrast with saying the objects are moving through space.

One final analogy may help you visualize why we say the Universe expands (space stretches) but galaxies do not. Imagine making a loaf of raisin bread (Figure 13.7). Before baking, the raisins are scattered around in the dough. Each raisin would be stuck in a particular position within the loaf of bread dough. As the bread bakes and rises, the volume of bread increases. And as in our other analogies, two raisins will see their distance from each other increase even though they have not moved “through” the dough (i.e. the "space"). Instead, they have been carried along by the expansion of the dough.

An important feature of the expansion of the Universe is that the galaxies themselves do not expand, only the space between them does. In the raisin bread analogy, the dough rises but the raisins stay the same size. In the Universe, the internal forces in galaxies, stars, people, and other objects overwhelm the expansion locally, so they remain the same size/distance while the expansion of space carries them farther away from one another on large scales.

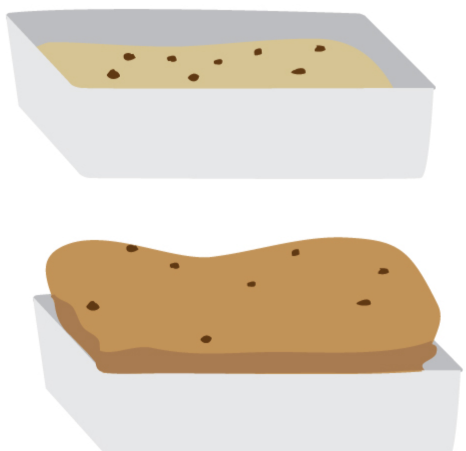


Figure 13.7: Another analogy for the expansion of space is a loaf of raisin bread baking; the raisins are like galaxies and the dough is like space. Credit: NASA/SSU/Aurore Simonnet

Each of these analogies is meant to help you visualize the three-dimensional stretching of space that takes place as the Universe expands. There is a major limitation shared by all of these analogies that we should mention: each of them features something expanding into space. In the actual Universe, space itself is expanding. It is not necessarily expanding into anything. So, for example, the Universe does not necessarily expand into previously existing empty space. The expansion is instead caused by creation of new space everywhere, constantly.

In the next activity, you will explore the Hubble Law and its implications for the expansion, or stretching, of space.

The Stretching of Space

In this activity, you will see representations of a set of galaxies in the Universe at different times. Between each time the Universe will stretch. You will be able to explore this expansion from the viewpoint of several different galaxies.

To view the Universe from a particular galaxy’s viewpoint, place the red circle over that galaxy.

Play Activity

A. Imagine you are an inhabitant of galaxy A

https://phys.libretexts.org/@go/page/31429

B. Now imagine you are an inhabitant of galaxy B

C. Center of the Universe

We have just seen that a homogeneous Universe that stretches according to a Hubble Law remains homogeneous. This means the stretching looks the same (on average) at every location. This property, homogeneity, does not hold for other possible explanations for the Hubble Law. The next activity gives an illustration of how homogeneity can be broken under one particular expansion scenario.

The stretching Universe also appears the same (on average) in all directions. We call this property isotropy, and we say that the Universe is isotropic. If you think about it you can probably convince yourself the homogeneity implies isotropy, but not the converse.

Consider the “ancient explosion” scenario that many students initially picture when thinking about how the Universe formed. If all of the galaxies we see are moving away from our location because of some ancient explosion, we should see a different pattern of recession velocities, as well as other consequences of the explosive event. For instance, we might expect that the density of galaxies would be higher toward the center of the explosion, while it would be lower far from the center. We thus would expect to see, at the very least, a difference in density of galaxies in different directions. We do not. In fact, one of the basic observational constraints we have is that the Universe appears the same (on average) in all directions—in other words, that the Universe is isotropic**.** Another basic observational constraint we have is that the average density of galaxies on large scales is constant (in space, not in time) everywhere—in other words, the Universe is homogeneous on large scales. Obviously, on small scales the density is not constant, but on scales much larger than galaxy clusters it is.

The fact that we observe the Universe to be both homogeneous and isotropic makes it essentially impossible to reconcile the Hubble Law with an explosion of matter out from a single point in space. We certainly would not see an isotropic pattern of galaxies on the sky unless we happened to be at the very center of the explosion. Even then, we would not expect to see the homogeneity that is observed. A better explanation, one consistent with the properties we see in the Universe, is that space is expanding, or stretching.

Explosion Cannot Explain the Observations

As you move the slider bar, you will see representations of a set of galaxies in the Universe at different times. This time there has been a central explosion that shoots matter outward.

Play Activity

13.3.2: General Relativity and the Expansion

If we wish to describe the space of the Universe using general relativity, we know that we will have to write down a spacetime interval that describes the separation between events in that space. But what would such a space look like? It would have to be homogeneous and isotropic, and it would have to expand (or contract) in time. The spacetime interval (s^2) for such a space (ignoring curvature for now), would look like the expression below.

$$s^2 = S^2(t)(\Delta d)^2 - (c\Delta t)^2$$

where s^2 is the interval, Δd is the spatial component, $c\Delta t$ is the time component, and $S(t)$ is a stretching factor called the **scale factor**. We use parentheses with $S(t)$ to emphasize that the scale factor is a function of time. The purely spatial part of the interval,

$(\Delta d)^2$ is defined as follows.

$$(\Delta d)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

The expression should look familiar from the Pythagorean theorem.

This spacetime interval looks exactly like the one for special relativity except that it has the extra term $S(t)$, the scale factor. The scale factor is a function that describes how space stretches (or contracts) with time. When the spacetime interval is written this way, the scale factor affects all directions equally. Therefore, since normal three-dimensional space is already homogeneous and isotropic, this spacetime interval will be as well.

This is a general form for a spacetime interval that describes a homogenous space which is not curved. We do not need to scale time since it is space that is homogenous and isotropic, not space and time. When we refer to a homogenous and isotropic spacetime we mean that only the space parts are homogenous and isotropic; the Universe is not the same in the past as in the present and in the future.

We still have not put any curvature into this equation, i.e. there are no curvature terms present. The equation describes a flat Universe, one that is not curved in any direction. The term "flat" may conjure up images for you of space being only two dimensional, but notice that we have included all three dimensions in our equation. In the context of cosmology and geometry, "flat" is a technical term and it simply means a lack of curvature.

To understand why our equation does not cause space to curve, we will examine what happens before and after space is stretched. Combining our equations, we have an overall expression:

$$s^2 = S^2(t)[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2] - (c\Delta t)^2$$

Consider a particular time, t_o. At that time the scale factor has a value that we could call $S_0 \equiv S(t_0)$. Plugging into the above equation, the spacetime interval is given by:

$$s^2 = [(S_0\Delta x)^2 + (S_0\Delta y)^2 + (S_0\Delta z)^2] - (c\Delta t)^2$$

We have definitely stretched the space, since our coordinates all have new values now: x has become $S_0 x$, y has become $S_0 y$ and z has become $S_0 z$. S_0 is the same number in all three cases. However, since we have stretched each direction by the same amount (the number S_0) everywhere, the space remains flat, i.e. it does not curve.

Since there is no curvature in the evenly-stretching interval we have written above, there is not any gravity either. We will add gravity later by ensuring that the interval satisfies Einstein's field equations. For the moment we are only considering how to describe a homogeneous, isotropic and expanding spacetime.

13.3.3: Describing an Expanding Universe: Comoving Coordinates

In order to more easily discuss the relationships among objects in a dynamic space that can be expanding or contracting, we introduce the concept of **comoving coordinates**. The core idea of comoving coordinates is to have a system to distinguish between an object moving through space versus an object sitting still in space while being carried away from other objects because space itself is expanding. Comoving coordinates remain fixed for an object as long as it does not move around through space.

We could draw such coordinates on a rubber sheet and then stretch the rubber sheet—the coordinates would remain fixed, but vertices on the sheet would move apart as the sheet itself expanded. Even though the physical distance between two points on the sheet would increase as the sheet stretched, the comoving distance (the coordinates of any point) would be unchanged. This is because the marks would still be in the same spots on the sheet, but additional space (surface area in this example) would be created between them.

Using comoving coordinates, it is easy to represent expansion or contraction in a uniform space. If you place a collection of stationary objects at different points in space, they will stay at the same comoving distance from each other no matter what the space does. The physical distance between them does change, however. This change can be related to their comoving distances with just a single parameter, the scale factor, *S*, introduced in the last section. This formulation suggests a convenient way to compare different times in the history of the Universe.

We do not know the absolute size of the Universe, and indeed it may be infinite. For this reason it is more convenient to consider the size history of the Universe in relative terms. We can define a scale factor such that $S = 1$ today and $S = 0$ at the moment that the Universe (the existence of space and time) began. With this definition, the physical distance $d(t)_{physical}$ between two objects at any time t is just their (unchanging) comoving distance $d(t)_{comoving}$ multiplied by the scale factor at that time $S(t)$.

$$d(t)_{physical} = d_{comoving}S(t)$$

In Figure 13.8, the "Universe" has expanded by a factor of two from the first panel to the second panel and by another factor of two from the second panel to the third panel. The physical distance between the two points shown depends only on the scale factor at the particular time of the measurement since the comoving distance between them does not change. Only S changes with time; the positions of objects in the comoving system do not.

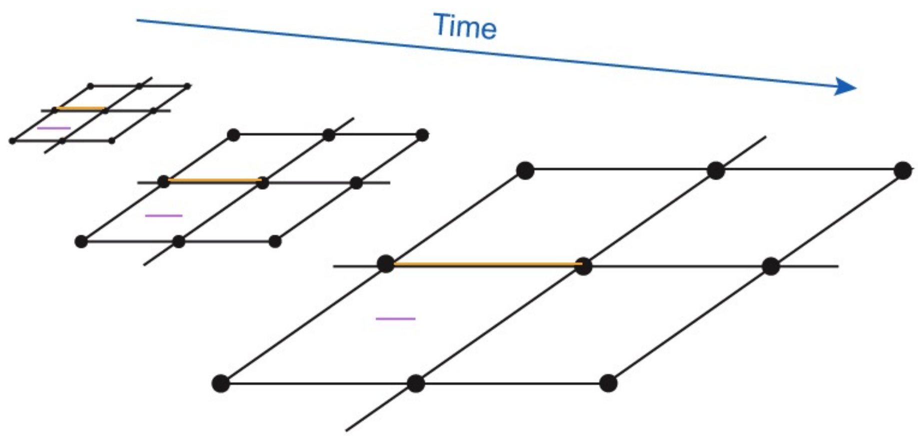


Figure 13.8 As a rubber sheet stretches, a coordinate grid drawn on its surface will also expand, carrying adjacent vertices away from each other as the surface area of the sheet increases. The grid blocks are comoving coordinates, which remain the same even as the space expands. The purple lines are lines of constant physical size. If the orange lines are measured in comoving coordinates, their lengths are always 1. However, their physical sizes grow, so their lengths become larger relative to the purple lines. Credit: NASA/SSU/Aurore Simonnet

We generally pick the reference time at which we set $S = 1$ to be the present time. This is just for convenience. Then the comoving coordinates are the same as the physical coordinates at the present time. We can find the physical coordinates at other times by multiplying the current physical coordinates by $S(t)$ for the suitable value of the time, t.

In the next two activities, you will explore the relationship between physical distances, comoving coordinates, and the scale factor.

Comoving Coordinates

In this activity, you will be able to adjust the scale factor for the Universe, measure physical distances, and measure comoving coordinates. Galaxies will be represented by different color dots in this activity.

- To adjust the scale factor, move the slider bar. A scale factor of 1 is the present, a scale factor of less than 1 is the past, and a scale factor of more than 1 is the future.
- To measure the comoving distance between two galaxies, count the number of grid blocks between them.
- To measure the physical distance between two galaxies, check the measure button, then click and drag the line between the two galaxies.

Play Activity

Find two galaxies that are 4 grid spaces apart. Call these galaxies A and B. (You may need to reload the activity if you don't see two galaxies that fit this criterion at first.)

A. The Present.

Set the scale factor to 1.0.

B. The Past.

Set the scale factor to 0.5.

Using the Scale Factor

Worked Example

1. Galaxies A and B are 50 Mpc apart today, when the scale factor is 1. This is their physical separation.

a. What is their comoving distance today?

- Given: $d_{physical}$ = 50 Mpc, S = 1
- Find: $d_{comoving}$
- Concept: $d(t)_{physical} = d(t)_{comoving}\ S(t)$
- Solution: 50 Mpc = $(d_{comoving})(1)$ → $d_{comoving}$ = 50 Mpc

b. What was their comoving distance apart at a time in the past when the scale factor was 0.5?

By definition, the comoving distance is always the same. In this case, it is 50 Mpc.

c. What was their physical distance apart at a time in the past when the scale factor was 0.5?

- Given: $d_{comoving}$ = 50 Mpc, S = 1
- Find: $d_{physical}$
- Concept: $d(t)_{physical} = d(t)_{comoving}\ S(t)$
- Solution: $d_{physical}$ = (50 Mpc)(0.5) → $d_{physical}$ = 25 Mpc

So, when the scale factor was half of what it is today, the physical separation between galaxies A and B was half of what it is today. However, the comoving distance is the same today and in the past.

You may have noticed from the previous activity that if a pair of objects are being separated by Hubble expansion, the ratio of their physical distance in the past to their physical distance today is equal to the ratio of the scale factor in the past to the scale factor today. This holds true for the future as well. Mathematically, this can be expressed as below.

$$\frac{d(t_1)_{physical}}{d(t_2)_{physical}} = \frac{S(t_1)}{S(t_2)}$$

where t_1 and t_2 are any two times in the past, present, or future of the Universe.

13.3.4: The Cosmological Redshift

When Edwin Hubble noticed that all galaxies were redshifted, he interpreted that observation as meaning that they were Doppler shifted, and that they moved away from him through space. We stated earlier that the redshift is actually correctly interpreted as being due to the expansion of space. With the spacetime interval for a homogenous and isotropic spacetime and the idea of comoving coordinates we are now ready to explore this idea further: these give us a notion of space itself stretching and carrying objects along with it. This cosmic stretching has a profound effect on light.

When we look at the spectrum of an astronomical object, the redshift, z, is defined as the wavelength shift in some spectral feature, divided by the unshifted wavelength.

$$z = \frac{(\lambda_{observed} - \lambda_{emitted})}{\lambda_{emitted}}$$

There are several processes that can produce wavelength shifts in the light emitted from astronomical sources. For example, the Doppler effect due to relative motion or the gravitational redshift required by general relativity. The stretching of space causes yet another kind, called the **cosmological redshift**.

At low velocities, Doppler redshifts and cosmological redshifts obey the following relationship between redshift and velocity:

$$z = \frac{v}{c}$$

where z is the redshift, v is the velocity, and c is the speed of light.

At high velocities, the picture changes. Cosmological redshifts still obey this relationship, but Doppler shifts do not. Special relativity requires a different formula for Doppler shifts, because relative velocities of objects through space must always be less than light speed $(z < 1)$. Cosmological redshifts, on the other hand, can be greater than one, because objects are being carried along by the expansion of space rather than traveling through space.

As the Universe expands and the scale factor grows, the stretching of space stretches light along with it. So we can write a relationship that expresses the stretching of the wavelength of light in terms of the stretching of the space itself:

$$\frac{\lambda_{observed}}{\lambda_{emitted}} = \frac{S(t_{observed})}{S(t_{emitted})}$$

Here, $\lambda_{observed}$ and $\lambda_{emitted}$ are the wavelengths of the light when it is observed and emitted, respectively. Similarly $S(t_{observed})$ represents the scale factor of the Universe when the light is observed, and $S(t_{emitted})$ is the scale factor at the time the light is emitted.

If we express the observed wavelength in terms of the emitted wavelength and the change in wavelength $\Delta\lambda$, we can relate the scale factor to the redshift, z.

$$1 + z = \frac{S(t_{observed})}{S(t_{emitted})}$$

This means that redshift is directly related to the change of the scale factor of the Universe over the time of travel for the light. In fact, the amount by which the Universe has expanded is given by the term $1 + z$. For example, for an object with a redshift of $z = 1$, the Universe has expanded by a factor of 2 since the time the light left that object. For $z = 0.5$, the Universe has expanded by a factor of 1.5, or 50%. This is the proper interpretation of the cosmological redshift. It is not a Doppler shift caused by the motion of galaxies *through space*. It is a stretching of light *caused by the stretching of space*. As time passes, space stretches. The cosmological redshift results from the wavelength of light stretching along with it.

These expressions are a direct result of general relativity and are motivated in **Going Further 13.2: Deriving the Cosmological Redshift**. In **Going Further 13.3: The Hubble Law and the Cosmological Redshift**, we show how the cosmological redshift leads to the Hubble Law.

The Stretching of Light

Light is stretched by the expanding Universe in exactly the same proportion as space. We can consider what effect this has on how we see astronomical objects.

Worked Examples

1. The calcium K line of Galaxy A is observed to be 9835 Å.

a. What was its wavelength when the light was emitted?

The rest wavelength of Ca K is 3934 Å, so that is the wavelength of the light when the galaxy emitted it.

b. What is the redshift of Galaxy A?

- Given: $\lambda_{observed}$ = 9835 Å, $\lambda_{emitted}$ = 3934 Å
- Find: z
- Concept: z = ($\lambda_{observed} - \lambda_{emitted}$)/$\lambda_{emitted}$
- Solution: z = (9835 Å – 3934 Å)/(3934 Å) = (5901 Å)/(3934 Å) = 1.5

c. By what factor has the wavelength of light stretched from the time it was emitted to the time it was observed?

Take the ratio $\lambda_{observed}$ / $\lambda_{emitted}$ = 9835 Å /3934 Å = 2.5. In other words, the wavelength of the light has stretched by a factor of 2.5, from ultraviolet when it was emitted, to infrared when it was observed.

d. By what factor has the Universe stretched since Galaxy A emitted the light that we are just seeing now?

- Given: z = 1.5
- Find: the ratio $S(t_{observed})/S(t_{emitted})$
- Concept: $1 + z = S(t_{observed})/S(t_{emitted})$
- Solution: $S(t_{observed})/S(t_{emitted}) = 1 + 1.5 = 2.5$
- The Universe has expanded by a factor of 2.5 since the light we are seeing now was emitted from Galaxy A.
- Equivalently, using the expression $\lambda_{observed}$ / $\lambda_{emitted}$ = $S(t_{observed})/S(t_{emitted})$ also gives a factor of 2.5.

For more examples, see Math Exploration 13.1.

Math Exploration 13.1

Going Further 13.2: Deriving the Cosmological Redshift

Imagine that light from a galaxy is emitted at some time, $t_{emitted}$ (t_e for short). At some later time, $t_{observed}$ (t_o for short), the light is observed at a second galaxy lying a comoving distance R from the first. The spacetime interval traveled by the light is described by the spacetime interval we introduced earlier for a flat, homogeneous spacetime. For this case it is better to write the interval in spherical-polar coordinates instead of Cartesian coordinates (you will see why next).

$$s^2 = S^2(t)[(\Delta r)^2 + (r\Delta\theta)^2 + (r sin\theta\Delta\phi^2] - (c\Delta t)^2 s^2$$

We can choose our coordinates so that the second galaxy, where the light is observed, is at the origin. That will make the path of the light completely radial. In this case there is no change in either θ or ϕ. Under this simplification the angular terms are zero and the interval becomes the following.

$$s^2 = S^2(t)(\Delta r)^2 - (c\Delta t)^2$$

Since we are imagining the path of a beam of light, we know that the total spacetime interval will be zero. This is always true for photons traveling in spacetime. Thus we can write the equation below.

$$0 = S^2(t)(\Delta r)^2 - (c\Delta t)^2$$

Now we can rewrite the equation so that it gives the comoving distance traveled, Δr, in terms of the corresponding time to travel that distance, Δt, and the scale factor, $S(t)$.

$$\frac{c\Delta t}{S(t)} = \Delta r$$

This expression is true for small parts of the path, but we cannot simply plug in the total comoving distance traveled, R, and the total time traveled. As the photon travels, the scale factor is constantly changing. We must find a way to take this change into account.

Is there a value of the scale factor that is appropriate for the entire journey? No, but for some tiny interval, where both Δr and Δt are very small, the scale factor is essentially constant. This suggests that one way to correctly evaluate our expression is to add up all such tiny contributions along the light travel path, as below.

$$R = \Delta r_1 + \Delta r_2 + \Delta r_3 + \ldots + \Delta r_i + \ldots \Delta r_N$$

In terms of the other side of the equation we get the expression below.

$$R = c\Delta t_1 S(t_1) + c\Delta t_2 S(t_2) + c\Delta t_3 S(t_3) + \ldots + c\Delta t_i S(t_i) + \ldots + c\Delta t_N S(t_N)$$

These sums can be written more compactly using summation notation. This notation is shorthand to indicate that we are going to add up some number of terms. We use the letter sigma (Σ), the Greek form of ‘s’ for “sum.” The individual terms are subscripted with \(i\) and the total number of terms in the sum N. So the sum will add up the terms from $i = 1$ to N^1.

The sum of all the Δr terms becomes the following.

$$\sum_{i=1}^{N} \Delta r_i \equiv \Delta r_1 + \Delta r_2 + \Delta r_3 + \ldots + \Delta r_i + \ldots \Delta r_N$$

The sum of the terms on the other side of the equation can be written as below.

$$\sum_{i=1}^{N} c\Delta t_i S(t_i) \equiv c\Delta t_1 S(t_1) + c\Delta t_2 S(t_2) + c\Delta t_3 S(t_3) + \ldots + c\Delta t_i S(t_i) + \ldots + c\Delta t_N S(t_N)$$

Combining these two, we have a way to write an overall expression, summing over the whole path, as the Universe expands.

$$\sum_{i=1}^{N} \Delta r_i = \sum_{i=1}^{N} c \Delta t_i S(t_i)$$

Such a sum will take into account the changes of the scale factor if each time element Δt_i is small enough so that the corresponding scale factor $S(t_i)$ of the i^{th} term is very nearly constant over that time interval. Adding the small contributions from all of the tiny terms in the sum we will give us the total comoving distance.

As we have written above, the term on the left of the equal sign is by definition the total comoving distance traveled, R. In addition, the speed of light, c, can be factored out of the summation on the right hand side, because it is a constant that occurs in every term. So we can simplify by writing as follows.

$$R = c \sum_{i=1}^{N} \Delta t_i S(t_i)$$

This expression gives the total comoving distance the light travels. It does this by adding up all the tiny distances $\Delta r_i = \Delta t_i / S(t_i)$ traveled in some tiny interval of time Δt_i. During these time intervals the scale factor of the spacetime (in other words, of the Universe) is $S(t_i)$ and is assumed to be approximately constant.

Now consider a single wavelength of light. The front end of the wave will be emitted from the source at a time t_e and arrive at the observer at a time t_o. The back end of the wave would be emitted by the source at a slightly later time, $t_e + \delta t_e$ and arrive at the observer at a time $t_o + \delta t_o$. The times δt_e and δt_o are both tiny, but not zero, and they are not necessarily the same. Since the comoving distance between the galaxies is the same for both light waves—remember that the comoving distance has no dependence on the expansion of the spacetime—the sum above applies to the comoving distance traveled in terms of either the front or the back of the wave.

We will now break this down and consider the sum from the front and the back of the wave separately. First we will write a sum over interval between the emission of the front of the light wave in the first galaxy and the observation of the front of the light wave by the observer in the second galaxy.

$$R = c \sum_{i=1}^{N'} \Delta t_i S(t_i)$$

For the back of the light wave we have as follows.

$$R = c \sum_{i=1}^{N''} \Delta t_i S(t_i)$$

Here the sum runs from the interval between the emission of the back of the light wave in the first galaxy and the observation of the back of the light wave by the observer in the second galaxy.

The sums are essentially identical, but with a subtle difference. We have used N' and N'' as the limits on the sums to remind ourselves of this difference, namely that the second sum is taken over a slightly different spacetime interval than the first sum. In both cases, the total number of terms (N' or N''), is large enough to make the individual time intervals in the sum small enough that the scale factor $S(t_i)$ remains constant during each time interval.

These two expressions are a bit complicated, so we should break them down. The first says that the total comoving distance traveled by the front of the light wave is given by summing up all the tiny distances traveled by the light during many tiny time intervals. The total time spans from when the front of the wave is emitted at the first galaxy to when it is observed at the second galaxy. The scale factor is different for each of these little time intervals, but remains constant during any given interval. This way of computing the total comoving distance allows us to take into account the differences in the scale factor over the light's journey. The second expression says exactly the same thing, except that now we are finding the comoving distance traveled by the back of the light wave, not the front.

Since the back of the light wave is emitted an instant after the front of the light wave, and it is observed an instant after the front is, the two sums are not identical. Using N' and N'' for the number of terms sums is a way of reminding ourselves of this subtle but important difference, though it otherwise has no mathematical significance.

The two comoving distances are the same because the expansion of the spacetime does not affect comoving distance: all information about the expansion is contained within the scale factor. Since the two expressions evaluate to the same value, R we can set them equal to each other.

$$R=c\sum_{i=1}^{N'}\Delta t_i S(t_i)=c\sum_{i=1}^{N''}\Delta t_i S(t_i)$$

Now consider these two sums carefully. They are nearly, but not quite identical. The first has an extra little bit during the interval δt_e before the second one begins, or in other words, before the back of the light wave is emitted. Likewise, the second summation has an extra little bit during the interval δt_0, after the first summation ends, due to the extra time required for the end of the light wave to reach the observer. We have used an N' as the limit for the second sum to remind us of this difference.

We can pull the extra term in each sum out of the summation notation and write it explicitly. We can also cancel the common factor of c. In that case we have the following.

$$\sum_{i=1}^{N}\Delta t_i S(t_i)+\delta t_e S(t_e)=\sum_{i=1}^{N}\Delta t_i S(t_i)+\delta t_0 S(t_0)$$

Now we use the same number of terms, N, in each sum because the sums are over exactly the same time spacetime interval. The extra bit in each sum, not contained in the other, is written out explicitly. The quantities in the square brackets are obviously identical, so after canceling them, the only terms left are the extra terms from the beginning and end of the sums. This relates the intervals for the emission and detection of the wave in terms of the scale factors at the times of emission and absorption.

$$\frac{\delta t_0}{\delta t_e}=\frac{S(t_0)}{S(t_e)}$$

The ratio of the intervals at the observer's end and the emitter's end is the same as the ratio of the scale factor at the time of observation and the time of emission. The time interval between the passage of the front of the wave and the back of the wave is the period of the wave. Remember that the frequency (f) of a wave is the inverse of the period (so $f=1/\delta t$ here). Writing our expression in terms of frequency we get the following relation.

$$\frac{f_0}{f_e}=\frac{S(t_e)}{S(t_0)}$$

This is the relation we were looking for. It expresses the observed frequency of light in terms of the emitted frequency and the amount by which the spacetime has expanded while the light was traveling. It is called the **cosmological redshift**. Note that it has nothing to do with Doppler shifts, but only with the fact that we are in a spacetime that is expanding; it is the stretching of spacetime itself that stretches the wavelength of the light.

If we wish, we can write the cosmological redshift in more familiar wavelength terms. We just have to substitute for frequency using $f=c/\lambda$.

$$\frac{(c/\lambda o)}{(c/\lambda e)}=\frac{S(t_e)}{S(t_0)}$$

Canceling the factor of c and rearranging, we have the equation below.

$$\frac{\lambda_0}{\lambda_e}=\frac{S(t_0)}{S(t_e)}$$

This form shows explicitly that the light is stretched by exactly the amount by which the spacetime has expanded while the light was traveling. Notice that we have not yet put gravity into our spacetime at all. It is not gravity but the stretching of spacetime itself that creates the cosmological redshift.

The equation above relates the wavelength of light from a source at the time it was emitted and observed to the scale factor of the Universe at the time the light was emitted and observed. If we want to rewrite this in terms of a redshift, we can use the definition of redshift.

$$z = \frac{(\lambda_0 - \lambda_e)}{\lambda_e} = \frac{\Delta\lambda}{\lambda_e}$$

where $\Delta\lambda$ is the change in wavelength. First, express the observed wavelength in terms of the emitted wavelength and the change in wavelength, then separate the terms, then finally substitute for redshift.

$$\frac{\lambda_0}{\lambda_e} = \frac{\lambda_e}{\lambda_e} + \frac{\Delta\lambda}{\lambda_e} = 1 + \frac{\Delta\lambda}{\lambda_e} = 1 + z$$

So we have the relation we seek.

$$1 + z = \frac{S(t_0)}{S(t_e)}$$

[1] In general, N could be finite, though perhaps large, or it could be infinite, in which case techniques from integral calculus might be used.

Going Further 13.3: the Hubble Law and the Cosmological Redshift

In the previous section we showed that the expansion of spacetime causes light to be stretched in the same manner that space is stretching. We will now see that this leads directly to the Hubble law.

Hubble showed that the redshift of a galaxy, which he interpreted as its velocity, is proportional to its distance. Redshift, z, is given by

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_0 - \lambda_e}{\lambda_e}$$

We can write this expression in a slightly different way.

$$z = \frac{\lambda_0}{\lambda_e} - 1$$

From the previous section we know that we can write the wavelength ratio in terms of the scale factor. We will use that relation to replace the ratio of wavelengths. We then have the expression below.

$$z = \frac{S(t_0)}{S(t_e)} - 1$$

Now we substitute for $S(t_0)$ using the definition of $\Delta S \equiv S(t_0) - S(t_e)$. Rearranging, we get the following.

$$S(t_0) = S(t_e) + \Delta S$$

Upon substituting into the expression for z, above, this is our result.

$$z = \frac{S(t_0) + \Delta S}{S(t_e)} - 1$$

Now consider a short time in which the spacetime does not expand by very much. Under these conditions the scale factor will be essentially constant ($S(t_0) \approx S(t_e)$) and we can simplify the expression above. Note that we are not saying S is constant, just that it is nearly so.

$$z \approx \frac{S(t_e) + \Delta S}{S(t_e)} - 1 = 1 + \frac{\Delta S}{S(t_e)} - 1$$

Or simply as follows.

$$z \approx \frac{\Delta S}{S(t_e)}$$

Note that we have switched from an equals sign to an approximately equal sign to remind us of the approximation here.

If the expansion over this time happens such that S changes by ΔS in a time Δt, then we can define the rate of change of the expansion as follows.

$$\dot{S} \equiv \frac{\Delta S}{\Delta t}$$

We can rewrite the redshift expression using this substitution.

$$z \approx \frac{\dot{S}\Delta t}{S(t_e)} = \frac{\dot{S}}{S}\Delta t$$

We now drop the t_e and use simply S from now on. Since we are assuming that S is nearly constant over the time under consideration, it is not necessary to consider its time dependence explicitly.

Next we can use the spacetime interval and the fact that it is zero for a light beam. We can then express Δt in a more useful way.

$$0 = (S(t))^2(\Delta r)^2 - (c\Delta t)^2$$

Rearranging the terms and taking the square root, we have the following.

$$c\Delta t = S(t)\Delta r$$

As above, with the assumption that Δt is small enough that $S(t)$ remains approximately constant, we can replace $S(t)$ with S. Now we can solve this for Δt.

$$\Delta t = \frac{S\Delta r}{c}$$

Substituting this expression in for Δt, the redshift can be written as below.

$$z \approx \left(\frac{\dot{S}}{S}\right)\left(\frac{S\Delta r}{c}\right)$$

But $S\Delta r$ is the physical distance traveled (not the comoving distance) by the light during this brief time interval. So if we call the physical distance $d\ (= S\Delta r)$, then we can rewrite the redshift expression. Multiplying through by c we get the following.

$$cz \approx \left(\frac{\dot{S}}{S}\right) d$$

The ratio in the parentheses is a constant; it is the fractional amount by which the spacetime expands per unit time. In other words, it is the Hubble constant H_0, and the expression is the Hubble law: velocity $v = cz$ is proportional to distance.

$$v \approx H_0 d$$

The subscript zero on the Hubble constant reminds us that we are only considering time intervals for which the scale factor is approximately constant, and in particular we consider the time period very close to the present one. In this short time, light will not have traveled very far. The relationship should only be valid for the “local” Universe, with the exact meaning of “local” depending upon how fast the scale factor changes.

If we begin to look at galaxies over larger distances (meaning the light has been traveling for a longer time) then the approximation of a nearly constant Hubble constant is not necessarily valid. In that case we might expect to see deviations from a strict proportionality. In fact, that is exactly what we do see, and the amount and nature of the deviations can tell us many things about the history of the Universe and its energy balance.

This page titled 13.3: The Universe is Expanding is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.4: The Age of the Universe

Learning Objectives

- You will be able to perform calculations and understand conceptually the relationship between the expansion rate and age of the Universe.
- You will be able to distinguish the concepts of Universe and observable Universe.

Age of the Universe

13.4.1: Olbers' Paradox - The Dark Night Sky Tells Us That Time Had a Beginning

It is possible to make a simple observation that shows that the Universe cannot be both infinitely old and completely uniform. This is the observable fact that the night sky is dark!

If the entire Universe was similar to the region around us, whether we define that region as a few hundred light-years of stars or a few hundred million light-years of galaxies, and if it was also essentially unchanging over time—with the same assortment of stars or galaxies forever—we would be cooked in a fraction of a second. In such a Universe, if you looked in any direction, your line of sight would eventually intersect a distant star. An average star has a surface temperature of several thousand kelvin, so the whole sky would glow with that mean temperature—just like the surface of a star. In fact, the Universe would be hotter than the hottest oven. Heinrich Olbers (1758–1840) pointed this out in 1823, though others had noted it before, with Thomas Digges (c.1546–1595) being the first. Nonetheless, this is referred to as Olbers' paradox, though it is only a paradox if one believes the Universe to be infinite in time and spatial extent.

One early attempt to remedy the paradox and permit an infinitely old and uniform cosmos was to propose that gas and dust between the stars were absorbing much of their radiation and therefore lowering the mean radiant temperature of the sky. However, that does not work. The gas and dust would soon be heated to the same high temperature as the stellar surfaces, and oven-like conditions would prevail. So either the Universe is not uniform out to infinite distances, or it cannot be infinitely old. It is easy to imagine—though difficult to account for—a Universe consisting of a finite set of stars like the Milky Way, surrounded by emptiness. Then, only when we looked exactly in the direction of one of its hundred billion stars would we measure a radiative temperature of thousands of degrees. Because stars subtend such tiny angles, the average temperature of the sky could be very low. We do not live in such a Universe. We see galaxies spread quite uniformly around us out as far as we can see.

A different solution to the paradox is immediately implied by the Hubble expansion—the Universe has a finite age. Even if it is infinite in extent, light has only had time to reach us from the closest galaxies and the average sky temperature must remain extremely low as a consequence. In addition, the stretching of space has increased the wavelength and therefore lowered the energy (and corresponding effective temperature) of the radiation coming from the most distant reaches of space.

13.4.2: Hubble Time

Imagine the Hubble expansion scenario playing like a movie in reverse. Instead of galaxies moving away from each other as time goes forward, galaxies would rush toward each other as time goes backward. Galaxies would be closer and closer together in the past, until at some time in the distant past the matter that makes up the galaxies would have been very close together. We can extrapolate back to this time, the beginning of the Universe. If we know the expansion rate for the Universe and assume that it has been constant, we can calculate how much time the Universe has been stretching.

The Hubble constant is an example of a stretching rate. The Hubble constant is generally expressed in units of km/s/Mpc due to how it is measured. However, both km and Mpc are units of distance and cancel out, so the Hubble constant, or any stretching rate, actually has units of 1/time. Again, assuming that the expansion rate has been constant, we therefore have an expression for the age of the Universe t.

$$t=\frac{1}{H_0}$$

In this expression for the age, t is called the Hubble time. It is computed by taking the reciprocal of the Hubble constant, H_0. This expression tells us that if H_0 is bigger, then t will be smaller and vice versa. A faster expansion rate will lead to a younger age for the Universe. A slower expansion rate will lead to an older age of the Universe. To go back to our stretchy band analogy, if the expansion rate is faster, it will take less time for the galaxies to reach their current distances from each other. If the expansion rate is slower, it will take them a greater amount of time.

This mathematical formula can also be expressed graphically, as in Figure 13.9.

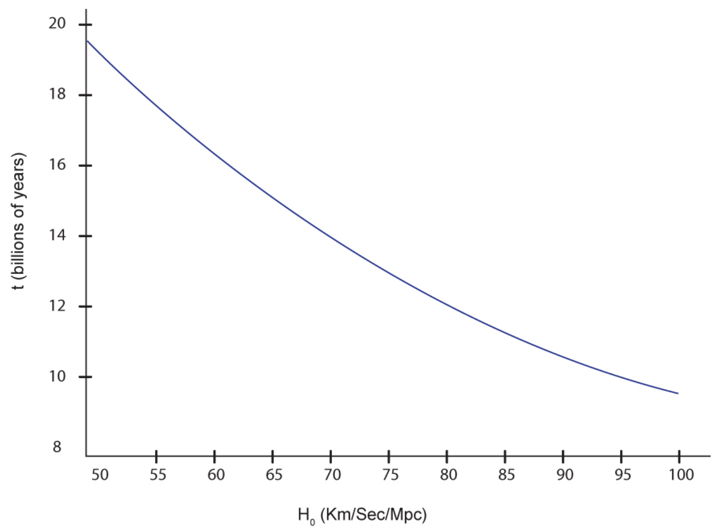


Figure 13.9: The current age of the Universe (t) is plotted vs. the current expansion rate (H_0). This is the graphical form of the equation $t = 1/H_0$. Credit: NASA/SSU/ Aurore Simonnet

Going Further 13.4: Another Way to Think About the Hubble Constant

Another way to look the Hubble law and the Hubble constant is from the perspective of elementary physics. We can find the distance traveled by an object moving at constant speed if we multiply the speed by the time traveled. Mathematically this idea is expressed as distance (d) equals speed (v) times time (t), or

$$d=vt$$

Hubble's Law has exactly this same form:

$$v=H_0d$$

This might not look like the same form at first, but lets rearrange terms a little bit. Dividing both sides by the Hubble constant we get:

$$d = vH_0 = v\left(\frac{1}{H_0}\right)$$

Now the first and the last equations look exactly the same. We just have to realize that the Hubble constant is the reciprocal of a time:

$$t = \frac{1}{H_0}$$

But what time is this? It is the time needed for any given galaxy to travel to its current distance, and of course, this time is the same for all galaxies. The faster galaxies travel farther during this time; the slower ones travel not as far. This is a consequence of what we reasoned before, but some students prefer to think about it in this manner.

The Expansion and Age of the Universe

In this activity, you will use the graph in Figure 13.9 to explore the relationship between the expansion rate today (H_0) and the current age of the Universe (t).

Converting the Expansion Rate Into Units of Inverse-Seconds

Because of the way it is measured, the Hubble constant has strange units: km/s/Mpc. However, km and Mpc are both units of distance, so we can cancel them and express the Hubble constant in inverse-seconds (1/s). You will practice doing so in this activity.

You will need the conversion between Mpc and km: 1 Mpc = 3.09×10^{19} km.

Worked Example

1. Convert H_0 = 70 km/s/Mpc to inverse-seconds (1/s) by canceling out the distance units.

To make the conversion, it is best to start with the units of H_0 written out carefully rather than in shorthand notation. So, 70 km/s/Mpc can also be written like this.

$$\frac{70\,\mathrm{km/s}}{\mathrm{Mpc}}$$

This form makes it easier to see how to convert the units.

$$\left(\frac{70\,\mathrm{km/s}}{\mathrm{Mpc}}\right) \times \left(\frac{1\,\mathrm{Mpc}}{3.09 \times 10^{19}\,\mathrm{km}}\right) = 2.27 \times 10^{-18}\,\mathrm{sec}^{-1}$$

Notice the km and Mpc on the top and bottom canceled out, leaving the answer in 1/s.

Questions

> 📌 The Relationship Between H_0 and T

As described in the text above, assuming the Universe has been expanding at a constant rate since its beginning, then its age is $t = 1/H_0$, where t is in units of seconds (s) and H_0 is units of 1/s. We will use this relationship to compute the age of the Universe for several possible values of H_0.

Worked Example

1. If H_0 = 70 km/s/Mpc, what is the age of the Universe in seconds?

- Given: Using the results of the previous activity, the Hubble constant in units of inverse-seconds is H_0 = 2.27E-18 1/s
- Find: t, the age of the Universe

- Concept: $t = 1/H_0$
- Solution: t = 1/(2.27E-18 1/s) = 4.41E17 s

2. Convert the age of the Universe to years. 1 year = 3.15 × 10^7 s.

t = 4.41E17 s x (1 yr / 3.15E7 s) = 1.4E10 yr

3. Convert the age of the Universe to billions of years. 1 billion = 1 × 10^9.

t = 1.4E10 yr x (1 billion / 1E9) = 14 billion yr

So, if the Hubble constant is 70 km/s/Mpc, then the Universe is 14 billion years old.

Questions

13.4.3: The Hubble Constant Over Time

The Hubble time, as we have just derived it, gives us a rough idea of the age of the Universe. Of course, we have no reason to expect that the stretching rate has been constant across the history of the Universe. So Hubble "constant" is another of those cases where a term in astronomy can be somewhat misleading. The Universe's stretching rate will be close to a constant value throughout space at any short instant in time, including today, so in that sense the "Hubble constant" is a constant. However, it is also likely to have varied over long time scales. To distinguish these two possibilities, astronomers often reserve the term Hubble constant, denoted H_0, only for its value today. They then use the term Hubble parameter, denoted H (without the subscript), when referring to the expansion rate at other times. But you should still be prepared to see the term Hubble constant in both contexts, because not everyone always uses the term Hubble parameter.

In later chapters we will look in detail at why we think the Hubble parameter might not actually be constant over the age of the Universe. For the time being you should simply keep in mind that the rate of expansion could change, and this will affect our estimate of the age of the Universe.

13.4.4: Independent Age Estimates

The Universe must naturally be older than all of its constituents. This constraint can serve as a consistency check for our physical theories: when objects in the Universe have estimated ages greater than our best estimate for the age of the Universe, something must be reconciled. Either our estimate of the cosmological age is in error or the estimate of the age of the objects in question is wrong.

When Hubble made his first determination of the Hubble constant in 1929, he made a number of understandable errors. These made the Hubble constant several times too large and the Hubble time correspondingly too small—only 2 billion years. That was in serious conflict with the ages of the oldest known stars calculated from our knowledge of nuclear physics, not to mention the age of Earth itself. Even in the late 20th century there were problems. Estimates of the Hubble time could still be as short as 10 billion years. At the same time, the oldest structures in our Galaxy, globular clusters, were calculated to be as much as 15 billion years old. There were large uncertainties in both values, around 2 billion years in the case of the age of the Universe, and 3 billion years for the ages of globular clusters. Still, the inconsistency between these ages pushed scientists to improve their theories of stellar evolution and to improve their cosmological measurements. As a result, today most estimates fit without conflict and with much greater precision, partly as a consequence of the discovery of new physics.

One exception is the value of the Hubble constant that we measure from expansion compared to its value as measured by the Cosmic Microwave Background. We will return to this conflict in the chapter on the background radiation.

13.4.5: The Observable Universe

Another important consequence of the finite age of the Universe is that we can only ever see part of it—that part from which light has had time to reach us. We do not know how big the entire Universe is, and it might be infinite. In either case, it is likely that the Universe is much bigger than the part from which we can receive information.

The size of the observable Universe is easily calculated. We have seen that the age of the Universe is 13.8 billion years. If the Universe was static, we would only be able to see objects 13.8 billion light-years away. Light from anything farther away would not have had time to reach us. Even though space has been stretching during all those years, we can still think in terms of the lookback times to the objects we observe. Consider the following thought experiment.

Think of your observable Universe as a sphere. This sphere is centered on you—as it would be around any other observer anywhere today. The sphere extends out to anything with a lookback time of 13.8 billion years; that is the size of the sphere because that is how long light has been traveling in the Universe. Anything farther away is not visible. For example, if there is a galaxy with a lookback time of 15 billion years away from you, you cannot see it right now because light has not had enough time since the beginning of the Universe to traverse the distance separating you from that galaxy. We say that it would be beyond your observable Universe, or cosmic horizon. To see such a galaxy, you would have to wait another 1.2 (15.0 – 13.8) billion years.

Observers somewhere in the Andromeda galaxy right now would see a similar spherical volume, but it would be centered on themselves instead of on us. So they would see a little sliver more of the Universe on the side of their sphere away from us, and a little sliver less on the side toward us. Observers way out toward the edge of our observable Universe would see our galactic neighborhood as being way out toward the edge of their observable Universe. In the opposite direction they would see a huge swath that we cannot see at all—and visa versa. Of course they would see us as we were many billion years ago, not as we are now, just as we see them as they were many billion years ago, not as they are now.

The Observable Universe

In this activity, we will explore the concept of observable Universe.

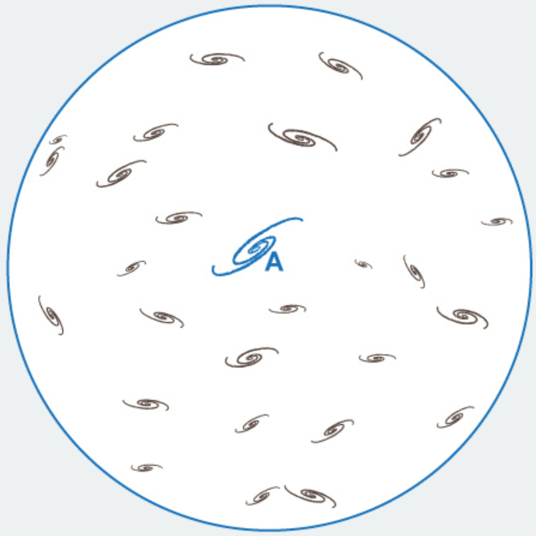


Figure A.13.2: The blue line represents the observable Universe for galaxy A. Credit: NASA/SSU/Aurore Simonnet

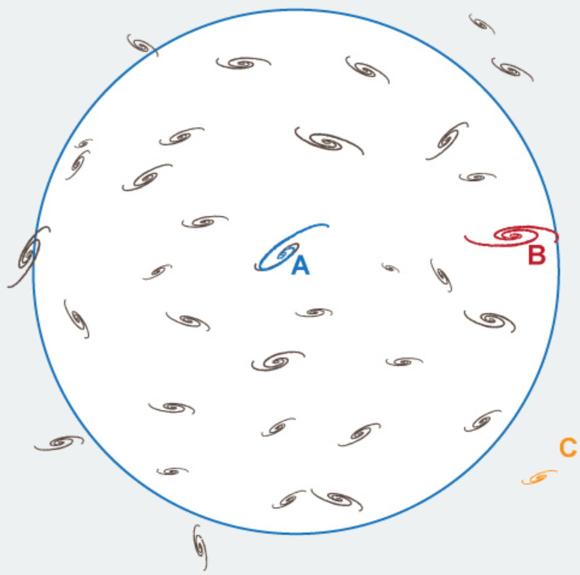


Figure A.13.3: The blue line represents the observable Universe for galaxyA. Credit: NASA/SSU/Aurore Simonnet

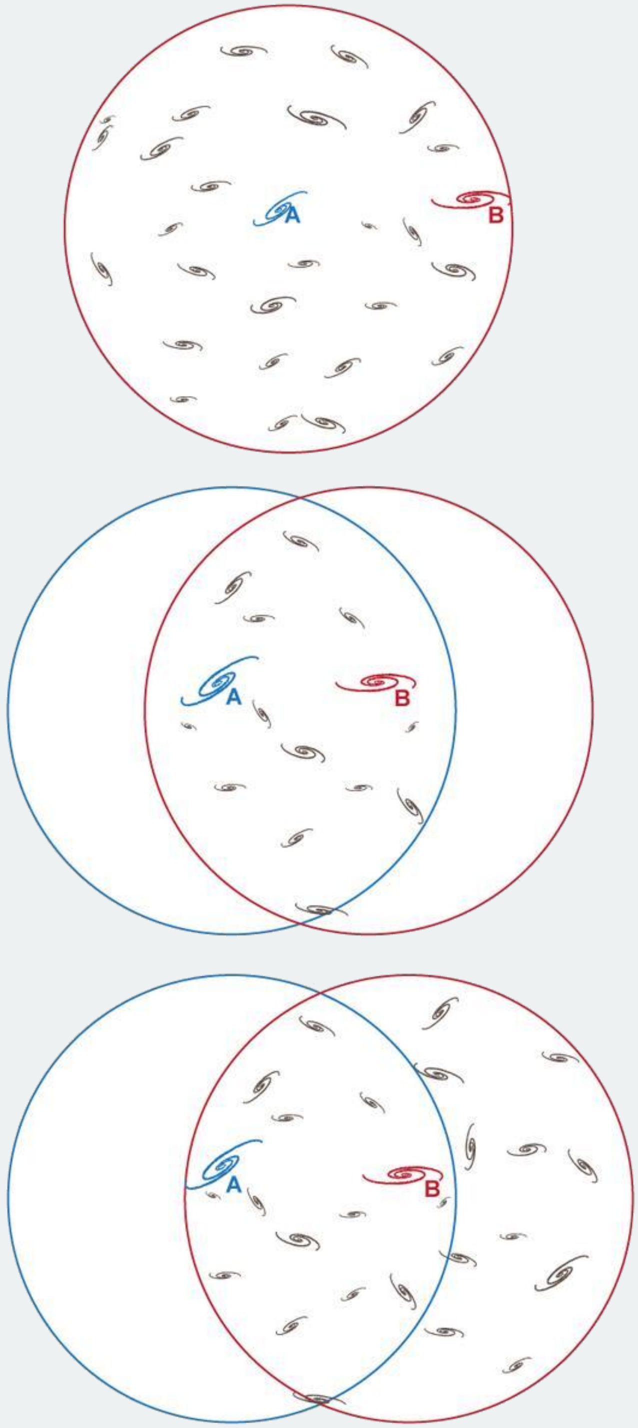


Figure A.13.4: The red circles represent three possibilities for the observable Universe and its contents for galaxy B. Credit: NASA/SSU/Aurore Simonnet

In the last activity, you should have found that the **observable Universe** for an observer is centered on that observer, and is a subset of the whole Universe. The finite age of the Universe and the finite speed of light combine to limit what we can observe of the Universe in several ways. There is a boundary to our observable Universe called the cosmic horizon. We cannot observe anything beyond that boundary because light from there has not had enough time to reach us—even 13.8 billion years has not been enough time. The size of the observable Universe gets bigger as the distance that light has traveled since the start of the Universe gets bigger. In other words, there may be objects beyond our current cosmic horizon that we cannot see right now, but as the Universe gets older, we will be able see more of them.

This page titled 13.4: The Age of the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.5: The Basic Big Bang Model

Learning Objectives

- You will know that the Big Bang theory features an expansion of space not an explosion of matter.
- You will know that the expansion of the Universe can be described by general relativity.
- You will know that the expansion leads to a change in the Universe over time, including temperature and density.

The Big Bang

13.5.1: Key Components of the Model

The best and most widely accepted current model for how our Universe changes over time is usually referred to as the standard Big Bang theory, or model. The Big Bang model is based on general relativity and other models from physics and is supported by much observational evidence. According to the Big Bang model, space is expanding and cooling, and every part of the Universe was once extremely hot and dense. The Universe—space, time, and all of the matter and energy in it—came into existence billions of years ago. All of the matter and energy that exists today has existed since the beginning; the Universe has been expanding, cooling, and changing ever since. These ideas lead to the following implications, some more obvious than others:

13.5.1.1: *Implication 1: The Expansion of Space*

On average, galaxies are moving away from each other, in a way that follows the Hubble law, due to the stretching of space. We emphasize that the galaxies are moving away from *each other*, that there is no special central point in the Universe. Finally, unlike objects we are familiar with in everyday life, such as balloons, that expand into space, the Universe itself does not need to be expanding into any other kind of space; instead, the creation of new space is what drives the expansion.

13.5.1.2: *Implication 2: Finite Age*

The Universe has a finite age related to its rate of expansion. None of the objects contained in the Universe can be older than the Universe itself. Current best estimates put the age of the Universe at 13.8 ± 0.1 billion years.

13.5.1.3: *Implication 3: Hot Dense Phase*

If the Universe is expanding, that means that at some time in the past, everything in the Universe was compressed extremely close together, in a state of high density. The high-density state would also have been very hot. The Universe was hot and dense everywhere and every part of it participated in the expansion, including right here. If we go back far enough, we can imagine the density and temperature of the Universe reaching extremely high values. This is different from an explosion, which starts at one point in space and scatters material through other regions.

13.5.1.4: *Implication 4: Big Bang Nucleosynthesis*

A hot, dense plasma as we imagine existed early in the history of the Universe would have undergone nuclear reactions, just as happens inside stars today. This is because, at some point, the conditions in the entire Universe must have been similar to the conditions at the centers of stars. This period would have been fleeting, with earlier temperatures too hot for stable nuclei to exist, and later temperatures too cool for nuclear fusion. However, during this period some of the same reactions that occur in stars today should have taken place throughout the Universe. We should be able to see evidence of this early epoch of Big Bang Nucleosynthesis.

13.5.1.5: *Implication 5: Relic Radiation from the Hot Dense Phase*

The imprint of the hot, dense phase should still be with us in the form of relic radiation. A hot, dense plasma creates a radiation field that is in equilibrium with the matter, so both will have the same characteristically high temperature. As the Universe expands and cools the matter and radiation will eventually come out of equilibrium, but the radiation field will remain. It will cool as the Universe expands, but it should still be in evidence today. We should therefore be able to detect this early relic radiation in every direction we look.

13.5.1.6: *Implication 6: Continual Evolution*

The Universe is continuing to evolve as it expands and cools. The Big Bang model makes predictions about the future evolution of the Universe based upon its current conditions and its contents.

We will explore these implications in the remaining chapters. Each of them lends itself to some remarkably detailed investigations of the cosmos, and each plays an important part in our understanding of the history and evolution of the Universe.

GOING FURTHER 13.5: BIG BANG—WHAT'S IN A NAME?

The Big Bang Model

Watch the two animations, which each show a region of space and the matter in it. In both animations, the grid lines represent space, and the dots represent matter. You can toggle back and forth between the animations with a pull-down menu. The animation begins when you click on the center dot. Reset the animation using the "reset" button. Answer the following questions.

Play Activity

Density is how much matter is present in a given volume of space. Use the animation you found to be correct in question #8 along with the aperture tool to figure out how the density of the Universe changes over time. To do this:

- Reset the animation.
- Click aperture control and draw a circle; this represents the volume of space at some early time in the history of the Universe.
- Count the number of dots in the circle.
- Click animation control again. The circle should stay on the screen.
- Run the animation until it stops; this represents some later time in the history of the Universe.
- Again, count the number of dots in the circle.
- If the density is higher, there will be more matter (dots) in a given volume (circle). If the density is lower, there will be less matter (dots) in a given volume (circle).
- (To try again with a different aperture, click outside the circle and the original circle will be erased. You can then draw a new circle.)

We can also relate the change in density of the Universe over time to the change in its temperature over time. Have you ever used "canned air" to clean a computer keyboard or other electronics? The can contains a certain amount of air confined to a small volume. When you press the nozzle and release the gas, it is now spread out over a much larger space. Temperature is a measure of the kinetic energy of the molecules—how fast they are moving and how much they are colliding. When the volume is increased, the molecules will have many fewer collisions because they are spread out. This reduces the temperature and makes the can and the air around the can feel cooler. Although the Universe is not exactly a gas in the same way as what is coming from the canned air, the idea is similar: as the volume of space expands, there are fewer collisions and thus the temperature decreases.

13.5.2: Explaining How the Universe Changes Over Time With General Relativity

General relativity was not developed for cosmological reasons. Rather, Albert Einstein was interested in generalizing special relativity to deal with objects accelerating with respect to each other; recall that that special relativity deals only with objects that have constant relative velocities. He also wanted to explain why our experience of gravity can feel essentially indistinguishable from acceleration. Like all physical theories, however, general relativity applies, or should apply, throughout the Universe. So as soon as it was available it was applied to cosmology in order to understand the properties of the Universe on the largest scales of space and time.

General relativity deals with space as an entity (as opposed to an absence of objects) and it describes how space is warped by mass and energy. The mathematical description is difficult to use for complex distributions of mass and energy, for example, black holes and galaxy clusters. However, it is actually simpler for the Universe as a whole. For that case we can consider mass and energy to be smoothly distributed on scales of billions of light-years.

Fortunately for many models of the Universe, using general relativity is fairly straight-forward. Time is not warped at all and space is only subject to some simple universal stretching—just what we observe in Hubble expansion. For a homogeneous distribution of matter and energy throughout the Universe, we find that we can use a single time that is valid everywhere (often called cosmic time) together with a curved and stretching three-dimensional space. The Friedmann equations are the solution to the Einstein equations for a homogeneous and isotropic Universe. The primary Friedmann equation relates the expansion history of the Universe and the matter and energy density to the curvature of space:

$$H^2 - \frac{8\pi G\rho}{3} = \frac{kc^2}{S^2}$$

(expansion)−(gravity)=(curvature)

This is basically the equation of motion for space, much like Newton's laws give us the equation of motion for an object. The Friedmann equation tells us how the Universe as a whole changes over time. First, we have an expansion term: the Hubble parameter, (H) describes how fast the scale factor of space (S) changes with time ($H = \Delta S/\Delta t$). Next, we have a term for things that affect the dynamics of spacetime: the density of matter and energy (ρ), which affects the stretching of space through gravity. Finally, we have a term that includes the curvature of spacetime (k). This equation embodies the essence of the Einstein Equation: matter and energy affect how spacetime bends.

Again, the interplay between the terms of the Friedmann equation is analogous to the interplay between gravity and the energy of motion that we saw when discussing Newtonian physics. In Newtonian physics, if you throw a ball in the air or launch a rocket, whether or not it will escape into space depends on the relative values of its kinetic energy (like the expansion term in the Friedmann equation) and gravitational potential energy (like the gravity term). The curvature term is like the total energy in a Newtonian system. The rate at which the Universe will expand (or contract) over time will depend on the relative values of the Hubble constant and the density of matter and energy. General relativity tells us that these parameters affect the curvature of space. Over the next several chapters we will see how the evolution of the Universe as a whole plays out according to the Friedmann equation, and how theory compares with observations.

As an example, we will do a short calculation that is analogous to the one we did for escape velocity in Math Exploration 13.2.

Math Exploration 13.2

The value of the density that causes the Universe to be flat (zero curvature) is called ρ_{crit}, the critical density. It depends on the expansion rate, and that as the expansion rate increases the critical density also increases. The condition is analogous to saying (from a Newtonian perspective) that the escape speed from a planet (analogous to the expansion rate) gets bigger as the mass of the planet (analogous to the density of the Universe) gets bigger. For a Hubble constant around 70 km/s/Mpc, the critical density is around 10^{-29} g/cm^3. If the density is higher than the critical value, then the expansion will not be fast enough; the Universe will eventually stop expanding and then re-collapse. This case is analogous to throwing a ball upward from a planet at less than the escape speed, in which case we know that the ball eventually falls back down. If the density is less than the critical density then the expansion of the Universe is so fast that even after an infinite amount of time the Universe will still be expanding—a case analogous to launching a ball upward with a speed greater than the escape speed. The critical density is therefore the boundary case in which the Universe expands forever but approaches zero expansion rate with time. This is just like launching a projectile at exactly the escape velocity. We will explore these ideas in much greater detail in later chapters.

13.5.3: Summary

In this chapter we have explored the most important discovery we have made about the Universe in the last century, that it is stretching and appears to have started from a hot, high-density state nearly 14 billion years ago. The space we occupy is a thing with properties of its own; it is not just an empty container in which events take place. The Universe is not matter exploding from some special place into some pre-existing structure, and space might be infinite, or might be finite but unbounded. In either case it has no center or edge, and we see only a fraction of all the space there is.

Perhaps the most important thing to realize about the Big Bang theory is that it provides a framework that connects many seemingly unrelated observations and physical principles. In this chapter we have seen our first piece of evidence: the Hubble expansion and some of its implications. In the remaining chapters, we will work through other pieces of evidence and how they help us understand the history and fate of the Universe. We will see how each piece of evidence is related to other pieces of evidence, so the elements of the model fit together and support each other. Given the huge amount and variety of data, it is remarkable that it is possible at all to fit together these various observations in a self-consistent way.

One final thing to keep in mind is that the Big Bang model does not say anything about what caused or created the Universe. It says that at early times the Universe was hot and dense, and that it has expanded and cooled since then, evolving as it does so according to the laws of physics. It does not say anything about how the Universe got to be in its initial state.

Going Further 13.6: the Steady State Theory—a Model Without a Beginning of Timeote

Does a Universe expanding according to Hubble's law require the Universe to extrapolate back to a beginning time? Not necessarily. In the 1950s, Fred Hoyle, Hermann Bondi (1919–2005) and Tommy Gold (1920–2004) developed a Steady State theory in which the Universe was eternal and unchanging. It required the continuous generation of matter everywhere to drive the observed expansion. The continuous emergence of new matter would constantly refresh the Universe, allowing new stars and galaxies to form without end. That is a very different Universe than that predicted by the Big Bang theory. The Big Bang predicts a period of rich star formation at some early time and a slow depletion of material available for new stars as the Universe evolves. In other words, the Universe should change in time.

By the late 1950s, Steady State models were already in serious trouble because observations were showing that the early Universe was very different from the Universe today. Recall that when we look at very distant objects we are seeing them as they were long ago when the light we measure left them, a time when the Universe was much younger than it is now. Observations of the early Universe (distant objects) showed that the density of galaxies was higher than today. What's more, the galaxies were both more active and different in structure than galaxies are now. In addition, the chemical abundance of all the lightest elements in the Universe seemed close to the predictions for the Big Bang theory, an idea we will explore in Chapter 14. In 1965, the Cosmic Microwave Background (CMB) was detected. Big Bang models had predicted just such a glow, coming from the time when the Universe was much denser and hotter than it is today. Steady State models, on the other hand, had no natural explanation for the background radiation, and Steady State accounts of the CMB seemed contrived, created only to fit the observations. From 1965 onward there has been a strong scientific consensus that the expansion we see implies that the Universe started in an ultra-dense, hot state.

This page titled 13.5: The Basic Big Bang Model is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.6: Wrapping It Up 13 - How Well Do We Know the Expansion Rate and Age of the Universe?

Learning Objectives

- You will be able to put everything together to demonstrate your understanding of the expansion and age of the Universe, including determining the accuracy of your measurements for these two important parameters.

You have already determined a best fit value for the expansion rate and age of the Universe based on your measurements of galactic distances and speeds. In this activity, you will use your data to determine how well you know the expansion rate and age. Make sure you have your data available in the *Graphing Tool*.

13.6.1: Part I: Accuracy of the Hubble Constant

13.6.1.1: A. Your Data

You may have noticed when you calculated your best fit line that it does not go through all of the data points on your graph. This is the nature of data; there is always some spread.

1. Plot the best fit line through (0,0) again. Recall that the slope is the Hubble constant, H_0. What is your best fit value for H_0?

km/s/Mpc

One simple way to estimate the accuracy of the value of H_0 is to draw the steepest reasonable line and the shallowest reasonable line on the graph, and measure their slopes. To do this using the *Graphing Tool*, click on the tab for Draw Line and adjust the slider bar for the slope.

2. What is the maximum value you get for the value of H_0?

3. What is the minimum value you get for the value of H_0?

You can express how well you know the value of your measurement in terms of the best fit value ± the accuracy. You calculate the "+" number by subtracting the best fit from the maximum value and you calculate the "–" number by subtracting the minimum value from the best fit. For example, if your measurement of the Hubble constant was 70 ± 10 km/s/Mpc, the minimum would be 60 km/s/Mpc and the maximum would be 80 km/s/Mpc.

4. What is the best fit value ± the accuracy of the Hubble constant for your data set?

You could also express how well you know the value of your measurement in terms of ± a percent. To do this, take your accuracy ("±" number) and divide it by the best fit value. For example, if your measurement of the Hubble constant was 70 ± 10 km/s/Mpc, the accuracy expressed in terms of a percent would be: 10/70 = ± 14%.

5. Express the accuracy of your measurement in terms of a percent.

13.6.1.2: B. Compare to the HST Key Project

The HST Key Project team measured a value of 72 ± 8 km/s/Mpc.

1. What is the maximum value of H_0 according to the HST team?

km/s/Mpc

2. What is the minimum value of H_0 according to the HST team?

km/s/Mpc

3. Does the range of values measured by the HST team overlap with the range that you measured? If the ranges overlap, the measurements are said to be consistent. Explain what factors may or may not contribute to a discrepancy.

4. Who measured the value of H_0 more accurately, the HST team or you? What factors in the *data* may have contributed to a discrepancy? (You are not allowed to use the professional or amateur status of the researchers as a reason.)

13.6.2: **Part II: Accuracy of the Age of the Universe**

13.6.2.1: A. Your Data

1. What is the age of the Universe from your best fit value of H_0 (in billions of years; you calculated this previously)?

billion years

2. How old would the Universe be for your maximum value of H_0 (in billions of years)?

3. How old would the Universe be for your minimum value of H_0 (in billions of years)?

4. Based on your data, does a bigger Hubble constant (faster expansion) correspond to a bigger (older) or smaller (younger) age of the Universe?

a. Bigger (older)

b. Smaller (younger)

5. What is the final best fit value and accuracy for the age of the Universe from your data? (For example, if you get a best fit age of the Universe of 14 billion years, with a range from 13 billion to 15 billion, you could express this as 14 ± 1 billion years.)

6. What is the percent accuracy for your measurement of the age of the Universe? (For example, if your best fit age is 14 ± 1 billion years then your accuracy expressed as a percent is: 1/14 = 7%.)

13.6.2.2: B. Compare to the HST Key Project

The HST Key Project team’s measurement of 72 ± 8 km/s/Mpc corresponds to an age of the Universe of $13.6^{+1.7-1.3}$ billion years.

1. What is the maximum value of t according to the HST team?

billion years

2. What is the minimum value of t according to the HST team?

billion years

3. Does the range of values measured by the HST team overlap with the range that you measured? If the ranges overlap, the measurements are said to be consistent. Explain what factors may or may not contribute to a discrepancy.

13.6.2.3: C. Comparing to Other Observations

Measurements that include the HST data, plus other lines of evidence, yield an age of the Universe of 13.8 ± 0.1 billion years.

1. Does this overlap with your range of values for the age? If so, your measurement is said to be consistent with this one.

a. Yes

b. No

2. How does the accuracy of your measurement compare to this one? What factors in the *data* may contribute to the difference?

This page titled 13.6: Wrapping It Up 13 - How Well Do We Know the Expansion Rate and Age of the Universe? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

13.7: Mission Report 13 - How Well Do We Know the Expansion Rate and Age of the Universe?

. Please rate on a scale of 1 to 5 how effective you think the “Wrapping It Up” activity for this chapter was in helping you understand the material. (1: not effective at all → 5: very effective)

*

1

2

3

4

5

B. What were the main ideas that you learned in conducting the “Wrapping It Up” activity for this chapter? Be specific and detailed in your response. Please address the following questions: What did you learn? How did you learn it? What is still unclear? (At least 150–200 words.)

*

C. If the “Wrapping It Up” activity for this chapter included measurements or data, please describe what factors influenced the accuracy of your results. (Do *not* include mistakes, only unavoidable measurement imprecision.) If you obtained any numerical values for the accuracy of your measurements during the activity, note those here. If there were no measurements or data, say so explicitly.

*

D. Questions to be graded for accuracy:

1. Would the Universe be older if the expansion rate was H_0 = 50 km/s/Mpc or if H_0 = 75 km/s/Mpc? Explain.

*

2. What is the age of the Universe if the Hubble constant is measured to be 75 km/s/Mpc? Express your answer in billions of years. Would this value be consistent with the range for the Hubble time from the HST Key Project? Explain.

*

3. On average, how do the velocities of far away galaxies compare to those of closer galaxies? Why?

*

4. Based on the measurements you made in this chapter, how do you know that the Universe is expanding (stretching) but the individual galaxies are not?

*

This page titled 13.7: Mission Report 13 - How Well Do We Know the Expansion Rate and Age of the Universe? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

14: The Growth of Structure

Chapter 14 traces the evolution of large-scale structure in the Universe back to the first generation of stars and galaxies. Observations are compared to computer models for the formation of structure at different hierarchical levels, including filaments and voids, clusters and groups of galaxies, and individual galaxies. The chapter concludes with an examination of the formation of the first stars and galaxies.

This page titled 14: The Growth of Structure is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

14.0: The Growth of Structure Introduction

Video Transcript

> ***Mapping Our Universe: Transcript***
>
> *Our story begins on a blue planet called Earth, in the United States, in southern New Mexico; home of Apache Point observatory, site of the Sloan Digital Sky Survey.*
>
> *Inside this structure is two-and-a-half meter telescope which is being used to construct the largest map of the Universe ever made. When the survey is completed in the next few years, it will provide a three-dimensional atlas of nearly a million galaxies. Let's take a look at what they have found so far.*
>
> *We begin our journey looking back at the Sun. Moving outward, we see thousands of nearby stars, many of them just like our own Sun. As we continue to pull away from the Sun, we see that these stars are part of a much larger spiral-shaped collection of a hundred billion stars, a galaxy called the Milky Way.*
>
> *Moving outward from the Milky Way, we begin to see the nearest galaxies mapped by the Sloan Digital Sky Survey. Galaxies come in a variety of shapes, sizes, and colors, from typical blue spiral galaxies like our own Milky Way to giant red elliptical galaxies.*
>
> *Now hundreds of millions of light-years from Earth, we see that galaxies are not randomly distributed throughout the Universe; they collect together in groups of different sizes ranging from clusters of hundreds of galaxies to huge web-like structures stretching across hundreds of millions of light-years.*
>
> *The survey is being carried out in thin slices across the sky like pieces of a watermelon. When completed, these slices will merge together to form a three-dimensional map of the Universe.*
>
> *Beyond the Sloan Survey, we reach the cosmic microwave background radiation as mapped by NASA's WMAP satellite. This radiation gives us a picture of the temperature of the temperature Universe when it was only 400,000 years old.*
>
> *By combining these observations, cosmologists are closing in on a consistent picture of how the Universe evolved from its earliest moments to the present day.*

Whether you live in a big city, a small town, or in the countryside, a view of the stars presents a reminder of the vast Universe in which we reside. You may be able to pick out some constellations, point in the direction of star clusters or even view the Milky Way stretch across the sky. It is human nature to try to understand the pattern of stars, grouping them into familiar constellations and inventing clever stories to explain the shapes we see.

With bigger and better telescopes, our view of the Universe has grown past these individual stars to the galaxies beyond. Of the billions of galaxies scattered throughout the observable Universe, some look similar to our own while others look quite different. On the largest scales, the Universe looks like a giant cosmic web with bright spots along stringy filaments, flat walls, and large bubbles of nearly empty space.

In the opening video we are introduced to the Sloan Digital Sky Survey, one of several surveys that have been mapping millions of galaxies and other objects. In the video, we take fly-through of the Universe as revealed by Sloan. We start from our vantage point at Earth, and then we visit galaxies, clusters, filaments, and places nearly devoid of galaxies. The survey is done in slices, and the completely blank areas are slices yet to be mapped. As we go farther out in into space we are also seeing farther back in time, almost to the edge of our observable Universe. This edge is depicted as a sphere centered on Earth. In the last few frames of the video, we see light from a time before galaxies formed, the cosmic microwave background. One thing that the video makes clear: the Universe of long ago is not identical to the Universe we live in now. There have been noticeable changes to its mean properties over its lifetime.

In this chapter we will see how the gravitational attraction of dark matter gave rise to galaxies, galaxy clusters, and the largest structures of the Universe. Combining powerful telescopic surveys and massive computer simulations based on the laws of physics, astronomers have created a detailed picture of the known Universe.

This page titled 14.0: The Growth of Structure Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

14.1: Large Scale Structure

Learning Objectives

- You will know how galaxy surveys are done
- You will understand visualizations of galaxy survey data
- You will know how galaxies are distributed: that the large-scale structure of the Universe is web-like or sponge-like
- You will understand that structure grows via gravitational attraction
- You will be able to compare simulations with various cosmological parameters and data to determine which model best describes the data
- You will be able to determine the matter density of the Universe

What Do You Think: Where Are the Galaxies?

14.1.1: Observations: Galaxy Surveys

The ability to measure distances is a critical step to developing an understanding the structure of the Universe. With the exception of a few very nearby galactic neighbors, every star we see with the unaided eye is within our own spiral galaxy, the Milky Way. We measure the distances to stars inside the Galaxy using stellar parallax or main sequence fitting. To measure the distances to nearby galaxies, we can use Cepheid variable stars as standard candles. For more distant galaxies, we use supernovae and a few other methods. These have all been discussed in Chapter 6: Measuring Cosmic Distances.

If you live in or have visited the Southern Hemisphere you may have seen the Small and Large Magellenic Clouds, our nearest galactic neighbors. From modestly dark areas of the Northern Hemisphere, if you have good eyes you can see another nearby object, the Andromeda Galaxy. These are just a few members of the Local Group, a collection of 40 - 50 gravitationally bound galaxies including the Milky Way. The Group is about 10 million light-years in diameter (Figure 14.1). This group itself is part of the Virgo Supercluster, a large grouping of galaxies containing around 100 such groups and spanning approximately 100 million light-years (Figure 14.2).

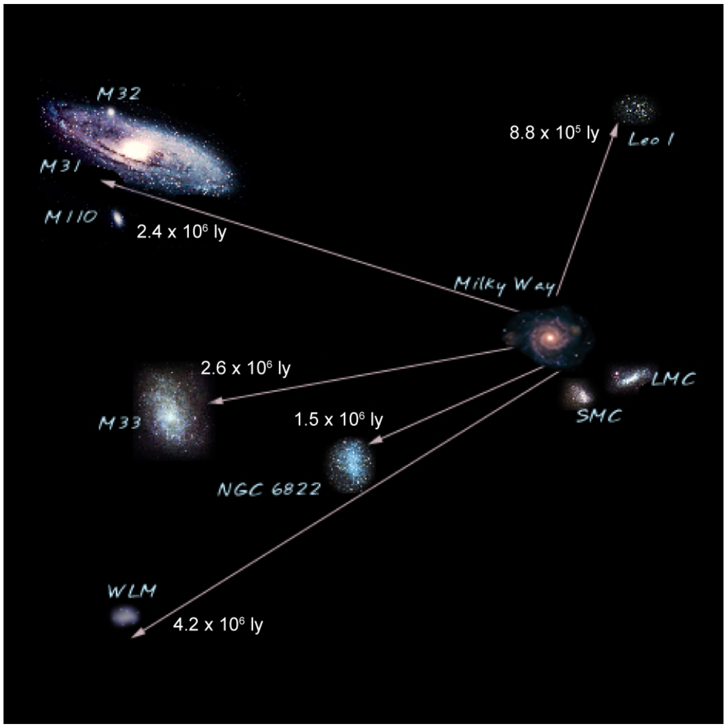


Figure 14.1: This image of the Local Group is a composite of real images of the actual galaxies placed in approximately the correct orientation but with distances between galaxies not to scale. Credit: NASA/SSU/Aurore Simonnet based on a figure by NASA/HEASARC/Maggie Masetti

We have not always known what galaxies are. This is because astronomers did not always know their distances, as we have discussed in Chapter 13: The Expansion of the Universe. For a long time, astronomers lacked any techniques that could be used to measure the distances to them. In the early 18th century, it was clear the Milky Way could be resolved by telescopes into individual stars; however, some astronomers debated whether or not other faint objects were within or beyond its reaches. In the late 18th century, William Herschel (1738 - 1822) and his sister Caroline Herschel (1750 - 1848) mapped the distances to stars in multiple directions based on a simple assumption; all stars are of the same intrinsic brightness. Today we know this is a grossly flawed premise. The intrinsic brightness of stars (their luminosity) ranges from a thousand times dimmer than the Sun to a million times brighter than the Sun. Nonetheless, the Herschels made progress on one idea that has been shown to be true: we live within a flattened disk of stars (Figure 14.3).

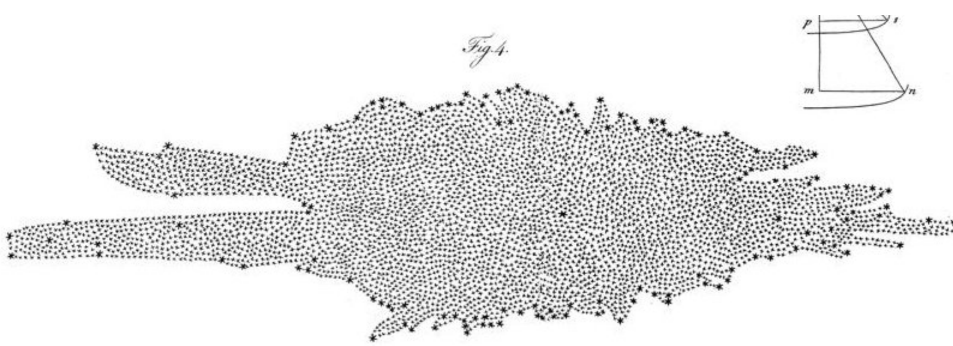


Figure 14.3: Herschel's drawing shows a section of the Milky Way from his 1785 paper, "On the Construction of the Heavens." Credit: Herschel, W. (1785) *Philosophical Transactions of the Royal Society of London*, 75, 213

It was another 130 years before Jacobus Kapteyn (1851 - 1922) could estimate distances to stars with statistical analysis of stellar parallaxes and proper motions of nearby stars. While the methods used by Kapteyn neglected the effects of intermediate gas and dust blocking light from distant stars, he did succeed in confirming the Sun is not in the center of the Milky Way. He also provided astronomers with the scale and extent of the Galaxy. The **Kapteyn Model** published in 1922 depicts the Milky Way as a flattened disk approximately 50,000 light-years in diameter and about 10,000 light-years thick. Today we know it to be 100,000 light-years in diameter and about 3,000 light-years thick.

Even with the size of the Galaxy known, there was still disagreement over other objects. In 1781, Charles Messier (1730-1817) had assembled his Messier Catalog of 110 celestial objects, a favorite among amateur sky watchers because of the accessibility of its objects to small telescopes. Initially developed to distinguish permanent objects from transient comets, this catalog contained numerous diffuse objects distinguishable from solitary stars. As the resolving power of telescopes increased, astronomers could identify individual stars making up many of objects, all called nebulae at that time. With this, the distance to Messier objects became hotly debated. Are they objects within the Milky Way or are they something beyond? The only way to settle this argument was to determine accurate celestial distances.

In 1925, Edwin Hubble was the first to show that one of the dim objects in this catalog, Andromeda, was actually a neighboring galaxy. Hubble did this using the Cepheid period-luminosity relationship that had recently been developed by Henrietta Leavitt - see the discussion in Chapter 13.1. We now know that some of the objects in Messier's catalog are other galaxies (outside our Galaxy), while others are star clusters and nebulae within the Milky Way.

As we have already seen, Hubble's Law relates the recessional velocity of a galaxy to its distance.

$$v = H_0 d$$

In this expression, v is the velocity of a galaxy, d is its distance, and H_0 is a constant of proportionality called the Hubble constant. This relationship is a consequence of the expansion of space, and it causes a cosmological redshift. The farther away a galaxy is, the faster its velocity and the greater its redshift. The relationship between velocity and cosmological redshift are given the equation below, as discussed earlier.

$$v = cz$$

In this expression, v is the velocity of a galaxy, z is its redshift, and c is the speed of light. This is valid when redshifts are small. Combining these two equations, we get:

$$cz = H_0 d$$

We can rearrange terms to obtain an expression for distance in terms of redshift:

$$d = \frac{cz}{H_0}$$

Once we have determined the value of the Hubble constant by using a sample of galaxies (as Hubble did), we can then turn around and use the relationship to determine the distance to any galaxy. We need only measure the redshift of the target galaxy! You will practice this technique in the next activity. This method does not give extremely accurate distance determinations for individual galaxies because there is a lot of scatter in the Hubble relation. Nonetheless, it does give us an idea of the distance to any galaxy from which we are able to obtain a spectrum.

Redshift and Distance

In this activity you will assume a value for the Hubble constant and use that to calculate the distances to galaxies given their redshifts.

Worked Example:

1. A galaxy has a redshift of 0.02. Assume the Hubble constant is 70 km/s/Mpc. What is the distance to this galaxy?

- Given: z = 0.02, H_0 = 70 km/s/Mpc, c = 3e5 km/s
- Find: d
- Concept: d = cz/H_0
- Solution: d = (3e5 ~~km/s~~) (0.02) / (70 ~~km/s~~/Mpc) = 85.7 Mpc

Questions:

1.

https://phys.libretexts.org/@go/page/31433

2.

3.

In the previous activity you learned how to compute a galaxy's distance from its redshift. To measure a galaxy's redshift, it is only necessary to take its spectrum. This is a much less labor intensive way to measure the distances to galaxies than other techniques. Redshift surveys have allowed astronomers to map the Universe in three dimensions on a massive scale.

Building a comprehensive map of distant objects requires broad surveys of galaxies near and far all over the sky. Modern surveys like the Sloan Digital Sky Survey (SDSS), the 2 Micron All Sky Survey (2MASS) (Figure 14.4), and the 6dF have measured redshifts for millions of galaxies.

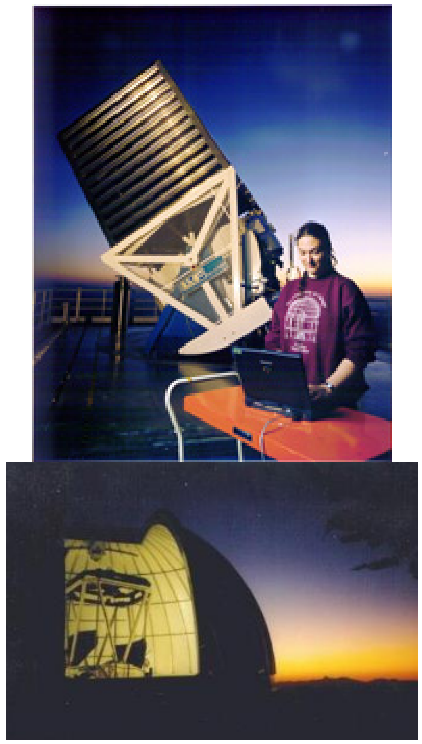


Figure 14.4: Pictures of SDSS (left) and 2MASS (right) telescopes, which carry out galaxy surveys. The SDSS telescope, which maps the sky in optical (visible) light, is located at the Apache Point Observatory in New Mexico. Prof. Connie Rockosi of UC Santa Cruz was part of the team that built the camera for the survey. The 2MASS telescope pictured here is located at Mt. Hopkins in Arizona; there is a second telescope located in Chile. Both of the 2MASS telescopes map the sky in infrared light . Credit: Image courtesy of Sloan Digital Sky Survey; Image courtesy of 2MASS/UMass/IPAC-Caltech/NASA/NSF

The SDSS canvassed more than 25% of the night sky for 8 years from 2000 to 2008. Since 2008, SDSS has continued to do astronomical observations with specific scientific goals that include deeper studies of the Milky Way galaxy, exoplanets, and cosmological measurements. To date (through 2013), SDSS has gathered multi-color images and spectra of over 1.8 million galaxies, and for more than 300,000 quasars. SDSS uses a 120-megapixel camera; it can snap a picture 8 times larger than the full Moon and record spectra for more than 600 objects at a time. This is an enormous amount of data. On a typical day, SDSS could process over 12 terabytes, or 12,000 gigabytes, of data.

Dust and gas in the plane of the Milky Way obscures the views of optical surveys like SDSS. To provide an unprecedented view into this "zone of avoidance," 2MASS charted the sky in dust-penetrating infrared light. Observations with 2MASS were taken between 1997 and 2001 in Arizona and Chile. Throughout its run, the survey collected nearly 25 terabytes of data. After image processing, the final database of 12 terabytes has been made available to the public.

With these surveys we see that on the largest observable scales, galaxies and clusters are found in thin, filamentary web-like structures forming walls with large bubbles mostly devoid of galaxies (Figure 14.5). Shown on top is one slice of galaxies in the SDSS survey. Earth is at the center of this map, and the distance away from us for these objects is represented here using redshift, z. Shown on the bottom is an all-sky map from 2MASS.

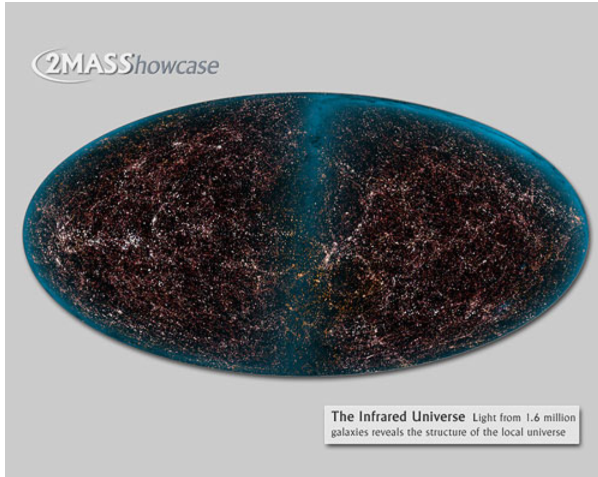


Figure 14.5: (Top) This is a slice through the 3-dimensional map of the distribution of galaxies from the SDSS survey. Earth is at the center and each point represents a galaxy, typically containing about 100 billion stars. Galaxies are colored according to ages of their stars. Red is old and green is young. (Bottom) The 2MASS all-sky catalog. Again, each point is a galaxy. The blue haze is the area obscured by our Galaxy. Both SDSS and 2MASS reveal that the large-scale structure of our Universe resembles a web. Credit: M. Blanton and the Sloan Digital Sky Survey; 2MASS/T.H. Jarrett, J. Carpenter, & R. Hurt

The Sloan Digital Sky Survey

In this video, you will see a rotating view of a three-dimensional plot of the galaxies in SDSS as of 2004. Each dot in the video is a galaxy. They survey is being conducted in large wedges looking out from Earth. As more data are collected the space between the slices will start to fill up.

1.

(A) (B)

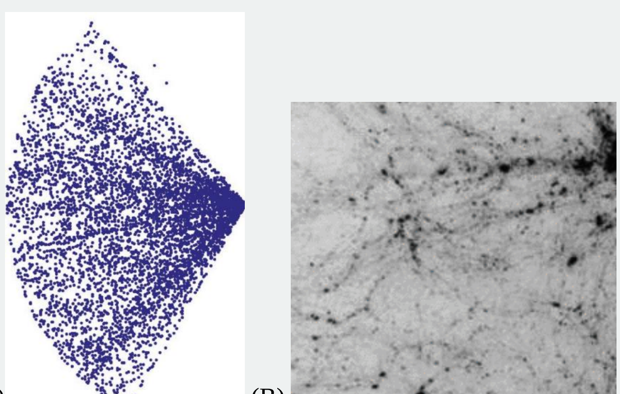
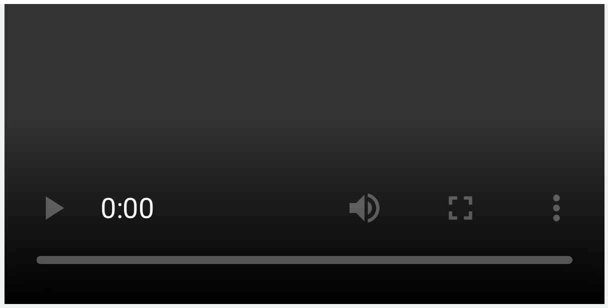


Figure 14.A.1: Some ideas for what the large-scale structure might look like. What do you think? Credit: Daniel Smith, South Carolina State University; Zoe Buck



Credit: Sloan Digital Sky Survey

After you have watched the video, answer the following questions:

2.

3.

4.

5.

6.

7.

From SDSS data, we can identify several features. Galaxy filaments are massive string like formations typically 150 - 300 million light-years long. Galaxy nodes sit at the intersection of galaxy filaments. Sheets of galaxy filaments form massive walls which in turn outline the bubble-like voids. The statistics of voids, including their size and distribution, are closely tied to cosmological parameters and the detailed physics of structure formation. Surveys and simulations show that the voids tend to be around 100 million light-years in diameter. Estimates put the total volume of voids at roughly 40% of the total volume of the Universe.

Large Scales of the Universe

Play Activity

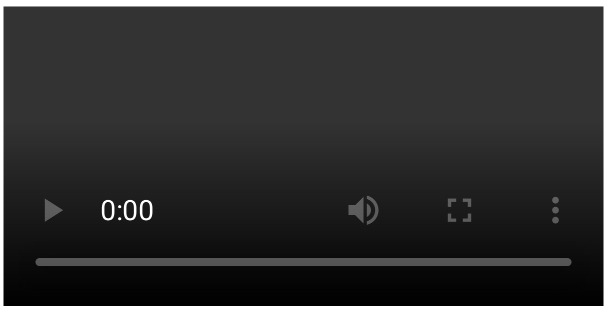
In the early part of the 20th century, scientists like Einstein were using the theory of general relativity to describe the behavior of the Universe. Astronomers studying the Universe made a simplifying assumption that is now known as the cosmological principle.

It states:

On the largest cosmic scales, the Universe is both homogeneous and isotropic.

Homogeneity means that there is no preferred location in the Universe. That is, no matter where you are in the Universe, if you look at the Universe, it will look the same. Isotropy means that there is no preferred direction in the Universe. That is, from your current location, no matter which direction you look, the Universe will look the same. When we say “the same”, we do not mean identical. Rather, we mean that if we analyze statistical properties that describe how matter is distributed, we will get the same results in any direction and at any location (when we analyze large enough areas). Can we test if this is true?

At first glance, diagrams like Figure 14.5 of the distribution of galaxies in the Universe seem to imply that the Universe is not homogeneous and isotropic. They make it seem that galaxies in one direction are not distributed in the same way as the galaxies in another direction. However, the galaxies in the SDSS slice shown only extend to a redshift of z ~ 0.15 and in the 2MASS map to z ~ 0.1. As we study objects to larger distances, the structure does smooth out and become more homogeneous on larger scales. In Animated Figure 14.6, we see layers of the 2MASS survey at increasing redshift that show greater homogeneity with increasing distance. When statistical analyses are made on luminous galaxies and quasars in the SDSS catalog out to redshifts of ~0.6 and 1.8 respectively, the Universe is found to be homogeneous on scales of 400 – 600 million light-years and greater. Thus, we currently find support for the cosmological principle in the distribution of galaxies in the Universe.



Animated Figure 14.6: This movie shows layers of the 2MASS survey at increasing redshift, or equivalently, increasing distance. On larger scales, the distribution of matter becomes more homogeneous. Credit: Movie courtesy of 2MASS/UMass/IPAC-Caltech/NASA/NSF

GOING FURTHER 14.1: LARGE QUASAR GROUPS

14.1.2: The Formation of the Largest Structures

Galaxies, clusters of galaxies, and super-clusters of galaxies make up the large-scale structures that we observe as part of the cosmic web. But how did these structures form?

The Universe is expanding, consistent with the Big Bang theory: the Universe was much hotter and denser in the past, and it changes as it expands and cools. Structure formation happens in this context. A complete cosmological framework should be able to explain the structures we see, not just the fact that the Universe is expanding, but that it contains galaxies and other structures. Furthermore, the model should naturally connect the early epochs in the history of the Universe to current conditions in a natural way.

Structure formation models therefore must begin when the Universe was a hot, dense plasma that was much more uniform than the lumpy Universe we see today. Initially tiny fluctuations present in the early Universe grew by gravitational attraction into the magnificent large-scale structures that we observe in galaxy surveys. We should see signs of this structure in all phases of the Universe's evolution, for instance, in the Cosmic Background Radiation.

When the Universe was about 500,000 years old, the density of matter in a region of space that would eventually form a galaxy was about 0.5% higher than the density of its surroundings. At that time, denser regions of space expanded more slowly than other regions of space because the pull of gravity was higher in the regions with a higher mass density. As a result, denser regions became even denser compared to surrounding regions. When the Universe was ~15 million years old, dense regions, like the one that eventually formed the Milky Way galaxy, were about 5% denser than surrounding regions. This enhancement continued. By the time the Universe was about a billion years old, typical collections of matter were about a factor of two denser than their surroundings. At the present time, the disk of the Milky Way is about a factor of 1000 denser than its halo, which is in turn about a factor of 1000 denser than the Universe on average.

At first, when the density was nearly uniform everywhere, with only small variations (10^{-4} to 10^{-5}), gravitational collapse proceeded in a straight-forward linear manner. As more matter collected in the denser regions, eventually the fluctuations became equal to the average density and the gravitational collapse became non-linear. The scenarios for how structure could have formed in this gravitational attraction framework are modeled with the aid of computers. One class of models works well for the so-called linear regime, when fluctuations were still small. These models start when the Universe was about 10,000 years old and continue through the formation of the cosmic microwave background at 380,000 years. They are able to generate statistics for the distribution of galaxies on the largest scales, and these can be compared to observations of the real universe. The second class of models, which will be the focus of this chapter, explores the non-linear regime of larger scale structure formation that follows.

The computer simulation in Figure 14.7 illustrates how large-scale structures form over time. They show an enhancement of density as time progresses, a result of gravitational attraction. Computer simulations of the growth of large-scale structure include the laws of physics, such as gravity and thermodynamics, and track the motions and energy of cold dark matter and regular (baryonic) matter. Most of the mass of the Universe is cold dark matter, not baryonic matter, so structure formation is dominated by dark matter; baryonic matter is attracted to and enhances the structures initially created mainly by the clumping of dark matter.

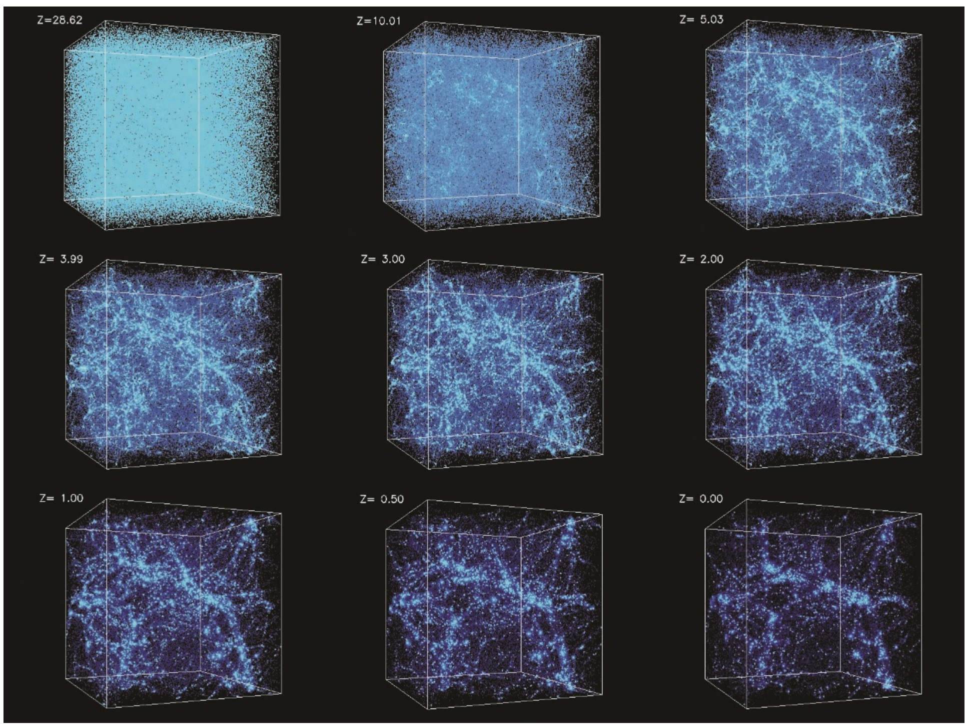


Figure 14.7: Simulation of the growth of large-scale structures. The Universe gets lumpier as time progresses due to gravitational attraction. This simulation ran from a redshift of z = 30, when the age of the Universe was about 100 million years old, to today (z = 0). The size of the box is 140 million light-years wide today. The coordinates are co-moving with the expansion of the Universe so that the same patch of sky is shown in each box, even though the size of the box is much larger at recent times. Credit: Simulations were performed at the National Center for Supercomputer Applications (NCSA) by Andrey Kravtsov (The University of Chicago) and Anatoly Klypin (New Mexico State University). Visualizations by Andrey Kravtsov.

How do the large structures of the cosmic web form if the key driving force is the gravitational pull of dark matter? Since gravity pulls inward in all directions, you might at first guess that large clouds of matter would remain spherical as they collapse. In fact, we find many other shapes on the largest scales. In the next activity, we will consider what happens to the more general shape of an ellipsoid.

Collapse of Clouds of Matter

A roughly spherical cloud of matter can be perturbed such that each of three axes is a slightly different length. The cloud will then be shaped like an under-inflated football that someone squished. The football-shaped object has a long axis (c) that passes through the pointy ends and two additional axes, one through the squashed face (b) and another perpendicular to it (a) as shown in Figure A.14.2.

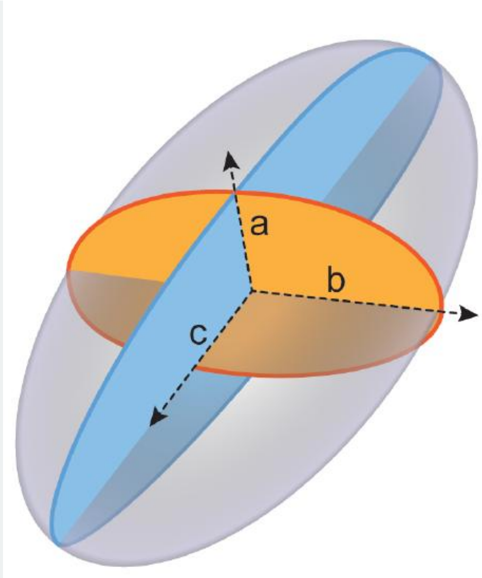


Figure A.14.2: An ellipsoidal cloud of matter. The shortest axis collapses faster than the rest, forming a disk. The next shortest axis collapses forming a filament. Finally, the third axis can collapse to form a node. Credit: NASA/SSU/Aurore Simonnet

The gravitational force between any two objects with mass is given by Newton's Law of Gravitation.

$$F = \frac{GM_1M_2}{r^2}$$

In this equation, r is the distance between two objects of mass M_1 and M_2. For illustrative purposes, we will consider four identical clumps of matter: one sitting at the center of Figure A.14.2 and one at the end of each axis, *a*, *b*, and *c* which we will call clumps *A*, *B*, and *C*, respectively.

Questions

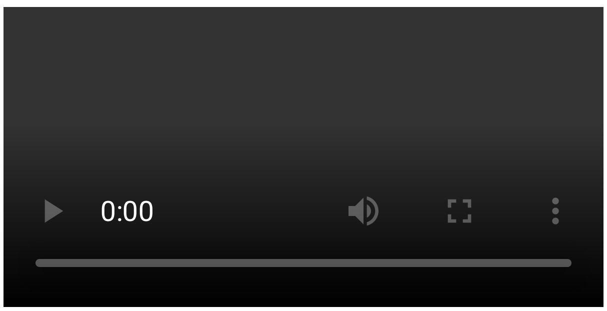
1.

2.

3.

From this activity we see that clumps of matter along the shortest axis will accelerate faster toward the center than matter along the largest axis. If we watch a simplified movie of gravitational collapse, this cloud of matter will first collapse from an ellipsoid to a pancake, then a rope, and finally a condensed ball. In this way, matter in the Universe formed into numerous massive sheets (walls), many of which further collapsed into cylinders (filaments), and within these large filaments, or at their intersections, we finally get compact nodes. We can get this great variety of shapes even without considering rotation.

Animated Figure 14.8 is a computer simulation showing the formation of the cosmic web. The simulation begins with a nearly uniform volume of gas at high redshift and runs through the present day. We then fly though the survey, zooming in on structures that might resemble the Local group and Virgo cluster. This simulation, and others like it, include the laws of physics, such as gravity and thermodynamics. They model the evolution of both regular and dark matter. The output of the simulations closely matches the cosmic web pattern observed in galaxy surveys.



Animated Figure 14.8: This simulation by the CLUES Team shows the formation of large-scale structure in the Universe and then zooms in on more local structures. The box size is 275 light-years in co-moving coordinates. This movie shows the evolution of dark matter, though the simulation also includes baryonic matter. Credit: CLUES team and C. Henze, N. McCurdy, J. Primack

Using these models, we can probe various cosmological parameters, or aspects in the models that tell us about the nature of the Universe. Model parameters include the amount and nature of the matter and energy in the Universe and the types and strength of initial fluctuations. For example, each panel in Figure 14.9 shows a different pattern for the distribution of galaxies because each uses a different set of parameters. By comparing simulations of differing parameters with data from galaxy surveys we can determine the best-fit values of the parameters; we can infer that the models with the best fit to the data tell us the properties of our Universe.

Of course, there are often many ways to adjust model parameters to obtain similar fits because different parameters can counteract one another. For this reason, it is not always straightforward to make conclusions based upon a single model, or even several models. Nevertheless, as our models increase in their level of sophistication, and different sets of models by different researchers give us complementary and consistent ways to understand the data, our confidence increases that many aspects of our views of the growth of structure must be correct, at least in general outline.

In the next activity you will run a simple cosmological simulation with only two parameters and find a best-fit match to SDSS data.

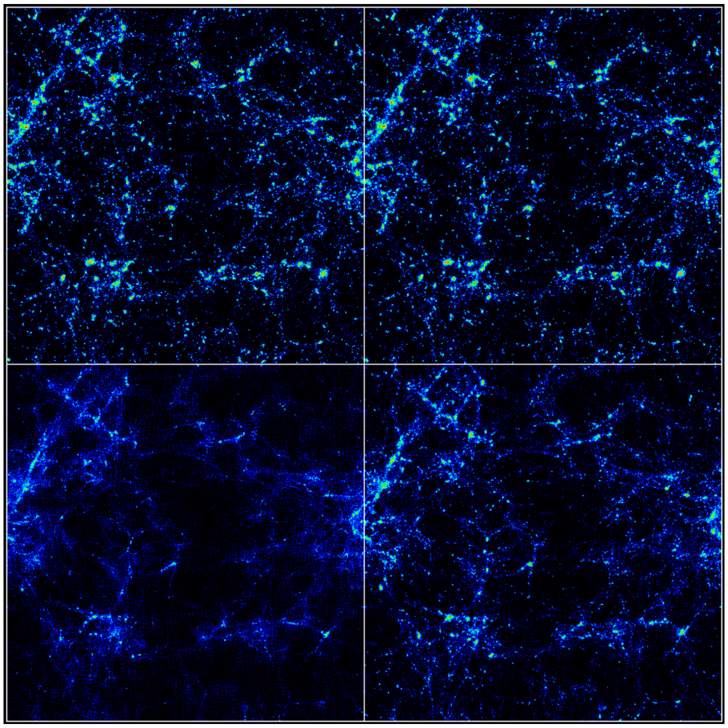


Figure 14.5, in order to determine the best fit values of the parameters. Credit: C. P. Ma, University of California, Berkeley

Structure Formation and Cosmological Parameters

In this activity, you are going to match the output of a large-scale structure simulation to data from galaxy surveys in order to determine the value of two cosmological parameters: the density of matter (Ω_m) and the amplitude of fluctuations (σ_8).

- The matter density, here written as Ω_m (Greek letter “Omega”), is the ratio between the matter density to the critical density; Ω_m is a number between 0 and 1.
- The amplitude of fluctuations, here written as σ_8 (Greek letter “sigma”), is a measure of how strong the fluctuations are on scales of 8 Mpc (26 million light-years).

Now go to the simulation website:

Play Activity

Click "Launch Interactive Lab" (on the picture in the middle of the page).

A. The effect of changing parameters.

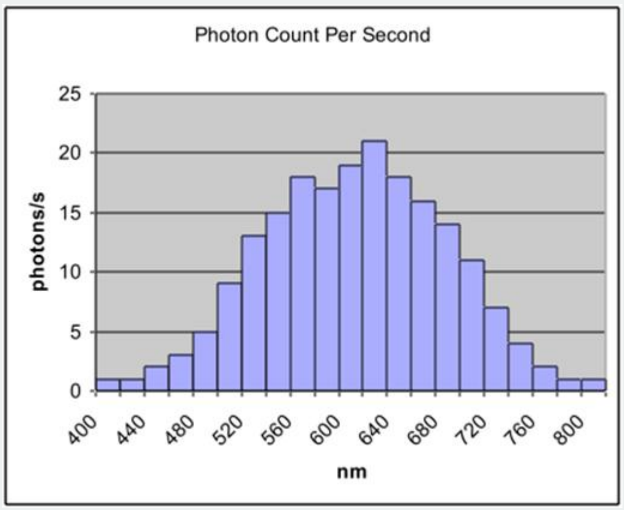
1.

2.

https://phys.libretexts.org/@go/page/31433

B. Best fit parameters.

Run the simulation with different values of Ω_m and σ_8 until you find the values that best match the observed image and histogram



Click “check answers” to see if you are correct.

Enter the best fit values for Ω_m and σ_8 here:

1.

2.

This page titled 14.1: Large Scale Structure is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

14.2: The Formation of Galaxy Clusters and Groups

Learning Objectives

- You will understand that galaxy clusters are composed of galaxies, hot gas, and dark matter
- You will understand how galaxy clusters form from the hierarchical merging of smaller structures

What Do You Think: Which Came First?

14.2.1: Observations of Basic Cluster Properties

Clusters of galaxies are the largest gravitationally bound structures. They contain hundreds to thousands of galaxies that can be seen with optical telescopes. They also contain hot gas between the galaxies (30–100 million K) that can be seen with x-ray telescopes. We also know that galaxy clusters contain exotic (cold) dark matter. The mass in stars makes up about 1–2% of the cluster mass, the hot gas makes up about 5–15%, and dark matter makes up the rest (up to 85%). In other words, the mass of the gas between galaxies is about 6 times more than the mass of the stars in the galaxies and the mass of the dark matter is in turn about 8 times more than that of the gas. Galaxy clusters range in mass from a little more than 10^{13} to 10^{15} solar masses, and typical sizes are from 3–30 million light-years. Figure 14.10 is a composite image of the cluster Abell 1689 in optical and x-ray light.

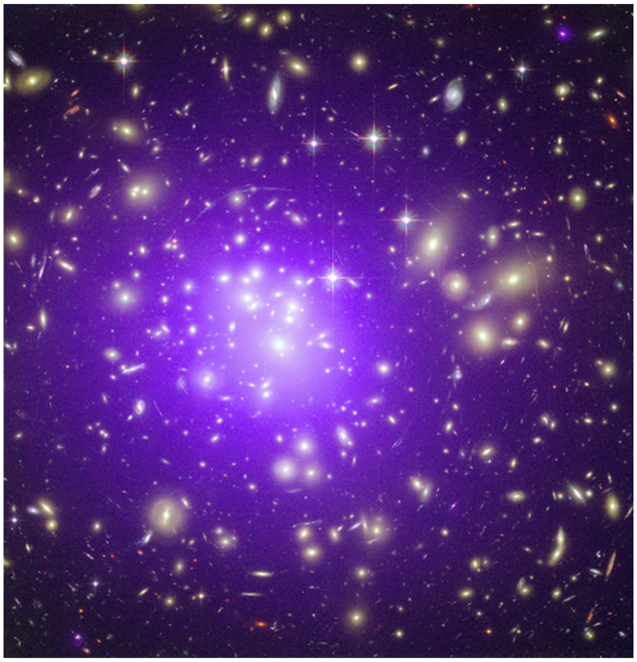


Figure 14.10: This composite image shows the galaxy cluster Abell 1689 in x-ray (purple) and optical (yellow, white, red) light. Hot x-ray gas in between individual galaxies contains more mass than all of the stars in the galaxies combined. The effect of gravitational lensing due to the total cluster mass (stars, gas, and dark matter) can be seen as arcs in the image. Credit: X-ray: NASA/CXC/MIT/E-H. Peng et al.; Optical: NASA/STScI

In many clusters, the gas and galaxies have settled into the same general area as the dark matter, a result of gravitational interactions. In some clusters, however, mergers are ongoing. In those systems the stars and gas have not yet settled into the dark matter potential well. One example is the Bullet cluster, shown in Figure 14.11. Two clusters have passed through each other. The system provides one of the few opportunities to study how exotic cold dark matter and normal baryonic matter behave differently in dynamical situations.

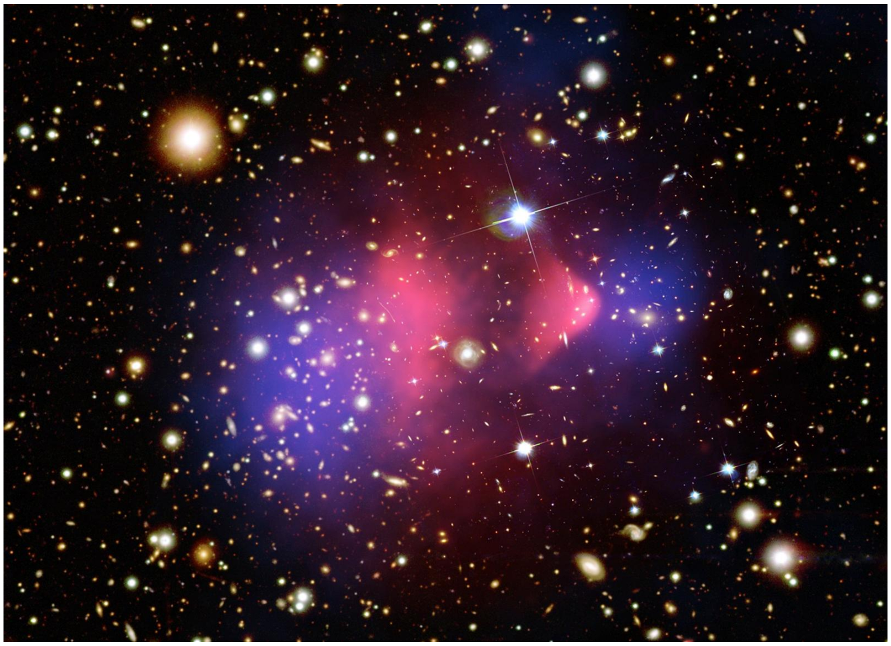


Figure 14.11: A composite image of the Bullet cluster. In addition to galaxies, which are seen in optical light, we see hot x-ray gas, here shown in pink. The blue areas are regions of dark matter, as determined from gravitational lensing. This is actually the result of two clusters that collided with each other. They are a distance of 3.4 billion light-years away from Earth and about 2 million light-years apart from each other. The gas, which is made of regular matter, has clumped in the middle; a shock front (shaped like an arrowhead) can be seen on the right. In contrast, the dark matter, which is collisionless, has passed through, and can be seen on the edges. Credit: X-ray: NASA/CXC/CfA/M.Markevitch et al.; Lensing Map: NASA/STScI, ESO WFI, Magellan/U.Arizona/ D. Clowe et al.; Optical Image: NASA/STScI, Magellan/U.Arizona/D. Clowe et al.

Exotic cold dark matter particles are collisionless, which means they can move past each other without interacting strongly. Gas particles, which are made of regular matter on the other hand, interact very strongly via the electromagnetic force. They can produce shock waves. In the Bullet cluster, we see clearly that the dark matter in the two clusters (displayed in blue) has passed through the collision, ending up on the edges. The gas, on the other hand, has interacted through the collision and has ended up in the center. This distribution is well explained by dark matter and cannot easily be explained by alternative theories of gravity.

14.2.2: Modeling the Formation of Galaxy Groups and Clusters

One aspect of clusters that can be modeled with simulations is whether clusters formed from the mergers of smaller structures or from larger structures breaking apart. In a bottom-up structure formation scenario, small structures formed first and stuck together to form bigger structures. In a top-down scenario, bigger structures formed first and broke into smaller pieces. Simulations show that the formation of structures proceeded in a bottom-up manner—that small structures formed first and gathered together to build small galaxies, then larger galaxies, then clusters of galaxies. In fact, observations show that while galaxies themselves are mainly finished forming today, many clusters and super-clusters are still assembling.

In Animated Figure 14.12 we zoom in on a computer simulation from large-scale structures to galaxy scales. In Figure 14.13 we see a simulation of the formation of a group of galaxies similar to the Local Group, in which the Milky Way is located. In Animated Figure 14.14 we see a computer simulation. It starts with a nearly uniform volume of gas around 13.7 billion years ago. It then evolves to form a cluster and ends at present day with a massive central galaxy similar to the Milky Way. In all of these simulations, galaxies, composed of gas, stars, and dark matter, are colliding and forming as part of the cosmic web. This is hierarchical structure formation in action.

Video

Animated Figure 14.12: This movie of the Millennium Simulation zooms in from the largest scales to galaxy scales. Credit: Millennium Simulation

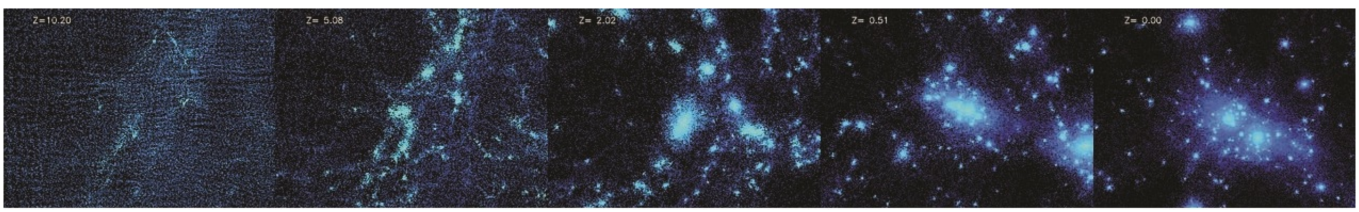
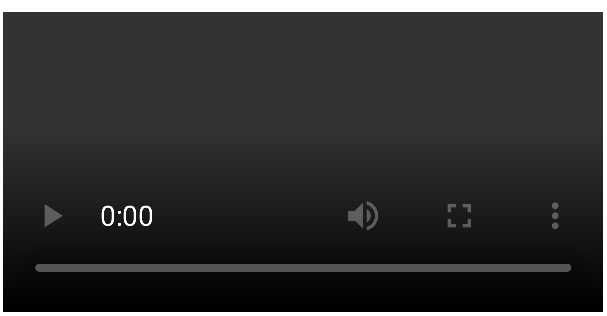


Figure 14.13: Simulation of the formation of a group of galaxies, from z = 30 to today. Earlier times are on the left and later times are on the right. Smaller structures form first and merge to form larger structures, beginning in earnest around a redshift of 5. The box size for this simulation is 14 light-years, in co-moving coordinates. Credit: Simulations were performed at the National Center for Supercomputer Applications (NCSA) by Andrey Kravtsov (The University of Chicago) and Anatoly Klypin (New Mexico State University). Visualizations by Andrey Kravtsov.



Animated Figure 14.14: This simulation shows the hierarchical formation of structure within the cosmic web. Here we see smaller galaxies merging to form larger galaxies and a galaxy cluster. Credit: NASA/Goddard Space Flight Center and the Advanced Visualization Laboratory at the National Center for Supercomputing Applications (NCSA)

In all of these simulations, smaller structures form first and merge to assemble larger structures in a bottom-up scenario. Again, matter becomes more densely clumped over time. Dark matter provides most of the mass, and baryonic matter is attracted to it via gravity. Dark matter dominates the gravitational mergers and accretion, providing the backbone of the structure. Much of the gas is heated to x-ray temperatures by compression and shocks during the gravitational collapse. Stars form from cooler gas, eventually going supernova and injecting energy into the cluster. The gas also feeds active galactic nuclei in the centers of the largest galaxies of the cluster.

Galaxy clusters are interesting objects because they are large enough to be of cosmological importance, yet detailed astrophysics must be used to model them accurately. Simulations of their formation have been extremely successful at reproducing their observed properties, but there is still work to be done to fully understand their inner cores.

This page titled 14.2: The Formation of Galaxy Clusters and Groups is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

14.3: The Formation and Evolution of Galaxies

Learning Objectives

- You will understand why the different galaxy types look the way they do
- You will know that galaxies collide
- You will understand how the Milky Way formed from the mergers of smaller galaxies

What Do You Think: Why Do Galaxies Look the Way They Do?

14.3.1: Galaxy Types

The Hubble Ultra-Deep Field (Figure 14.15) combines 800 exposures of an unassuming spot in the sky from NASA's Hubble Space Telescope (HST). Hold your index finger out at arm's length. The angular size of the Moon is about half of the width of your finger. The spot Hubble looked at for just over 11 days is only a tenth of the size of the Moon. Nearly every single blob of light you see in the image is an entire collection of stars, dust, and dark matter with a black hole in the center bound together by gravity —in other words, a galaxy**.** From this image we can start to classify types of galaxies.

Figure 14.15: In 2004 astronomers pointed NASA's Hubble Space Telescope at an empty patch of sky near the constellation Fornax and collected light for over 11 days. This image shows approximately 10,000 galaxies of various ages, sizes, shapes, and colors. Credit: NASA, ESA, S. Beckwith (STScI) and the HUDF Team

Classifying Galaxies

Classification is often the first step toward scientific discovery. When confronted with a collection of data they do not understand, scientists try to organize the data by searching for patterns or trends that can give clues about the physical processes that affect the data.

In this activity you will analyze images of 16 nearby galaxies taken by the Hubble Space Telescope. You will attempt to organize them according to their shapes, colors, and one more characteristic of your choosing.

Play Activity

A. Shape

In this part you will examine the 16 images and classify them according to their shapes.

1.

2.

3.

B. Color

In this part you will examine the 16 images again and classify them according to their colors.

1.

2.

https://phys.libretexts.org/@go/page/31435

3.

C. Your choice

In this part you will examine the 16 images again and classify them according to a characteristic of your choosing. **Do not use size!**

1.

2.

3.

4.

5.

D. Comparing categories

1.

2.

The structure or morphology of galaxies can be categorized as being either spiral, elliptical, or irregular. You might have chosen to use those terms or something similar in the previous activity. Edwin Hubble was the first person to classify galaxies in much the same manner. He noticed a correlation when comparing galaxy morphology and color. A correlation does not necessarily imply that one thing caused the other, but it does hint that the two properties are related. Elliptical galaxies tend to be red, while spirals and irregular galaxies tend to have more blue stars. Hubble developed a detailed morphological classification scheme known as the Hubble tuning-fork diagram shown in Figure 14.16.

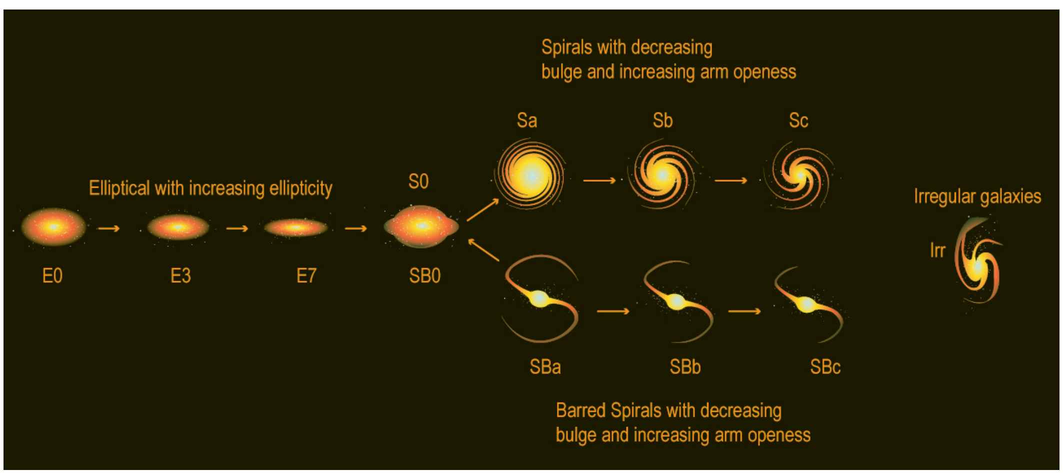


Figure 14.16: Edwin Hubble’s tuning fork diagram divides galaxies into three groups: elliptical, spiral, and irregular galaxies. Elliptical galaxies are classified by how round or flat they look. Spiral galaxies are classified by how tightly the arms are arranged and whether or not the galaxy appears to have a central bar. Irregular galaxies are neither spiral nor elliptical and can have any number of shapes. Credit: NASA/SSU/Aurore Simonnet

These three broad categories can be further subdivided based on visual appearance. Hubble noticed that some spiral galaxies have a bright line, or bar, running through them. He called them barred spiral galaxies. Those without bars he simply denoted as spiral galaxies. Spiral galaxies are further classified by how tightly their arms are wound and the brightness of the central bulge. Elliptical galaxies are classified by how round or elliptical they appear. A transitional galaxy type, named lenticular is somewhere between highly squished ellipticals and disks. These lenticular galaxies have a central bulge but no spiral arms.

https://phys.libretexts.org/@go/page/31435

Hubble believed galaxies evolved from left to right in his tuning-fork diagram. We now understand this is wrong. There is no demonstrable way to make an elliptical galaxy spin up enough to form spiral arms. There is, however, a possibility that multiple spiral or irregular galaxies could collide and form a massive elliptical galaxy. Galaxy collisions are a natural consequence of gravitational interaction. In fact, astronomers have seen many examples of interacting galaxies (Figure 14.17).

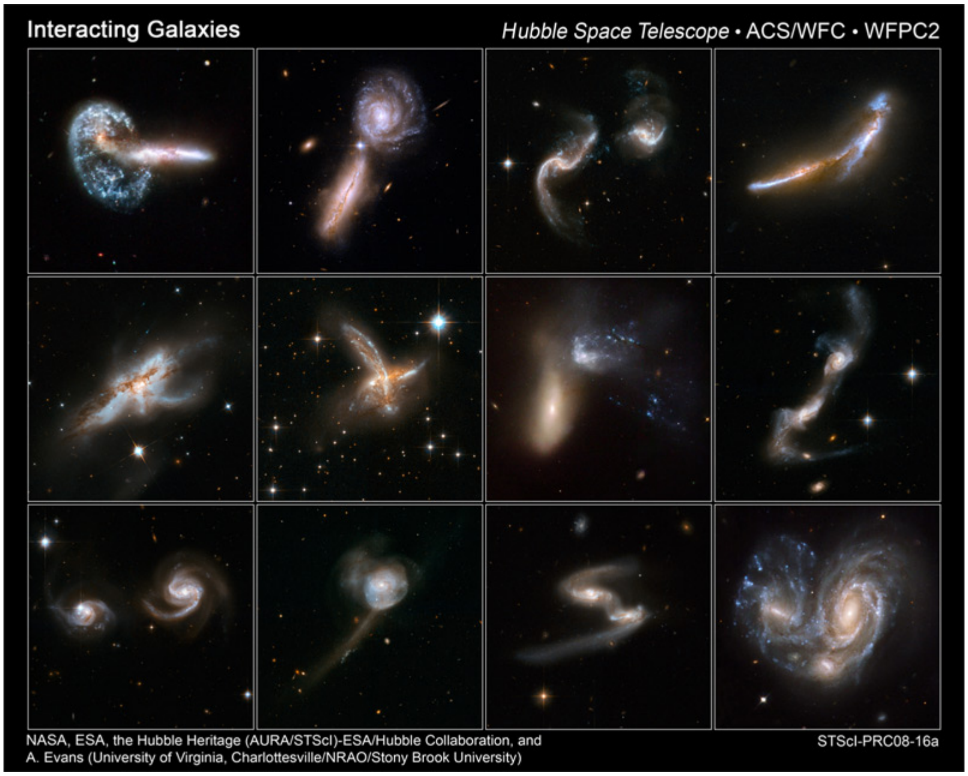
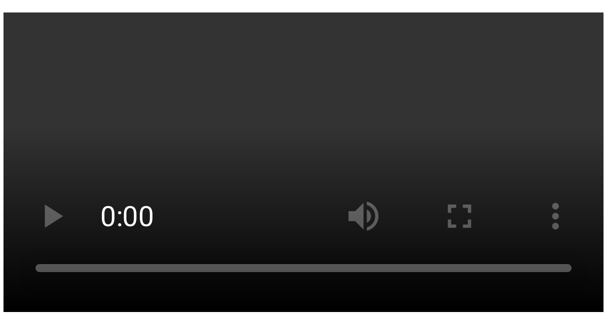


Figure 14.17: These groups of galaxies are all undergoing gravitational interactions. Some of these may one day end up as elliptical galaxies. Credit: NASA/ESA/AURA/STScI/A. Evans

Galaxy collisions can be modeled with simulations such as the one in Animated Figure 14.18. In this simulation, two spiral galaxies pass through each other, yielding a single larger elliptical galaxy at the end of the simulation. In galaxy collisions, the stars and dark matter in each galaxy interact gravitationally. Individual stars are so far apart that they do not actually collide. In fact, a better word for this process than "collision" is "merger." The galaxies do not actually collide during a galaxy collision, but the two interacting systems do eventually merge into one as they settle down. However, there is one component that, if present, does collide: the gas. Collisions between the gas components of the merging galaxies usually triggers huge bursts of new star formation during the interaction.



Animated Figure 14.18: This computer simulation follows two spiral galaxies as they collide and orbit around each other before finally merging. Credit: John Dubinski

The colors of galaxies can be explained by the types of stars they contain. As a galaxy forms, small density perturbations within the gas lead to regions where stars of various masses form. The color of a main sequence star ultimately depends on its mass. Blue stars tend to be more massive than 2 M_{Sun}, stars about the size of the Sun are yellow, and stars much smaller are red. These color differences are a direct result of their surface temperature. Color, temperature, mass, and radius are all intrinsically linked for main sequence stars due to the physics involved; blue stars are hotter, more massive, larger, and die quickly. Red stars are cooler, less massive, smaller, and live longer.

Since blue main sequence stars are also intrinsically brighter. They are so bright, in fact, that they tend to outshine all the red ones within any given group. It is not until a stellar population ages and the blue stars die out that we finally get to see the dimmer red glow of the small long-lived stars. Therefore, the average color of a galaxy is an excellent indicator of how much time has passed since new stars formed. Because stars form from gas, it also tells us how much free gas is available to condense into stars.

Irregular galaxies tend to have the youngest stellar populations. Vast quantities of gas provide the perfect stellar nursery for red, yellow, and blue stars alike. As one star cluster ages and turns more red, another nearby cluster starts anew, causing the galaxy to shine bright blue once again. While irregular galaxies contain stars of all colors, the young blue stars outshine all the others.

Elliptical galaxies are, for the most part, old and devoid of gas. Their reddish colors are dominated by long-lived, red main sequence stars. A majority of the regular matter is locked up in stars. As a result, there is very little ongoing star formation, so only the older, long-lived stars are still present.

Spiral galaxies show a range of star-formation histories and therefore a range of colors. The composite image shown in Figure 14.19 depicts what galaxies similar to the Milky Way would look like farther back in time - so at higher redshifts/lookback times. Notice in the middle, around 9.4 billion years ago, galaxies like ours started to show hints of yellow and red as star formation slowed and their population of massive blue stars began to die off.

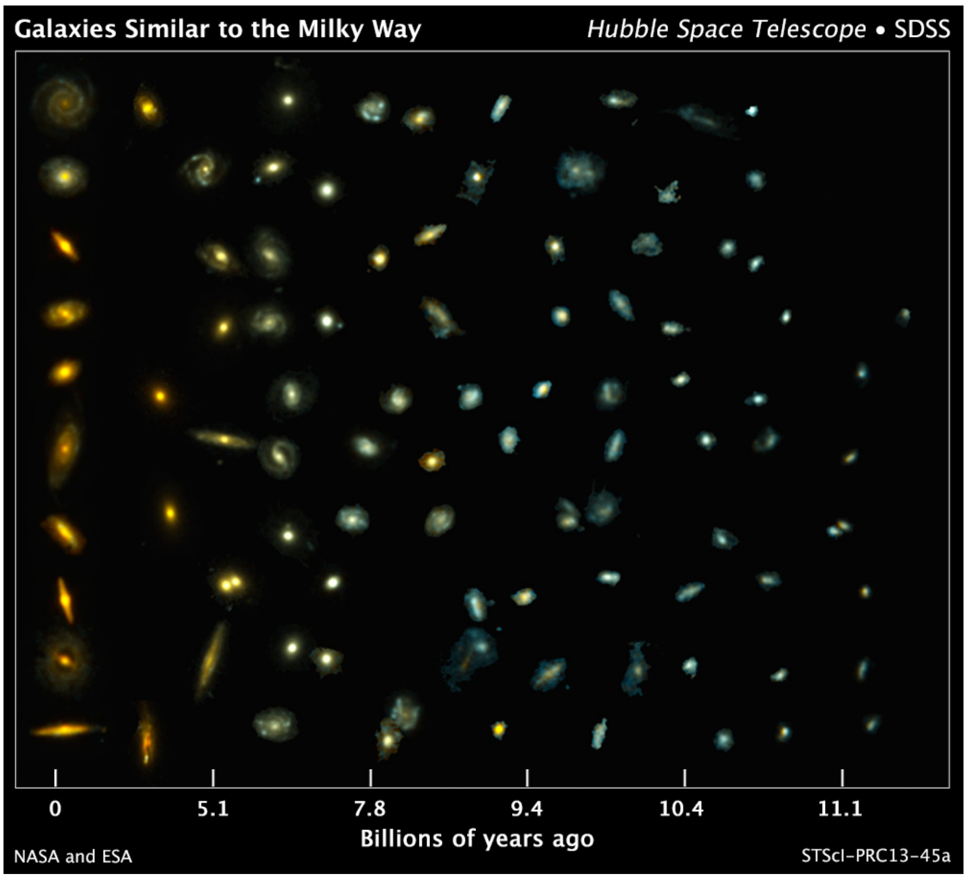


Figure 14.19: This composite image shows examples of galaxies similar to our Milky Way at various stages of construction over a time span of 11 billion years. The galaxies are arranged according to time. Those on the left reside nearby while those on the far right existed when the Universe was about 2 billion years old. The bluish glow from young stars dominates the color of the galaxies on the right. The galaxies at the left are redder from the glow of long-lived red stars. Credit: NASA, ESA, P. van Dokkum (Yale University), S. Patel (Leiden University), and the 3D-HST Team

The Colors of Galaxies

In this activity you will examine a cluster of stars within a galaxy. The slider on the bottom allows you to advance forward in time as the cluster evolves.

Play Activity

1.

2.

3.

4.

5.

https://phys.libretexts.org/@go/page/31435

6.

7.

14.3.2: Modeling the Formation of Galaxies

Once again we can compare detailed observations with computer simulations to understand how galaxies, particularly our own Milky Way, formed. From this we know our Galaxy built up over time due to the mergers of many satellite galaxies. These smaller dwarf galaxies collided with the Milky Way and were subsumed into it. In the distant future, the Small and Large Magellanic Clouds, as well as Andromeda, will merge with our Galaxy (Figure 14.20).

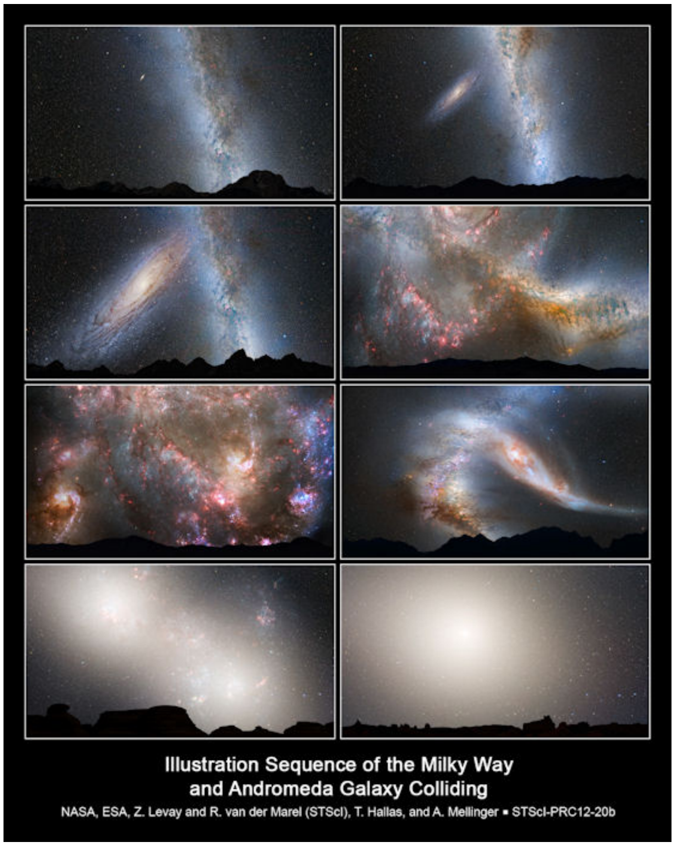


Figure 14.20: This series of photo illustrations shows the predicted merger between the Milky Way and Andromeda as seen from Earth. The first frame is the present day; the last frame is 7 billion years from now. Credit: NASA, ESA, Z. Levay and R. van der Marel (STScI), T. Hallas, and A. Mellinger

We see relics of cannibalized galaxies in the night sky as streams of stars all moving in the same direction or with the similar distances and velocities. Figure 14.21 shows one such observation from the Sloan Digital Sky Survey.

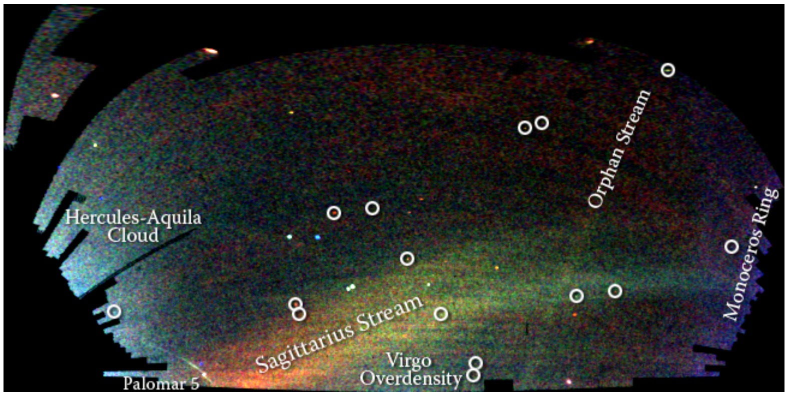
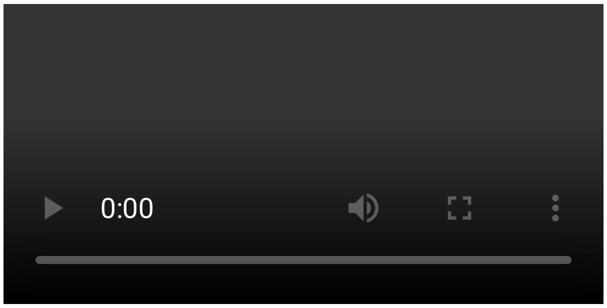
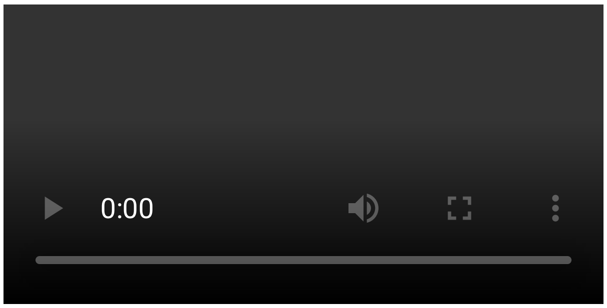


Figure 14.21: This image is a map of stars in the outer regions of the Milky Wave covering about one-quarter of the night sky, as observed by the Sloan Digital Sky Survey. The trails and streams that cross the image are stars torn from disrupted Milky Way satellites. The color corresponds to distance, with red being the most distant and blue being the closest. The circles show the location of several dwarf galaxy satellites. Credit: Vasily Belokurov, SDSS-II Collaboration

The goal of running galaxy computer simulations is to recreate these observations. Astronomers fine tune parameters like the number of dwarf satellites, how often mergers occur, the timing of star formation, and how much star formation affects the overall environment. Combining all of these effects allows astronomers to understand the underlying physical phenomena. The simulation in Animated Figure 14.22 shows one such model run at the University of Zurich. Many small clumps of gas pass by and eventually fall in to the galaxy. In Animated Figure 14.23, we can see what the arrival of the Small and Large Magellanic Clouds to our Galactic neighborhood would look like if you had very sensitive eyes. In fact, galaxies have such low surface brightness that even the several nearby dwarf galaxies very close to the Milky Way appear to us as indistinguishable from the stars that are part the general Milky Way star field.



Animated Figure 14.22: This movie shows the formation of a spiral galaxy like our own Milky Way. Credit: University of Zurich



Animated Figure 14.23: The movie begins with a picture of the Small and Large Magellanic Clouds, our nearest galactic neighbors, and helps us visualize what they would look like if we could see their dark matter halos. The simulation traces the formation of the Milky Way, highlighting these two dwarf satellite galaxies. Credit: Busha, Kaehler, Marshall, and Wechsler

This page titled 14.3: The Formation and Evolution of Galaxies is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

14.4: The First Stars

Learning Objectives

- You will know the evidence that the first stars and galaxies formed when the Universe was ~400 Myr old.
- You will know the evidence that the first stars contained no heavy elements, making them different from later generations of stars
- You will understand the physical conditions under which the first stars formed

What Do You Think: When Did the First Objects Form?

Thus far we have discussed the formation of the largest-scale structures, clusters, and galaxies, but what do we know about the first stars? What were they like? What was their relationship to the larger structures, and what were conditions like in the Universe at that time?

14.4.1: Evidence for When the First Stars Formed

Stellar evolution is the processes involved with the formation, the lifetimes, and the deaths of stars. From the study of stellar evolution we know that stars form from clouds of gas that contract due to gravity and then heat up until nuclear fusion begins. During their lifetimes, stars fuse heavier and heavier atomic nuclei. These nuclei are then dispersed into the surrounding interstellar medium when a tar either explodes as a supernova or sheds its outer layers as a planetary nebula or red giant star. Through these processes, successive generations of stars are made with an increasingly higher abundance of heavy nuclei. Because none of this had happened until stars formed, we expect the first generation of stars to contain only hydrogen, helium, and trace amounts of lithium. These elements were produced in the first minutes after Big Bang itself. All heavier nuclei were fused in the cores of stars and were thus not available to form any part of the first generation of stars.

No direct detection of stars free of heavy elements has been made. Not yet. Astronomers hope to detect them with the upcoming James Webb Space Telescope. Though we have not been able to detect the first stars directly, there is much indirect evidence that tells us about their properties and how they formed. We can see the effects they must have had, much like the fabled three bears discovering that the child Goldilocks had visited their home.

Because the first stars were made entirely of light elements and devoid of heavy elements, they were distinct in their properties compared with later generations of stars. Though they fused hydrogen into helium like massive stars today, the first stars could not use heavy elements as catalysts in the process, so they had to be hotter. As a result, they emitted higher energy radiation (for a given stellar mass) than stars do today. This radiation from the first stars influenced the surrounding inter-galactic medium, causing the Universe to become ionized (separating electrons from the nuclei of their atoms). Finding this ionization signature will tell us when the first stars must have formed.

Constraints on when the first stars formed come from different lines of evidence. Observations from the WMAP satellite, which are sensitive to a small percentage of the Universe being ionized, have shown that the first stars must have turned on around a redshift of $z \sim 10$. This corresponds to a time when the Universe was about 400 million years old. Observations of systems of hydrogen gas clouds in the SDSS are a sensitive probe of the time when there was only a small percentage of neutral hydrogen remaining, after most of the gas had been ionized. This occurs around a redshift of $z \sim 6$, when the Universe was about 1 billion years old.

The first stars, then, were major players on the cosmological scene between 400 million and 1 billion years into the history of the Universe—such small objects influenced conditions in the Universe from the smallest to the largest scales! The ionization at this time was called reionization because this was the second time the Universe was ionized; the first was from its birth through the first 380,000 years (when the cosmic microwave background formed). Though the first stars may not resemble today's stars exactly, the data show that reionization was indeed caused by stars or star-like objects and could not have been caused by other energy sources such as quasars.

14.4.2: Models of Formation of the First Stars

What does it take to form a star from a cloud of gas? It may not seem obvious, but gas will not condense into a star unless it starts out cold, on the order of 200–300 K (about 0 °F). To understand why that is, we will examine the competing forces involved.

Primordial clouds of hydrogen sit embedded in large, roughly spherical halos of dark matter. Due to gravitational interactions alone, the gas particles must collapse toward the center of the halo. Just as your hands heat up when you rub them together, a gas will heat up as atoms get closer together and start interacting and colliding more frequently. This heat leads to an outward thermal pressure, which competes directly with the inward gravitational force (Figure 14.24). When the effect of pressure gets high enough, collapse halts. To learn more about the mass required to form a star, see **Going Further 14.2: Jeans Mass**.

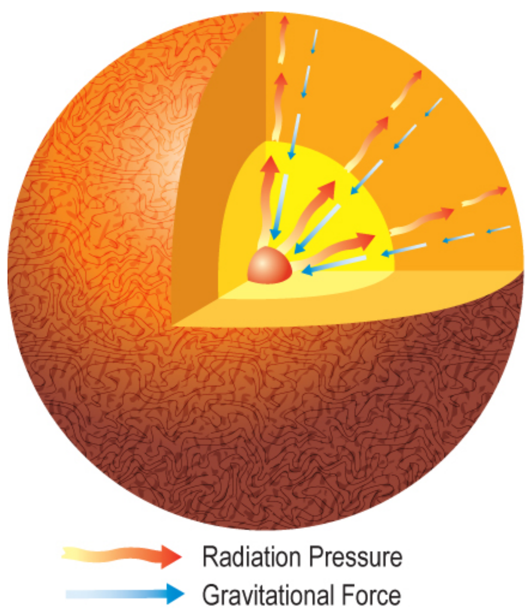


Figure 14.24: As a cloud of gas condenses it gradually heats up. When matter becomes warm, it radiates energy. This outward radiation force opposes the collapse of gravity if it is too strong. Credit: NASA/SSU/Aurore Simonnet

The only way to get rid of excess pressure is for the gas to cool. A single atom of hydrogen is limited in the amount of energy it can shed—the maximum is 13.6 eV (which will ionize the atom, completely removing the electron from the nucleus). Heavier elements can emit energy through many different electron ionizations and state changes, and molecules can also lose energy through vibrations and rotations. Since the first stars were composed of only hydrogen and helium and no heavy elements, the avenues for cooling were limited. The Sun formed from gas composed of much heavier elements, so it cooled more efficiently than the first generation of stars did. To learn more about cooling, see **Going Further 14.3: Cooling Gas**.

Once the gas is dense enough to ignite hydrogen fusion, a star is born. Computer simulations have shown that the first stars could form only from very large gas clouds. If a gas cloud is large enough, the gravitational attraction will win out over the pressure, even with inefficient cooling, and the star will collapse. Because the Sun was able to cool more efficiently, it was able to form from a much smaller cloud. In trying to determine the masses of these early stars, some researchers have hypothesized that these stars formed exclusively in the range of 100 solar masses or greater. Their argument is based upon the previously discussed cooling problem.

An important property of stars with no heavy elements is that they have higher surface temperatures than stars with compositions like that of the Sun. The production of nuclear energy at the center of a star is less efficient without heavy elements, and the star would have to be hotter and more compact to produce enough energy to counteract gravity. This would explain why we see no stars free of heavy elements today—because they would have died long ago (massive stars have short lifetimes). The problem with this theory of only massive first stars is that simulations show that these hypothesized stars and associated supernovae do not actually lead to the observational properties of the Universe that were used as evidence to support their existence in the first place. Determining the masses of the first stars is therefore a subject of ongoing study and great research interest.

GOING FURTHER 14.2: JEANS' MASS

GOING FURTHER 14.3: COOLING GAS

14.4.3: The Relationship Between the First Stars and the First Galaxies

At the time when the first stars formed, galaxies were also starting to form. From computer simulations we know that galaxies were much less massive in the past, probably only about a million solar masses at the time of the first stars (compare this to the Milky Way which is about a trillion solar masses today). One outstanding question is whether these early galaxies could have sustained the first stars. Galaxy formation and star formation are inter-related because stars are dependent on galaxies; the stars from from the gas collected in the galaxies. Furthermore, supernovae can trigger star formation by compressing gas, or halt it by blowing the gas away. Some calculations show that massive galaxies are necessary to sustain star formation. For example, galaxies should have a mass of at least 100 million solar masses to sustain star formation. But galaxies of such large mass would have formed only later. The relationship between the first stars and the first galaxies is therefore also a subject of ongoing study.

To further probe the growth of structure during the reionization era, a number of research groups have used radio arrays to investigate the ratio of neutral to ionized hydrogen present in the early Universe. Neutral hydrogen emits a faint glow at a wavelength of 21 cm, corresponding to a frequency of 1420 MHz, in the radio band. This radiation is used to map out hydrogen clouds in our Galaxy and other galaxies locally, but it can also be used to probe the general presence of neutral hydrogen in the Universe when it was only 400 million to 1 billion years old.

As the Universe becomes more and more ionized due to ultraviolet light from the first stars, the glow from the neutral hydrogen becomes fainter because there is less neutral hydrogen and more ionized hydrogen. By this technique, it is possible to determine the amount of neutral hydrogen at different eras; recall that the emission is redshifted to different frequencies due to the expansion of the Universe. So, for example, emission coming from neutral hydrogen when the Universe is 400 million years old will have a longer wavelength and lower frequency, compared to emission coming from a time of 1 billion years. Each frequency "channel," then, gives a snapshot of how ionized the Universe was at each era.

These radio measurements have a problem equivalent to that of needing a dark sky in optical light. Because of the redshift of the radio waves, the radio arrays actually observe at lower frequencies than the rest frequency of 1420 MHz; they must observe at frequencies between 50 and 200 MHz. This band contains channels 2–13 on broadcast television! In order to escape interference from TV broadcasts, the measurements are being done from the most remote parts of the globe, including areas of China and Australia. They cannot be done from the United States at all.

The redshift range of $z \sim 6$–10, which corresponds to times when the Universe was between 400 million and 1 billion years old, is now the critical epoch for theorists and observers for the next decade. It represents a time in the history of the Universe that we are just on the verge of being able to directly observe. What we can say right now is that star-like objects were present and becoming important in the Universe at $z \sim 10$ or so, and that these first stars likely died by $z \sim 6$. There are plenty of exciting theories, computer simulations, and tantalizing data so far, but much more is unknown. However the arrival of much more relevant data is imminent, so stay tuned for the next 10–20 years!

Formation of Structure Timeline

In this activity you will use an interactive timeline to answer questions about when various structures formed. The timeline is divided into three time periods: the Early Universe, the Matter Dominated Universe, and the Modern Universe. The period when matter dominated the Universe is when much of the structure formation that we have discussed in this chapter occurred. To learn more about an event , first click the “Matter Dominated Universe” tab, then click the tab for the event.

Play Activity

1.

2.

3.

This page titled 14.4: The First Stars is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

14.5: Wrapping It Up 14 - Map the Universe

Learning Objectives

- You will use galaxy survey data to explore some typical objects and a 3D map of the Universe.

In this activity you explore some typical objects as well as make a three-dimensional map of the Universe with actual data from the Sloan Digital Sky Survey.

14.5.1: Part I: The Coordinates and Distances of Galaxies in the SDSS

The SDSS database is hosted publicly online. Launch the SkyServer. The menu on the left hand side of the webpage lists the types of searches we can do to pull information out of the database.

- Click on “Visual Tools” from the left hand navigation menu.
- At the very top of the left hand menu is a small row of links. Be sure to click on “Navigate”.
- In the “Parameters” box, enter a right ascension (RA) of 180 deg and a declination (DEC) of 0 degrees.
- In the “Drawing options” box select “Objects with spectra .”
- Click “Search.”

Now we can begin!

1.

2.

3.

Use your browser back button to return to the search page. Enter the following coordinates: RA: 130 deg, DEC: 11 deg. Again click the “Objects with spectra” check box and hit “Search.”

4.

5.

6.

We can make a three-dimensional map of galaxies by transforming their RA, DEC, and distance coordinates to familiar x, y, and z space.

7.

These calculations you just performed are just what astronomers do to make the incredible maps of large-scale structure. In this next section a computer program will take care of calculating the distances and (x, y, z) coordinates for 100 stars, galaxies, and quasars and plot them on an interactive data viewer.

14.5.2: **Part II: Mapping the Galaxies**

Play Activity

What you are looking at is the SDSS data viewer. It allows you to load a saved data file downloaded from the Sloan Digital Sky Survey database and allows you to view, rotate, and zoom around positions of galaxies and other distant objects. Earth is located at the intersection of the three axes. A simulated dataset has already been provided for you.

- To rotate your view of the data, click on the axes and drag them around.
- To read the distances to objects (in Mpc), click "3D Cursor" and adjust the sliders.

1.

2.

3.

4.

5.

6.

7.

8.

9.

This page titled 14.5: Wrapping It Up 14 - Map the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

14.6: Mission Report 14 - Map the Universe

A.

B.

C.

D. Questions to be graded for accuracy. Show your work!

1.

2.

https://phys.libretexts.org/@go/page/31477

3.

4. You will need to use the SDSS homepage to answer this question. We can only plot distances to objects when we have their redshifts. In Data Release 10, what types of ob

This page titled 14.6: Mission Report 14 - Map the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

15: The Cosmic Microwave Background

Chapter 15 describes the observations and our understanding of the Cosmic Microwave Background (CMB), arguably the most important evidence that supports Big Bang theory. CMB observations are described in detail, including the measurements of the tiny variations in temperature that are the seeds of structure formation. The chapter concludes with a comparison of the CMB observational data to predictions made by varying models of the Universe.

This page titled 15: The Cosmic Microwave Background is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.0: The Cosmic Microwave Background Introduction

This video contains no audio

The opening movie begins with a perspective on our place in the Milky Way Galaxy, in the spiral arms of the disk. Looking out at the night sky, we see the disk of our Galaxy as a band of stars. The movie then shows a grid of coordinates, the Galactic equivalents of latitude and longitude. These coordinates can be spread out into an ellipse, called a Mollweide projection, so that all positions on the sky can be seen at once. Next, the movie shows us views of the sky in different wavelengths, from visible light, to infrared, to microwave. At microwave wavelengths, there is a nearly uniform glow coming from all directions in the sky. In the final segment of the movie, the contrast on the map is dialed up, so that we can see the tiny variations from place to place in the microwave sky.

Since its discovery in 1965, the cosmic microwave background (CMB) has provided some of the most important observational constraints for our theories of the nature and evolution of the Universe. The CMB is a glow of microwave light coming from every direction in the sky. Together with the abundances of the lightest elements and the Hubble expansion, it is one of the key observational pillars supporting the Big Bang theory. We can make several types of measurements of the CMB, each of which we will discuss in detail in this chapter. To date, there have been over 50 experiments either completed or ongoing to observe the CMB. The instruments used to measure the CMB feature components similar to devices we use in everyday life. They have features in common with radio receivers, satellite dishes, televisions, cell phones, and microwave ovens. In fact, if you want to observe the CMB yourself, find a TV with an antenna and switch to a channel on which no station is broadcasting; a small fraction of the "snow" pattern is the CMB. You might need an older-model TV for this, as newer models generally display a blue screen on any channel without a signal. If you have an FM radio, a small fraction of the static between stations is also the CMB.

This page titled 15.0: The Cosmic Microwave Background Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.1: Observations of the CMB Spectrum

Learning Objectives

- You will know that the CMB is nearly uniform and coming from all directions
- You will know that the CMB spectrum is a 3K blackbody

What Do You Think: Temperature of the Universe

We can use the light from astronomical objects to measure their temperatures. The CMB is coming from every direction in the sky, so we can use this light to take the temperature of the Universe as a whole. But first, what exactly is the cosmic microwave background?

Arno Penzias and Robert Wilson did not set out to discover the CMB. They were working with a new kind of detector at Bell Labs in New Jersey in 1964. In the course of making very careful measurements and re-checking their equipment, they realized they had detected a source of "noise" in their antenna. It was coming from all directions in the sky and could not be attributed to any known source.

Neither Penzias nor Wilson knew what to make of their antenna noise. But then Arno Penzias learned of a paper by Robert Dicke, Jim Peebles, and David Wilkinson, all cosmologists at nearby Princeton University. In the paper, which was still in draft form, they discussed relic radiation that should have been created in the early stages of a hot dense Universe. After reading the paper, Penzias invited the Princeton scientists to come to Bell Labs and have a look at the antenna (Figure 15.1), along with his and Wilson's results. Together they decided to publish simultaneous papers announcing the discovery of the background radiation predicted by the Big Bang theory. The Princeton group would write about the theoretical underpinnings of the radiation, and the pair from Bell Labs would write about their discovery. The papers were published back-to-back in *Astrophysical Journal Letters* in 1965. Penzias and Wilson won the 1978 Nobel Prize in physics for the discovery.

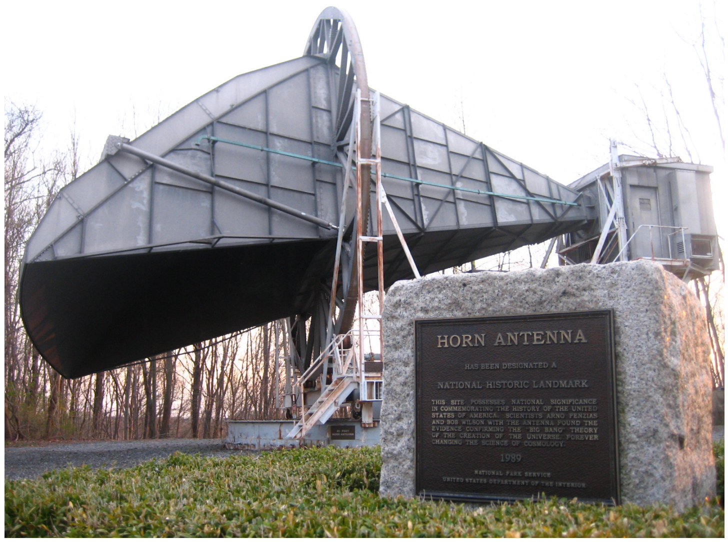


Figure 15.1: This horn antenna was used by Penzias and Wilson at Bell Labs in their discovery of the CMB. Credit: Wikimedia Commons

At first glance the CMB is an almost completely uniform glow in the entire sky as seen in microwaves. It is similar to the blue glow seen in the sky on a cloudless day; there are almost no discernible features. Figure 15.2 illustrates the uniformity of the temperature of the CMB across the entire sky compared to a map of temperatures on Earth. Since the CMB is observed with microwave telescopes rather than visible light, color is typically used to represent temperature, not wavelength, in maps of the CMB. Furthermore, CMB maps are usually shown in the Mollweide projection so that all positions on the sky can be seen at once. An example of what a map of Earth would look like in a Mollweide projection is shown in Figures 15.2 (bottom panel) and 15.3.

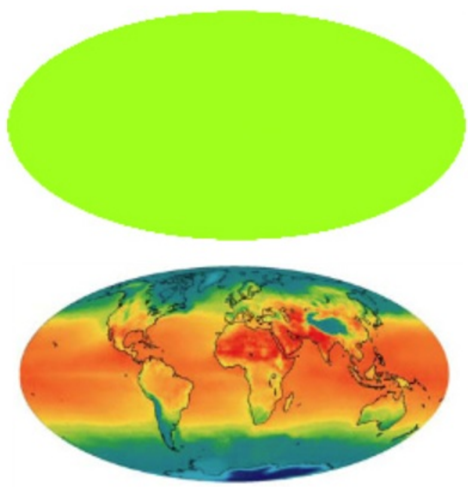


Figure 15.2: The uniformity of the temperature of the CMB (top panel) compared with a map of Earth on the same temperature scale (bottom panel). The temperature of the CMB is much more uniform than the temperature across Earth. Both maps use a projection such that the entire sphere of the sky (in the case of the CMB) or globe (in the case of the Earth) can be represented at once (as in Figure 15.3). Credit: NASA/WMAP Science Team

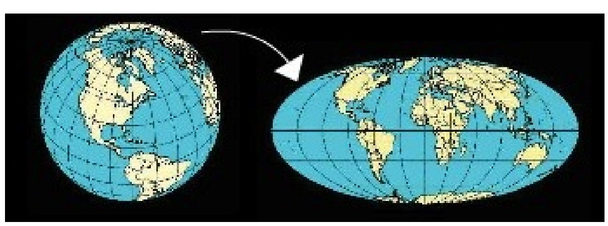


Figure 15.3: A Mollweide map projection of Earth. The advantage to this projection is that it allows all positions on the globe to be seen at once. There are still distortions relative to a spherical globe, but they are less than those of a rectangular (Cartesian) map. Credit: NASA/WMAP Science Team

Recall that we can measure the temperature of an object from its spectrum—a plot of wavelength (horizontal axis) vs. their intensity of emission (vertical axis) at that wavelength. The most common type of continuous spectrum is called a blackbody spectrum (or Planck spectrum) and it has a characteristic shape. We also learned that the peak wavelength of the blackbody spectrum corresponds to the temperature: the hotter the object, the shorter the wavelength at the peak and the higher the curve at all wavelengths.

In 1989, the COBE satellite was launched, with a goal of measuring the spectrum of the CMB over the entire sky. The COBE team included dozens of scientists and engineers. Hundreds of other people helped make the mission a success. The team leaders for the project, John Mather and George Smoot, won the Nobel Prize in 2006 for discoveries made by the COBE team.

COBE contained an instrument called the **Far-InfraRed Absolute Spectrophotometer**, or FIRAS. The FIRAS instrument measured the intensity of the CMB at multiple wavelengths and determined that it has a blackbody spectrum with a temperature of 2.725 ± 0.002 K. This is the best example of a blackbody spectrum that we know of in the Universe; it is a more perfect blackbody than any oven, charcoal briquet, or lamp we have ever created. Figure 15.4 shows the spectrum of the CMB, as measured by FIRAS. The data and the model agree to high precision; the uncertainty in the data points is smaller than the width of the line used to plot the model fit.

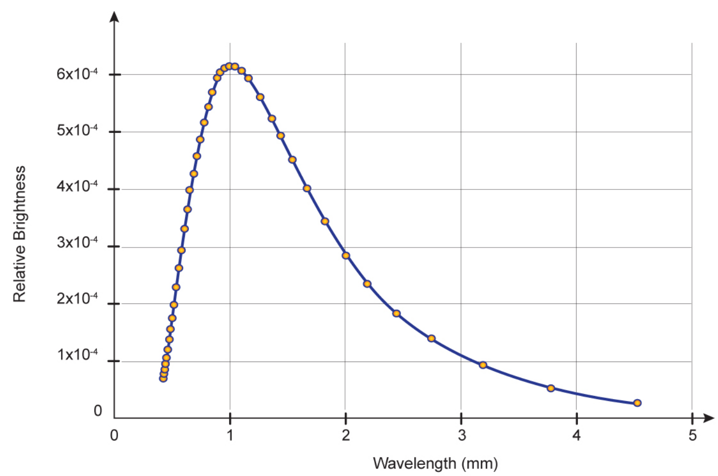


Figure 15.4: Spectrum of the CMB as measured by the FIRAS instrument on the COBE satellite. The CMB is the most perfect blackbody known. It has a temperature of about 3 degrees above absolute zero, which corresponds to a peak wavelength of about a millimeter. Theory and observation agree to better than the width of the line in the graph. Credit: NASA/SSU/Aurore Simonnet based on COBE/FIRAS data

Taking the Temperature of the Universe

In this activity you will use the COBE/FIRAS data to determine the temperature of the Universe. The COBE scientists fit the entire curve, but you will try to estimate only the position of the peak. It is not as accurate this way, but it gives the gist of how the temperature is measured.

1. First you will measure the peak wavelength of the blackbody (Planck) curve from the data. To do this, use the following graph:

USE GRAPH

a.

2.

Since the CMB is radiation coming from everywhere in the sky, you have just measured the overall temperature of the Universe!

This page titled 15.1: Observations of the CMB Spectrum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.2: Implications of the CMB Temperature and Spectrum

Learning Objectives

- You will understand why the CMB supports the Big Bang Theory

What Do You Think: Evidence for the Big Bang Theory

One piece of evidence in support of the Big Bang theory is the **Hubble expansion**. On average, we see galaxies moving away from each other as the space between them stretches. Extrapolating this expansion back to earlier times in the history of the Universe, the Big Bang theory predicts that the Universe was once much denser, and as a result, hotter. The Big Bang theory naturally explains the CMB as the glow left over from the hot, dense early state of the Universe.

At about 380,000 years after the Universe came into existence (more exactly 375,000 ± 3000 years according to current best estimates), the temperature everywhere was about 3000 K. The entire Universe was still a fairly dense soup of simple nuclei, free electrons, and photons. The photons before this time could not get very far because they were constantly running into electrons and being scattered, similar to the way water droplets scatter light in a fog. The electrons were still separate from the nuclei because in the high temperature conditions they had too much energy to stay bound. As the temperature dropped, however, the electrons were eventually able to bind to nuclei, forming neutral atoms, mostly hydrogen and helium. Neutral hydrogen and helium gas do not interact strongly with 3000 K photons, and so the Universe became mostly transparent at this time. The light could travel freely for great distances without interacting with anything. The CMB is the light that was set free at this time, and it has been traveling across the Universe for nearly 13.8 billion years—perhaps only stopping when it hits the antenna on your television. When we are looking at an image of the CMB, such as that observed by the Planck satellite in Figure 15.5, we are seeing a *direct* picture of the Universe in its infancy.

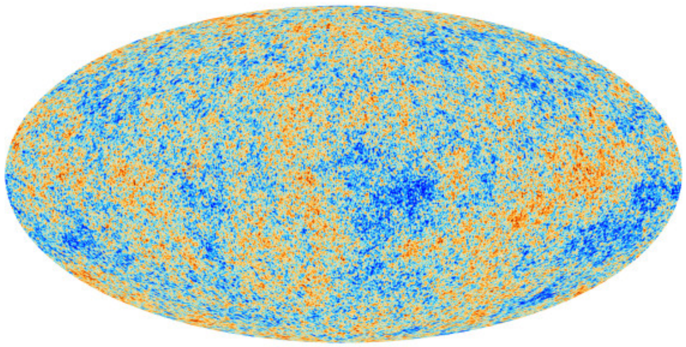


Figure 15.5: Map of the CMB as measured by ESA's Planck Satellite in 2013. This radiation was emitted when the Universe was 375,000 years old. The contrast has been enhanced compared to the CMB map in Figure 15.2 such that the red spots are slightly hotter than the blue spots by about 1/10,000th of a degree Kelvin . The map is shown in the Mollweide projection so that all positions on the sky can be seen at once. Credit: ESA and the Planck Collaboration

The temperature of the Universe at the time when photons from the CMB were set free (about 3000 K) corresponds to infrared light, since it is blackbody radiation. The expansion of the Universe has caused the light to redshift to a wavelength more than 1000 times longer. That is why we see it as microwave light today.

In an expanding Universe, the wavelength of a photon will be stretched along with the space in which it is embedded. The scale factor of the Universe is related to the redshift:

$$\frac{\lambda_o}{\lambda_e} = 1 + z = \frac{S_o}{S_e}$$

where λ_o is the wavelength of CMB photons, z is the redshift, and S is the scale factor of the Universe. The subscript e describes when light was emitted and the subscript o describes when we observe it today. Using the relationship between wavelength and temperature for a blackbody spectrum ($T = 2.9 \times 10^{-3} / \lambda_e$), we can also relate the scale factor and redshift to temperature:

$$\frac{\lambda_o}{\lambda_e} = 1 + z = \frac{S_o}{S_e}$$

where T is the temperature of the CMB. This means that at early times, when the scale factor is smaller, the Universe is hotter.

In the following activity, we will determine how much the Universe has stretched since light from the CMB was emitted and how the temperature has changed.

✔ The Redshift of the Cmb and the Temperature of the Universe

Observations of the CMB tell us that it formed at a redshift of z = 1100. The observed temperature of the CMB today is 2.73 K.

Worked Examples:

1. By what factor has the Universe stretched since light from the CMB was emitted?

- Given: z = 1100
- Find: S_o/S_e
- Concept: S_o/S_e = 1+z
- Solution: S_o/S_e = 1101 times

This means the Universe has stretched by a factor of 1101 since light from the CMB was emitted.

2. How much hotter was the temperature of the Universe when light from the CMB was emitted?

- Given: z = 1100
- Find: T_e/T_o
- Concept: T_e/T_o = 1+z
- Solution: T_e/T_o = 1101 times hotter

3. What was the temperature of the Universe when light from the CMB was emitted?

- Given: $z = 1100$, $T_o = 2.73$ K
- Find: T_e
- Concept: $T_e/T_o = 1+z$
- Solution: $T_e = (T_o)(1+z) = (2.73\text{ K})(1101) = 3006$ K

Now compare these temperatures and stretch factors to those for some of the objects you learned about previously.

Questions:

1.

2.

This page titled 15.2: Implications of the CMB Temperature and Spectrum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.3: Observations of Variations in the CMB

Learning Objectives

- You will understand what the spots are on CMB anisotropy maps
- You will understand how CMB measurements are made

What Do You Think: Variations in the CMB

When scientists look at the spatial distribution of the CMB in detail, they notice that superimposed on the nearly uniform background are tiny (about 10 – 100 parts per million) variations in temperature. This departure from uniformity, called anisotropy, was first observed by the Differential Microwave Radiometer (DMR) instrument on the COBE satellite in 1991 (Figure 15.6). This was important because for the first time it showed that the microwave background was not completely even. Since COBE's early result, fluctuations in the CMB have been measured with increasing precision by dozens of instruments.

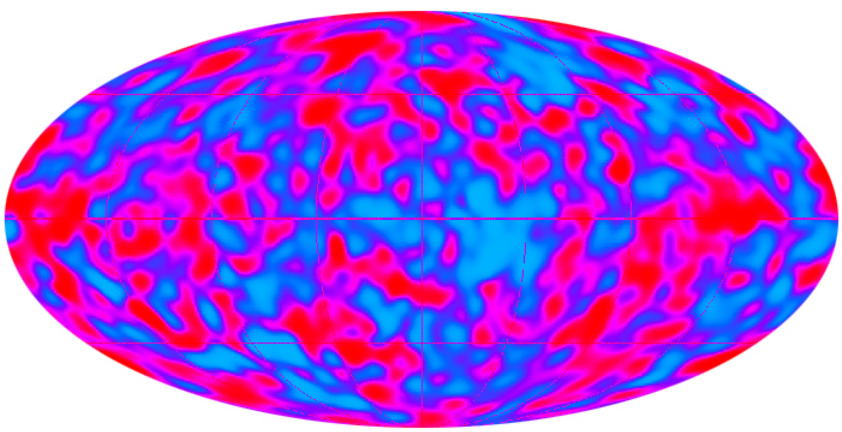
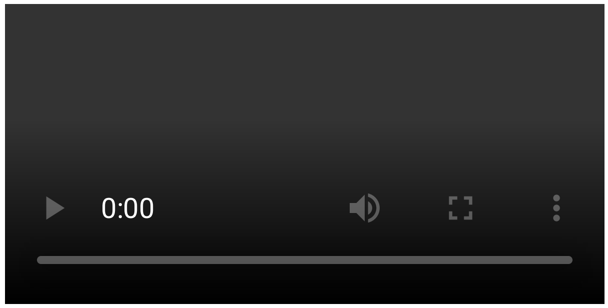
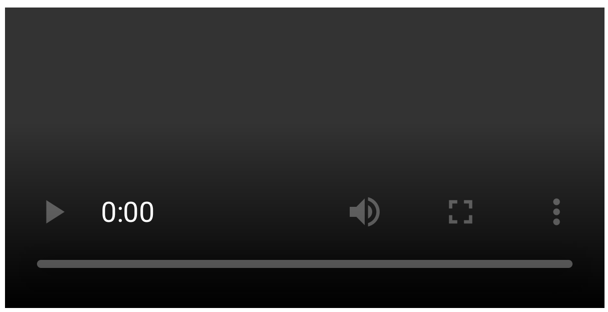


Figure 15.2 such that the red spots are slightly hotter than the blue spots by about 1/10,000th of a degree Kelvin.Animated Figure 15.7 shows a map of the CMB as observed by COBE and then later by the WMAP satellite at better resolution . It is possible to distinguish smaller angular sizes in the WMAP data . The red spots in the map correspond to slightly hotter regions and the blue spots correspond to cooler regions. The contrast between the hotter and colder regions is greatly exaggerated in order to see the effect; their typical temperature differences are about 1/10,000th of a degree!



Animated Figure 15.7 CMB anisotropy was first observed by the COBE satellite in 1991 at large angular scales (initial image in animation). Since then, it has been observed with better resolution by dozens of other instruments, including the WMAP satellite (final image in animation). The contrast has been enhanced compared to the CMB map in Figure 15.2 such that the red spots are slightly hotter than the blue spots by about 1/10,000th of a degree Kelvin. The WMAP image has better angular resolution than the COBE/DMR image. The difference between DMR and WMAP is similar to the difference between a near-sighted person without their glasses and a person who does not need glasses trying to recognize faces in a crowded room. The person who does not need glasses can see more detail than the person who does. Credit: NASA/WMAP Science Team

In both COBE and WMAP data, microwave signals from our Galaxy and other galaxies have been removed from the map, leaving only an image of the CMB. This is possible because the CMB has a unique blackbody spectral signature. Other sources of microwave emission, such as dust in our galaxy, have different spectral signatures. By observing the sky at several different frequencies, we are able to isolate and then subtract the sources other than the CMB. This process is illustrated in Animated Figure 15.8, with data from the Planck satellite.



Animated Figure 15.8 Microwave radiation from our Galaxy and other galaxies has been removed from the data in order to make a map of only the CMB. Credit: ESA and the Planck Collaboration

Another source of microwave light that can be found in CMB maps is a signature from the interaction of CMB photons with galaxy clusters, known as the **Sunyaev-Zel'dovich (SZ) effect**. Because of observations of the SZ effect, we know the CMB must be cosmological, that is, coming from all over the Universe, not just a nearby source (such as our Galaxy). Light from the CMB is distorted in the vicinity of galaxy clusters, which means it must be coming from places beyond the galaxy clusters (which are themselves far away) in order to be affected. The SZ effect can be seen at great distances, because it does not follow an inverse square law. This makes it ideal for finding galaxy clusters. Several groups are involved in efforts to measure the SZ effect and use it to survey galaxy clusters, as described in **Going Further 15.1: The Sunyaev-Zel'dovich (SZ) Effect**.

Going Further 15.1: The Sunyaev-Zel'dovich Effect

The Sunyaev-Zel'dovich effect is caused by interactions of photons with charged particles. We have already discussed on several occasions how photons can be scattered by charged particles, typically electrons. Sometimes this scattering merely changes the direction of the photons without changing their energy. That is what happens to sunlight as it passes through Earth's atmosphere, and the preferential scattering of short wavelengths over longer causes the sky to appear blue on a clear day.

However, it is also possible for the scattering of light to transfer energy to or from the photons. If the population of scattering electrons has on average more energy than the population of photons, then energy is transferred to the photons, which appear hotter as a result. On the other hand, if the typical energy of the electrons is lower than the energy of the photons, then the light will transfer energy to the electrons. The light cools at it looses energy to the electrons.

The microwave background began as light with a typical temperature around 3000K. Since the Universe has grown by a factor of about a thousand since the CMB was created, the CMB light now is characterized by a temperature of around 3K, and of course, it was cooling from 3000K to 3K over its entire journey. However, some of the microwave photons have encountered galaxy clusters as they traveled through the cosmos. These photons have had their temperature modified by the scattering described previously.

In addition to containing galaxies, galaxy clusters also contain immense amounts of hot gas. The gas, which actually dominates the total baryonic mass of the clusters, has a temperature around 10 million kelvin, much higher than that of the CMB. This is because it sits deep in the gravitational well of the cluster. When CMB photons encounter the hot gas they gain energy from its particles because the CMB is so much cooler than the typical electron in the cluster gas. As a result, the photons passing through a cluster have a slightly increased temperature than CMB photons that don't pass through a cluster. The spectrum is distorted and shifted to higher energy. The CMB photons mark the positions of the clusters on the sky because they cause the clusters to stand out as small spots in the microwave background. At higher wavelengths than the peak of the CMB, these spots are hotter, and at longer wavelengths (lower energies), these spots are colder, because the photons that would normally be present have been shifted to a different part of the spectrum. An example of an SZ cluster observation is shown in Figure B.15.1.

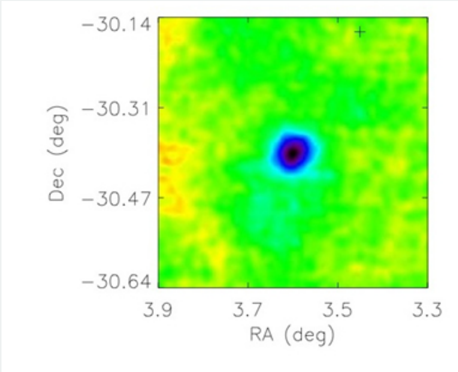


FIGURE B.15.1: SZ EFFECT. Galaxy cluster Abell 2744 as seen by the South Pole Telescope (SPT). Light from the CMB interacts with the hot gas in the cluster, causing a cold spot at the wavelengths that SPT observes. This is called the Sunyaev-Zel'dovich (SZ) effect. Credit: SPT Collaboration.

The SZ effect is important at high-l values, or small angular scales, because galaxy clusters are not very large on the sky. The small SZ temperature fluctuations are not caused by any primordial cosmological parameters, they are modifications of the original CMB spectrum by intervening mass concentrations as its photons travel to us. These fluctuations must be accounted for in the cosmological analysis of the CMB since they introduce a spurious signal (or noise, if you prefer) on the true cosmological perturbations.

However, the SZ effect is interesting in its own right because it provides a way to understand how galaxy clusters have formed and evolved over the history of the Universe. Because the CMB comes to us from essentially the beginning of the Universe, it probes as far as can be probed by photons. What's more, the SZ effect depends only on the angular size of the clusters and the temperature of their gas, it is not affected by the $1/r^2$ dimming law that affects many astronomical measurements. Thus SZ provides a powerful means to study structures anywhere along the line of sight. One of the many telescopes studying the SZ effect is the South Pole Telescope. Animated Figure B.15.2 shows this telescope being built.

The South Pole Telescope (SPT) is one of several telescopes studying the SZ effect, which can be used to find galaxy clusters. Credit: SPT Collaboration.

In the next activity, you will examine several CMB maps to get a sense of the temperature and angular scales of variations in the CMB.

Understanding Cmb Anisotropy Maps

Figure A.15.1 presents maps of CMB anisotropy as observed by the COBE, WMAP, and Planck satellites. Use these maps to answer the questions below.

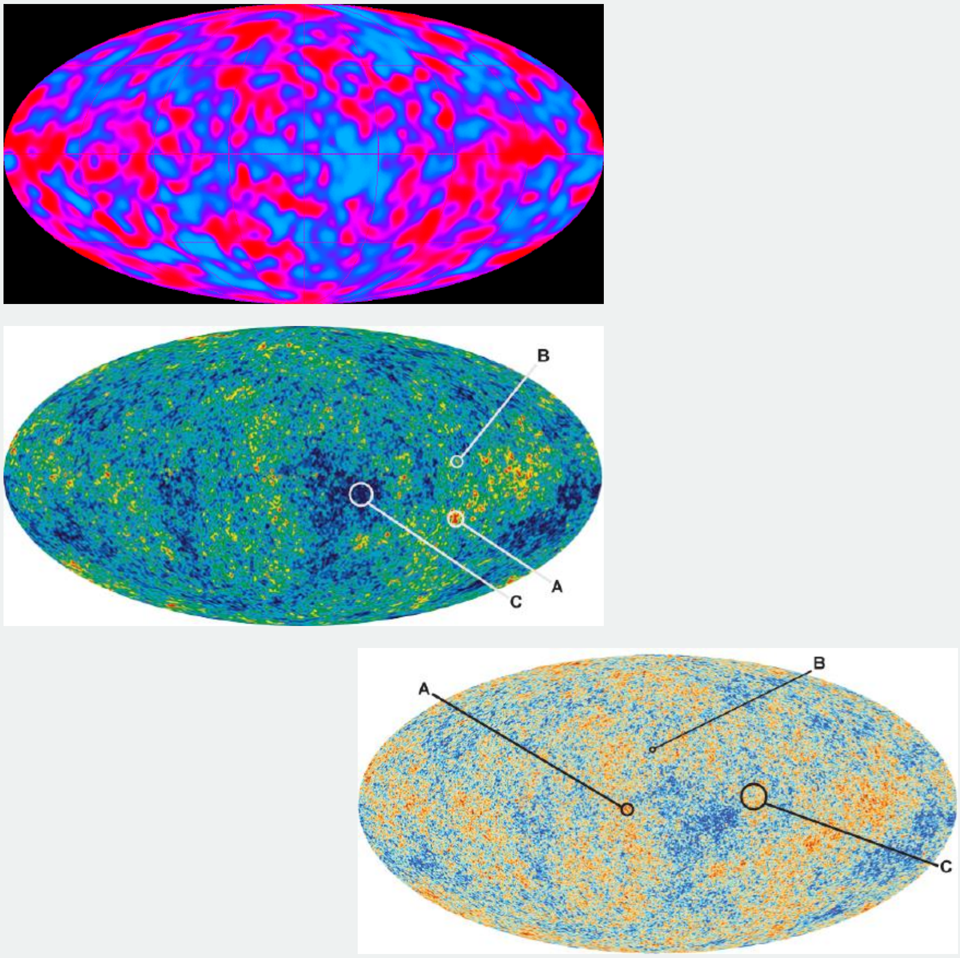


Figure A.15.1: Maps of the CMB, as observed by three different satellites. (Top) COBE. (Middle) WMAP. (Bottom) Planck. Credits: NASA/COBE Team, NASA/WMAP Science Team, ESA and the Planck Collaboration

A. Resolution and Angular Scale

1.

2.

3.

B. Temperature

1.

2.

C. Comparing With Known Objects.

1.

2.

3.

As with many astronomical measurements, CMB observations are often conducted from locations inhospitable to humans. The chosen locations tend to be high and dry in order to reduce the interference from Earth's atmosphere. Though the CMB is everywhere, we go to these great lengths and exotic locations in order to get the most precise measurements. To date, there have been three satellites, over a dozen balloon-borne experiments, and more than three dozen ground-based experiments that have studied the CMB, with more to come in the future.

The advantage of satellite missions is that they can map the whole sky with high precision and without interference from the atmosphere. The disadvantages are that they often have a long development period and are costly. Experiments carried by long duration balloons around Antarctica fly at an altitude of about 120,000 feet (three to four times higher than a commercial airplane), above most of Earth's atmosphere. A plethora of experiments have detected or constrained CMB anisotropy from the ground. The South Pole is an especially attractive site for CMB measurement because it has a stable atmosphere, has a high altitude (~10,000 feet), and is the driest place on Earth. Examples of other favorable locations include the California desert and the mountains of Chile.

Conducting CMB observations today is a team effort. It is often done in large collaborations where various group members make highly specialized contributions to the larger effort. Some team members build the telescopes or detectors, others make the observations, and still others are responsible for the data analysis in which the observations are compared with theoretical models. Figure 15.9a shows some of the telescopes and their instruments from recent CMB experiments and Figure 15.9b shows some CMB telescopes in action.

Figure 15.9a: A few of the many instruments used in recent CMB experiments: a close-up of the ACBAR feed horns and electronics, the BOOMERANG telescope , and the WMAP satellite. To date there have been more than 50 experiments designed to measure the CMB. Credits: Acbar Collaboration, NASA/COBE Team, BOOMERANG Collaboration

Figure 15.9b: Several recent CMB telescopes in action: the launch of the balloon-borne BOOMERANG telescope from the coast of Antarctica, DASI at sunset at the South Pole, the CBI dome in the mountains of Chile, and the launch of the WMAP satellite. Prime locations for microwave telescopes are high and dry, above as much of Earth's atmosphere as possible. Credits: BOOMERANG Collaboration, DASI Collaboration, CBI Collaboration, NASA/WMAP Science Team

To help you visualize what the CMB would look like to you if you could see what some of these telescopes see, Figure 15.10 is a fanciful picture super-imposing a false-color map of the CMB on the launch site for the BOOMERANG balloon-borne telescope.

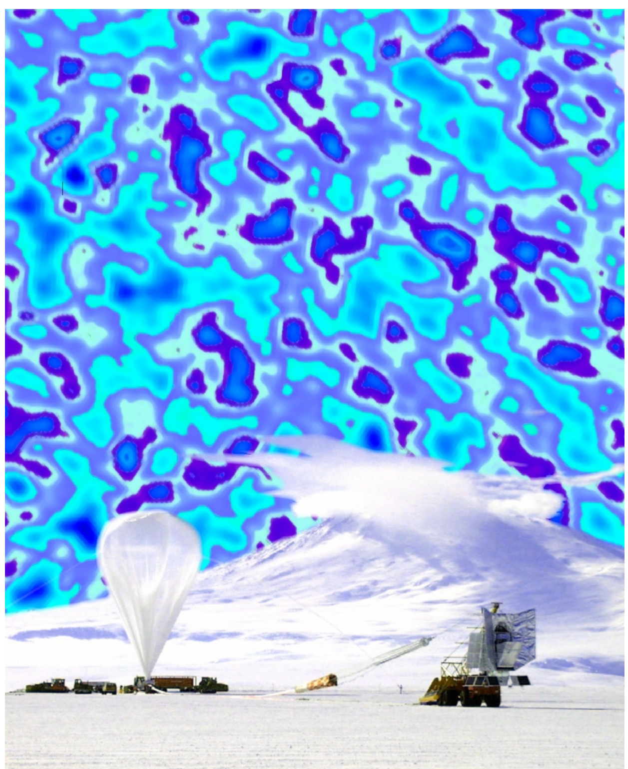


Figure 15.10: Fanciful image of what you might see in the sky if you could see what the BOOMERANG microwave telescope saw. The BOOMERANG map of the CMB is super-imposed to scale on the launch photo. Credit: BOOMERANG Collaboration.

This page titled 15.3: Observations of Variations in the CMB is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.4: Understanding the Variations in the CMB

Learning Objectives

- You will understand why the universe became transparent when the CMB formed
- You will understand acoustic oscillations

? What Do You Think: What Causes the Variations in the CMB

Out of the relatively smooth early Universe formed the lumpy Universe that we see today, with its galaxies, clusters and stars. How did this happen?

The Big Bang theory doesn't tell us how structures formed in the Universe, but it does provide a context in which they can form. We discussed how, once the Universe had cooled enough so matter and light stopped interacting strongly, the photons were free to travel freely through space. At the same time, the atoms were freed as well. For the first time, baryonic matter was able to begin clumping into what would become the structures we see today, falling into and enhancing the dense regions already occupied by the dark matter.

The CMB is the link between the smooth early Universe and the lumpy older one. As we saw in Section 15.3, while the CMB is overall remarkably uniform, there are slight variations in its temperature, called **anisotropy**. These temperature variations correspond to variations in the density of matter in the Universe at the time that the CMB formed. In the CMB anisotropy maps we have shown, blue spots correspond to colder, denser regions, and red spots to hotter, less dense regions. The temperatures, and therefore the densities from place to place are only different from each other by a tiny fraction: 10–100 parts per million.

The seeds for the variations in the CMB were produced early in the history of the Universe as a by-product of the fact that the vacuum of space is not really empty. Energy is constantly appearing and disappearing in the form of particles that pop rapidly into and out of existence, a concept confirmed by laboratory experiment. These quantum fluctuations very early in the history of the Universe could have been the tiny seeds of structure that became the variations in the CMB and eventually the large-scale collections of galaxies that we see today. The tiny energy fluctuations could have been expanded to very large proportions during an early epoch of rapid expansion.

Before the formation of the CMB at an age of 380,000 years, before electrons combined with nuclei to form neutral atoms, the Universe was a soup of baryons (protons and neutrons), leptons (electrons and neutrinos), and photons. Oscillations much like sound waves in this primordial fluid were created by the pressure of photons resisting gravitational compression of the plasma. Variations in the density of the primordial soup from one place to another affected the temperature of the plasma, and therefore the light; regions of compression and rarefaction at the time the CMB formed correspond to cold and hot spots. As described above, the reason the denser regions appear colder in a CMB map is because denser regions have a stronger gravitational field. The CMB photons lose some energy escaping from that gravitational pull, and therefore have a longer wavelength, which corresponds to a lower temperature.

The exact anisotropy pattern produced depends on cosmological parameters, much like the tone quality of a musical instrument depends on the materials from which it is composed. For example, the amount of baryonic matter in the Universe, including both the luminous and dark baryonic matter, cannot by itself produce sufficient gravitational attraction to allow the small variations in density to grow to the density of galaxy structures. For this reason there must be non-baryonic matter, specifically cold dark matter, present. Without the cold dark matter we cannot explain the level of anisotropy in the CMB.

If the Universe was made only of baryonic matter we would see a different anisotropy pattern in the CMB. The reason for this is that while the baryonic matter and light were interacting strongly, the pressure of the light prevented the baryons from forming dense structures. This was not true for the dark matter because it does not interact with light. Dark matter was able to begin to collapse even while the light and regular matter were strongly coupled—assuming its temperature was cold enough for it to do so. Thus the seeds of our current structures were planted even while atoms were prevented from creating any structures. By examining the observed CMB anisotropy pattern, we can place limits on the amount of baryonic matter and cold dark matter in the Universe.

One of the ways we can examine the CMB is with a map, but another technique allows us to measure temperature differences at each angular scale. This is done by computing a power spectrum of the data (or map). The power spectrum is like a filter on the map. It describes how common variations in temperature with different angular sizes are on the sky. Scientists could express this in terms of the angle itself, but mathematically it is more efficient to give each angular scale a multipole number, l. Small values of l correspond to large angular scales (some greater than a degree) and larger values of l correspond to small angular scales (some much less than a degree). Animated Figure 15.11 shows how a map can be filtered at different angular scales to create a power spectrum.

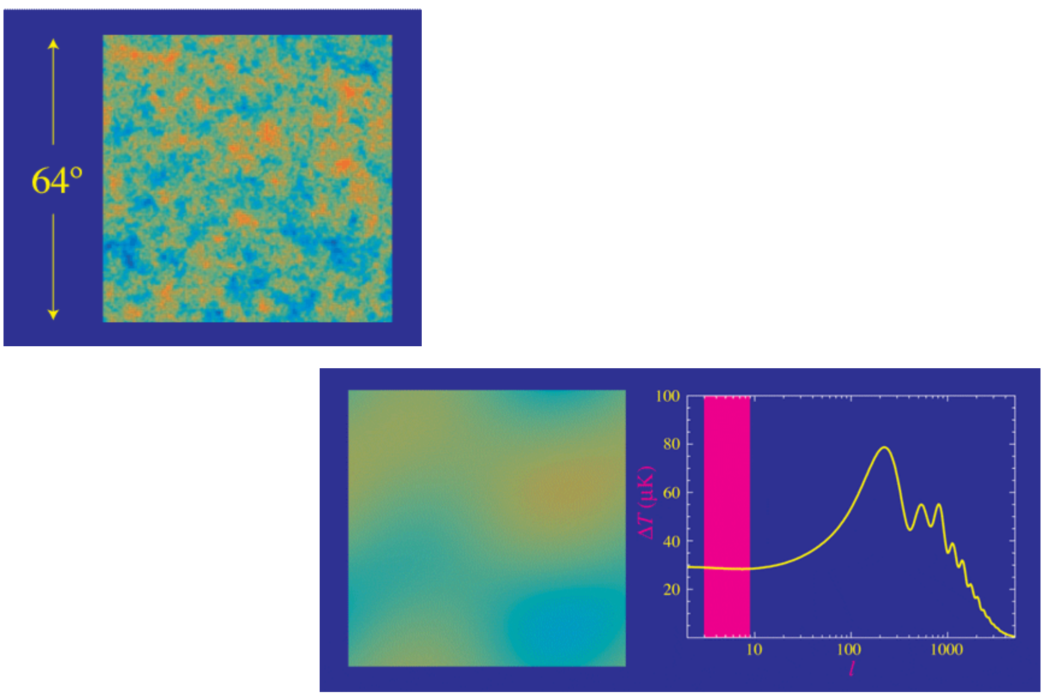


Figure 15.11: A map of the CMB can be decomposed into different angular scales, to create a power spectrum. Credit: Wayne Hu

The measured power spectrum in the sky is then compared to model power spectra that are computed for different values of cosmological parameters; some have more or less matter, others have different global geometry for the Universe, or they have varying values for the initial density fluctuations, etc. By varying all the different parameters it is possible to create a model with an impressively good fit to the data. Figure 15.12 shows the power spectrum and model fit for the Planck data, taken in 2013. The power spectrum peaks at an l of about 200 or an angular scale of about 1 degree. That angular scale is the most prominent scale in the Planck map, i.e. most of the spots are about a degree in size.

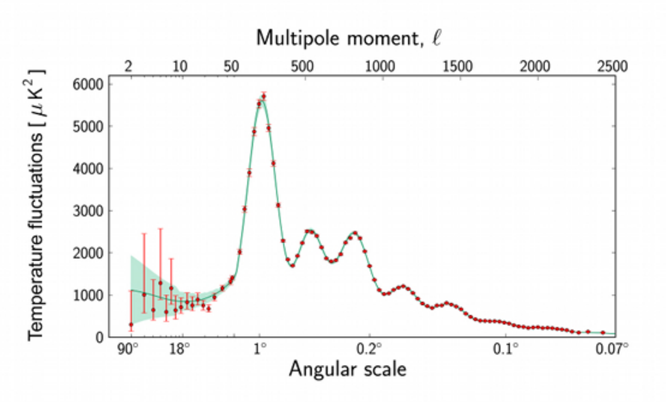


Figure 15.12: Power spectrum of the Planck anisotropy data. The data points are shown with error bars in red and the best fit model is shown in green. The Planck data are exquisitely precise and cover a wide range of angular scales. Credit: ESA and the Planck Collaboration

How Cosmological Parameters Affect the CMB

Going back to our musical analogy, the angular scales (l number) in the power spectrum are like musical tones. The power spectrum for the CMB is different depending on the contents of the Universe, similar to how the note "C" sounds different on a trombone and saxophone, because they are made of different materials.

In this activity you will explore how changing the amount of regular matter and dark matter affects the CMB power spectrum. When you play the activities, adjust the amounts by using the slider.

A. Regular Matter

1.

Play Activity

2.

https://phys.libretexts.org/@go/page/31442

B. Cold Dark Matter

1.

Play Activity

2.

3.

In the previous activity, you should have noticed several effects on the CMB power spectrum when changing the amount of regular (baryonic) matter and (exotic) cold dark matter (CDM). These changes can be explained by differences in how these types of matter interact and how structures grow. Baryons and CDM are similar in that they both have mass and therefore gravitational attraction. But only baryons interact electromagnetically. Therefore baryons, but not CDM, are part of the acoustically-oscillating fluid, along with photons and electrons.

Structures can begin to grow via gravitational attraction when they are smaller than the cosmic horizon scale. The cosmic horizon is the radius of the observable Universe, or how far light can travel based on the age of the Universe at that time. Essentially this is a size scale on which different parts of the Universe can exchange information with one another; within the cosmic horizon, all parts of the structure are able to influence each other via gravity and energy exchange. The acoustic mode corresponding to the first peak in the CMB power spectrum is the scale of the cosmic horizon size at the time the CMB formed. This marks a dividing line between what are considered "large" and "small" scales.

An important point in the history of the Universe, known as matter-radiation equality, is depicted in Figure 15.13. At this time, the density of radiation (photons) was equal to the density of matter. Before this time, radiation dominated the energy density of the Universe. Depending on the model, matter-radiation equality occurs when the Universe is about 50,000 years old. In a model with more matter, this time occurs sooner than if the amount of matter is lower.

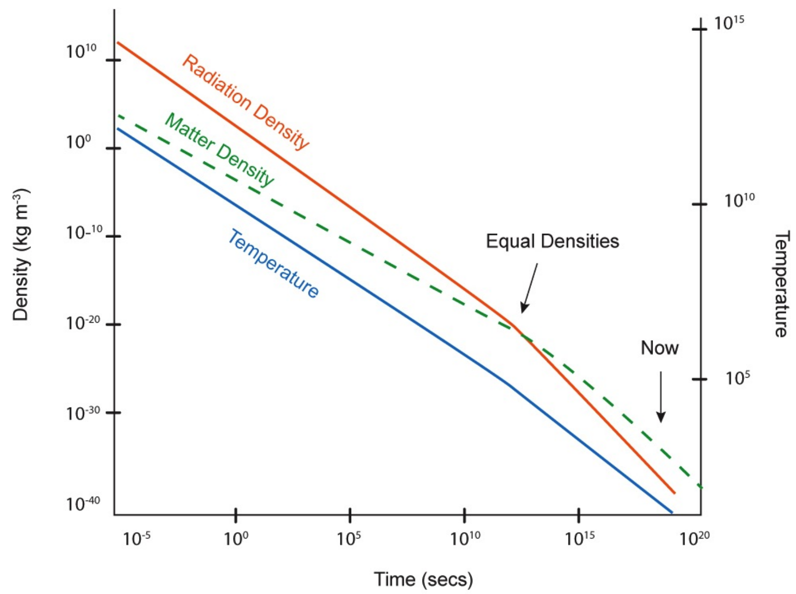


Figure 15.13 Matter-radiation equality equality occurs when the Universe is about 50,000 years old, when the densities of matter and radiation (photons) are exactly equal. Matter-radiation equality occurs before the time when neutral atoms form and the CMB begins free streaming at 380,000 years. Credit: NASA/SSU/Aurore Simonnet

So, for example, in a model with more dark matter, matter dominates earlier, meaning photons become less important at a relatively early time. The result is a lower amplitude seen in the peaks of the CMB power spectrum. You should have observed this effect in the last activity; more CDM meant a lower power spectrum. We can explain the effects of changing the baryon fraction on the power spectrum in terms of the oscillating baryon-photon fluid. First, if we add more baryons, we notice an overall increase in the amplitude of oscillations. We also notice that odd peaks are enhanced relative to even peaks. Odd peaks in the power spectrum are associated with regions of greater density (compression). Baryons are gravitationally attracted to regions of greater density. So, if there is a greater number of baryons, odd peaks will be relatively bigger and even peaks will be relatively smaller compared to the case of fewer baryons.

Finally, you may have noticed some subtle shifts of the power spectrum at large l values (small scales) where the power is dropping off. This part of the power spectrum is known as the damping tail. Damping of the oscillations occurs on scales smaller than the distance photons travel during the short time the CMB formed. In the activity, you may have noticed that increasing the baryon fraction shifts the damping tail to smaller scales (bigger l). Increasing the amount of cold dark matter shifts the damping tail to larger scales (smaller l) by increasing the age of Universe at the time neutral atoms form.

For further reading on the physics of how the CMB works, see Wayne Hu's CMB Tutorials.

Going Further 15.2: CMBPolarization

Light is a wave. It is made up of oscillating electric and magnetic fields. These fields have both a strength and a direction. In electromagnetic waves, the electric and magnetic fields always point in a direction perpendicular to the direction the wave travels. They also are perpendicular to each other. See Figure B.15.3 for an illustration of the geometry.

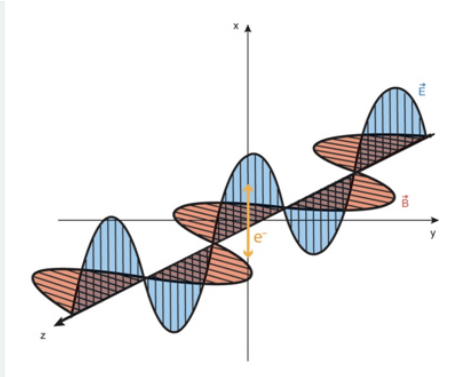


Figure B.15.3: Schematic of an electromagnetic wave. An electromagnetic wave traveling in some direction, taken here to be the z-axis, has an electric field (E, in blue) that points along the x-axis and a magnetic field (B, in brown) that points along the y-axis. The fields oscillate up and down but always remain perpendicular to each other and to the direction the wave travels. If an electron is sitting near the origin, it will feel a force as the wave passes, oscillating up and down. Credit: NASA/SSU/Aurore Simonnet

In Figure B.15.3 there is an electron depicted sitting near the origin. When a light wave passes by as shown, the electron will feel a force from the electric field in the wave. It will alternately be pushed up and then down along the x-axis, oscillating with the passing wave. An electron bouncing up and down will radiate electromagnetic waves - accelerated charged particles always emit electromagnetic waves. The emitted waves will have the same frequency at which the electron oscillates, and in this case that is the same frequency as the incident wave. The energy radiated by the electron must come from somewhere, and the only source of energy the electron has is the incident light wave. So what the electron is doing is absorbing some of the energy of the incident wave and re-radiating it in different directions. We call this scattering.

In the case that we have described here, scattering preserves the frequency of the incident wave, merely directing some of the wave's energy into new directions. Scattering does not always preserve the frequency of the incident light- it can increase or decrease the frequency depending on the details of the scattering material (whether it is hot, cold, in motion, etc.). We discussed this type of scattering in **Going Further 15.1: The Sunyaev-Zel'dovich (SZ) Effect**.

Now we will examine the direction of scattered light. The intensity of the electromagnetic waves radiated by the electron (in other words, scattered by the electron), will depend on the amplitude of the motion of the electron as seen by an observer in that direction. For instance, if you are an observer sitting somewhere in the y-z plane, you will see the electron move up and down to its full extent, and the intensity of the emitted radiation will be maximal. If you are somewhere above or below the y-z plane you will see a smaller amplitude for its motion since some of the motion is directly toward or away from you, and you cannot see that. In the extreme case that you sit on the x-axis directly above or below the electron, you will not see it move at all. For such an observer the intensity of the emitted radiation is zero. In other words, there is no radiation emitted directly along the x-axis for the situation depicted.

Light is said to be polarized if the electric field only oscillates in one direction. We will now describe how scattering can polarize light. Imagine that you are an observer watching the scattering of light from the electron, and you are positioned such that your line of sight to the electron makes a 90 degree angle with the direction the wave travels, which puts you somewhere on the x-y plane. If you are aligned along the x-axis you will not see any of the scattered light from the electron because its motion is directly toward and away from you. If you are on the y-axis you will see maximum intensity for the scattered light. In other places in the x-y plane you see intermediate intensities. The geometry is depicted in Figure B.15.4.

https://phys.libretexts.org/@go/page/31442

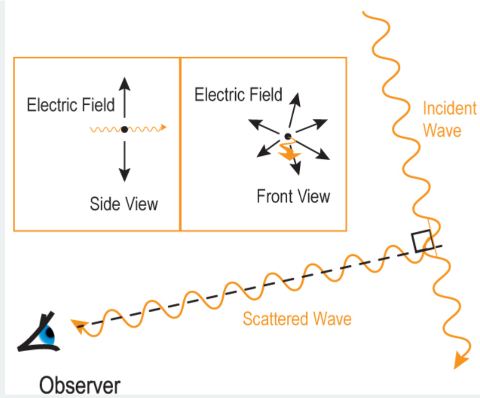


Figure B.15.4: An electromagnetic wave, incident from the lower right, is shown in this image. Part of the wave is scattered by an electron through a 90 degree angle toward an observer. From the point of view of someone looking toward the incident wave, all orientations of the electric field are visible, as shown in the inset, at right. However, to an observer of the scattered wave, only those directions of the electric field that are perpendicular to the sightline are visible. As a result, the observer of the scattered wave sees light polarized in a direction perpendicular to the direction of travel of the incident wave. Credit: NASA/SSU/Aurore Simonnet

What is interesting about the geometry described is that the light you see will always have its electric field along the x-axis. Since the incident light is propagating in the z-direction, there is no component of the electric field for the incident ray in that direction, so the electron does not move in that direction. It can move along the y-direction, but if you are on the y-axis you cannot see that motion. The scattered light you see therefore will have only one direction for its electric field, no matter what the direction of the electric field of the incident light. Another way to think about this is that it does not matter how the x and y axes are aligned with respect to the incident light wave: as long as you can see some light it will be polarized.

You can see this effect in the daytime sky on Earth, just look at the sky in a region about 90 degrees away from the Sun using polarized sunglasses. Tilt your head slightly. You will see the intensity of the sky change as the polarized filters in the glasses block more or less of the light, depending on the orientation of the filters and the electric field in the scattered sunlight.

When light his a preferential orientation for its electric (and magnetic) field, we say it is polarized. The process of scattering tends to polarize light. Of course, there is a lot of scattered light in Earth's atmosphere, so when you look at a point in the sky there will generally be light coming from different directions, with only a fraction coming directly from the Sun. That is why the sky does not become totally dark when you look at it through polarized sunglasses. However, the incident light itself is not polarized. Different waves have different orientation of their electric fields, and it is only the process of scattering that polarizes the light.

This same scattering process occurred early in the Universe as the CMB was being produced. In general, at any given point there was light coming from all directions. This is a different situation from the case of sunlight in the atmosphere, where there is a preferential direction toward the source (the Sun). The polarization of light from one direction with be cancelled by the polarization of light coming from directions 90 degrees away, and so no net polarization occurs under this circumstances. However, there is a possible way to break the symmetry and get a net polarization of the scattered radiation.

To understand where this net polarization comes from, imagine that the fluid containing the scattering electrons is moving relative to the frame of the CMB. In the direction of the motion of the fluid the photons will be blueshifted, and in the direction opposite the motion the radiation will be redshifted by the same amount. In intermediate directions the redshift and blueshift will be less, with no shift occurring in directions perpendicular to the direction of motion of the fluid. This introduces an asymmetry in the radiation field. Alone, it does not produce a net polarization of scattered radiation. The incoming radiation from the forward direction, which has higher than average intensity due to its blueshift, causes scattered radiation with higher than average intensity. On the other hand, the incoming radiation from the backward direction, which has equally diminished intensity because of its redshift, creates scattered radiation with lower than average intensity. The two scattered components combine to give a net intensity that is identical to the average, and so no net polarization is seen. The same is true for directions oblique to the direction of motion: radiation from opposite sides of the sky combine to give no net polarization.

Now consider what happens if the electron is in a place where the gas on one side is falling toward it and the gas on the opposite side is also falling toward it (Figure B.15.5). In perpendicular directions the gas can be stationary, moving away or moving inward at a slower speed . Given any of those scenarios, the scattering electron will see hot spots in two opposite directions and (relatively) cool spots in directions 90 degrees from those. Now the light incoming from the hot spots does not

https://phys.libretexts.org/@go/page/31442

average out to give the typical temperature since the radiation from both directions has a higher intensity than the average. Similarly, light coming in from the cool places gives a lower intensity. Thus an observer looking at this region will see a polarization signal with either a slightly higher than average temperature or a slightly lower than average temperature, depending on how the gas was moving when the microwave light was emitted and the relative direction to the observer.

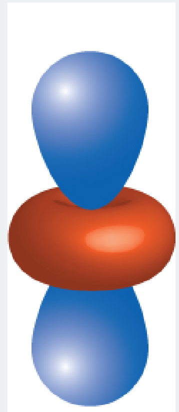


Figure B.15.5 CMB Polar Producer. This gives an idealized view of what the sky looks like to an electron that could produce polarized light in the CMB. The electron is imagined to be in the center of this shape. Along two directions opposite in the sky, the blue bubbles here, the temperature would appear hotter than average. In perpendicular directions, illustrated with a red torus, the sky would be cooler than average. This is one example of a temperature configuration, from the perspective of the electrons scattering the light, which would cause a net polarization in the scattered waves. Credit: NASA/SSU/Aurore Simonnet

This is a powerful technique. It allows astronomers to determine if the gas that created the CMB was collapsing in some places and expanding in others, and it gives them detailed information about how that collapse was occurring in different parts of the sky and on different angular scales. Since our theories of structure formation depend on gas collapsing into high density regions to form the structures we now see as galaxy clusters and filaments, careful observations of the polarization of the CMB give us important checks on these theories.

The polarization signal is quite weak. Power spectra of CMB polarization data are at levels about 100 times lower than CMB temperature anisotropy, where fluctuations are already at a level of 1 part per 10,000. Operating from the South Pole, DASI was the first group to detect polarization in the CMB in 2002. CMB polarization has since been detected by several other groups.

Finally, the polarization of the CMB is also sensitive to a small percentage of the Universe being ionized once stars started forming. This tells us that the first stars must have turned on around a redshift of around 10, when the Universe was about 400 million years old.

Going Further 15.3: the Steady State Theory Cannot Explain the CMB

In searching for your own understanding of the Universe, you can weigh the observational evidence for the Big Bang plus structure formation model presented in this module.

Here we will compare the explanatory power of the Big Bang theory with that of the Steady State theory. The Big Bang theory incorporates a beginning to the Universe as well as its evolution. An alternative, the Steady State theory, was proposed in 1948 by Hermann Bondi, Thomas Gold, and Fred Hoyle. The Steady State theory also incorporates the expanding Universe. Clusters of galaxies move farther apart over time, and the spaces between them enlarge. To keep the Universe looking the same, new matter, which condenses into new galaxies, must continually be created within the spaces. We cannot rule out such creation with laboratory tests. To fill the space left by the expansion of the Universe requires that three or four new atoms appear per year per cubic kilometer . A cubic kilometer of air at sea level contains 1037 atoms, rendering the few new ones completely undetectable.

How can we discriminate between the two theories? The Big Bang theory predicts the cosmic microwave background as the cooled remnant of the hot, dense phase of the early Universe. The Universe therefore appears to be changing with time. This conclusion is also supported by the observed evolution of galaxies and quasars as we look to distant and therefore younger reaches of the Universe. Though to be fair, Steady State models generally did not accept the common interpretation of the cosmological redshift, and so the evolution of the quasar population could be circumvented, though in a way that failed to convince most astronomers.

Steady State cosmologies had particular difficulty explaining the CMB, as there is no reasonable way to explain the perfect blackbody spectrum of the radiation observed in all directions. The CMB was thus seen as the death knell for Steady State cosmological models. Since that time evidence has accumulated for the evolution of the galaxy population and large scale structures in the Universe, both of which are incompatible with an unchanging Universe. Essentially no one takes Steady State models seriously anymore.

This page titled 15.4: Understanding the Variations in the CMB is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.5: Comparing Models and Data - The CMB and the Curvature of Space

Learning Objectives

- You will be able to compare models for the effect of the curvature of the Universe on CMB maps and power spectra and choose which one best fits the data

What Do You Think: Curvature of the Universe

There are several possibilities for the geometry of the Universe within the theoretical framework of the Big Bang. General relativity predicts that matter and energy curve spacetime. This idea of curvature applies to not just the space around individual objects but also the Universe as a whole; the entire matter and energy content of the Universe will bend the spacetime of the Universe as a whole. The global curvature of the Universe can be described by the Friedmann equation, which includes a term (called k) describing the global curvature of the Universe as well as terms for the expansion and density of the Universe. The global curvature term k is one of the things that affects the anisotropy pattern in the CMB. By comparing models of different curvatures with CMB data, we can determine the geometry of the Universe.

Recall for a moment the subject of Euclidean geometry, the geometry of the Pythagorean theorem. This is also the geometry in which parallel lines never intersect, angles of a triangle add up to 180 degrees, straight lines are the shortest distance between two points, and standard trigonometry applies. Space with this type of geometry is often referred to as flat. Flat in this case does not necessarily mean two-dimensional like the surface of a table or flat like a pancake, it just means that normal Euclidean geometry applies and that there is no overall curvature. A flat space in this sense can have three or even more dimensions.

Although most situations we are used to involve flat space, or at least a close approximation to it, there are a number of situations in which this is not the case. For example, if we consider large enough distances, Earth's surface is definitely not flat because Earth is a sphere. Although for short distances a flat geometry works pretty well to describe the relationship between events on Earth (think of a standard city map), for large distances we need to use a curved geometry if we want to accurately describe the relationships. In a spherical geometry, such as the surface of Earth, parallel lines converge and the angles of a triangle can add up to more than 180 degrees. A space with a spherical geometry is said to have a positive curvature.

Another curved geometry is a hyperbolic geometry, in which space is saddle shaped. In such a space parallel lines diverge and the angles of a triangle add up to less than 180 degrees. This type of space is said to have negative curvature.

The shortest distance between two points is not a straight line in curved spaces. The more general name for the shortest distance between two points is a **geodesic**. In a flat space, a geodesic is a familiar straight line. An example of a geodesic on a curved surface is a great circle on the surface of Earth—like lines of constant longitude. A great circle is the route an airplane flies between two cities because it is the shortest distance between them. That is why flights from North America to Europe generally pass over the Arctic. Have a look at an Earth globe and you will see that the shortest distance between the United States and Europe does indeed pass over the Arctic. To map out a great circle route, put a string on a globe at two points (say, your hometown and a distant place you might like to go on vacation—Rome, maybe, or Tahiti). On a globe the string will stretch directly from the beginning to the end of your journey, but the route will look curved when projected onto a flat map of Earth such as the one in the in-flight magazine on your trip. Figure 15.14 illustrates spherical, flat, and hyperbolic spaces.

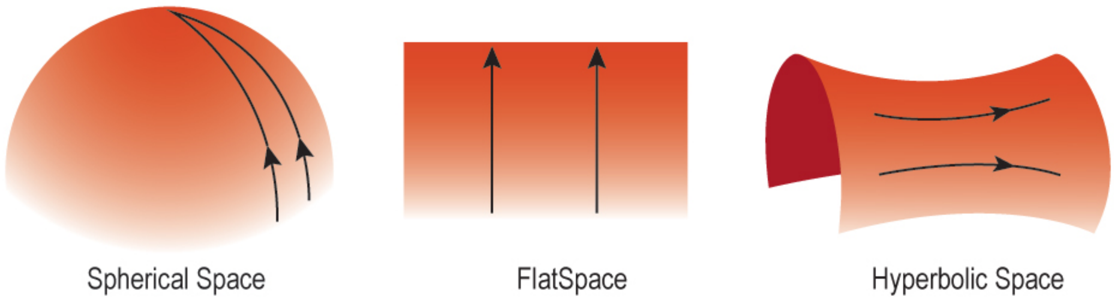


Figure 15.14: Geometry in spherical, flat, and hyperbolic spaces. In a spherical space, lines that start off as parallel eventually intersect. In flat space, parallel lines stay parallel. In a hyperbolic space, lines that start off as parallel eventually diverge. Credit: NASA/SSU/Aurore Simonnet

If spacetime is curved, light from an object will follow the curvature of spacetime; it will travel along a geodesic. Thus an object will appear to be a different angular size depending on how space is curved. This is analogous to looking at yourself in a fun house mirror or a car mirror that says "objects may be closer than they appear," both of which have curved surfaces that distort the images they produce. Unlike the case of a mirror, in which the light travels straight paths and only the curvature of the reflecting surface gives us the distortions we see, when space itself is curved it is the curved path of the light as it travels that causes the distortions. If space is positively curved, an object will appear bigger to us than the same object seen in a flat space. If space is negatively curved, an object will appear smaller than if space is flat. Figure 15.15 illustrates the path of light in spaces of various curvatures.

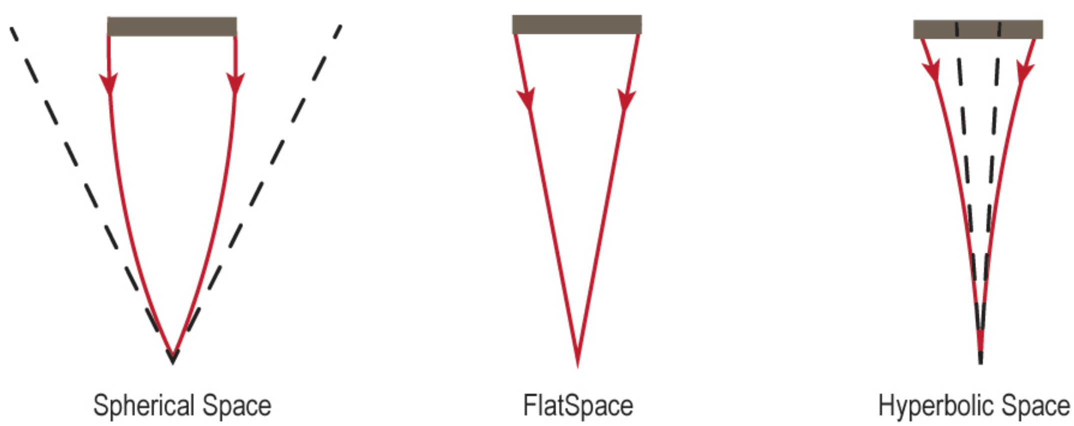


Figure 15.15: The path of light from an object to an observer in (a) space which is positively curved, (b) space which is not curved and (c) space which is negatively curved. An object of a given size appears bigger in a positively curved space and smaller in a negatively curved space than in a space that is not curved. Credit: NASA/SSU/Aurore Simonnet

One of the ways we can measure the curvature of the Universe is through anisotropy in CMB; the hot and cold spots in the CMB would appear bigger in a spherically curved Universe than in a Universe that is not curved. In 2000, the BOOMERANG Collaboration used this technique to determine the geometry of the Universe. In the next activity, you will compare their data to three theoretical possibilities in order to determine the overall curvature of the Universe.

https://phys.libretexts.org/@go/page/31483

> Determining the Geometry of Space

In this activity you will compare CMB data from BOOMERANG with theoretical models in order to determine whether the geometry of the Universe is spherical (positive curvature), flat (zero curvature), or hyperbolic (negative curvature).

A. Comparing Maps

Figure A.15.2 shows three theoretical possibilities for how the CMB anisotropy pattern might appear in a map of the sky depending on the curvature of the Universe.

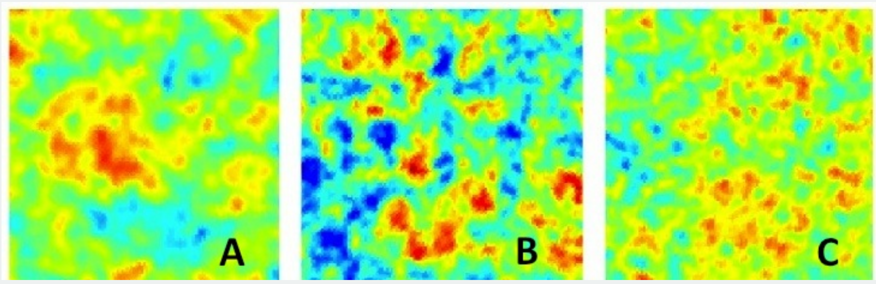


Figure A.15.2: Simulated maps of the CMB differ based on the geometry of the Universe. Credit: BOOMERANG Collaboration.

1.

2.

3.

4.

Figure A.15.3 shows the BOOMERANG data. We can compare it with the models.

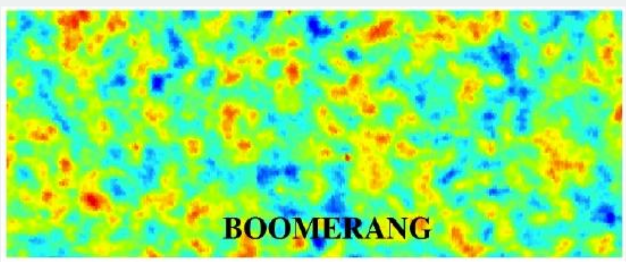


Figure A.15.2 to determine the overall curvature of the Universe. Credit: BOOMERANG Collaboration.

5.

https://phys.libretexts.org/@go/page/31483

B. Dependence of the Power Spectrum on Curvature

Figure A.15.4 shows three theoretical possibilities (plotted as lines) for how the CMB anisotropy power spectrum might appear depending on the curvature of the Universe.

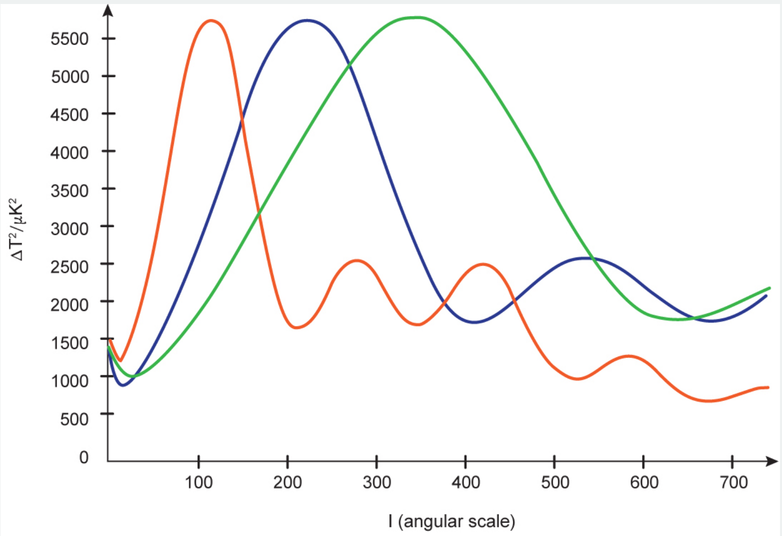


Figure A.15.4: The CMB power spectrum changes depending on the geometry of the Universe; in particular, the angular scale of the first peak shifts. Larger I values correspond to smaller angular scales. Three theoretical models are shown: spherical (pink), flat (blue), and hyperbolic (green). Credit: NASA/SSU/Aurore Simonnet. Models run with CAMB software through NASA's LAMBDA interface.

1.

2.

3.

4.

C. Comparing Power Spectra

Figure A.15.5 shows the CMB power spectrum as measured by BOOMERANG.

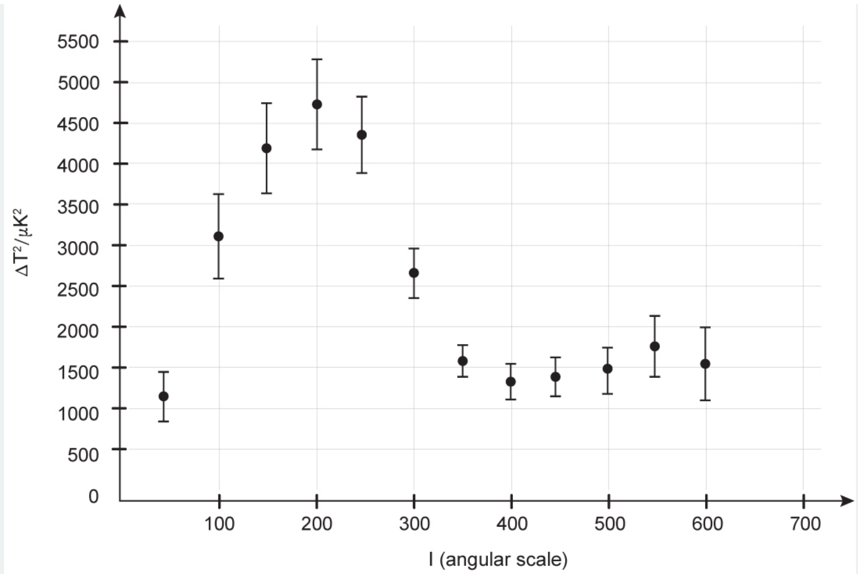


Figure A.15.5: BOOMERANG measurement of the CMB power spectrum . Credit: NASA/SSU/Aurore Simonnet based on BOOMERANG data from deBernardis et al. (2000), *Nature*, 404, 955.

1.

2.

In the last activity, you should have determined that overall, the Universe is not curved, it is flat. The BOOMERANG result of zero curvature has since been confirmed by a number of other observations, including some made with much more sensitive instruments, to better than 0.5% precision! Again, the technical term for “not curved” is “flat.” This does not mean two-dimensional, like a pancake; the Universe has at least three spatial dimensions. “Flat” also does not mean thicker in any particular dimension, it just means not curved. Stars, galaxies, and other objects still curve space locally, but these CMB observations indicate that globally, space is not curved.

This page titled 15.5: Comparing Models and Data - The CMB and the Curvature of Space is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.6: Wrapping It Up 15 - Using CMB Data to Determine Cosmological Parameters

Learning Objectives

- You will be able to use CMB anisotropy power spectra to compare to data with models and determine parameters (baryon fraction, cold dark matter, curvature) to demonstrate their understanding of how models compare with data. Parameters will be adjusted individually rather than jointly in this activity. You will further explore the telescopes used to study the CMB.

> *"We know of an ancient radiation*
>
> *That haunts dismembered constellations*
>
> *A faintly glimmering radio station..."*
>
> *– Cake*

15.6.1: I. Cosmological Parameters

In the first part of this activity you will combine what you have learned about the CMB power spectrum in order to determine the values of several cosmological parameters: the amount of regular matter (the baryon fraction), the amount of cold (exotic) dark matter (CDM), and the curvature of space overall.

You will do this by comparing theoretical models to data from the most recent satellite, Planck. Though scientists actually vary all of the parameters at once in their analysis, you will compare each parameter individually in this activity for now.

For simplicity, we have not shown all of the models. We have provided models for only a few values of the parameters, but scientists can compute additional models for greater precision and better fits.

15.6.1.1: A. Curvature

Play Activity

1.

2.

15.6.1.2: B. Amount of Baryonic Matter

Play Activity

1.

2.

3.

15.6.1.3: C. Amount of Cold Dark Matter

Play Activity

1.

2.

3.

15.6.2: **II. CMB Observations: A Scavenger Hunt**

In this chapter you have learned about several key things astronomers measure using the CMB. NASA keeps a list of CMB experiments, with basic information about each, as well as links to websites of each of the CMB groups. Browse the list and answer the following questions.

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

11.

12.

This page titled 15.6: Wrapping It Up 15 - Using CMB Data to Determine Cosmological Parameters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

15.7: Mission Report 15: Using CMB Data to Determine Cosmological Parameters

Learning Objective

- Pull together key concepts from the wrapping it up activity and chapter.

A.

B.

C.

D. Questions to be graded for accuracy. Show your work!

1.

2.

3.

4.

This page titled 15.7: Mission Report 15: Using CMB Data to Determine Cosmological Parameters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

16: The Early Universe

Chapter 16 travels back in time to the earliest moments of the Universe. The chapter starts with the formation of the lightest chemical elements, and continues back to the extreme temperatures and densities at which the first subatomic particles formed. The last sections of the chapter discuss cosmic inflation and the limits of our knowledge at the very start of the Universe’s existence.

Thumbnail: Artist's interpretation of the Big Bang, with representations of the early universe and its expansion. (Public Domain; NASA's Goddard Space Flight Center/CI Lab)

This page titled 16: The Early Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

16.0: The Early Universe Introduction

Video Transcript

> ***Recreating The Big Bang: Transcript***
>
> *The Universe came into existence 13.7 billion years ago with the event known as the Big Bang. We have been able to see back, close to the beginning, using telescopes to capture distant light emitted billions of years ago. The famous Hubble Space Telescope has allowed us to look back when galaxies were still in the process of forming.*
>
> *The WMAP provided us with an amazing view of the cosmic microwave background – the afterglow of the Big Bang. Before this point, the Universe was too hot and light was unable to shine through the dense fog of charged particles. This prevents telescopes from being able to see beyond this point. However, there is a way to recreate the conditions of this early Universe right here on Earth.*
>
> *The conditions of the early Universe were of extreme heat and energy, which lead to the creation of strange forms of matter and physics. For decades, particle physicists have been recreating these early conditions with particle accelerators, smashing particles together to create similar levels of heat and energy. Particle detectors are then employed to capture these events to help us learn more about what the Universe was like in these early times.*
>
> *ATLAS is such a detector at the Large Hadron Collider. Late in 2009, the LHC and the ATLAS detector broke world records by creating and detecting the highest energy collisions ever made. In the years to come, it will reach its full capacity and push to even higher energies and further back in time. ATLAS will be discovering new kinds of particles, such as the Higgs, and unlocking new realms of physics that may include new dimensions of space and particles of dark matter.*
>
> *This frontier of knowledge will continue to be pushed to help us learn more about the origins of our Universe and what happened a fraction of a second after the Big Bang.*

Credit: ATLAS Experiment/CERN

By studying large-scale structure and cosmic microwave background, we have traced the history of the Universe all the way back in time to when the light left the farthest objects we can see with our best telescopes. We saw that there is a limit to our lookback time, because until about 380,000 years after the Universe began it was opaque. The video shows that we can investigate much earlier times by reproducing their conditions on a tiny scale here on Earth; for this we use huge particle accelerators.

The very early Universe was extremely hot and dense. Remember that temperature is just a measure of the kinetic energy of particles, so it is a measure of their speed. Particle colliders like the Large Hadron Collider (Figure 16.0.1) can use pulsing electromagnetic fields to accelerate charged particles, usually protons or electrons, to close to the speed of light. Such speeds correspond to those typical of the entire Universe around 10^{-14} seconds after the Universe came into existence. At that time the temperature was 100 million billion (10^{17}) kelvin! Although we only look at the characteristics of a few particles at a time using our colliders, we are confident that we can use what they show us to understand the Universe at early times. Complex structures like molecules and atoms cannot exist at high temperatures, so the physics of the early Universe would have been very simple. Collisions between elementary particles would have dominated - a situation analogous to what we can study with accelerators.

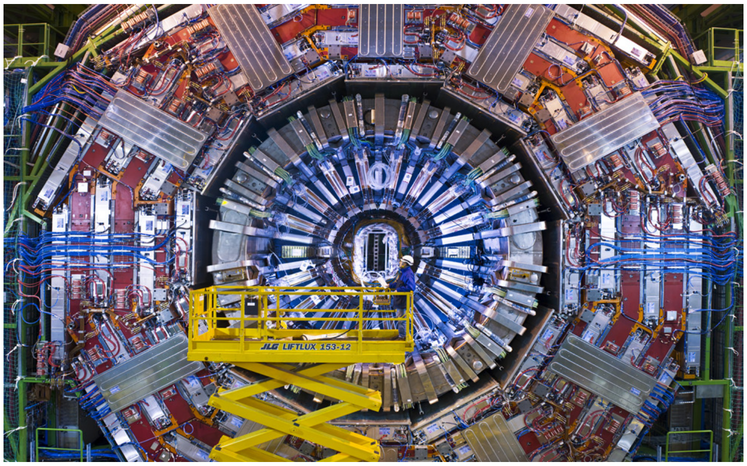


Figure 16.0.1: The CMS (Compact Muon Solenoid) experiment at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland. CMS is an experiment at the Large Hadron Collider (LHC), currently the most powerful particle accelerator in the world. Credit: CERN/Maximillien Brice

Going Further 14.1: Studying the Early Universe Using Particle Colliders

There are two basic kinds of charged particle accelerators. One kind was so common a few years ago that many ordinary people owned them! The cathode ray tube of old televisions is an example of an electrostatic accelerator. It uses the force from an electric field to accelerate electrons. The electrons then hit a screen and release the photons that we see.

Much higher energies can be achieved using oscillating electric fields. In these machines, the charged particles are accelerated for a short distance and then the electric field changes direction to give them an extra kick. As they pass through a hole in a plate, the electric polarity is switched so that the plate repels them and they are accelerated toward the next plate. In order to accelerate particles to near the speed of light, microwave cavities are used instead of simple plates, because the oscillation rate of the electric fields becomes higher in higher energy machines.

Oscillating field accelerators have two subtypes. Linear accelerators send their particles in a straight line. The particle paths can be up to several kilometers long. Circular accelerators, called synchrotrons, curve the particle tracks using powerful magnets. Synchrotrons take their name from the fact that the electromagnetic field oscillations must be synchronized to the particle orbital motions so that the particles receive a boost in energy each time they come around their circular path.

Only the largest machines are used for particle physics research; less powerful versions have many medical and industrial applications. The goal of the research machines is to produce the simplest kinds of interactions at the highest possible energies by colliding the accelerated particles with a target. Depending on the application, the target might be static or a beam of particles approaching from the opposite direction. Elementary particle physicists primarily use machines with beams of electrons and positrons, or protons and anti-protons. Nuclear physicists and cosmologists often use beams of bare atomic nuclei from lead or other heavy atoms that have been stripped of electrons. These heavy ions are better suited to investigating the structure, interactions, and properties of the nuclei themselves. They can also reveal the properties of condensed matter at extremely high temperatures and densities, such as might have occurred in the first moments of early Universe.

The largest linear accelerator (LINAC) built thus far was at the Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory in Menlo Park, CA. The LINAC, shown in Figure B.14.1, operated from 1966 until 2012. Its main accelerator was 2 miles long, and it accelerated electrons and positrons. These were made to collide in a storage ring built near the end of the LINAC. SLAC's experiments were critical in illuminating the physics of the electroweak interaction as well as probing the structure of quarks inside nucleons.

Figure 16.0.1: An aerial view of the SLAC linear accelerator. The accelerator passes under Interstate 280, which can be seen in the middle distance. Credit: SLAC National Accelerator Laboratory

Linear accelerators are well-suited to collide low mass particles such as electrons and positrons. Ring-shaped colliders, such as the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Lab on Long Island, New York, the Tevatron at Fermilab Fermilab near Chicago, and the Large Hadron Collider (LHC) at CERN near Geneva, Switzerland, all shown in Figure B.16.2, can accelerate particles for many circuits of the ring, making them better suited to accelerating heavier particles. Heavier particles require many circuits of the ring to accelerate to near light speed. Very high energies are achieved in the process.

The continuous deflection in circular accelerators causes the particles to lose energy by emitting photons. The radiation emitted is called synchrotron radiation and takes its name from the accelerator type. Radiation losses can be decreased by making the circular rings larger but cannot be completely avoided. Thus there is an upper energy limit on the particle energies that can be produced in accelerators.

Figure 16.0.1: AERIAL VIEWS OF RHIC, FERMILAB, AND LHC. (Top) In the Relativistic Heavy Ion Collider (RHIC), collisions between heavy atomic nuclei are used to study fundamental particle interactions and conditions in the early Universe. (Middle) The Tevatron, at the time of its construction in the late 1970s, was the largest particle accelerator in the world. It collided beams of protons and anti-protons to achieve energies of 1 TeV (10^{12} eV), hence its name. It was shut down on September 30, 2011. (Bottom) The Large Hadron Collider (LHC) accelerates protons and heavy atomic nuclei in a circular synchrotron, achieving energies of 7 TeV. It is currently the most powerful accelerator in the world. Credit: Brookhaven National Lab, Wikimedia Commons, CERN/Maximilien Brice.

The highest energy machines are usually complicated. They employ linear accelerators to inject particles into circular synchrotrons, and storage rings are used to maintain particles at their high energies. The machines are typically buried underground to provide shielding against the particles generated by the collisions. Huge and complex arrays of detectors are arranged around the points where the collisions occur. These are able to both track the resulting particles and to absorb them to measure their energy.

Large machines like the ones pictured can accelerate particles to energies that were typical in the Universe in its first fraction of a second; the current generation of machines are probing back to about the first hundred trillionth of a second (10^{-14} seconds).

This page titled 16.0: The Early Universe Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

16.1: The Formation of the Lightest Elements

Learning Objectives

- You will know the overall abundances of the elements.
- You will know how we measure abundances (especially of the lightest elements).
- You will be able to determine which abundances are primordial.
- You will be able to explain how the neutron to proton ratio = 1/7 causes basic proportions of ¾ H and ¼ He.
- You will understand the effect of baryon-to-photon ratio on the abundances of the light elements.
- You will know that the measured abundances of the light elements can all be explained by Big Bang Nucleosynthesis.

What Do You Think: Where Did the Elements Come From?

We have seen that the Universe is hotter as we look further back in time. Figure 16.2 shows how the temperature of the Universe changes with time. In this chapter, we will see how conditions in the early moments of existence are shaped by this change in temperature.

https://phys.libretexts.org/@go/page/31445

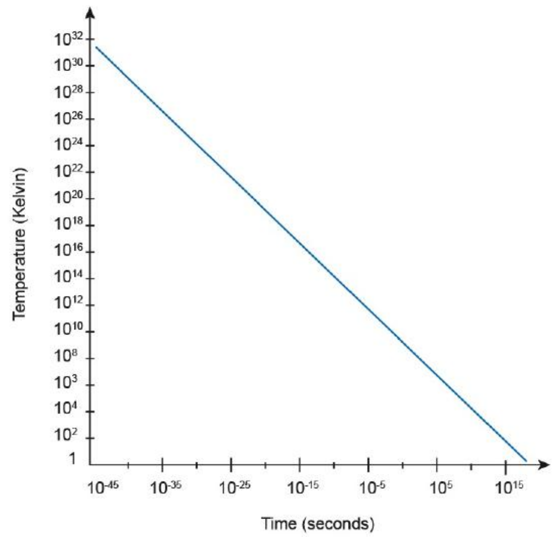


Figure 16.2: As the Universe ages, it cools. At early times, it was extremely hot everywhere. Credit: NASA/SSU/Aurore Simonnet

During a period from about 0.001 seconds to about 3 minutes, when temperatures were billions to millions of kelvin, our ideas of what conditions in the Universe were like are still on fairly firm ground. The relevant physics at these temperatures is well-understood from laboratory experiments, theoretical models, and astronomical observations. We know that the conditions in the first few minutes of the Universe were just right for nuclear fusion to form the very lightest elements: hydrogen, helium, and lithium.

16.1.1: Observations

As we gaze out into the night sky, we wonder where the stuff we are made of came from and whether the rest of the Universe is made of the same stuff as we are. Remarkably, it is possible to answer these questions through observations. We can use spectroscopy to determine the chemical composition of planets, stars, and other objects in the Universe. We accomplish this by looking at their light, using the fact that each chemical element has a unique spectral signature.

As we study the distribution of the chemical elements in the Universe, we can also ask where the elements came from, and why there are different amounts of each. From studies of meteorites, the atmospheres of other planets in the Solar System, the Sun and other stars, as well as the dust and gas between stars and even between us and distant quasars, we have determined the abundance of the various elements in the Universe. Figure 16.3 summarizes our findings in terms of the numbers of each element relative to the others. Today, hydrogen makes up about 74% of the mass (and 92% of the number) of the atoms in the Universe. Helium makes up about 24% of the mass of atoms (8% of the number). The rest, only 2%, is all of the other elements.

In contrast, Earth is made primarily of heavier elements; hydrogen and helium are only tiny constituents in Earth's composition, the surface of which is mostly silicon and oxygen. There are also significant amounts of aluminum, iron, calcium, and a few others. Earth's atmosphere is primarily nitrogen and oxygen, and most living organisms are substantially composed of carbon and its compounds, along with water (hydrogen and oxygen). So hydrogen is only a small fraction of Earth's mass, and the same is true of helium. Chemically speaking, Earth is highly atypical of the Universe as a whole.

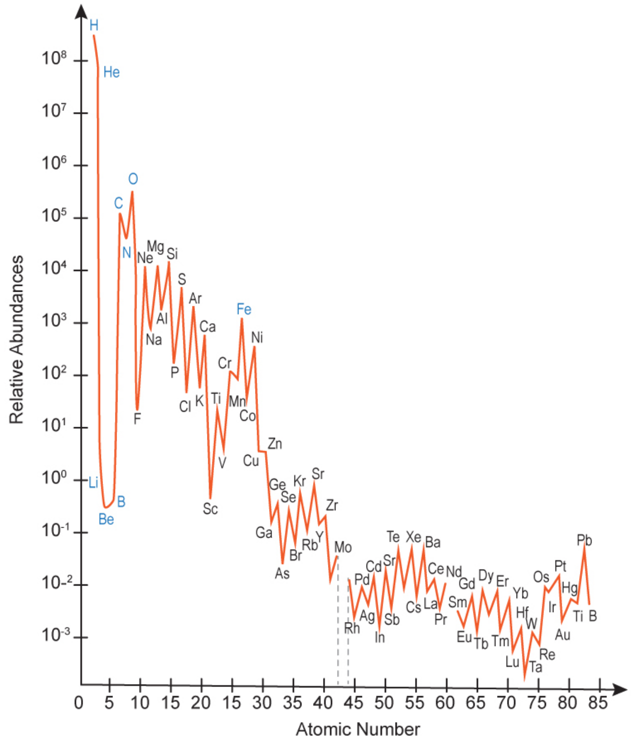


Figure 16.3: The relative abundances of the elements vs. their atomic numbers. The chemical symbols for the elements are given next to their abundance levels. Hydrogen and helium are by far the most abundant elements in the Universe, in contrast to carbon, nitrogen, and oxygen, which are found in abundance on Earth. Elements highlighted in blue are discussed in this chapter. Credit: NASA/SSU/Aurore Simonnet

Some of the most intriguing celestial observations are of the lightest elements: hydrogen (H), helium (He), lithium (Li), and the isotope of hydrogen known as deuterium (^{2}H; 1 proton and 1 neutron, often denoted D). We never measure less than about 23% of the mass of atoms being in ^{4}He, whether it is measured in stars, nebulae, or galaxies. We also always see some deuterium- at least 1 part in 100,000 (10^{-5}) compared to regular hydrogen, whether it is measured in the molecules of Jupiter's atmosphere, the local interstellar medium, meteorites, or distant quasars - despite the fact that deuterium is destroyed inside stars, not produced there. We measure the ^{3}He abundance in nebulae, the oldest meteorites, lunar soil, and the solar wind, and it also has an abundance about 10^{-5} compared to hydrogen. The same process that destroys deuterium in stars creates ^{3}He, so we have a check on our model for deuterium. The observations of these elements lead us to believe they were formed somewhere other than in stars.

We can also compare the abundance of light elements (such as lithium, beryllium, and boron) in stars to the abundance of the heavy elements (such as iron) in stars. Stars create heavier and heavier elements throughout their lives as the result of nuclear fusion in their cores. When they die, they eject the elements into the surrounding interstellar medium and new stars (and planets) form from the enriched material. This means that stars with a lower abundance of heavy elements must have been created at earlier epochs in the Universe, before many generations of stars had sprinkled the interstellar gas with the products of their nuclear fusion. Stars with a higher abundance of heavy elements must have been formed later.

Figure 16.4 shows the spectra of several stars, each from a different generation. The abundance of heavy elements increases over time; thus, the abundance of heavy elements can be used as a marker for time.

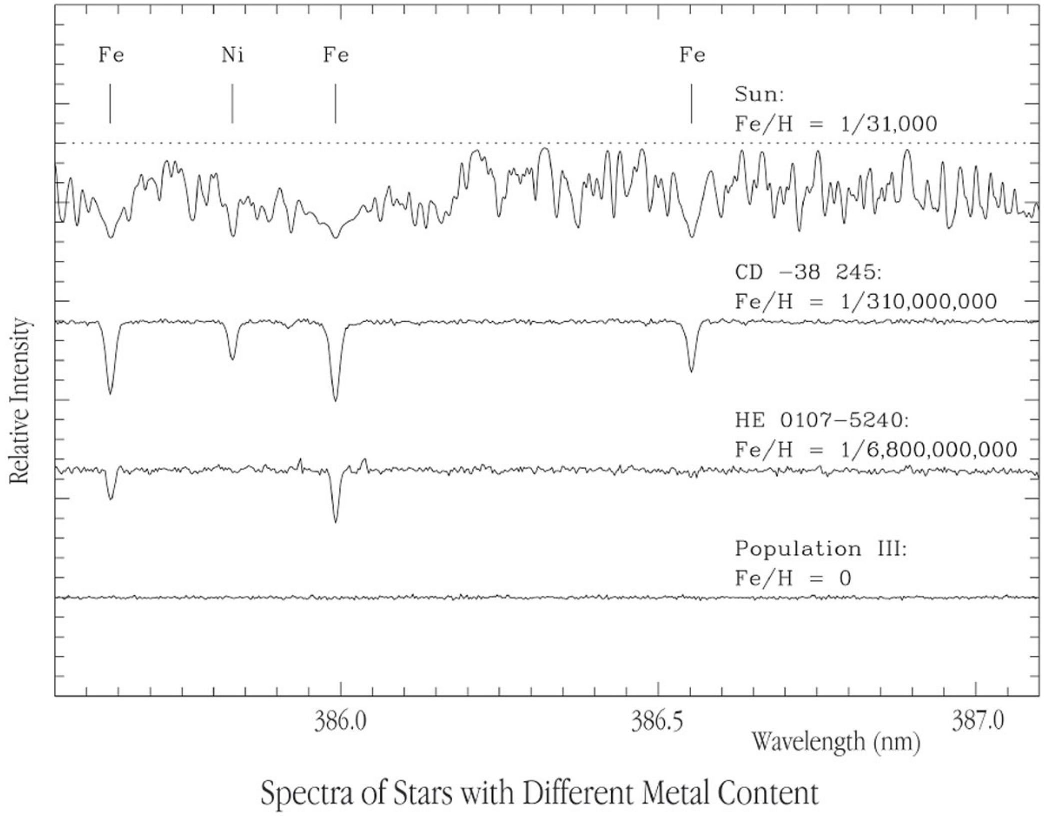


Figure 16.4: Stars like the Sun, shown at top, contain many heavy elements. These elements have been produced by earlier generations of stars and are incorporated into subsequent generations. The elements can be detected by their absorption features in the stars' spectra. Earlier generations of stars have far fewer heavy elements, as the lower panels show. In principle, there might be stars with no heavy elements at all. No such star has ever been seen, and the lowest amounts of heavy elements measured in stars are a few thousand times smaller than those measured in the Sun. Credit: European Southern Observatory Lithium is the atom after helium in the periodic table. It has three protons, so ^{7}Li is an isotope of lithium that contains four neutrons.

Measurements of solar-like stars (as well as of meteorites and the local interstellar medium) imply an abundance today of about 1 part in a billion (10^{-9}) for ^{7}Li. Starting in the 1980s, astronomers started looking at the abundance of light elements in very old stars, that is, stars with a low abundance of heavy elements such as iron. Stars with fewer heavy elements (indicating older generations of stars), have less ^{7}Li- up to a point. The abundance plateaus at a value of $\sim 10^{-10}$ and does not drop as iron abundance continues to drop. The next heavier elements in the periodic table, beryllium and boron, do not show such a plateau, as indicated in Figure 16.5. This result implies that at least some ^{7}Li was present before the oldest stars formed, but heavier elements, such as beryllium and boron were not.

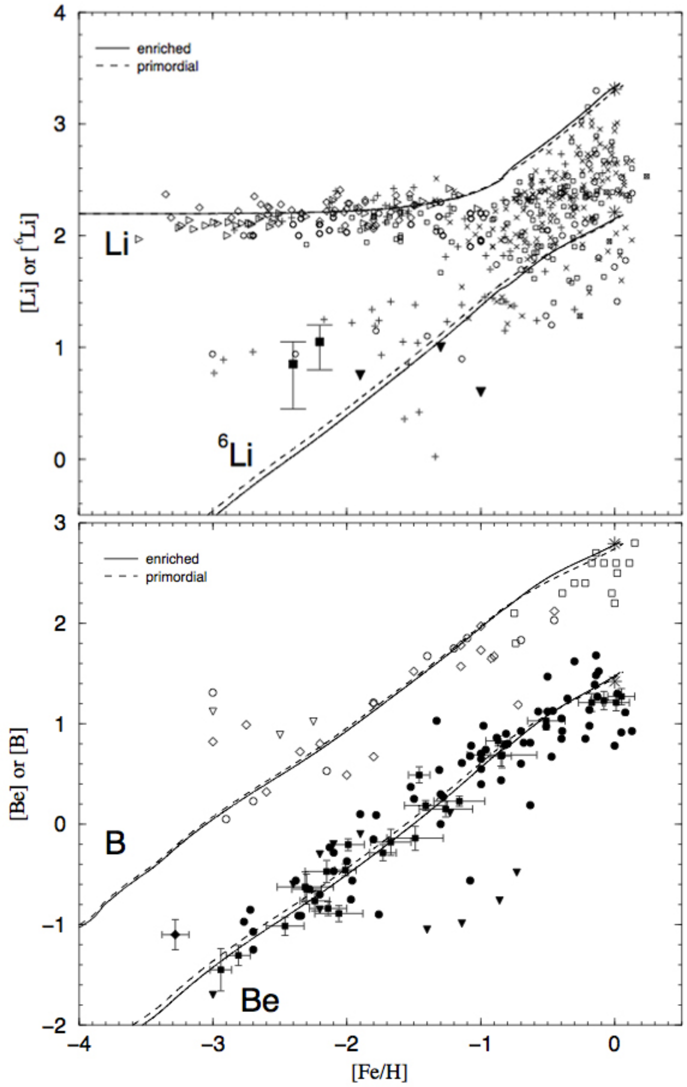


Figure 16.5: Abundances of the elements lithium, beryllium and boron as a function of the abundance of heavy elements in stars. Each point represents the abundance in one star. Down on the vertical axis corresponds to lower abundance. Left on the horizontal axis corresponds to lower heavy element abundances. From models of how elements are created in stars, we know that stars with lower abundances of heavy elements were born at earlier times in the Universe, so left on the horizontal axis also corresponds to earlier times. The plateau in the ^{7}Li abundance at early times indicates that there was a small amount of primordial lithium present before the formation of the first stars. Any such plateau for beryllium or boron is too low to be seen with current data . Credit: Alibés, Labay, and Canal (2002), *Astrophysical Journal* , 571,326

Taking all of the data into account, the primordial abundances by number of atoms of the elements present before the formation of first stars were: 92% hydrogen (^{1}H), 8% ^{4}He, ~10^{-5} deuterium, ~10^{-5} ^{3}He, and ~10^{-10} ^{7}Li. By mass, this corresponds to about 75% hydrogen (^{1}H) and 25% helium (^{4}He), and only a tiny fraction of the mass in other isotopes.

16.1.2: Models

The observed amounts of the lightest chemical elements, such as hydrogen, deuterium, helium, and lithium, imply that these elements were created early in the history of the Universe before stars formed. This can only have been in the first few minutes, when conditions were hot and dense enough for nuclear fusion to occur. The abundance measurements are consistent with expectations given the relative densities of protons, neutrons, and photons, and the temperatures and densities of the Universe at these early times.

The creation of the lightest elements in the first few minutes of the Universe through fusion reactions is known as **Big Bang nucleosynthesis** (BBN). The principles that BBN is based upon are well understood from nuclear physics; these are principles that can be carefully tested in laboratories. Since the Universe was expanding and cooling rapidly, the BBN era did not last very long. It

https://phys.libretexts.org/@go/page/31445

produced isotopes through lithium, but heavier elements came a few hundred million years later when the first stars created the conditions in which the heavy elements could form.

The relative abundance of elements depends on both the overall temperature and the weak nuclear force. Hi temperatures are needed to overcome the electric forces that repel protons from one another. The weak nuclear force is responsible for radioactive decay in unstable nuclei, and it affects the basic physics governing the protons and neutrons which must be combined to create heavier elements. Recall that protons and neutrons release energy, called binding energy (Figure 16.6), when combined to form nuclei during nuclear fusion. The overall peak of the binding energy curve occurs at iron, but at low atomic weights the most tightly bound nucleus is ^{4}He. It is therefore favored for production compared to other light isotopes during BBN.

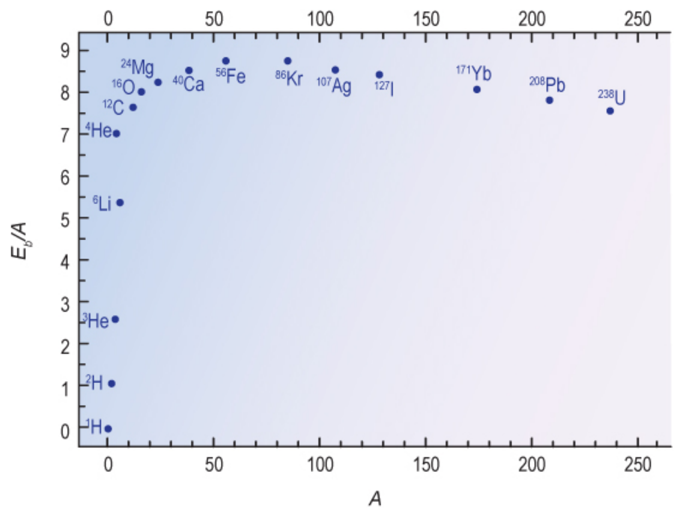


Figure 16.6: The binding energy of nuclei is plotted against atomic weight. The peak of the curve occurs at ^{56}Fe, the most tightly bound nucleus that exists. At low atomic weights the most tightly bound nucleus is ^{4}He. Credit: NASA/SSU/Aurore Simonnet

One limitation on fusion reactions in the early Universe is that a free neutron will decay into a proton (plus an electron and an antineutrino) even at low temperatures - this is a spontaneous reaction, a result of the weak nuclear force. On the other hand, a proton can only be turned into a neutron (plus a positron and electron neutrino) at high temperature and pressure. Energy input is needed for the reaction to run in that direction. For the first second in the history of the Universe (when the temperature is greater than 10^{10} K), protons and neutrons could change back and forth easily because there was ample energy to convert protons back to neutrons. After that, the temperature of the Universe dropped and there was not enough energy in collisions to turn protons into neutrons. The balance of these reactions, set by the mass difference between the neutron and proton, fixed the ratio of protons to neutrons at about 7:1. This ratio in turn dictated what elements could form.

Since normal hydrogen consists of one proton, and ^{4}He consists of two protons and two neutrons, we can see from the ratio of neutrons to protons how we would get roughly the observed abundance of each of them. In the next two activities, you will determine these by mass and by number.

Amount of Hydrogen and Helium by Mass

How do we go from a proton-to-neutron ratio of 7:1, to 75% of the atomic mass in the Universe being hydrogen and 25% being helium? We can count the number of atoms and then figure out the mass.

To do so, you will need the following information:

- To make a ^{1}H nucleus, 1 proton is required.
- To make a ^{4}He nucleus, 2 protons and 2 neutrons are required.
- When BBN started there were 7 protons for every 1 neutron.

Figure A.16.1 shows protons and neutrons in the correct proportions at the start of BBN as well as how they must be grouped to create ^{4}He.

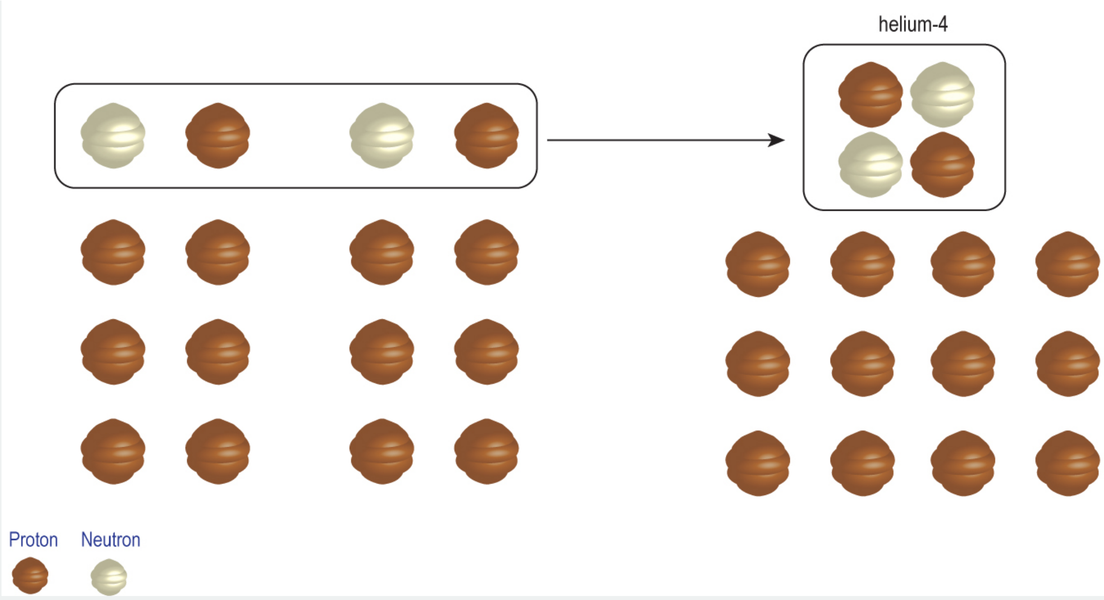


Figure A.16.1: In the early Universe, there were seven protons for every neutron. Nuclear reactions in the first few minutes after the Big Bang converted some of these particles to helium, leaving us with the amount of hydrogen and helium we see today. Credit: NASA/SSU/Aurore Simonnet

1.

2.

3.

Each particle has a mass unit of 1 amu (atomic mass unit).

4.

5. Take your answer to question 3 divided by question 1 to determine the hydrogen fraction by mass. What percentage is this?

Amount of Hydrogen and Helium by Number

In the previous activity, we worked out the mass fractions of hydrogen and helium in the Universe given a proton:neutron ratio of 7:1. Another way to discuss atomic abundances is in terms of numbers of atoms.

1.

2.

3.

When quoting abundances in terms of number instead of mass, each atom, rather than each constituent particle, counts as 1.

4.

5.

In the previous activities, you saw why the abundance is roughly three-fourths (75%) hydrogen and one-fourth (25%) helium by mass (that is, about 75% of the mass of atoms in the Universe is in hydrogen). If we count by number instead of mass, these percentages are 92% for hydrogen and 8% for helium. Scientists who study BBN use computer codes that include all of the possible reactions between neutrons, protons, and the light elements they produce. These codes give ranges for the yields of various elements at the end of the BBN era, and these yields must be compared to the observed abundance, as we will see.

Of course, elements besides helium were also produced. Isotopes such as deuterium were created by the collisions of protons and neutrons. At very high temperatures, additional collisions with other protons and neutrons, or especially with photons, tended to break deuterium apart. However, as the temperature dropped, collisions were no longer energetic enough to destroy deuterium. Deuterium atoms could then survive long enough to react with themselves and with protons and neutrons to make both ^{3}He and ^{4}He. Collisions could also form the radioactive isotope of hydrogen called tritium (^{3}H or T). The first reactions that occurred are shown schematically in Figure 16.7.

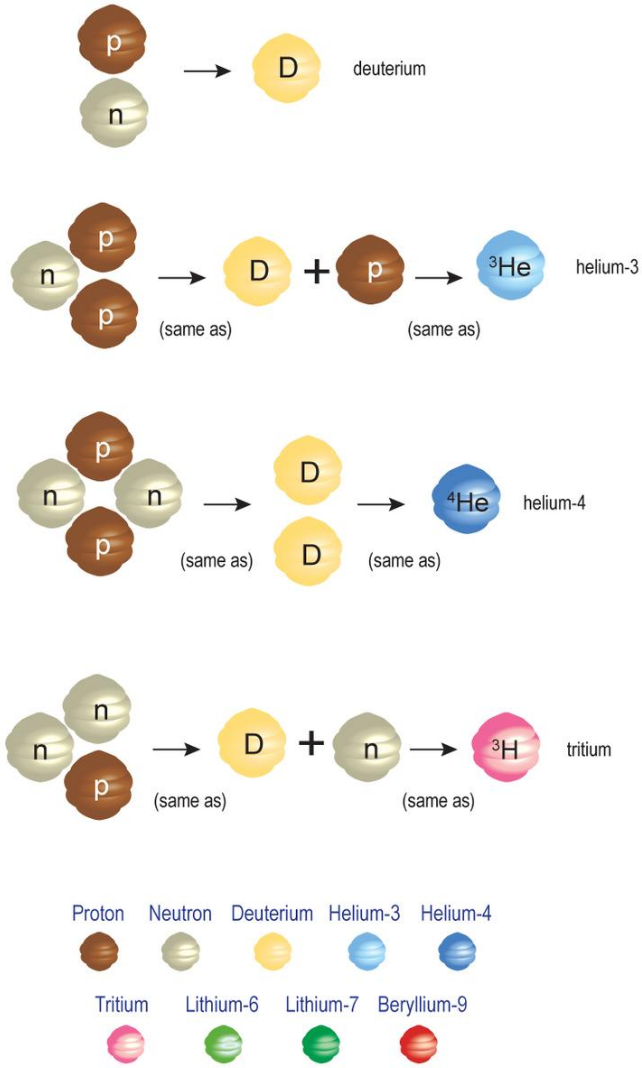


Figure 16.7: Combining the fundamental nucleons produced compound nuclei like helium and the heavy isotopes of hydrogen - deuterium and tritium. Credit: NAS A/SSU /Aurore Simonnet

In principle it is possible to build all the way up the curve of binding energy to iron, which has the highest binding energy of any nucleus. In fact, there was only time for the chain to build up to lithium (^{7}Li), and possibly a tiny amount of beryllium (^{9}Be). Protons have positive charge, and there are no negative charges in the nucleus, so it takes energy to get them to come together. Once they are very close, within about 10^{-15} m, the strong nuclear force can bind them together. Before that can happen they must be moving fast enough to approach positively charged nuclei containing one or several protons. This means that heavier elements require higher temperatures and more time in order to overcome the greater electric repulsion caused by their greater number of charges.

But the Universe was expanding, lowering both the temperature and the density. At around 3 or 4 minutes, when lithium was produced, the density and temperature had dropped too low to allow much of anything else to form. Therefore, only hydrogen, helium and lithium were available when the earliest stars formed several hundred million years later. As we have seen, almost all of this material was ^{1}H and ^{4}He. The reactions that could form the lightest elements beyond hydrogen and helium are shown in Figure 16.8. The formation of beryllium is speculative, as we do not have evidence that it was actually produced by BBN. What's more, any that was produced could well have been destroyed inside stars since then.

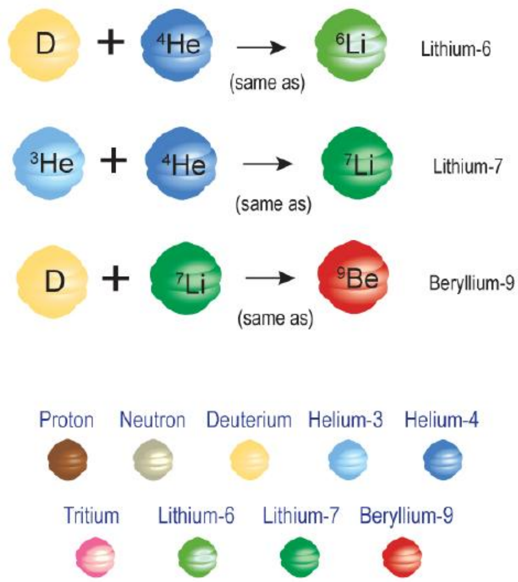


Figure 16.8: Once the lightest nuclei beyond hydrogen formed it was possible to combine them with other particles (and themselves) to create still heavier ones. This is how ^{7}Li and ^{4}He were produced. Perhaps even a tiny amount of beryllium in the form of ^{9}Be was made, though there is no sign of it anymore. A sampling of the possible reactions that occurred are shown in this figure, though others are also possible. Credit: NASA/SSU/Aurore Simonnet

It was not time alone that weighed against building elements beyond lithium. As shown in Figure 16.9, a serious bottleneck results from that fact that all isotopes with mass numbers 5 and 8 are unstable. Even stars have difficulty getting past these gaps. However, what stars have that the early Big Bang did not is time. At a minimum, stars have millions of years to work with, and they manage to build heavier elements by combining three helium nuclei. That is a slow process, and there was no time for it in the first few minutes of the early Universe. Even given the extra time, the elements between helium and carbon are very rare, an indication of their fragility. Consider the abundance of each as plotted in Figure 16.3.

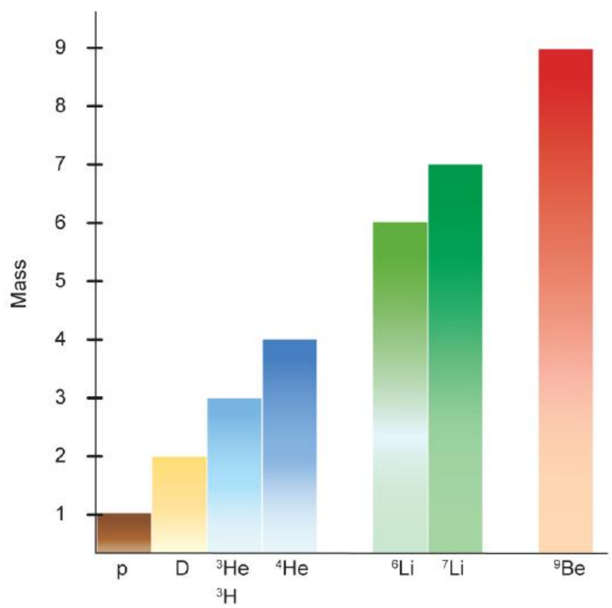


Figure 16.9: The lightest elements were formed in the first few minutes of the Universe. Only elements up to beryllium would have been produced because gaps at masses of 5 and 8 create roadblocks that slow nuclear reactions. It is possible that a very small amount of ^{9} Be was made, though a primordial abundance has not been detected. After about 4 minutes, the Universe had cooled and the density had dropped to the point that no further reactions were possible. Credit: NASA/SSU /Aurore Simonnet

https://phys.libretexts.org/@go/page/31445

Building Atoms Through BBN

In this activity you will build atoms, simulating the process of Big Bang nucleosynthesis.

You will have 4 minutes to create as many atoms as you can, just as the Universe did! See how many you can make. As you do the activity, here are a few things to keep in mind:

- Protons (brown) and neutrons (white) will move across the screen for you to use as raw material.
- To combine any two particles, click on each of them once.
- You score more points for creating heavier elements.
- Watch out for photons-they are flying by and can break apart atoms, causing you to lose points.
- If particles are too far away, they will not combine with each other.
- If a nuclear reaction is forbidden, the particles will not combine.
- Unstable atoms will fall apart instantly.

Play Activity

In the previous activity, you probably saw that your ability to create elements depended on how many protons and neutrons were available and how many photons were breaking the elements apart. Detailed computations show that the abundance of deuterium, ^{3}He, and ^{7}Li are very sensitive to the density of baryons (particles made of three quarks - protons and neutrons for our purposes), from which they are made. The dependence is generally stated in terms of the baryon-to-photon ratio, η (pronounced “eta”)-the ratio of the number of baryons to the number of particles of light. A higher density of baryons (larger η) means reactions go faster because there are more particles around to combine. More photons, on the other hand (smaller value of η), means nuclei like deuterium are more easily photo-dissociated (split apart by the photons). That makes building heavier elements more difficult. The measured abundance of these elements provides a way to probe the actual value of η, and hence conditions in the early Universe.

The abundances relevant to determining η are those present in the early Universe, before stars began to modify them with their own nuclear fusion. The pattern of elemental abundance that was in place just after Big Bang nucleosynthesis ended is called the primordial abundance. It can be difficult to measure because stellar processes have modified this pattern over the intervening billions of years. For instance, deuterium is easily fused to form helium inside stars. So if we look at a young star like the Sun we will measure a lower deuterium abundance than was present in the early Universe. The solar deuterium abundance is even lower than it was in the material that formed the Sun. This is because stars destroy deuterium, they don’t create it.

Similar problems hamper the measurement of other light elements’ primordial abundance. However, by choosing to study very old stars we can minimize the effects of stellar evolution. Better still, if we look at diffuse clouds in intergalactic space, far away from any galaxies (and thus stars), we can measure the elemental abundance that should be truly primordial. That is what has been done for deuterium - we measure its absorption features in the absorption spectra of distant quasars (see Figure 16.10). The absorption features in these spectra are not due to the quasar itself, which just serves as a background light source. The absorption is caused by gas clouds located between us and the quasar, and most of these are in the space between galaxies. They are thus far away from any stars that would have polluted them with the products of stellar nuclear fusion.

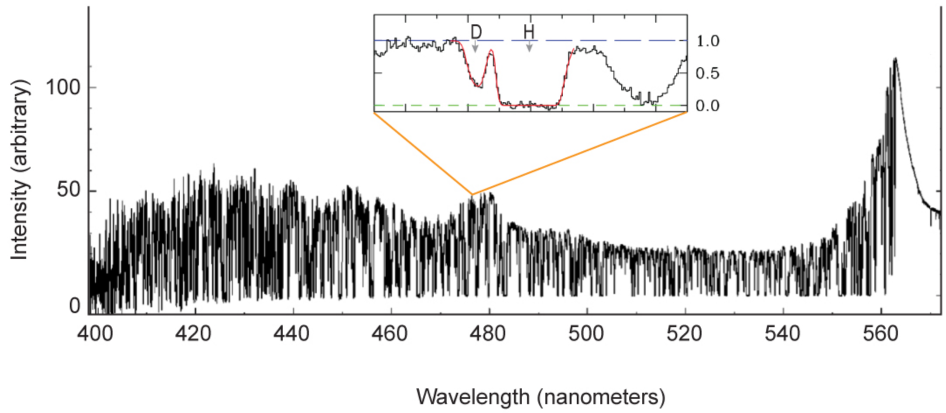


Figure 16.10: Quasar spectra show many hydrogen absorption systems from clouds that lie between us and the quasar. In the main part of the figure is part of a spectrum showing the hydrogen emission from Lyman-alpha, with a rest wavelength of 121.5 nm. It has been redshifted by cosmic expansion in this system to around 560 nm, a redshift of around 3.5. The many absorption features at shorter wavelengths are the same Lyman-alpha feature seen in absorption by intervening clouds of gas . Inset at the top is a detail from one hydrogen absorption system. The deuterium absorption (D) can be seen slightly offset from the main hydrogen (^{1} H) feature, an offset caused because there is a small difference in the energy levels of deuterium atoms as compared to normal hydrogen atoms. Credit: NASA/SSU /Aurore Simonnet , adapted from Rauch (1998), *Annual Reviews* , 36, 267 (large image) and Pettini and Cooke (2012), *MNRAS* , 425, 2477 (inset).

Measurements of the primordial abundance of the light elements constrain the baryon fraction, that is, the fraction of normal matter in the Universe. A comparison of the measured abundance and the abundance computed from BBN has important implications for the viability of the Big Bang theory and the composition of the Universe. We can plot the abundance in terms of either baryon fraction or η, as in Figure 16.11. This figure is somewhat compact and requires a brief explanation to be understood. The vertical axis, which relays the fraction of each element, has a broken scale. This is done so that all these isotopes can be shown together, given that they cover a vast range of abundance. We will break each part down separately.

The top part of Figure 16.11 shows the mass fraction of ^{4}He, just as we calculated above. Measurements place this value at between 24% and 25%, similar to our previous estimate. The abundance of ^{4}He is not very sensitive to the baryon fraction, so it remains high for all of the values of baryon density on the plot. This is because the ^{4}He nucleus is extremely stable. It has the highest binding energy of any light atom, as can be seen in Figure 16.6. As a result, most of the free neutrons at the beginning of the BBN era are combined with protons to form ^{4}He. The faster the reactions run (the higher the baryon density) the faster this happens.

Below ^{4}He is shown the abundance for deuterium, ^{3}He, and ^{7}Li. They are all shown in terms of number of atoms of each, per atom of hydrogen. So, for instance, deuterium is measured to be a few times 10^{-5} as abundant as hydrogen. That means that out of every fifty thousand or so hydrogen atoms, one will be deuterium. ^{3}He is about half as abundant, though only upper limits are given; this is indicated by the downward-facing arrow. Finally, ^{7}Li is measured to be a few times 10^{-10} (note the break in the vertical axis between ^{3}He/D and ^{7}Li), so for every ^{7}Li atom in the early Universe there were about 5 billion hydrogen atoms.

Each of these measurements has uncertainty associated with it, and these are shown by the boxes on the plot; they indicate the range over which the measurements are able to constrain the abundance of each isotope. Because the abundances for each isotope are independently measured, the uncertainty boxes are not all the same size, and in fact for ^{3}He there are only limits on the abundance: we know that it must be below about 10^{-5} of hydrogen.

The favored status of ^{4}He is apparent in Figure 16.11, where we see that as its abundance goes up, the abundance of ^{3}He and ^{2}H (deuterium) go down. The other light isotopes are converted to ^{4}He at higher baryon fractions because nuclear reactions can proceed more quickly, and it is this effect that causes the graphs to trend as they do.

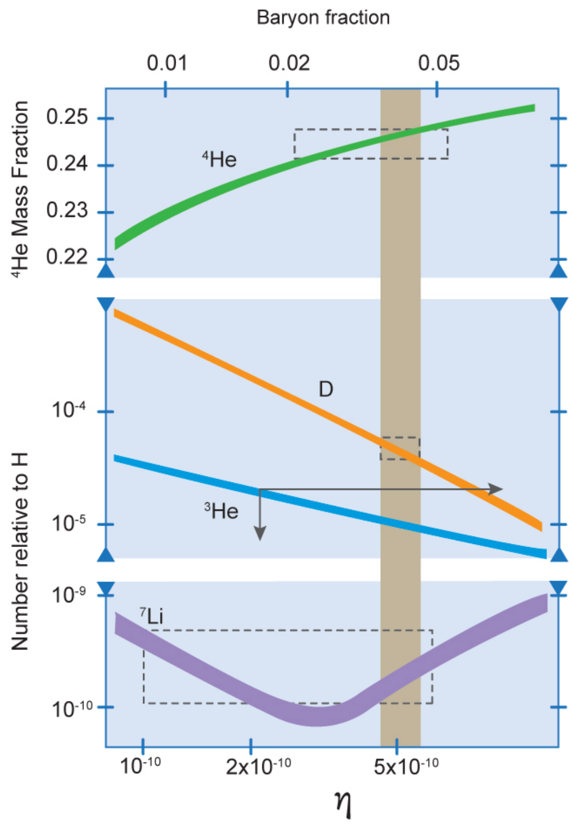


Figure 16.11: The solid, colored lines show the abundance (by mass) of ^{4}He, ^{2}H (D), ^{3}He, and ^{7}Li predicted to have formed during Big Bang nucleosynthesis. The boxes and horizontal bars are the constraints from observations. Abundance is plotted as a function of baryon fraction (top horizontal axis) and baryon-to-photon ratio, η (bottom horizontal axis). Where the vertical tan stripe intersects the horizontal axis is the range of values for the baryon density of the Universe, because that is the density for which all of the measurements agree with each other. Credit: NASA/SSU /Aurore Simonnet adapted from K. Nollett

Figure 16.11 shows a range of possible abundance for each element, both from theoretical calculations and from the measurements of each. The tightest limits are given by the abundance of deuterium; the other abundance measurements are consistent with the deuterium measurement. Taking them together we find the baryon fraction is 0.04 +/- 0.004. This value for the baryon density is consistent with the value we found using CMB measurements. It is several times larger than that found from luminous matter alone, indicating that there are many baryons that do not shine.

It is remarkable that the abundance of light atoms predicted by BBN are so well borne out by observations. The predictions for various isotopes range over 10 orders of magnitude, and yet all agree within the uncertainties of the measurements - which are considerably smaller. Deuterium in particular sets very stringent limits on the baryon density, and it is right in line with what is predicted by other isotopes. The strong agreement between these several isotopes is one of the lynch pins of modern cosmology. The agreement provides highly compelling evidence in favor of the standard hot Big Bang model of the early Universe.

Going Further 14.2: Expanding Gases Cool

In this chapter, and throughout much of the module, we have repeatedly stated that the Universe was denser and hotter when it was younger. Or equivalently, we have said that the Universe must have cooled as it expanded. Why must this be the case? If you have had a course in chemistry or physics you know the answer, but if not (or maybe even if you did) you might find these statements confusing. We will now explain why gases exhibit this property.

Consider a gas inside a closed container from which it cannot leak out. If you imagine that the container is a cylinder with one end free to move, then you could push or pull on the cylinder to change the volume occupied by the gas, and thus its density. A hand bicycle pump is an example of this sort of device. What effect would this have on the gas?

From experience you probably know that you must push on such a cylinder in order to compress it, and the more you manage to squeeze the gas, the harder it becomes to compress it further, sort of like trying to squeeze a tube of toothpaste with the cap on. The reason for the increasing difficulty is two-fold.

The first cause of your increasing effort is that you are expending energy to compress the gas. That energy is not lost: energy is always conserved. The only place for the energy to go in this case is into the gas. The energy you expend pushes on the gas particles and imparts a higher kinetic energy into them, so they move faster because you have sped them up. Since the particles are moving faster, they strike the walls of the container with a greater force. That increases the pressure on the walls and is the first reason you must push harder.

The second reason you have to exert more effort to compress a canister of gas is because the gas volume is getting smaller as you push inward. That means that the particles do not have to travel as far when they bounce between one wall and the opposite wall; they strike each wall more often. Because the particles are moving faster, they travel that shorter distance in a shorter time than would would be true if they moved at a lower speed. This also causes the particles to hit the walls more frequently.

The net result is that by compressing the gas you speed up the particles and cause them to hit the walls more often and with greater force. The fact that the gas particles are moving faster is generally conveyed by saying their temperature (T) has increased. That they are confined to a smaller space means that their volume (V) has decreased, and these changes have led the pressure (P) to increase. The total number of particles (N) in this case has remained the same because no gas has leaked out of the cylinder. We could also think about what happens in terms of the number density, N / V, which is increasing in this case.

The pressure, volume, number of particles, and temperature of a gas are related by the following equation.

$$PV = NkT$$

The lowercase k is called Boltzmann’s constant, which we have already seen. The equation is known as the ideal gas law. You might be familiar with this equation from a chemistry course. It is often written as below.

$$PV = nRT$$

Now, n is the number of moles and R is the ideal gas constant. Chemists like to use moles, physicists use the number of particles instead, but they are essentially the same. According to the ideal gas law, as the density or temperature of an ideal gas goes up, the pressure also tends to go up. Don't let the word "ideal" confuse you. Most gases follow this law quite closely. It fails to work when the particles interact strongly with one another, but at that point you are probably no longer dealing with a gas.

Conversely, if we let the volume of our cylinder expand, the gas inside would have to push against an outside pressure, expending energy to do so. The energy would have to come from the internal energy of the gas itself, thus lowering its temperature, density and pressure.

That is why gases cool as they expand and heat up when they are compressed. It is the basis of technologies like refrigerators and air conditioners. In the case of the Universe, it expands not against an external pressure, but against its own self-gravity. The result is the same: the gas within the Universe cools as a result of the expansion.

This page titled 16.1: The Formation of the Lightest Elements is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

16.2: Particle Soup

Learning Objectives

- You will know the fundamental particles.
- You will know the fundamental forces.
- You will be able to calculate the ambient temperature corresponding to the mass-energy of particle reactions.
- You will know that at different temperatures, the Universe undergoes various transitions: electroweak transition, quark confinement, proton-antiproton annihilation, neutrino decoupling, and electron-positron annihilation.
- You will know that the electromagnetic, weak, and strong forces were once unified.

What Do You Think: particles

Big Bang nucleosynthesis is a triumph of modern physics. Its ability to predict what conditions were like in the early Universe is remarkable. Still, it can only take us back to a time when the Universe was a bit under a second old, and it describes the overall evolution of the Universe for several hundred seconds thereafter. What occurred at earlier times, when the Universe was much less than a second old, BBN cannot tell us. It might seem ridiculous to ask what the Universe was like at such early times, but just as we can use our understanding of physics to paint a picture of the BBN era, we can also use it to try to predict what conditions would have been like at earlier times. However, when we go further back in time the conditions become more extreme and the physics becomes less certain.

Using our knowledge of physics as our guide, the next step takes us back to times when the Universe was just a tiny fraction of a second old and things were packed more than a billion times closer together than during the BBN era. The Universe was a sea of subatomic particles because temperatures were much too high for atomic nuclei to be stable.

16.2.1: Elementary Particles

To get a feel for how scientists classify particles based on their properties and interactions, you will classify some fictional particles in the following activity. You will use criteria of your own choosing.

Particle Classification

The purpose of this activity is to help you understand how scientists classify the fundamental particles.

Listed in a table, you will see 13 hypothetical particles and some of their properties. These properties include mass, electric charge, and two quantum properties called “spin” and “color charge.” For both mass and color charge, the possible values are either “yes” or “no." Electric charge can be either positive or negative and have a variety of different values. Spin is either an integer or half-integer number.

Play Activity

In the last activity you classified a number of hypothetical particles. The actual fundamental particles in our Universe are shown in Figure 16.12, arranged into a few categories. Each particle also has a corresponding anti-particle, which has the same properties as the particle except its charge is reversed.

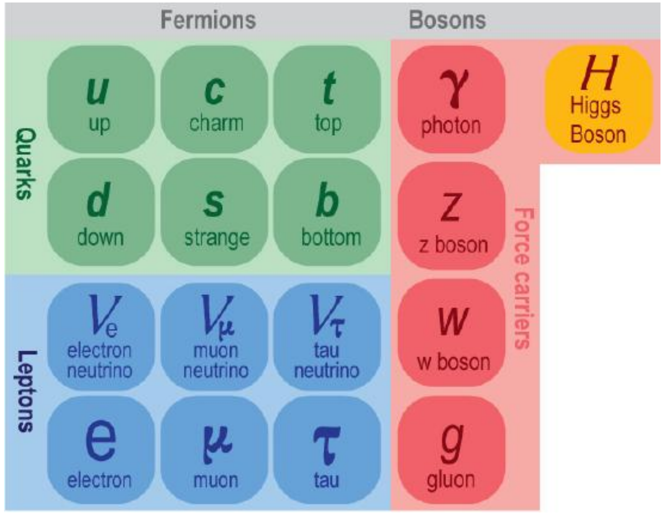


Figure 16.12: All matter is composed of combinations of quarks and leptons. Each of these particles has a corresponding antiparticle with opposite charge. In addition, there are the force-carrier particles by which the matter particles interact with one another. These are shown in the right-hand column. The recently discovered Higgs particle is also a fundamental particle. Credit: NASA/SSU/Aurore Simonnet

All matter is composed of combinations of the fundamental particles shown in Figure 16.12. All of the atoms from which we are made are composed of combinations of protons and neutrons (which together are called nucleons) and electrons. Protons and neutrons are each composed of three quarks. The sub-structure of an atom is shown schematically in Figure 16.13. You can see that protons and neutrons each contain one down and one up quark. It is the third quark differentiates the two particles.

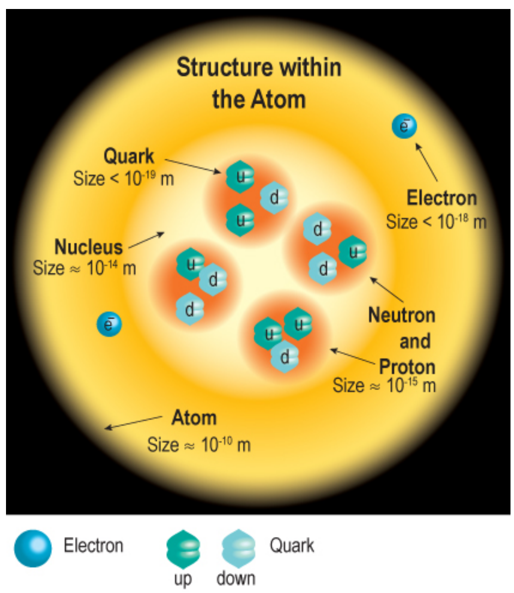


Figure 16.13: The nuclei of atoms are composed of protons and neutrons, which are in turn composed of quarks. Electrons orbit the nucleus in a large cloud. Credit: NASA/SSU/Aurore Simonnet

There are four additional quarks, called charm, strange, top, and bottom, making six quarks total. The names are whimsical and are not intended to convey any of the properties of the particles themselves, though charm, strange, top, and bottom do have slightly different properties from up and down, in addition to having higher masses. None of the heavier quarks are stable. In the current conditions in the Universe they all decay, often through complex reaction chains, into lighter particles that are stable. There are hundreds of (mostly short-lived) particles that do not combine to form chemical elements (atoms). These are built from various combinations of two or three quarks.

Particles made of three quarks are called baryons, and particles made of two quarks are called mesons. The general name for particles made of quarks is hadron. Quarks have fractional electric charge. They combine to make hadrons of integer charge. For example, an up quark has a charge of +2/3 and a down quark has a charge of -1/3, so a proton, which is made of two up quarks and one down quark, has a charge of 2/3 + 2/3 – 1/3 = +1. A neutron, which is two down quarks and one up quark, has a net charge of 2/3 -1/3 - 1/3 = 0.

In addition to the quarks, there are lighter particles, similar to the electron, but with higher mass. These are the muon and the tau particle. You can think of these as heavier versions of the electron. And like the electron, each has its own version of the neutrino. The light particles are collectively called leptons; there are six leptons total, just like the quarks.

From these particles, together with the force-carrier particles and the Higgs particle, nearly all of the physical properties observed (so far) in the Universe can be explained.

16.2.2: Four Fundamental Interactions

The four fundamental interactions (or forces) are: the **electromagnetic interaction**, which keeps electrons bound to atomic nuclei and thereby determines the structure of atoms; the **strong nuclear interaction**, which holds the nucleus and the particles in it together; the **weak nuclear interaction**, which is important in radioactive decay; and **gravity**, which dominates on large scales.

Each of these interactions is mediated, or carried by, associated particles called field bosons. An analogy for a carrier particle would be a ball that two people are passing back and forth; as long as the two people are passing the ball, they are interacting. Fundamental properties of the interactions depend on the properties of the carrier particles involved. To take one example, the particle that carries the electromagnetic interaction is the photon, a particle of light. The electromagnetic force has infinite range because the photon is massless and has no charge. For the strong nuclear interaction, an exchange of gluons, which are also massless, holds quarks together to form larger particles. Because gluons interact strongly with each other, the force acts over only extremely short ranges, comparable to the size of an atomic nucleus. The weak interaction is mediated by W and Z bosons, which are heavy and short-lived. This limits weak interactions to an even shorter range.

Though gravity has so far defied this concept of carrier particles, in a quantum theory of gravity, the gravitational force would be mediated by massless gravitons. This would imbue the gravitational interaction with infinite range. The quantum view is very different from the geometrical framework of general relativity. Thus far, the two systems have yet to be mathematically unified, and research into a theory of quantum gravity is ongoing and robust.

We have described electromagnetic and gravitational forces extensively in other chapters. They are both important for many cosmological phenomena. The weak and strong nuclear forces are not obviously at work in the present day Universe, though of course both are critical. But because neither operates on macroscopic scales, they were not even known to exist until well into the 20th century, when scientists began to probe the structure of atomic nuclei. Nonetheless, they become especially important in the early Universe, when conditions everywhere were similar to, or even more extreme than, the conditions in nuclear reactors and particle accelerators.

For example, the weak interaction can convert a neutron to a proton and vice versa (Figure 16.14). It is fundamentally different from the electromagnetic interaction—it follows a different set of rules and involves a different set of particles. For example, weak interactions do not act only on charged particles. In the conversion of a neutron to a proton, two charged particles and two neutral particles are involved. The interaction converts one of the down quarks in the neutron into an up quark, an electron, and an anti-neutrino. The electron and anti-neutrino fly away. The up quark remains in the nucleon, which is now a proton (uud) instead of a neutron (ddu). At a more fundamental level, there is an exchange of W bosons with one of the down quarks in the neutron that mediates the conversion.

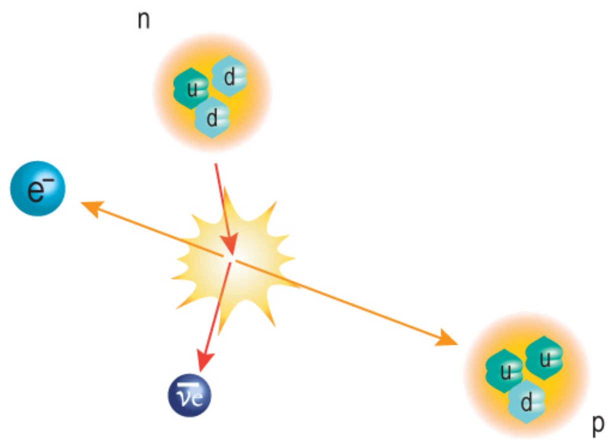


Figure 16.14: When a neutron decays, one of the down quarks in the neutron turns into an up quark, and thus the neutron (udd) turns into a proton (uud). In this weak interaction an electron and anti-neutrino are also produced. The electron is necessary for total electric charge to be conserved , which all known interactions do. Credit: NASA/SSU/Aurore Simonnet

The weak interaction is called weak for historical reasons, but that is a bit of a misnomer. The interaction is rare in a sense, but that is not because it is weak; it is because the interaction has a short range. Whereas electromagnetic and gravitational interactions have an infinite range that falls off like $1/r^2$, the weak interaction seems to have a finite range. Weak interactions occur only if particles are about 10^{-16} m from each other. This is about a million times smaller than the diameter of an atom—or about 10 times smaller than the diameter of a proton. Particles don't often approach one another so closely. That is why the weak interaction seems weak.

Like the weak force, the strong force also has a finite range, about 10^{-14} - 10^{-15} m, the size of atomic nuclei. At these scales, the strong force is about 100 times stronger than the electromagnetic force. This difference in strength is important for overcoming the repulsive force between protons in nuclei, which are positively charged. The strong force binds protons and neutrons to each other inside atomic nuclei. It also binds the quarks within nucleons together.

Quarks have a property, known as color charge, that determines how they combine to form hadrons. Color charge has nothing to do with colors of the electromagnetic spectrum. It was given that name because of the way the quarks must combine. There are three types of color charge: red, green, and blue, along with three complimentary colors for the anti-quarks: anti-red, anti-green, and anti-blue. To make a particle, the color combination must be color-neutral. For example, to make a baryon, which consists of three quarks, one red, one green, and one blue quark must be used. Just as a combination of red, green, and blue light makes white light on a computer screen, one red up quark, one green up quark, and one blue down quark can make a proton. To make a meson, which consists of two quarks, a red and anti-red, or green and anti-green, or blue and anti-blue combination must be used.

If you would like to learn more about particle physics, visit the Particle Adventure website.

16.2.3: Matter-energy Conversion and the Temperature of the Universe

In the conditions of the very early Universe, matter and photons exchanged energy with each other and could be described by a single temperature, similar to the interiors of stars today. However, conditions throughout the Universe in the first fraction of a second were far more extreme than exist in any star. Temperatures and densities were both much, much higher than any star can produce.

We know that energy and mass can be converted back and forth into each other ($E = mc^2$). Subatomic particles can be turned into one another (following the rules of the force interactions) as long as there is enough energy for the reaction to run. At high enough temperatures, this happens spontaneously. The energy can come from the environment if the temperature is high enough. In that case, collisions between particles are able to furnish the required energy.

For instance, when temperatures are high enough, a photon can spontaneously convert to an electron—positron pair upon interacting with another particle (Figure 16.15). The additional particle is necessary to conserve momentum; a photon converting on its own is not able to simultaneously conserve both energy and momentum. An electron and a positron can also combine to create two photons in the opposite process, which is known as annihilation. In general it is necessary to solve equations for both

https://phys.libretexts.org/@go/page/31446

conservation of momentum and conservation of energy simultaneously in order to describe these interactions. However, our purpose here is to gain insight into how energetic photons must be in order for the energy conversion process to be viable.

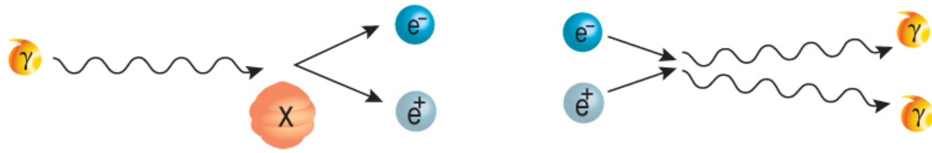


Figure 16.15: At high enough energies it is possible for a photon to spontaneously convert its energy into particles. The process must involve another particle, which must carry away some of the momentum of the photon. The opposite process, annihilation, occurs when a positron and electron combine to form a pair of photons. Each photon carries the energy of the rest mass of one of the original particles, thus conserving energy. However, two photons must be created in order to conserve the momentum of the original particles. Credit: NASA/SSU/Aurore Simonnet

As another example, in the early Universe, protons were able to turn into neutrons (Figure 16.16) because the environment supplied an ambient amount of energy corresponding to the difference in their masses.

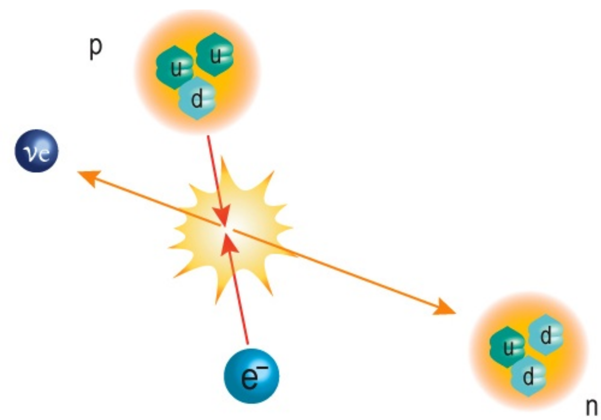


Figure 16.16: Protons and neutrons are made of different combinations of quarks. If the collision between a proton and electron is energetic enough, some of the energy of the collision can convert an up quark in the proton into a down quark, thus turning the proton into a neutron. The interaction also produces an electron anti-neutrino and occurs because of the weak force. Credit: NASA/SSU/Aurore Simonnet

In order to determine how much energy the environment can supply, we must relate the typical kinetic energy of particles (*E*) in a gas to the temperature of the gas (*T*). From thermal physics, we know that the relationship between the two is as follows.

$$E \sim kT$$

The *k* here is the Boltzmann constant, 1.381 x 10^{-23} J/K. This means that when the temperature of a gas is higher, it has more energy because the particles in it are moving faster. This fundamental relation tells us that temperature is a way of describing the typical energy of the particles in a system in equilibrium.

Combining this expression between energy and temperature with the relationship between energy and mass, we can find the ambient temperature needed in the environment for a given particle reaction to occur:

$$T \sim \frac{\Delta mc^2}{k}$$

In the following activity, we will explore the relationships between mass, energy, and ambient temperature.

Particle Reactions and the Temperature of the Universe

Worked Examples: Electron-Positron Creation and Annihilation

If we imagine that a photon converts its energy into an electron and a positron, then we can ask what the energy of the photon must be and at what temperature this conversion would be likely to occur. We'll find the energy first.

a. The rest mass of a positron is the same as that of an electron: 9.1 x 10^{-31} kg. What is the rest energy of these two particles?

- Given: m = 9.1E-31 kg
- Find: E
- Concepts: $E = mc^2$
- The rest energy of a positron is the same as that of an electron since both have the same mass. We need a photon with twice this amount of energy in order to create both an electron and a positron.
- Solution: the total energy of the photon is: $E = 2mc^2 = (2)(9.1 \times 10^{-31}\,\mathrm{kg})(3 \times 10^{8}\,\mathrm{m/s})^2 = 1.637 \times 10^{-13}\,\mathrm{J}$

This is a tiny amount of energy by everyday standards. However, we can compare it to energy scales more relevant to atomic and subatomic systems. We have already learned that the energy of visible light is around 10^{-19} J, an even smaller number. We generally speak of the energy of light in units called electron-volts (1 eV= 1.602 x 10^{-19} J), and a visible light photon typically has energy of around one or two eV. This is a better way to describe the energies we are talking about because we do not have to worry about inconvenient factors like 10^{-19}.

b. Find the energy of the photon needed to create an electron-positron pair in terms of electron-volts.

$$E = (1.637\text{E-13 J}) \times (1\text{ eV}/1.602\text{E-19 J}) = 1.02\text{E6 eV}$$

To put this into perspective, this photon is about a million times more energetic than visible light, and also about a million times more energetic than the bonds holding the outer shell electrons in most atoms. We would classify it as a gamma-ray photon. Though not particularly energetic by gamma-ray standards, it would still be about a thousand times more energetic than the x-rays one encounters at the dentist or doctor's office. Note that when we perform calculations we must always use the same set of units; in this case we have chosen SI. We only use electron-volts for convenience in understanding energies that would be extremely small in terms of joules, which are the SI unit of energy.

c. Now that we know roughly how energetic a photon must be to spontaneously create an electron-positron pair, we can ask how hot a gas containing such photons must be. What is the temperature of the gas?

- Given: E = 1.67E-13 J, energy of the photon
- Find: T, temperature of the gas
- Concepts: We know that the typical kinetic energy of the particles in a gas is given by $E \sim kT$, where k is the Boltzmann constant, k = 1.38 x 10^{-23} J/K, and where the temperature T must be measured in kelvin . To find the temperature give the energy, we will use $T \sim E/k$.
- Solution: $T \sim E/k$ = (1.637E-13 J) / (1.38E-23 J/K) = 1.2E10 K

This is about 10 billion kelvin, an incredibly high temperature. You might recall from our discussion in the last section that it is the same temperature scale that existed during the BBN era. That means that while all of the neutrons in the Universe were being consumed to make helium, the photon background—at that time gamma-rays, not microwaves as it is today—was busily converting between e^+e^- pairs and back again. These were the conditions throughout the Universe when it was a second or so old.

When we compute the temperature as in the previous example we are merely getting an idea of how hot a gas must be for pair-conversion to occur. In fact, gases contain particles with a range of energies, and the temperature describes only the typical (the average) energy of a particle in the gas. Some particles (photons included) will have higher than the typical energy, so we expect that pair production will occur in somewhat cooler gases than we have found with our estimate. For an example exploring the inter-conversion between protons and neutrons, see Math Exploration 16.1.

Math Exploration 16.1

Questions: Proton—Antiproton Creation and Annihilation

Looking to still earlier times, we expect the temperature to be ever higher. At some point in its history, the temperature was high enough for photons to create proton—antiproton pairs (Figure A.16.2). Now you will estimate the temperature at which this process can occur.

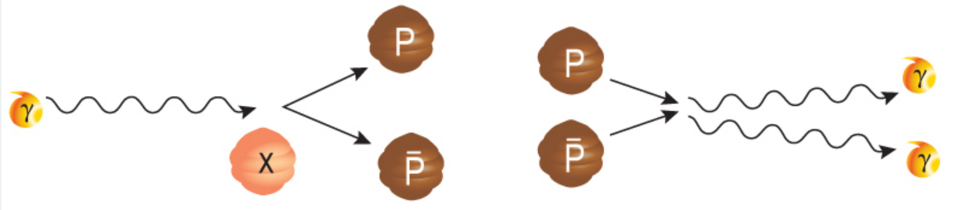


Figure A.16.2: At high temperatures photons spontaneously produce a proton and an antiproton, as shown at left. The particles will quickly combine with other protons or antiprotons and annihilate, as shown at right, creating two photons. These two processes keep each other in balance until the temperature drops to the point that photons are no longer energetic enough to produce the proton—antiproton pairs. Credit: NASA/SSU/Aurore Simonnet

1.

2.

After the temperature dropped below the critical point you just computed, there was not enough ambient energy in the environment to run the reaction both ways. Protons and antiprotons annihilated and turned into photons, but the reverse no longer happened.

In the previous activity, you learned how to calculate the ambient temperature of the Universe corresponding to the energy required for several particle reactions. These processes have important consequences for how the early Universe changes over time. We can calculate which particles can exist and in what form at different times based on the temperature of the Universe at those times.

16.2.4: Transitions in the Early Universe

In the last section, we saw how the temperature of the Universe corresponds to energies of possible particle reactions and therefore dictates how the Universe takes on different characteristics over time. Now we will look at some of the key events and the corresponding conditions in more detail. As we will see, the temperature also affects the fundamental interactions. We will consider these events by running time forward from around 10^{-12} seconds to about 1 second, from higher temperatures to lower temperatures. The events are summarized graphically in the timeline of Figure 16.17.

https://phys.libretexts.org/@go/page/31446

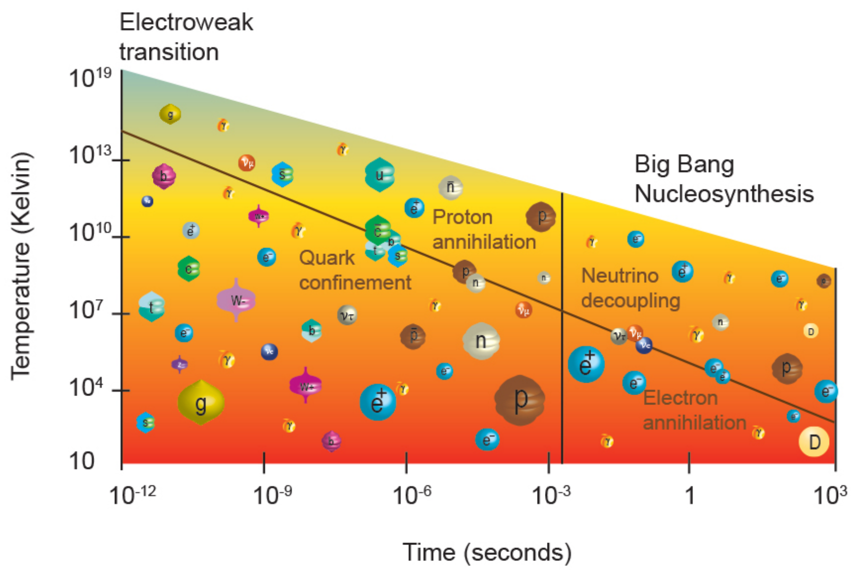


Figure 16.17: Conditions in the early Universe depend on its temperature. When the ambient temperature is greater than or equal to the energy of a particle reaction, it can run both ways, but when the temperature drops, the reaction typically only runs one direction. This dictates which particle structures are present at what times. When the temperature is high, structures are simpler; more complex structures are built up as the Universe cools. Credit: NASA/SSU/Aurore Simonnet

16.2.4.1: *Quark-gluon plasma, $t < 10^{-12}$ s, $T > 10^{15}$ K*

At early times, conditions were even more extreme than during BBN. There were no complex structures, atoms, or even nucleons (protons and neutrons), because the collision energies were so high that atoms were unstable: they were more likely to be destroyed by collisions than to be created by them. Instead, matter consisted of a zoo of sub-atomic particles constantly converting between matter and energy. Temperatures were high enough to generate every kind of known subatomic particle, and they all existed together in approximately equal numbers in a bizarre dense "soup" called a **quark-gluon plasma**. Free quarks and gluons, as well as photons, electrons, neutrinos, and exotic particles like W and Z bosons and Higgs bosons, the heavy quarks and leptons, and perhaps other particles as well, were all spontaneously formed and annihilated with their antiparticles.

This state of matter is usually referred to as the quark-gluon plasma, not because it contains only quarks and gluons, but because those are the primary components of matter at such extreme temperatures and densities. Scientists are currently attempting to recreate the quark-gluon plasma in laboratories, where giant particle accelerators smash large atomic nuclei—not just protons and antiprotons - into each other at near the speed of light. In fact, it is from these sorts of collisions between large hadrons that machines like the Large Hadron Collider (LHC) and the Relativistic Heavy Ion Collider (RHIC) take their names. Hadrons are particles composed of quarks, so they include baryons and also atomic nuclei, which are composed of baryons.

16.2.4.2: *Electroweak transition*, $t = 10^{-12}$ s, $T = 10^{15}$ K

At early times, when the Universe was dense and hot, with a lot of ambient energy available in the environment, the distinction between the electromagnetic and weak interactions was not apparent. They were unified into the electroweak interaction**.** The discovery of this unification was a drawn-out process, and physicists have only come to fully understand it over the past several decades.

The unification is similar in spirit, if not in detail, to that achieved by James Clerk Maxwell in the 19th century when he unified the until-then distinct electric and magnetic forces into the electromagnetic force. Maxwell was able to write all electromagnetic interactions in terms of a single mathematical framework (theory), and since that time physicists have wondered if it might be possible to develop a single theory that describes all four of the known forces in a similar way. In the 1970s, physicists Sheldon Glashow (b. 1932), Abdus Salam (1926–1996), and Steven Weinberg (b. 1933) developed the mathematical framework for the unification of the electromagnetic and weak interactions.

The electroweak unification is only apparent at very high temperatures, when typical particle energies are above 100 GeV (10^{11} eV). These are energies that are easily attained by particle accelerators of the past several decades, and so the physics of the electroweak interaction has been well—studied in laboratories. To find a time when the entire Universe could be described by such energies, when the temperature was more than 10^{15} K, we have to go back to the first 10^{-12} seconds of its existence.

16.2.4.3: Quark confinement, $t = 10^{-6}$ s, $T = 10^{13}$ K

By about 10^{-6} seconds, quarks and antiquarks had cooled and slowed sufficiently that they could bind together in groups of three to produce protons and neutrons and their antiparticles. There was not enough energy to break the new compound particles apart, so this is when almost all of the protons and neutrons in the Universe were made, including those that make up your mass. Protons have stayed intact for 13.8 billion years, and they will stay intact for trillions of years more. Quark confinement is the first time the Universe creates composite (non-elementary) particles, and it is the first step toward the great complexity we see now in the Universe.

16.2.4.4: Baryon annihilation, $t = 10^{-4}$ s, $T = 10^{12}$ K

When the Universe was about 10^{-4} seconds old and the temperature dropped below about 10^{12} K, it was no longer possible to spontaneously produce protons and antiprotons from light. So soon after all the protons and neutrons were made they collided with their antimatter equivalents, composed of anti-quarks, producing huge numbers of photons. In every cubic centimeter of the Universe 100 million tons of matter and 100 million tons of antimatter converted into pure energy. To put this annihilation event into perspective, just the annihilation of the matter in your hand would produce the same amount of energy as a 10 megaton hydrogen bomb.

The great surprise—and a puzzle—is that a small number of protons and neutrons survived this carnage. In producing proton–antiproton pairs, standard particle theory plays no favorites. The photons should create just as many antiprotons as protons, and when the two annihilate they should convert completely back to photons. There should be no protons (or neutrons) left. However, there must have been an imbalance, because the Universe is overwhelmingly dominated by matter (and does not contain much antimatter). If there was a lot of local antimatter, we would see many energetic annihilation events all around us. If there were distinct matter and antimatter regions in the Universe, we would see annihilation events at the boundaries of these regions.

Therefore, there must have been a time early in the Universe when certain reactions produced slightly more “matter” than “antimatter” particles. It is from this excess of matter that all current matter is made. Out of every few billion particles, a single proton emerged that did not find an antimatter proton with which to annihilate. That number is easy to determine because annihilations produce photons, so the baryon-to-photon ratio tells us the ratio of survivors to annihilations. Without this tiny matter–antimatter asymmetry, the Universe would be empty of matter. This asymmetry resulted in a Universe today that is not just photons, but contains enough surviving protons and neutrons to make stars, galaxies, and us.

16.2.4.5: Neutrino decoupling, $t = 1$ s, $T = 10^{10}$ K

Neutrinos are elusive particles. They interact with protons, neutrons, electrons, and positrons via the weak force, as described in Section 14.2.2. Because they only feel gravity and the weak nuclear force, neutrinos can easily travel through immense thicknesses of lead without interacting with anything. Before the Universe was about 1 s old though, the matter density of the Universe was high enough for them to interact often. After 1 second the Universe was no longer dense enough for neutrinos to react, and they decoupled in an event similar to the one that occurred for photons about 380,000 years later - when they formed what is now the CMB. Since that time, neutrinos have been free to fly across the Universe unhindered. Some of them are passing through you now. If we can ever develop sensitive neutrino detectors, we will be able to obtain a picture of the Universe when it was 1 second old, just as the CMB reveals an image of the Universe at an age of 380,000 years.

16.2.4.6: Electron–positron annihilation, $t = 10$ s, $T = 4 \times 10^9$ K

By an age of 10 seconds, the Universe had cooled to a temperature of 4×10^9 K. At higher temperatures and earlier times, the Universe had enough ambient energy to create electron–positron pairs. Photon collisions produced electron–positron pairs, and when electrons and positrons collided, they produced photons. After this time there was no longer enough energy in the photons to produce electron–positron pairs and annihilation became a one-way process. Electrons and positrons annihilated, leaving huge numbers of photons and slightly more matter (electrons) than antimatter (positrons). The energy released at this time was equivalent to a 5000-megaton nuclear bomb in every cubic centimeter of the Universe.

The photons produced by matter–antimatter annihilations scattered off of the remaining particles (mainly free electrons) until about 380,000 years later when the Universe became transparent and the CMB was formed.

Electrons and neutrinos are key ingredients in the reaction that turns protons into neutrons via the weak force, so both electron–positron annihilation and the neutrino decoupling that happens right before it have important implications for Big Bang nucleosynthesis.

https://phys.libretexts.org/@go/page/31446

Particle Creation Timeline

In this activity you will combine elementary particles to create more complex ones, and you will determine where each step fits into a timeline of the early Universe.

Play Activity

A. Combining quarks

First you will see a plasma of up and down quarks of various color charges. Combine the quarks to create nucleons (protons and neutrons).

- To combine particles, click on them.
- A proton is made from 2 up quarks and 1 down quark.
- A neutron is made from 1 up quark and 2 down quarks.
- Particles made of quarks (hadrons) must be color-charge neutral, that is, composed of one red, one green, and one blue quark.

B. Combining nucleons

Combine the protons and neutrons to make all of the particles in the list at the bottom of the screen. The particles will light up when you have created the first one—you only need to make one of each particle type in order to proceed to the second part of the activity.

C. Timeline

Click and drag the particles to place them on the appropriate location on the timeline.

16.2.5: The Gut Time

Before the electroweak era, our understanding of the physics of the early Universe becomes much less certain. Thus, our picture of the Universe at those earlier times becomes correspondingly fuzzy. For instance, we think that, just as electromagnetism and the weak interaction combined at temperatures of 10^{15} K and energies around 100 GeV, it should be possible for the electroweak interaction to combine with the strong interaction at an even higher energy and temperature. This Grand Unified Theory (GUT) of the strong, weak and electromagnetic interactions is thought to become applicable at extremely high energies, although, at the present time, we do not know the exact value of the energy at which this “grand unification” occurs.

Present calculations indicate that the GUT threshold will be found somewhere around 10^{15} GeV. This is far beyond the capabilities of current particle accelerators and exceeds the energies of those likely to be built in the foreseeable future. The GUT threshold is an enormous energy that is equivalent to about 100 kilojoules *per particle* in the plasma. To put this energy into perspective, recall that the energy of a visible light photon is about 1 eV. To use a different comparison, a 90-mph fastball has about 120 joules of kinetic energy, so at the GUT energy scale, the *typical* particle energy is one thousand times higher than that of a fastball. The next activity asks you to convert this energy to its corresponding temperature scale.

The Gut Scale

At the GUT era of the Universe, the energies of particles are calculated to be at least 10^{15} GeV.

1.

2.

In this last activity, you determined the approximate temperature at which the strong, weak, and electromagnetic forces would all be unified into a single force. The transition to this GUT era would have occurred around a time of 10^{-35} seconds. GUT theories cannot now be experimentally tested because we cannot currently build accelerators that reproduce such high energies. However, the strong, weak, and electromagnetic forces can all be consistently explained within the framework of quantum mechanics.

So far, and within the framework of the Big Bang theory, we have probed the history of the Universe back to the first fraction of a second as described by well-tested laws of physics, including quantum mechanics, particle physics, nuclear physics, thermodynamics, and general relativity. But could there be additional phenomena, ones that are not explained by the basic Big Bang theory? How early in the history of the Universe can we probe? That is what we will examine next.

This page titled 16.2: Particle Soup is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

16.3: Inflation

Learning Objectives

- You will know that magnets are never seen as single poles and that and explanation for this led to a theory with cosmological implications.
- You will know that the Universe underwent a period of rapid, exponential expansion at early times.
- You will understand how inflation can explain the observed homogeneity and isotropy of the Universe.
- You will understand how inflation can explain the observed flat geometry of the Universe.
- You will know that inflation is driven by a release of energy from the field of a particle known as the inflaton.
- You will know that inflation makes predictions about structure formation, which can be tested.

What Do You Think: Inflation

The Big Bang theory explains a lot of what we see in the Universe very well. However, it cannot be complete. For example, the Big Bang theory does not explain why the Universe is isotropic and homogeneous on large scales. It also fails to explain why the geometry of the Universe is flat. Clearly there was something early in the history of the Universe that set up these conditions, but the original framework of the expansion model does not say what that was. Surprisingly, it was not from astrophysics that the best explanation for these observational puzzles arose. The explanation of the appearance of the Universe on the largest scales came from an attempt to understand a puzzle from other branches of physics. To understand how this happened, we will take a brief excursion into the physics of electromagnetism.

16.3.1: Magnetic Monopoles: A Puzzle From Particle Physics And Electromagnetism

By now you are familiar with the particle constituents of atoms: protons, neutrons, and electrons. You also know that protons and electrons have electric charge, whereas neutrons do not. In the classical theory of electricity, an electric field can be created by charged particles, so both the electron and the proton create an electric field in the space around them (Figure 16.18). On the other hand, the neutron, lacking an electric charge, does not create an electric field.

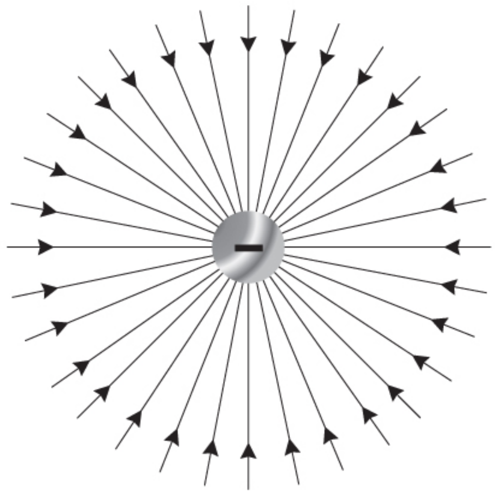


Figure 16.18: The electric field configuration for single negative electric charges like electrons. The electric field lines are directed radially inward for a negative charge (as indicated by the arrows in the figure). If the charge is positive, the field lines are directed radially outward. The arrows indicate the direction of the force that the field would exert on a positive charge placed in a given position. For a negative charge placed there, the force would be in the opposite direction, so opposite the direction of the electric field. Credit: Aurore Simonnet

For electric fields, the simplest type of field, as in Figure 16.18, is formed at a charged point-particle and flows radially inward or outward from that charge. Such a configuration is called a **monopole field** because only a single positive or negative charge (a single pole) is required to produce it. If two charges are close together, their fields interact to produce a **dipole field**, as shown in Figure 16.19. In this case there are two poles rather than one.

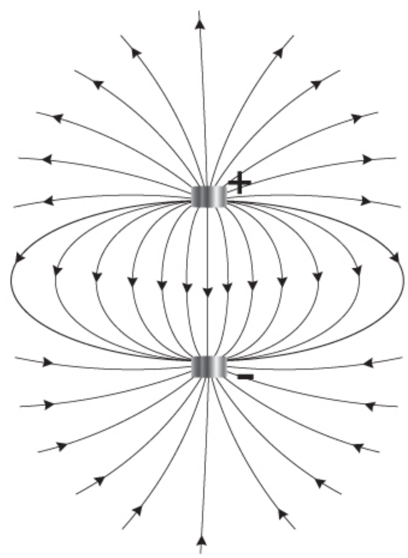


Figure 16.19: Electric dipole fields are created when two charges are in close proximity. Here we show the field for a dipole composed of opposite charges. As above, the arrows indicate the direction of the force exerted by the field on a positive charge placed at a given position. Credit: Aurore Simonnet

On the other hand, a magnetic field is formed in a very different way. Magnetic fields are created when an electrically charged particle moves through space. For example, magnetic fields form around wires carrying electrical currents. Currents are just streams of charged particles (generally electrons) moving in the wire. However, there are no individual "magnetic charges" as there are electric charges, so we do not see magnetic fields that look like the field shown in Figure 16.18. There is no known fundamental reason why there cannot be magnetic charges. In fact, some basic theories of particle physics suggest that such particles should exist. Still, none has ever been seen.

The simplest field configuration for a magnetic field always has two poles, called north and south, in reference to Earth's magnetic field, shown in Figure 16.20. Since there don't seem to be any individual magnetic charges, or if you like, no magnetic monopoles, it is not possible to create a magnetic monopole field.

The simplest magnetic field is a dipole field, shown in Figure 16.21. Note that the geometry of this field is the same as the geometry of the electric dipole shown in Figure 16.19. Only the geometry is the same here, the fields are not. One is an electric field, the other is a magnetic field. These are quite different, and they are formed in different ways, but related, ways. Unlike the electric field, the magnetic field is never created by isolated magnetic poles, and this lack of magnetic monopoles has had a profound influence on our understanding of the Universe.

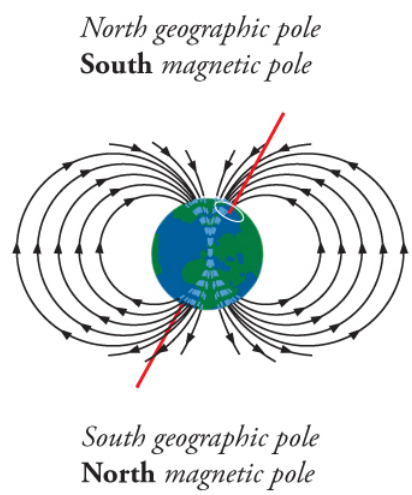


Figure 16.20: The magnetic field of Earth has a characteristic shape called a magnetic dipole. The field has two poles, one near the north geographic pole and one near the south - though the magnetic poles wander around somewhat on the surface. This field configuration is the simplest kind of magnetic field observed. Credit: Aurore Simonnet

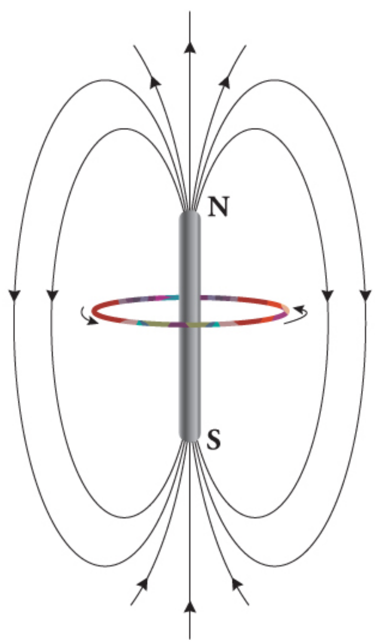


Figure 16.21: A magnetic dipole. Magnetic fields cannot have monopole configurations because they are formed by moving electric charges; there are no individual magnetic charges analogous to electric charges. Breaking a magnet in half does not result in isolated north and south poles. It creates two smaller magnets, each having its own north-south pole pair. Credit: Aurore Simonnet

Physicists long puzzled over why there are no magnetic monopoles. From pure symmetry grounds it seems that there should be. The equations describing the electric and magnetic fields are precisely symmetrical, or would be if there were magnetic monopoles analogous to electric charges. Such a particle would appear to be an isolated north or south pole. But in actual magnets the north and south poles always come in pairs. It is as if electrons and protons always came in pairs that could not be broken apart, so we only could see the electric dipole field of the pair, never the two separate charges.

In the 1970s and 1980s GUT theories were predicting that huge numbers of stable magnetic monopoles should have been made in the Big Bang, yet none had ever been detected. Such a large number magnetic monopoles were predicted in these theories that small loops of wire with a steady current would have been measurably disturbed by the close passage of a monopole about once every day. A couple of claims for detection were made, but none were convincing. Where, then, were all of the expected monopoles? This was a question that particle physicists set out to answer.

Of course, it was always possible the theories predicting monopoles were simply wrong, but other possibilities were also explored. It is these explorations that to the idea of this section: Inflation.

16.3.2: The Inflationary Paradigm

An interesting solution to the so-called monopole problem was devised independently by several theoretical physicists in the late 1970s and early 1980s. All of these solutions employed the concept of **vacuum energy**, the energy of “empty” space. Such vacuum energy has a negative pressure associated with it, something like the tension in a taut cord. In general relativity, this produces a repulsive effect. This repulsive force is the essence of the solution to the monopole problem. As it happens, it also solves many puzzling aspects of Big Bang cosmology as well.

Imagine a region of space that is empty of all matter. The region, though lacking matter, will still contain various fields. According to quantum physics, the presence of these fields will cause the region to have an energy associated with it. The energy results from a sea of virtual particles that constantly pop in and out of existence. To create additional empty space, it is necessary to furnish the energy in the vacuum it contains. This has the effect of producing a resistance to the creation of more empty space. This very resistance produces a negative pressure. Thus any small volume of space will repel adjacent volumes, paradoxically causing the creation of additional space. The energy of the newly created space is the same as that of the original, so the process of expansion is ongoing. In fact, as more space is created, it tends to cause the space to expand ever faster.

But how can the energy of this expansion arise? It can come from the decay of one of the fields in space. Just as an excited atom can decay when an electron drops to a lower energy state (releasing energy in the process) an excited field can also decay. If it does, then something quite extraordinary can happen. As the field decays, the energy released can create new particles. It can also cause space to expand, and by an immense factor.

It is now thought by many cosmologists that just such an expansion took place in the earliest moments after the Big Bang, and this brief period of rapid expansion is now generally referred to as Inflation. A young theoretical physicist named Alan Guth (b. 1947, Figure 16.22) stumbled upon this discovered around 1979 while he was trying to understand something else entirely: why no magnetic monopoles ever showed up in physics experiments. Early breakthroughs also came independently by Alexei Starobinsky (b. 1948) and Andrei Linde (b. 1948) in the Soviet Union. However, it was Guth who first saw the importance of the idea for cosmology.

Figure 16.22: A recent picture of Alan Guth. Credit: Betsy Devine / CC-BY-SA-3.0

Guth's discovery suggested a reason for the lack of monopoles: even if the Universe had been filled with monopoles at its birth, a phase of rapid expansion as envisioned above would so dilute them as to make it seem as though none existed at all. In this brief period of Inflation, the Universe expanded by at least a factor of 10^{20}, and possibly by a much larger factor (Figure 16.23). That means that two particles originally a micron apart at the beginning of the Inflationary epoch would end up 10^{20} times farther away, or at least 10^{14} m apart at the end of Inflation. This is a size several times larger than the solar system. The rapid expansion would extend across the entire part of the Universe in which the inflationary field was decaying, and any monopoles would be swept across the horizon as space rapidly expanded.

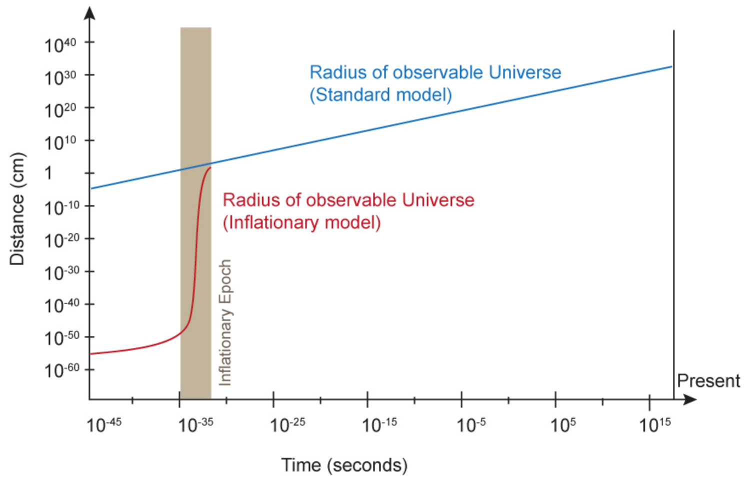


Figure 16.23: During a brief period of inflation, the size of the Universe expanded exponentially. Credit: NASA/SSU/Aurore Simonnet

Monopoles might or might not exist, so having a mechanism that would make them extremely rare is interesting, but not really compelling. However, Guth soon realized that Inflation would also have striking effects on appearance of the Universe as a whole, addressing key issues that the basic Big Bang theory does not.

16.3.3: Inflation Solves the Horizon Problem

So far we have seen that the Universe on the whole is homogeneous and isotropic. If we look in one direction in space we can see objects that are now tens of billions of light years away. When we look in the opposite direction, we see other objects similarly distant. Both regions look strikingly similar. Why should that be? They are so far from each other that nothing (not even light) can have passed between them to coordinate conditions. They are beyond each other's cosmic horizons (the distance that light has traveled since the Universe began). This mystery, on which a simple Big Bang theory is silent, is called the horizon problem.

The horizon problem is most striking in the cosmic microwave background. It has essentially the same temperature to a tiny fraction of a kelvin all around the sky. How did this condition come to be if these parts of the Universe were never able to exchange information or energy? Inflation provides a way.

Imagine a handful of patches of space close together only an instant after the expansion of the Universe began, but before the onset of Inflation. The regions will experience two effects. The expansion will stretch them away from each other, but as time passes they will also be able to see things farther away as light has more time over which to travel. They might all start with unique properties. Do they have the opportunity to share those properties or are they stretched away from each other too quickly?

In the standard Big Bang model they are always stretched away too quickly. The solutions of the Friedmann equation for radiation and matter dominated Universes all have a stretching rate that approaches infinity as we approach t = 0. Even though the horizon distance for each patch of space will grow as light speed times the elapsed time, that is no match for the stretching rate, and the patches are never able to exchange energy.

If, however, the Universe had an episode of Inflationary growth early in its history, the stretching could have started slowly before accelerating to a high rate. That would provide more time for sharing. In addition, the tiny regions that were able to share energy were inflated to regions much larger than we can currently observe. So even if the entire Universe has diverse properties on very large scales, on scales the size of our visible Universe it is homogeneous. Thus inflation solves the horizon problem, creating the homogeneous and isotropic conditions we observe.

16.3.4: Inflation Solves the Flatness Problem

In a Universe described by general relativity, space can have an overall, or global, curvature. It can be flat (zero curvature) and follow the familiar Euclidian rules of geometry (parallel lines never meet, triangles have three internal angles that always sum to 180 degrees, the circumference of a circle is π times the diameter, etc.), but it can also be curved and follow other rules. There is nothing in the basic Big Bang theory that requires the curvature of the Universe on the largest scales to have any particular value. But observations of the CMB have allowed us to measure the curvature to a precision of less than one percent, and they show it to be flat.

Why should the curvature be close to that special value? There are far more ways to create a Universe with high curvature than one that is precisely flat. Why should the Universe appear to be "fine tuned" in this way? Perhaps there is a good reason for it, but there is nothing in the standard Big Bang theory that says it should be flat. It seems a strange and unsettling coincidence. This is called the flatness problem.

Einstein's equations of general relativity tie the curvature of the Universe as a whole to its average matter and energy density. A flat geometry results only if the Universe has the critical density, with a precise balance between runaway expansion and runaway collapse. This seems extremely unlikely. But the Universe does seem to be in this special state. For the overall density to be equal or close to this special density today means it must have been even closer at earlier and earlier times. If the density is even a little bit off from the critical value at some time, at any latter time it will be further off. For more on why this is so puzzling, see Going Further 16.3: Why the Flatness Problem Is a Problem.

Inflation provides a way out of this puzzling situation. If the Universe expanded by a huge amount early in its history, then any curvature would be greatly diminished on the scales we can measure. As an analogy, consider Earth's surface. Globally it has a geometry resembling a sphere. However, standing on the ground you do not notice its curvature. Standing on a tall building in Chicago and looking around you, Earth's surface still seems to be flat. It will even appear flat if you stand on a high mountain above the California coast and look out to sea. Even if you look out the window of a jet aircraft cruising aloft, Earth appears flat. None of these perspectives lets you discern Earth's global curvature. The size of the planet is so vast that its global curvature is completely hidden on such small scales.

A similar situation applies to the Universe after Inflation. Even though the Universe might have some global curvature, Inflation expanded it so much that we can currently see only a tiny fraction of it. On these relatively tiny scales, and even on larger ones, the Universe appears flat. This is notwithstanding its global geometry. That is how Inflation solves the flatness problem: it inflates it away, as is depicted in Figure 16.24.

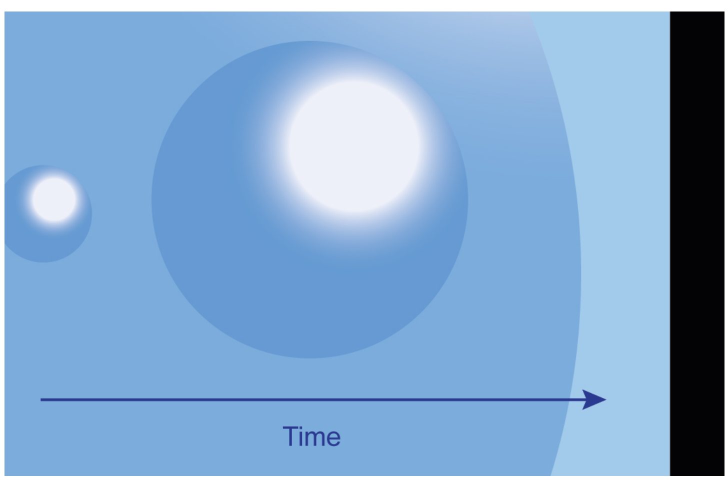


Figure 16.24: As the scale factor of the Universe grows exponentially during the time of inflation, any curvature becomes less significant, to the point of looking flat. Credit: NASA/SSU/ Aurore Simonnet

Going Further 16.3 Why the Flatness Problem Is a Problem

The coincidence of a flat Universe, or in other words, one with a density equal to the critical density, is quite special. Since there is no particular reason that the Universe should be close to the critical density, it is strange that it does lie so close.

If the density of the Universe is equal to the critical density, the expansion rate exactly equals the rate needed to make the Universe flat, so the Universe would expand forever but at a velocity always slowing towards zero. Understanding why this is the case is difficult without mathematics, so let’s try a simplified illustration.

We will use an analogy to the Newtonian notion of the escape speed from a planet. For illustrative purposes we can replace the expanding Universe with a single object leaving a planet at escape velocity, depicted in Figure B.16.3. The physics will be qualitatively the same, but the mathematical treatment is much simpler in the Newtonian case. If we reduced that object’s velocity by 1% it would no longer escape but would turn around and fall back. For example, if we launch a rock off of Earth’s surface at 99% of the escape speed, the rock would get almost all the way to the Moon before it stopped and fell back to Earth.

If we fancifully imagine that the planet at earlier times was compressed to smaller radii (the way the size of the Universe is smaller at earlier times), then we will find a striking effect. Launching our rock with the same 99% of the escape velocity for a compressed planet will hardly raise it off the surface at all before it falls back. For example, to get to the Moon’s orbit with a compressed Earth, we would have to launch the projectile with a higher velocity, greater than 99% though still slightly less than the escape velocity. The following activities show why this is so.

In this Newtonian example, the velocity of the object gets closer and closer to the escape speed as a planet gets smaller, analogous to the density of the Universe getting closer and closer to the critical density as the size of the Universe shrinks at earlier times.

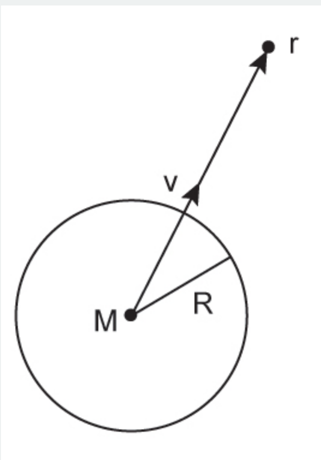


Figure B.16.3: Object Launched from a Planet Surface. If an object with mass m is launched at some velocity v from the surface of a planet with mass M, how far up does it go before it stops and falls back down? Credit: NASA/SSU/Aurore Simonnet

If an object of mass m is thrown upward with speed v from the surface of a planet it will have a total energy (E_{tot}), which is a combination of kinetic and potential energy, given by the equation below.

$$E_{tot} = \frac{1}{2}mv^2 - \frac{GMm}{R}$$

Here M is the mass of the planet and R is its radius. If we assume that the velocity is less than the escape velocity and the projectile eventually rises to the point r, then the total energy can also be written as follows.

$$E_{tot} = -\frac{GMm}{r}$$

Setting these two expressions equal to each other we can relate the distance traveled to the size of the planet.

$$\frac{GMm}{r} = \frac{1}{2}mv^2 - \frac{GMm}{R}$$

The escape velocity is found by setting the total energy to zero. (In that case the object will eventually have no kinetic energy, and its potential energy will be zero since r will become infinite.) But we are not looking for the escape speed here, we wish to know how high the object travels before it stops and falls back.

We can cancel the common factor of m from each term, and then solve for r.

$$\frac{GM}{r} = \frac{1}{2}v^2 - \frac{GM}{R}$$

https://phys.libretexts.org/@go/page/31447

Rearranging we get the following relation.

$$r = \frac{-GM}{\frac{1}{2}v^2 - \frac{GM}{R}}$$

This expression can be greatly simplified by factoring out the GM/R from the denominator.

$$r = \frac{-GM}{\frac{GM}{R}\left(\frac{Rv^2}{2GM} - 1\right)}$$

Now we can cancel the common term of *GM* from the numerator and denominator and raise the factor of $1/R$ from the denominator into the numerator. In addition, we see that there is a factor of the escape velocity in the denominator: $v^2_{esc} = 2GM / R$. Thus we can simplify as follows:

$$r = \frac{-R}{\left(\frac{v}{v_{esc}}\right)^2 - 1}$$

Now we can tidy up a bit by canceling a factor of *-1* from the numerator and denominator of the right-hand side, and we can divide through by *R* to get a final expression for how high an object rises given a planet's size and the object's velocity.

$$\frac{r}{R} = \frac{1}{1 - \left(\frac{v}{v_{esc}}\right)^2}$$

We can use this expression to determine how high an object will rise if launched from Earth, as in the following worked examples.

1. How far will an object rise off the surface of a planet (relative to the planet's radius) if it's launched at 90% of the escape velocity?

- Given: $v = 0.90v_{esc}$
- Find: r/R
- Concept: $r/R = 1/[1 - (v/v_{esc})^2]$
- Solution: $r/R = 1/[1 – (0.90)^2] = 5.26$

This means it will rise to 5.26 times the planet's radius.

2. How far will an object rise off the surface of a planet (relative to the planet's radius) if it's launched at 99% of the escape velocity?

- Given: $v = 0.99v_{esc}$
- Find: r/R
- Concept: $r/R = 1/[1 - (v/v_{esc})^2]$
- Solution: $r/R = 1/[1 – (0.99)^2] = 50$

This means it will rise to 50 times planet's radius.

So the object will rise to about fifty times the radius of the object from which it is launched.

3. How far is this for Earth, which has a radius of 6400 km?

- Given: R = 6400 km, r/R = 50
- Find: r
- Concept: r = (r/R)(R)
- Solution: r = (6400 km)(50) = 320,000 km

This is nearly the distance to the moon.

We can also plot the distance reached (*r/R*) versus launch velocity (v/v_{esc}). To create the graph, we can calculate *r/R* for several values of v/v_{esc}, as in Table B.16.1.

Table B.16.1. Velocity and peak height of an object launched from a planet.

v/v_{esc}	r/R
0.0	1.0
0.2	1.04
0.4	1.19
0.6	1.56
0.8	2.78
0.9	5.26
0.95	10.26

From this table, we use the *Graphing Tool* to graph r/R vs. v/v_{esc} in Figure B.16.4.

Figure B.16.4 height and escape speed

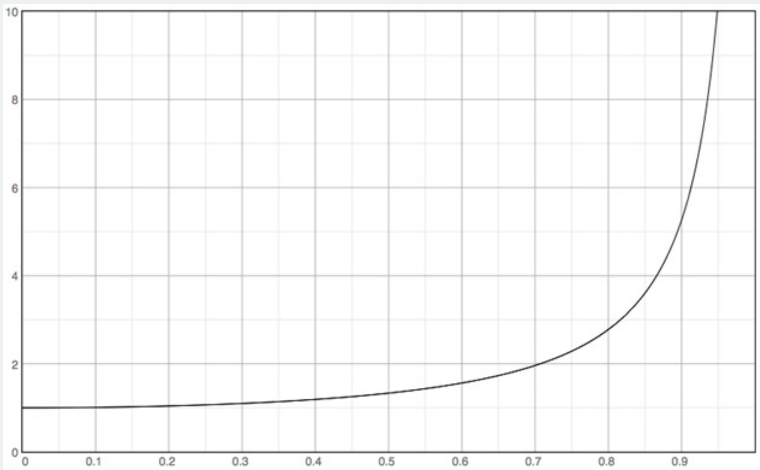


Figure: The ratio of the height attained to the radius of the planet as a function of velocity in terms of escape speed. Credit: NASA/SSU

From this graph, we can answer the several questions. You will see that the ratio of launch velocity to escape velocity becomes much more sensitive to deviations from 1.0 as an object shrinks. This is analogous to the expansion of the Universe: the Universe must be very exquisitely balanced near its critical density in order for it to be so close to that density today.

1. Why is the height attained equal to 1 when the launch velocity is zero?

   Since the projectile begins at $r = R$, if it has zero initial velocity it never leaves the surface, so $r/R = 1$ at its highest point.

2. Explain why the height seems to rise without bound as the launch velocity (in terms of escape velocity) approaches 1.

   When $v = v_{esc}$ the projectile is able to escape to infinity, thus as the launch speed approaches that value the height attained approaches infinity.

3. How does the height reached by a projectile depend on the speed at which it is launched?

   As the speed becomes a larger fraction of escape speed the height attained increases, regardless of the size of the planet. The answer does not depend on the size of the planet because both axes on the graph are in terms of the planet's size. The escape speed includes all of the relevant information on the planet's mass and radius, and the height attained is in terms of the planet radius, so any dependence on planetary parameters are hidden by this method of plotting.

4. As the planet gets smaller, what must happen to the launch velocity in order to cause the physical height reached by the projectile to remain constant?

If the projectile is to reach a constant value of r as the size of the planet (R) decreases, then its velocity as a fraction of escape speed must increase. Thus the absolute launch speed must also increase.

5. In this exercise, escape velocity is analogous to the critical density of the Universe. How does this exercise demonstrate that, if the density of the Universe today is close to the critical density, it must have been even closer to critical density in the past?

   The escape speed of a planet of given mass will increase as the planet is made smaller. That means that if an object is near escape speed on a large planet, it must be even nearer escape speed on a smaller planet of the same mass in order to reach a similar height. The ratio v/v_{esc} must approach 1 as the planet gets smaller in a way similar to the way that ρ/ρ_{crit} must approach 1 as we look further into the past when the scale factor of the Universe was smaller than it is today.

16.3.5: Possible Mechanisms for Inflation

The basic starting point for Inflation is the Friedmann equation of general relativity. It predicts that the expansion of a volume of space should slow due to the braking effect of the mass–energy present. However, unlike Newtonian gravity, it also permits a term due to the energy inherent in the vacuum fields of “empty” space. These have an opposite, accelerating effect—a topic we will look at in more depth later.

According to Inflationary theory, most of the energy in the Universe was initially concentrated in an energy field called the inflaton field. Fields are associated with many elementary particles, and usually the energy in these fields is zero. Inflation posits that the field of one particle—the inflaton—got stuck on a non-zero quasi-stable value during one of the force-separating transitions. The inflaton field permeated space and had a large amount of energy stored in the form of vacuum energy. In this case, the vacuum was not a true vacuum, but corresponded to an excited state of the field, called a **false vacuum**.

When the field began to decay, it released its energy, and since the vacuum energy was tied to space itself, as space expanded, the inflaton field caused the Universe to expand more and more rapidly. As more space was created by the expansion, the total vacuum energy also grew, making the expansion accelerate, and so on exponentially. The size of the Universe doubled over and over in very short periods. We do not know exactly how many doublings there might have been, though the minimum must have been enough to flatten and smooth the Universe on the largest visible scales. During the inflationary stage, the Universe expanded by at least a factor of 10^{20} or perhaps as much as $10^{100000000000}$ (1 followed by a quadrillion zeros).

The Universe did not remain in such an unstable state for long. Eventually the field decayed to its ground state, a true vacuum. Its energy was converted into other forms, such as particles, and the exponential expansion ceased in favor of a more steady expansion. The energy of the inflaton field may have decayed in what is known as a "slow roll," in which it stayed at a high value for a comparatively long time (though 10^{-35} seconds seems fast to most of us!) as if rolling down a hill through honey. Once it was part way down the hill however, the decay accelerated. When the value reached zero, inflation ended. The energy of the field went into repopulating space with energetic particles in a process called reheating. After reheating, the Universe looked almost the same as predicted by non-inflationary models—except all the troubling coincidences (such as the horizon and flatness problems) now make sense and the Universe is much bigger! The possible evolution of the energy of the inflaton field is shown in Figure 16.25.

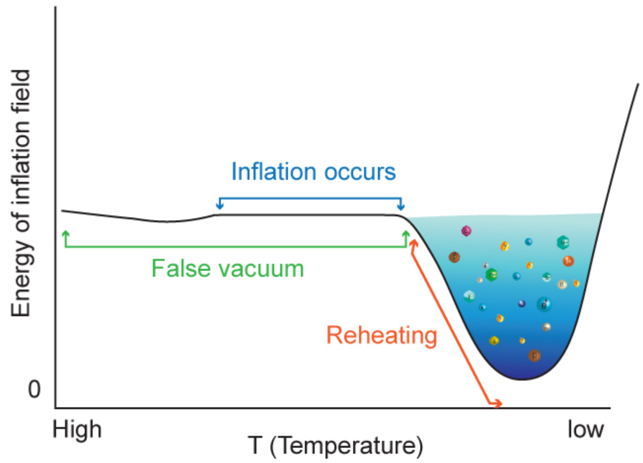


Figure 16.25: The inflation field started out stuck in a high-energy state called a false vacuum. This state required a long time (by these standards) to decay. While the decay of the field happened slowly, the Universe inflated. When the evolution of the field quickened and its energy dropped rapidly, reheating occurred until the energy of the field hit a true vacuum at zero, ending inflation and creating the mass/energy we see today. Credit: NASA/SSU/Aurore Simonnet

There are many possible theories varying the details of Inflation. In some, the field comes into play at the GUT scale. In others, its effects appear at higher energy scales. Inflation may have begun as early as 10^{-43} seconds after the beginning of the Universe, lasting about 10^{-35} seconds, or begun as late as 10^{-35} seconds, lasting about 10^{-32} seconds. Whenever the inflationary epoch was, the exponential expansion it caused gave us the flat, homogeneous, and isotropic Universe we now inhabit.

16.3.6: Additional Observational Tests

As always in science, we insist that ideas like Inflation, which explain so many features of the Universe, must make testable predictions. Inflation makes many such predictions, and we now have the ability to test some of them.

The basic Big Bang theory cannot explain either why the Universe is so homogeneous on billion light-year scales, or why we see any structure at all in the Universe. Why are there galaxies and clusters of galaxies? Why is the CMB slightly lumpy? In the basic Big Bang theory there is no reason for the expansion to be other than perfectly smooth. But a smooth expansion would not have the lumpiness to produce galaxies from gravitational collapse. There must be some mechanism for providing the “seeds” of these structures early in the history of the Universe, so that they can then grow via gravity.

Inflation suggests an answer to this problem too, by providing seeds of inhomogeneity of just the right size. These seeds stem from the inevitable fluctuations on the quantum scale. Inflation would have magnified and frozen-in whatever random set of fluctuations were happening on quantum scales at the moment the rapid expansion commenced. The fluctuations then provided the seeds for the structures we now see on large scales.

Inflationary models also make predictions about the patterns of structures in the Universe, and we can test some of these already. Inflation can provide theoretical seeds of structure that produce models that match the observations of the power spectrum of the CMB to high precision. Furthermore, most inflationary models use fundamentally random seeds to produce structure and the results of these models agree with the structures observed.

The clearest signal of inflation would be a specific signature pattern that would result from the interaction of light with gravitational waves. Inflation is such a dramatic event that it shakes the fabric of space. Gravitational waves created by this shaking would affect the CMB light as it travels through space, imposing a pattern on its polarization.

We have seen that electromagnetic waves have an electric field and a magnetic field perpendicular to each other and to the light’s direction of travel. Most light sources emit light with the electric and magnetic fields oriented randomly, but there are many processes that can favor particular orientations, producing light that is polarized. The effect on light by the gravitational waves generated by inflation should produce a specific pattern of polarization. This effect was tentatively reported by the BICEP team in March 2014. However, upon further analysis with higher quality data from the Planck satellite, the polarization results did not hold up. It is still possible that CMB polarization of the kind predicted by Inflation will be found in the future, but so far (as of 2021 summer), the search for this signal continues.

This page titled 16.3: Inflation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

16.4: The Beginning

Learning Objectives

- You will know that the early Universe was extremely dense and that changes happened rapidly.
- You will know that the Planck time is the earliest era we can probe.
- You will know that at the Planck time all four forces are thought to be unified.
- You will know that all of space and time were created in the beginning.

What Do You Think: Beginning of the Universe

16.4.1: Thinking About the Early Universe

Using mathematical models, we can push backward to times of high temperature and density in the Universe, and we can predict what happened within the first moments in its history. Extrapolations like these become less certain the earlier back we go. Eventually, the physics we currently understand need no longer be valid at all. So what can physics tell us, if anything, about conditions at these earliest times?

Two things sometimes bother students when thinking about conditions in the early Universe. One is the extraordinary density at these times. At $t = 10^{-4}$ seconds the density is that of an atomic nucleus. At earlier times, the density would have been higher still. How can anything be denser than an atomic nucleus? The answer is that even atomic nuclei are mostly empty space. We have already looked at the structure of atoms and seen that they consist of tiny nuclei containing protons and neutrons, surrounded by clouds of electrons. The ratio of the volumes of nuclei to atoms exceeds that of a flea to a football stadium. Atoms are almost entirely empty space. But nuclei are themselves also mostly empty space. The protons and neutrons occupy only a tiny fraction of the total volume, and each of them consists of just three quarks inside a volume trillions of times larger—in fact, it could be that the quarks inside nucleons have no well-defined size at all, similar to electrons, neutrinos, and photons.

Perhaps even more difficult to grasp is how it is possible that all the astonishing events we have considered in this chapter occurred in just a few minutes. As we contemplate the entire sweep of cosmic history, it is clear that big changes happened much more quickly at early times, and change has been far more gradual recently. The concentration of events early on is astonishing. However, at those times everything was much closer together and the relative energies and speeds of the individual particles were far higher. Even given that, it is difficult to comprehend how all those events were able to happen so quickly.

If we step back and consider a few common, but quite unfamiliar, timescales, it might help to put this all into perspective.

Our perceptions of the world are limited by our speed of thought, which evolved to allow us to negotiate our everyday world. Thinking is a slow process compared to many of the processes that occurred in the early, hot Universe. In order to think, trillions of neurons, each containing trillions of atoms, have to perform long sequences of complex operations. Thoughts require times of the order seconds to complete. A second is an immensely long time on the subatomic scale. Molecules can vibrate many billions of times in one second. During each single vibration, electrons orbit nuclei perhaps a million times. Protons may orbit the center of a nucleus a million times for each electron orbit. Inside protons and neutrons, the constituent quarks complete a million orbits for each proton orbit. Another way of looking at it is that many elementary particle reactions reach balance (equilibrium) in as little as 10^{-18} seconds. So, for an elementary particle, a second is like many billions of years for a human. There is plenty of time to get things done.

We will now discuss the limits of our current scientific understanding of the early Universe.

16.4.2: The Planck Time

The earliest time we can speculate about is the Planck time. We learned of this timescale already in Chapter 9 when we considered black holes and the failings of general relativity near the singularity at their centers. In Chapter 9 we introduced Planck units as a somewhat crude way to think about such singularities. Planck units are constructed from combinations of physical constants in ways that give characteristic scales (of length, time, mass, etc.) at which we expect quantum effects to be important. One of these was a unit of time given by the expression below.

$$t_{planck} = \sqrt{\frac{hG}{c^5}}$$

In this equation, h is Planck's constant, G is Newton's gravitational constant, and c is the speed of light in vacuum. Numerically, the Planck time is 10^{-43} seconds. On timescales of this size or smaller, the effects of quantum mechanics become dominant. Our classical theories no longer work. When the Universe had an age smaller than the Planck time, we cannot expect any of our physical laws to give us an accurate picture of its behavior. This is because the random effects of quantum physics will completely dominate its dynamics.

We can think of this in another way: We can construct the Planck length, the distance over which a photon could travel in a Planck time.

$$l_{planck} = ct_{planck} = \sqrt{\frac{hG}{c^3}}$$

Numerically, this works out to about 4×10^{-35} m, a tiny distance. When the observable Universe was of roughly this size, quantum effects would have dominated. We are not able to make any predictions regarding its behavior when everything we can see now was compressed into the size of the Planck length.

Thus, the beginning of time and space as we know it starts near the Planck time, $t = 10^{-43}$ seconds. The temperature of the Universe was around 10^{31} K and the density was near 10^{92} g/cm^3. Before this time the known laws of physics break down and time and space have no meaning. As far as we know, this is when time and space began, at least as far as our current conception of them is concerned.

16.4.3: Unification of the Forces

We learned how at a time of 10^{-12} seconds and a temperature of 10^{16} K, the electromagnetic and weak force combined into the electroweak force. We also learned how at a time of about 10^{-35} seconds and a temperature of 10^{27} K, these forces could have merged with the strong force, at least as described by GUT theories. Near the Planck time, we speculate that such high temperatures and densities may have had the effect of unifying all of the four fundamental forces of nature: gravity could combine with the GUT force. Physicists suspect that at the Planck time, all four forces would have been in perfect symmetry, unified in a **Theory of Everything** (ToE). The energies at this stage are far beyond our ability to reproduce in laboratory experiments, so these ideas are highly speculative. Today the strengths of the four forces have a tremendous range—a factor of 10^{38}—but at the extreme temperatures of the Planck time the strengths are expected to converge toward a single value. Figure 16.26 illustrates the time and temperature scales for the unification of the forces.

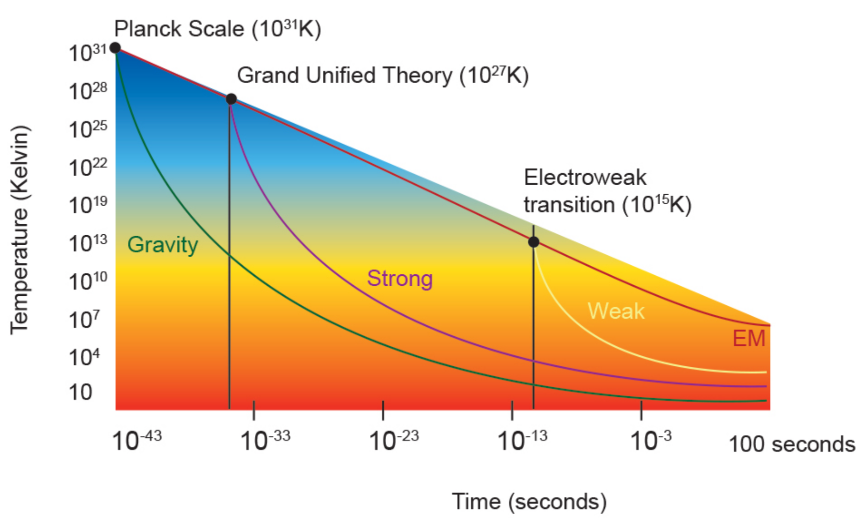


Figure 16.26: Physicists think that all four basic interactions between particles, the electromagnetic, weak, strong, and gravitational, might have all unified as a single interaction in the early Universe. As the Universe expanded and cooled, each force broke off from the others as the energy density in the Universe dropped. First gravity, then the strong interaction broke off. Finally, at 10^{-12} seconds, the weak force separated from electromagnetism, giving us the four interactions we have now. Credit: NASA/SSU/Aurore Simonnet

Much of what we can say about force unification is speculative, but there is some evidence that we are on the right track. We do observe the electroweak unification directly in particle accelerators, and it matches the theoretical predictions very well.

Unfortunately, both the GUT and ToE unifications are well beyond the energy ranges (temperatures) we can produce with current technology. We will have to await future experiments to know if our ideas about times before electroweak unification are correct.

One of the challenges in understanding the ToE unification is that gravity is described by general relativity, whereas the other three forces are described by quantum mechanics. Furthermore, the Big Bang model is built on general relativity. General relativity is what is called a "classical" theory; it takes no account of quantum structure on the smallest scales. We know that on small scales quantum effects dominate the dynamics of the Universe. In the earliest moments of cosmic expansion, the entire observable Universe would have been small enough to be completely dominated by quantum effects—the roiling and bustling of virtual particles popping in and out of existence. What would these effects have done to the Universe when it had a size comparable to these wild quantum fluctuations? To fully answer that question, we need a quantum theory of gravity. Only such a theory will allow us to properly understand the behavior of the Universe at the earliest times. However, a quantum theory of gravity remains elusive.

One way to view the separation of a single force into the four we now experience is as the emergence of the basic laws of nature during these critical early moments. Physicists often consider these laws in terms of what they call invariance or symmetry principles, which tell us whether some property of the Universe is affected by some type of change. Does the behavior change if I do it over there or over here? What if I rotate my perspective? If not, the property is said to have translational invariance, or symmetry, in the first case, and rotational symmetry in the second. If the property appears to operate the same way whether I watch it in normal time sequence or with the 'film' played backwards, it has time symmetry/invariance.

Starting in 1915, the German mathematician Emmy Noether (1882–1935, Figure 16.27) demonstrated a fundamental correspondence between such symmetries and the older concepts of conservation laws in physics. Invariance under translation (whether the process occurs over here or over there), for example, requires the law of conservation of momentum. Invariance under rotation requires the law of conservation of angular momentum, and invariance with time implies the law of conservation of energy. There are many quantum symmetries that require other conservation principles (such as those for charge, and the number of various types of particles). Symmetries are an expression of the laws of nature.

Figure 16.27: Amalie Emmy Noether was a German mathematician who made important contributions to topics in algebra and theoretical physics. Her work on symmetries in nature, called Noether's theorem, showed that such symmetries lead to conserved quantities. Credit: Science Source/Photo Researchers, Inc.

At earlier and earlier times, the Universe was in a sense simpler and simpler. The complexity now expressed in all our physical laws has emerged through a sequence of broken symmetries. They occur naturally as the Universe cools. Physicists suspect that one of them, early on, caused the single ToE force to separate into a GUT force and gravity. Later, when the Universe cooled further, another symmetry was spontaneously broken causing the GUT force to separate into an electroweak force and the strong nuclear force. At that point there were three particle interactions: gravity, the strong, and the electroweak. Finally, another symmetry was broken and the electroweak force became the electromagnetic interaction and weak nuclear interaction that we observe today.

The particular symmetries that were broken in the early Universe have to do with the form of the equations that describe particle interactions, but they work similarly to the way rotation does in mechanics, at least until they are broken. The interactions of gravity in a ToE might be related to a concept called supersymmetry, the idea that for each force carrier particle there ought to be a corresponding mass particle and for each mass particle there should be a force particle.

https://phys.libretexts.org/@go/page/31448

For example, the fast-moving massless photon, which interacts strongly with matter, would have a slow-moving massive counterpart called the photino. It would not interact strongly with matter. The collection of the massive particles that include the photino are called **weakly interacting massive particles** (WIMPS). They have never been detected, but they represent one possible variety of dark matter.

In this scenario, all but the lightest supersymmetric particle would be unstable. They will have decayed by the present day. The surviving (stable) lightest WIMP is a popular candidate for the dark matter particle that apparently dominates the mass of galaxies and galaxy clusters. The decay of the heavier, unstable WIMPS is analogous to the way free neutrons decay to form more stable protons.

So far supersymmetry is unverified. The lack of success in detecting any of the particles it predicts at accelerators like the LHC have almost ruled it out. Perhaps it is not the correct view of the world after all. But it is just one avenue being explored by physicists as they try to understand matter in its most basic forms.

GOING FURTHER 16.4: SYMMETRY BREAKING

16.4.4: The Beginning of Expansion and the Meaning of T = 0

The Planck time is the earliest moment about which we can make scientific statements. But the basic Big Bang theory includes a moment when there was infinite density—a singularity, at $t = 0$. Can we go all the way back to $t = 0$? Whenever a scientific model has a singularity, it is an indication that something is wrong, that the model has broken down and we need something better as we approach the singularity. In the case of our simplest Big Bang model, we know it is failing at $t = 0$.

When thinking about the beginning of the Universe, many people wonder what started it all. That is outside the scope of the Big Bang theory, which describes how the Universe changes *after* the Planck time, through today, and into the future. That is not necessarily a fatal problem. Just as there is a South Pole and nothing further south on Earth, the Universe may have “begun” in a hot, dense state, asymptotically approaching $t = 0$, and there was never any $t < 0$ to worry about.

Georges Lemaître (1896–1966, Figure 16.28) was among the first scientists to argue from the observed Hubble expansion that the Universe must have developed from a super hot, dense era with no obvious antecedent. In fact, he had derived Hubble’s law from general relativity two years before Hubble published his results on the expansion of the Universe. Lemaître called the start of space and time “the day without yesterday.”

Figure 16.28: Georges Lemaître was a Belgian cosmologist - an ordained Catholic Priest who was also a professor of physics. He deduced the cosmic expansion from Einstein's equations of general relativity before Hubble announced his discovery. Lemaître also devised the first theory based on general relativity and the expansion of the Universe that suggested the cosmos originated in a hot, dense state. Image © Shutterstock, Inc.

In this chapter, we have explored the earliest moments of the history of the Universe. The very first seconds of the Universe are not only intrinsically fascinating, they produced many structures that have remained unchanged ever since—like the protons in your body. These early moments began a continuing process of nested structure assembly, permitting more and more complexity—quarks to protons to nuclei to atoms to molecules to living organisms and planets, stars and galaxies. It is because these structures bear the imprints of earlier times that we can make inferences about the Universe's beginnings.

Our current understanding of physics allows us to understand times back to the first trillionth of a second in surprising detail. We can even push back, somewhat tentatively, to times when the Universe was only 10^{-35} seconds old. However, our current state of comprehension does not allow us to probe all the way back to the beginning. We require advances in both theoretical and experimental physics before we can confidently describe the earliest moments. At a minimum, such an understanding requires a more complete knowledge of gravity and its presumed quantum nature on small scales. For the moment, this understanding remains beyond our reach.

This page titled 16.4: The Beginning is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

16.5: Wrapping It Up 16 - Timeline of the Early Universe

Learning Objectives

- You will use a timeline to pull together their understanding of conditions and transitions in the early Universe. Transitions occur due to temperature, which corresponds to energy, and which decreases over time.

In this chapter you learned about transitions the Universe goes through in a short time as its temperature changes. In this activity, you will use an interactive timeline to answer questions about these transitions and the resulting conditions the early Universe. The timeline is divided into three time periods, including the early Universe. To learn more about an event in the early Universe, first click the "Early Universe" tab, then click the tab for the event.

Play Activity

16.5.1: **I. Transitions in the Early Universe**

16.5.1.1: A. Force transitions

1.

2.

https://phys.libretexts.org/@go/page/31449

3.

16.5.1.2: B. Matter and antimatter produce photons

1.

2.

3.

4.

16.5.1.3: C. Building structures

1.

2.

3.

4.

16.5.1.4: D. Other events

1.

2.

16.5.2: **II. Cosmological Eras**

In the opening video of this chapter you saw a timeline of the early Universe divided into different eras. The names of the eras reflect the dominant particles and forces at different times in the history of the Universe. The transitions that we have discussed mark the boundaries of these eras.

1.

2.

3.

4.

5.

6.

After several billion years, a mysterious thing called dark energy began to dominate the matter-energy density of the Universe, starting the dark energy era that continues today.

This page titled 16.5: Wrapping It Up 16 - Timeline of the Early Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

16.6: Mission Report 16 - Timeline of the Early Universe

A.

B.

C.

D. Questions to be graded for accuracy. Show your work!

1.

2.

3.

4.

This page titled 16.6: Mission Report 16 - Timeline of the Early Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

CHAPTER OVERVIEW

17: Dark Energy and the Fate of the Universe

Chapter 17 focuses on dark energy and its role in determining the ultimate fate of the Universe. The relatively modern discovery of the accelerated expansion of the Universe is described and different possibilities for the nature of dark energy are discussed. Measurements of dark energy and dark matter and of normal matter and energy are combined within the cosmic concordance model, and the implications of this model are discussed. The chapter concludes with a summary of our current knowledge of the state of the Universe.

This page titled 17: Dark Energy and the Fate of the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.0: Dark Energy and the Fate of the Universe Introduction

The discovery that the Universe is expanding, though unexpected, was not a complete surprise. General relativity had been known to allow this as one of several possibilities since Einstein wrote down his equations in 1916. In this final chapter, we will explore how we have come to understand the changes in the expansion rate of the Universe over time, and what that might say about its future.

This page titled 17.0: Dark Energy and the Fate of the Universe Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.1: Evidence for Dark Energy

What Do You Think: Does the Hubble Constant Stay Constant?

17.1.1: The Expansion Rate of the Universe Over Time

What does the expansion of the Universe look like as time passes? That is, if we could watch the Universe expand over its entire history, what would it look like? First we will look at the expansion rate today. We can do that using a Hubble diagram (Figure 17.1) for nearby galaxies. A Hubble diagram shows the velocities of galaxies plotted vs. their distances, as is discussed in Chapter 13.

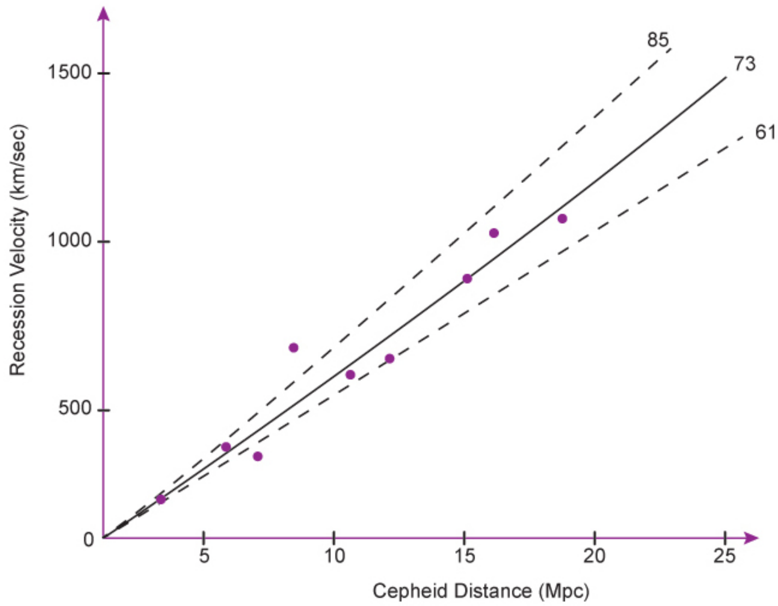


Figure 17.1: A Hubble diagram by the HST Key Project team. The velocities and distances for a sample of nearby galaxies are plotted. The slope of this line is the Hubble constant, which is measured to be 73 ± 7 km/s/Mpc. This means that for each megaparsec farther away in distance , the velocity of a galaxy is faster by 73 km/s. Credit: NASA/SSU/Aurore Simonnet based on data from W. L. Freedman et al. 2001, Astrophysical Journal, 553, 47

In this diagram, the expansion rate of the Universe is constant, as evidenced by the straight-line fit to the data. However, these galaxies are relatively nearby (so relatively close to us in lookback time as well). If we look at galaxies that are farther away in distance, we can measure the value of the Hubble parameter as we go farther back in time. Recall that the finite speed of light means greater distance means larger lookback time.

In the following activities, we will explore what a Hubble diagram looks like if the expansion rate is faster or slower or if it is increasing or decreasing. One thing to keep in mind is that the expansion history can be a little tricky to describe because of the early incident of Inflation. That event basically erased any earlier evidence of what the expansion might have been before. To avoid the ambiguity associated with the pre-inflation Universe, we will limit our inquiry to only the time after inflation occurred.

Hubble Diagrams and Expansion

A. The slope of graphs

Rank the following graphs by their slopes.

Play Activity

B. What if the expansion rate is faster or slower?

The slope of each graph in Figure A.17.1 is the Hubble parameter. On the left is a Hubble diagram with a slope of 70 km/s/Mpc. On the same scale are two other Hubble diagrams.

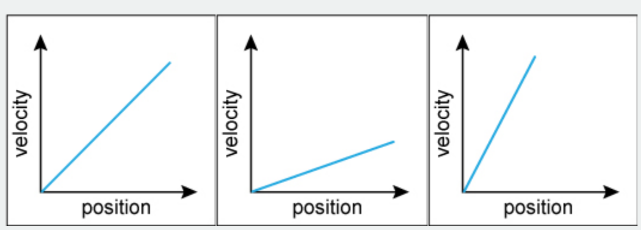


Figure A.17.1:Hubble diagrams with different values for the Hubble constant. On the left, the slope of the line is 70 km/s/Mpc. The diagrams in the center and on the right are plotted using the same scale. Credit: NASA/SSU/Aurore Simonnet

1.

2.

3.

4.

5.

6.

7.

C. What if the Hubble parameter is not constant?

Figure A.17.2 shows four possible Hubble diagrams.

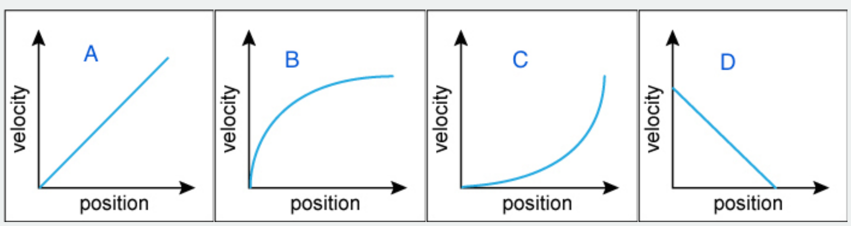


Figure A.17.2: Hubble diagrams, all plotted using the same scale. Credit: NASA/SSU/Aurore Simonnet

1.

2.

3.

4.

5.

6.

7.

https://phys.libretexts.org/@go/page/31452

8.

9.

10.

11.

12.

In the opening video for this chapter, we made an analogy between the expansion of the Universe and a ball in motion because in both cases we are dealing with the interplay of gravity and the energy of motion. In the next several activities, we will explore what these situations look like graphically.

The Motion of Balls

1. Imagine rolling a ball at constant speed away from you. What will its motion look like? (We assume that there is no friction in this example.) A graph of its position vs. time will resemble Figure A.17.3.

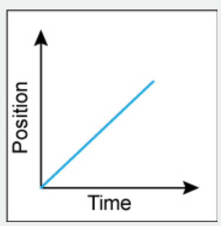


Figure A.17.3: A motion diagram for a ball rolling away from a person. On a position vs. time graph, the velocity is the slope of the line. Credit: NASA/SSU/Aurore Simonnet

1.

2.

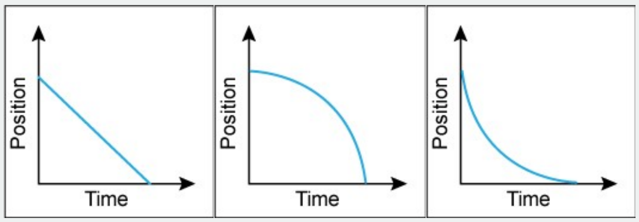


Figure A.17.4: A ball is dropped. On a position vs. time graph, the slope of the line is the velocity. Credit: NASA/SSU/Aurore Simonnet

3.

4.

5.

https://phys.libretexts.org/@go/page/31452

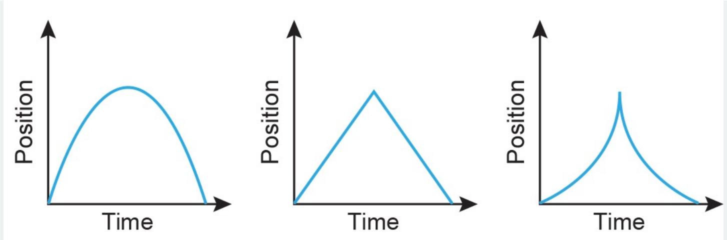


Figure A.17.5 A ball is thrown straight up. On a position vs. time graph, the slope of the line is the velocity. Credit: NASA/SSU/Aurore Simonnet

6.

7.

Imagine we throw a ball upward from Earth's surface as in the last activity. At first the ball will be moving quickly upward. Under the influence of gravity the speed of the ball will slow over time, generally to the point of stopping. The ball will then reverse its motion and fall back to the surface with increasing speed.

This scenario is not the only one that we might see. It is possible to give the ball so much speed at the outset that it never slows enough to fall back to Earth. This speed is called escape velocity , and for an object launched from Earth's surface it is about 11 km/s. But even if we throw the ball upward with escape velocity, or greater for that matter, it slows over time. It just does not slow fast enough for gravity to eventually halt its motion.

Given our intuitive understanding of gravity, we might expect that objects under an attractive gravitational influence will have their motion slowed if they initially are moving away from one another. Therefore, we might also conclude that the expansion of the Universe should have been slowing over time since all galaxies attract all other galaxies. But is this the correct scenario? In the following activity, we will examine how the scale factor of the Universe might change over time.

The Scale Factor of the Universe

The scale factor of the Universe, S, describes how much the Universe has expanded (or contracted) over time, and a graph of S vs. time tells us how the rate of expansion has changed, if at all. If the expansion rate remains constant over time, a graph of S vs. t will look like the one in Figure A.17.6.

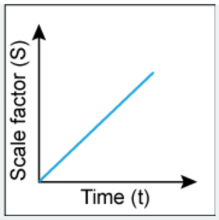


Figure A.17.6: The scale factor (S) vs. time (t). Credit: NASA/SSU/Aurore Simonnet

1.

You may have noticed that this diagram of scale factor vs. time resembles the Hubble diagrams in Figure 17.1 and Figure A.17.1, which show the position of galaxies on the x-axis and the velocity of galaxies on the y-axis. However, in the Hubble diagram, distant galaxies (and hence earlier times) are on the right, whereas here, earlier times are on the left. Both a scale factor vs. time and a Hubble diagram can be used to describe the expansion of the Universe.

In Figure A.17.7, we explore other possibilities for what the scale factor might do as time passes.

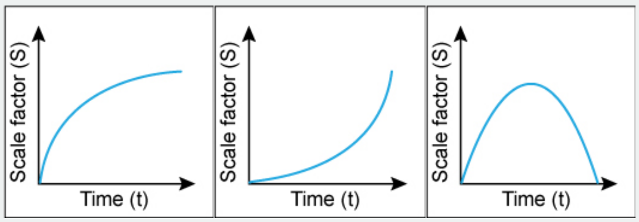


Figure A.17.7: Possibilities for how the scale factor (S) might behave over time (t). Credit: NASA/SSU/Aurore Simonnet

2. For which of the graphs does the scale factor (only) get bigger as time increases? (Check all that apply.)

Lef

3.

https://phys.libretexts.org/@go/page/31452

4.

5.

6.

7.

8.

> Constant, Slowing Down, or Speeding Up

We have seen two different ways to graphically represent the expansion of the Universe: a Hubble diagram (velocity vs. distance) and a diagram of scale factor vs. time. Place each of the following graphs in the correct bin, depending on whether it describes an expansion that is constant, slowing down, or speeding up.

Hint: Recall that when we are looking at far away galaxies, we are looking at earlier times in the history of the Universe.

Play Activity

17.1.2: The Expansion of the Universe is Accelerating

If we wish to measure the expansion of the Universe at large distances (early times), we must have a way to determine both the distance to objects and their recession speeds independently. The redshift is an easy way to determine a galaxy's recession from us, but determining distances on large scales is more difficult. It requires a calibrated method that allows distance determinations independent of redshift—either a standard candle or a standard ruler is needed.

Type I supernovae turn out to be excellent standard candles; each can be calibrated and its luminosity determined to high precision. This particular kind of supernova occurs when a white dwarf star (of a particular composition) accretes enough material from a binary companion to become gravitationally unstable. The instability sets in when the white dwarf reaches 1.4 solar masses. Because the resulting explosions come from objects that are similar, with only slight variations, they all have nearly the same luminosity. This means that, once a measurement of their apparent brightness is made, their distance can be determined by comparing it to their intrinsic brightness.

Another property of Type I supernovae that suits them to studying the early expansion of the universe is their extreme brightness. They are so bright, in fact, that they can be seen to very great distances, and thus to very early times in the history of the Universe. This means that astronomers can use them to create a Hubble diagram that spans a large fraction of the past. So white dwarf supernovae are an ideal tool to show us whether the expansion rate of the Universe has been constant over time.

The quantities we actually observe for the supernovae are flux (or magnitude) and redshift. From these observations we must use some framework or model to interpret the meaning of our measurements. One model is the inverse-square behavior of light. Using this law, we can determine the distance (d) to a supernova from its measured flux (F) and the known luminosity (L).

$$F = \frac{L}{4\pi d^2}$$

Astronomers often work with the magnitude system instead of direct fluxes. Magnitudes are related to the logarithm of the flux. To learn more about magnitudes, see **Going Further 3.6: The Magnitude System.** The following activity explores how the brightness of a supernova (in magnitudes) depends on its distance.

> Supernova Distances

Traditionally, the brightness of objects in optical light is expressed in terms of apparent magnitude (m) instead of flux (F). Correspondingly, the absolute magnitude (M) is used instead of luminosity (L). Like flux, the apparent magnitude relates to how bright an object appears to us. The absolute magnitude is a measure of its true brightness in absolute terms, like the luminosity. Specifically, M is the apparent magnitude an object would have if it were placed at a distance of 10 parsecs. As the distance increases the apparent brightness of an object will decrease, and so *m* will increase. (One counterintuitive aspect of the magnitude system is that fainter objects have larger values when measured in magnitudes.)

In terms of magnitudes, the inverse-square law becomes:

$$d = 10(m - M + 5)/5$$

From this expression, we can see that the distance depends on the quantity *(m-M)*, the difference between the apparent magnitude and the absolute magnitude. This relationship can be expressed graphically:

USE GRAPH

Using the graph, answer the following questions.

1.

2.

3.

Another model we use is general relativity. In particular, we assume that Universe is homogeneous and isotropic. That assumption allows us to connect distance and redshift in a way that depends on cosmic parameters like the Hubble parameter and the amount and type of material in the Universe. The following activity will demonstrate how the content of the Universe affects the apparent brightness of objects as their redshift varies.

Brightness, Redshift, and Cosmological Parameters

In this activity, you will vary the amount of matter in the Universe (matter fraction), along with the Hubble constant. In addition, you can change the amount of dark energy, a strange effect that causes the Universe to expand faster over time. We will look in more detail at what dark energy might be later in the chapter. You will notice that each of these quantities has an effect on the apparent brightness of a given object at a certain redshift (z). Again, the brightness is shown in terms of the difference in apparent magnitude (m) and absolute magnitude (M); larger (m–M) values mean fainter objects.

To create a graph, set the parameters to the values you want and choose "Graph Data."

Play Activity

A. Effect of the Hubble constant

Set the parameters to the following and plot a graph:

- Hubble constant = 70
- Mass fraction = 1.0
- Redshift = 3
- Dark energy fraction = 0.0

1.

2.

3.

4.

Now set the Hubble constant to 140 while keeping the other parameters the same. Make a graph.

5.

B. Effect of the matter fraction

Clear the graph, set the parameters to the following, and plot a graph:

- Hubble constant = 70
- Mass fraction = 1.0
- Redshift = 3
- Dark energy fraction = 0.0

While keeping the other parameters the same, change the matter fraction to 0.5 and plot the graph. Also plot a matter fraction of 0.25.

1.

2.

C. Effect of dark energy

Clear the graph, set the parameters to the following, and plot a graph:

- Hubble constant = 70
- Mass fraction = 1.0
- Redshift = 3
- Dark energy fraction = 0.0

While keeping the other parameters the same, change the mass fraction to 0.5 and the dark energy fraction to 0.5 and plot the graph. Also make a graph with the mass fraction = 0.25 and the dark energy fraction = 0.75.

1.

2. I

In the previous activity, you saw how changing various cosmological parameters affected the expansion of the Universe. In particular, you should have seen that adding dark energy term to the energy content increased the magnitude of objects (i.e., decreased their apparent brightness) relative to a model that had no dark energy. This is because dark energy causes the expansion rate to speed up, which in turn changes the relationship between distance and velocity. The differences are quite small at low redshift but start to become apparent for objects at redshifts approaching or exceeding z ~ 1. Objects at such a high redshift are immensely far away. That means we must observe standard candles that have an extraordinarily large brightness, like supernovae, in order to see the effects of dark energy.

In the 1990s two groups of astronomers, the Supernova Cosmology Project and the High-Z Supernova Search, began to exploit Type I supernovae as standard candles. They worked independently and by the end of the decade both had enough supernovae in their sample to draw some conclusions about the expansion of the Universe. In an unexpected result, both groups found that the Universe is expanding faster over time, or in other words, that some sort of dark energy is present. Figure 17.2 combines the published results of both the teams. You can see that the supernovae, despite having quite a bit of scatter, favor the fainter (top) curve, the one including a dark energy fraction of 0.7 and matter fraction of 0.3. These results, from 1998 and 1999, were the first indication that the Universe contains dark energy. In fact, they suggest that the dark energy is the dominant component of the energy in the Universe, though its effects are currently only apparent on very large size scales.

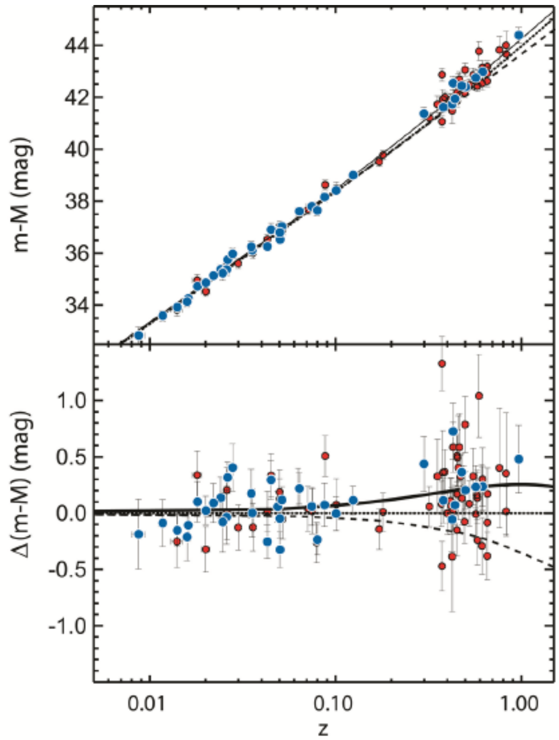


Figure 17.2: Data from Type I supernovae imply that the expansion rate of the Universe is accelerating. The blue dots are data from the High-Z Supernova Search Team and the red dots are data from the Supernova Cosmology Project. The lines are from different cosmological models. The solid dark line is a model in which there is dark energy and an accelerating expansion rate. The thin solid line is a constant expansion. The dashed line is a model with a slowing expansion rate. The top panel is a plot of magnitude (a measure of brightness and therefore of distance if the supernovae are good standard candles) vs. redshift (z). This is another way of making a Hubble diagram. The bottom panel shows the deviations from a constant expansion rate more clearly. Credit: Figure from the High-Z Supernova Search team, data from the High-Z Supernova Search Team and the Supernova Cosmology Project.

For their initial discovery of dark energy, Saul Perlmutter of the Supernova Cosmology Project and Adam Riess and Brian Schmidt of the High-Z Supernova Search Team were jointly awarded the 2011 Nobel Prize in Physics. We will explore additional evidence for dark energy later in the chapter.

This page titled 17.1: Evidence for Dark Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.2: Candidates for Dark Energy

WHAT DO YOU THINK: WHAT IS DARK ENERGY?

Some students are discussing a recent news report about dark energy.

- **Wendy:** Did you hear the news on the radio this morning? It said that the Universe is made mostly of dark energy.
- **Xiang:** I thought the Universe is mostly dark matter .
- **Yan:** Aren't dark matter and dark energy the same thing?
- **Zack:** No, dark energy is the energy that comes from dark matter.
- **Addy:** Actually, I think we call it dark energy because we're in the dark about what it is.

Do you agree with any or all of these students and, if so, whom?

Wendy

Xiang

Yan

Zack

Addy

None

Explain.

Current measurements indicate that dark energy is about 70% of the matter-energy contained in the Universe. The remaining 30% is matter, including both baryonic (i.e., normal matter) and cold dark matter. The discovery of dark energy has given theorists much to think about, and they are busy creating models for what it might be. There are several forms the dark energy could take. Even now, more than two decades after it was first detected, scientists are still not certain which of the possible forms is the right one. As an indication of their ignorance in this matter, they have chosen a somewhat confusing placeholder name: dark energy. The name suggests to many people that there is a connection between dark energy and dark matter. Despite the superficial similarity in their names, this is not the case. The only thing dark energy and dark matter have in common is that we can detect the presence of each, but we are not certain of the form that either takes. Otherwise, there is no relationship between dark energy and dark matter. We will now discuss some of the possibilities for what dark energy might be.

17.2.1: COSMOLOGICAL CONSTANT

The idea that some aspect of the Universe acts repulsively, rather than attractively like gravity, is not new. It goes back almost 100 years to the beginning of general relativity. Albert Einstein himself proposed a cosmological constant to patch up what he believed was a flaw in his new theory.

At the time that Einstein proposed the general theory of relativity, in 1915, little was known about the Universe. In fact, astronomers of the time did not know whether what we now know are galaxies were inside or outside of the Milky Way, nor did they understand the processes that powered the stars. In 1915, the necessary physics was not yet understood, and the observational evidence of the immense size of the Universe was still more than a decade away.

In the absence of evidence to the contrary, most scientists believed that the Universe was infinite in space and time. There was no physical basis for this belief, it was just a bias that scientists brought to their understanding of the cosmos. Einstein was not immune to the biases of his time, so he shared the idea that the Universe must be unchanging and ageless.

This bias created a problem for Einstein. A static Universe is not compatible with general relativity. According to its equations, a Universe that started out static would collapse under the effect of the gravitational attraction of all the objects within it. Clearly, the Universe was not collapsing if (according to belief) it was steady and unchanging. So, Einstein assumed that his equations must be incomplete with the true nature of the world. To fix them, he added a term that counteracted the gravitational attraction: in essence he created a term that allowed space to hold up the galaxies so they would not all come crashing into one another.

Einstein's original description of gravity was quite simple and stated that the curvature of spacetime, $\mathbf{G}$, was related to the energy and mass present, represented by a term $\mathbf{T}$.

$$\mathbf{G}=\frac{8\pi G}{c^4}\mathbf{T}$$

Do not confuse Newton's gravitational constant, the italic G on the right, with the spacetime curvature, the bold $\mathbf{G}$ on the left. Einstein's equation and the Friedmann equation, which is derived from it, describe the relationship between the expansion, mass-energy, and curvature of the Universe. To cancel the mass-energy and make $\mathbf{G}$ zero (so that the equations no longer predicted that the Universe would collapse), Einstein added a term to the right-hand side of his equation.

$$\mathbf{G}=\frac{8\pi G}{c^4}\mathbf{T}-\mathbf{\Lambda}$$

The extra term $\mathbf{\Lambda}$ (the Greek letter "lambda") is called the **cosmological constant**. It serves merely to cancel the mass-energy term represented by $\mathbf{T}$. So, the cosmological constant works in the opposite direction to the gravity (curvature) caused by the mass-energy term. The cosmological constant has the effect of making space expand rather than contract. We cannot go into the mathematics of this equation—the terms $\mathbf{G}$, $\mathbf{T}$, and $\mathbf{\Lambda}$ are tensors and require advanced mathematical methods to manipulate.

If we could explore the math, we would soon realize that $\mathbf{\Lambda}$ defined this way has a big problem. Although in principle it can cancel the gravity of the mass-energy in the Universe, it must be finely tuned to do so. Any small fluctuation to larger or smaller values will cause the Universe to collapse (if $\mathbf{\Lambda}$ is too small) or to expand (if $\mathbf{\Lambda}$ is too big). The problem is something like trying to balance a sharp pencil on its point: possible in principle, but try doing it and see what happens. So, a Universe as described by this modified Einstein equation is unstable, and it ends up collapsing (or expanding) after all. Collapse is what Einstein was attempting to avoid in the first place.

This problem soon fixed itself. Edwin Hubble announced in 1929 that the Universe is expanding. The problem of a static Universe became moot. Einstein quickly realized his mistake, and he is said to have referred to his introduction of the cosmological constant as his "biggest blunder." He understood that if he had been more trusting of his mathematics and less reliant on commonly held, but unfounded, notions of the nature of the Universe, he could have *predicted* the cosmic expansion more than a decade before any observational evidence for it was in hand. But general relativity is a theoretical tour d'force, and even Einstein was not willing (or perhaps not able) to make such a leap of faith where his new theory was concerned. One can hardly fault him for that.

The cosmological constant was mostly forgotten after the discovery of the cosmic expansion. The Hubble expansion did not disprove the existence of $\mathbf{\Lambda}$, it just made it unnecessary, at least for nearly a century. It has been invoked once again now that observations have shown the expansion to be speeding up. Perhaps Einstein's intuition was correct after all, even if his motivation was not sound. A cosmological constant could certainly cause a cosmic speedup like the one observed.

17.2.2: VACUUM ENERGY

In quantum mechanics, space can never be completely empty. That is because there is always a background bubbling and churning with virtual particles, popping in and out of existence. Any given virtual particle (or particle pair) will exist only as long as allowed by the uncertainty principle. Still, any small volume of space has many pairs coming and going all the time. As a result, if we measure the energy contained in that volume, we will not measure the classical value of zero. Instead, we will measure a small, but finite, energy content, even in "empty" space. This vacuum energy could contribute to the mass-energy of space, and it provides a possible physical reason for the cosmological constant.

In general relativity, any form of energy (not just matter) affects the curvature and expansion of spacetime. Imagine that we have a small container surrounding some volume in space, one end of which is a piston that can be used to expand or contract that space. We cannot actually perform this experiment because when we say "expand" or "contract" here, we mean that we are able to create more space, or shrink it, with our piston. We do not merely change the interior volume of the container in existing space—an important distinction.

If we use the piston handle, as in Figure 17.3, to expand our space, we will feel resistance. The resistance is directly related to the notion that the energy of the vacuum is non-zero; we must provide the energy needed to create any additional empty space. We feel that effort as a force that resists our expansion efforts. In other words, we feel that the vacuum has a negative pressure.

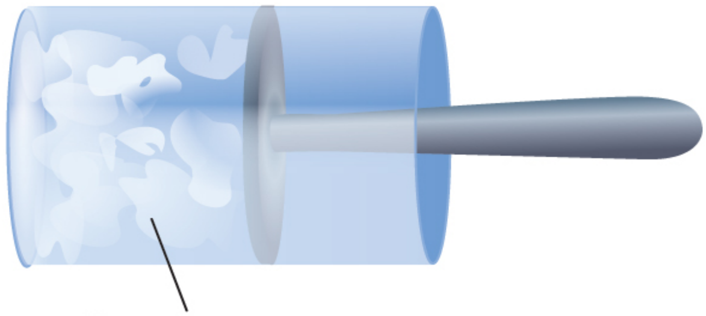


Figure 17.3 If we imagine a piston that can be used to expand or contract space, then we know that in creating additional vacuum, we create additional vacuum energy. Such a piston would resist our attempts to expand it because it would require that we furnish the additional energy in the vacuum we create. Credit: NASA/SSU/Aurore Simonnet

Pressure is different in one fundamental aspect from gravity: it can be attractive or repulsive. We generally think of pressure as being capable of imposing a push, but not a pull. That is not true for all materials. In our familiar world, gases do not produce negative pressures (tensions), but solids (like springs) can.

The existence of the negative vacuum pressure has extremely interesting gravitational ramifications. If we put the pressure into the Einstein equations, it will affect the curvature of space. Since the pressure of the vacuum is negative, not positive as in a gas, the gravity of the vacuum pressure has the opposite effect on the curvature that a gas pressure would have: it is repulsive. That means empty space tends to repel other empty space and the matter within it. This is the type of action we described earlier for the dark energy. Only general relativity provides a means to understand this kind of action. It is a completely non-Newtonian effect, and there are no Newtonian approximations or analogies we can use to help us make sense of it.

17.2.3: QUINTESSENCE

The presence of a cosmological constant or a vacuum energy is not the only possible explanation for an accelerating expansion. An accelerating expansion, known as inflation, is thought to have occurred early in the first fraction of a second of the Universe's existence. Though the inflationary period lasted for only an instant, something like 10^{-32} seconds, it expanded the scale factor of the Universe by at least 20 orders of magnitude. The current acceleration is clearly much less dramatic. It has been going on for billions of years and has had only a barely discernible effect. But could the acceleration we see now be in some way similar to the earlier inflationary event? Some cosmologists think so.

Rather than posit the existence of a constant vacuum energy that exists everywhere in space for the entire history of the Universe, an alternative is to suggest that the Universe contains an energy field, similar to the one that caused inflation. However, rather than decaying in 10^{-32} seconds, the current acceleration is caused by a field with a much slower rate of decay, one that is taking billions of years to evolve. Just as with inflation, the field is currently decaying from a quasi-stable state of false vacuum to its ground state. As it does so it mimics a cosmological constant, except it is changing in strength. To distinguish this decaying field model for the acceleration from the cosmological constant model, it is given its own name: **quintessence**.

Quintessence models are attractive for several reasons. First, the field is changing over time and we might be able to discern this change if we look at higher redshift parts of the Universe. The cosmological constant does not change in time, so its effects should not depend on redshift at all, but quintessence would. Quintessence is also more physically motivated than a cosmological constant, which is literally just a constant number added to the equations of relativity. An aspect of quintessence that makes it more attractive than vacuum energy has to do with a disturbing inconsistency. Quantum theory can be used to compute how big we expect the vacuum energy to be, but when those computations are done we find to our great surprise and embarrassment that the calculations disagree with the measured value by 120 orders of magnitude, or in other words, by a factor of 10^{120}. Quintessence models alleviate some of this problem, but we are going to need to learn a lot more about the Universe, on both its largest and smallest scales, before we solve this puzzle.

17.2.4: OTHER IDEAS

The leading ideas for understanding the apparent acceleration of cosmic expansion are the cosmological constant (perhaps in the form of vacuum energy) and quintessence. However, some other explanations of an even more exotic nature have also been put forth.

One idea is the existence of energy fields, distinct from the quintessence idea, that couple to other fields. These so-called chameleon particles might interact with the recently discovered Higgs field (or other fields) in a manner that would create an effect similar to the cosmological constant.

Another idea is that on large scales, general relativity is not a correct description of gravity and the apparent acceleration is an artifact of using an incorrect theory of gravity that produces the wrong redshift—distance relation for supernovae. However, a self-consistent theory of modified gravity has proven elusive so far.

Some cosmologists do not accept the notion of accelerating expansion at all. Instead, they suggest, the entire idea is due to an illusion caused by the fact that we reside in a region of the Universe with slightly lower than typical density, a condition that could mimic the effects of acceleration on the supernova data that provided the primary evidence of the acceleration in the first place.

Even interactions of our Universe with speculative higher dimensions in which it might be embedded have been suggested as the origin of the acceleration.

Which of these ideas, if any, is the correct one? We still do not know. For this reason, most scientists refer to whatever is causing the apparent acceleration of the expansion as dark energy. Only further observational and theoretical studies, and probably some new physics, are likely to lead to an eventual understanding of the cause and nature of dark energy.

GOING FURTHER 17.1: PARAMETERIZING DARK ENERGY

For the various models of dark energy, there is a mathematical relationship between the pressure (P) and density (ρ) called the equation of state, which has the form below.

$$\rho = Pwc^2$$

As usual, c is the speed of light. The parameter w is a different number for each type of dark energy model. From the equations of general relativity, we can find a relationship between the overall density (ρ), the scale factor (S), and w:

$$\rho \sim S^{-3(1+w)}$$

In terms of redshift (z), this can be written:

$$\rho \sim (1+z)^{3(1+w)}$$

For the case of constant dark energy (such as a vacuum energy) we would have $w=-1$, but other values of w are possible for different hypothetical forms of dark energy. For example, w could be larger or smaller than -1. Those cases lead to radically different evolutionary paths for the Universe. It is not even known if w remains constant in time. There is no reason that it should do so, but making the assumption of a constant w leads to simpler models. Lacking any compelling observational reason to adopt a more complex scenario, cosmologists usually assume that w is constant. We will do the same. In the future, if observational evidence warrants, it might be necessary to abandon this assumption. Table B.15.1 shows the equation of state and density as a function of scale factor or redshift for various substances.

TABLE B.17.1 EQUATION OF STATE FOR VARIOUS SUBSTANCES

Substance	w	Density and Pressure	Density and scale factor	Density and redshift
Cosmological constant (vacuum energy)	-1	P ~ -ρ	ρ = constant	ρ = constant
Matter	0	P ~ 0	$\rho \sim 1/S^3$	$\rho \sim (1+z)^3$
Radiation	+1/3	P ~ ρ/3	$\rho \sim 1/S^4$	$\rho \sim (1+z)^4$
Quintessence	-2/3	P ~ -2ρ/3	ρ ~ 1/S	ρ ~ (1+z)

This page titled 17.2: Candidates for Dark Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.3: The Friedmann Equation and the Fate of the Universe

> *Some say the world will end in fire,*
>
> *Some say in ice.*
>
> *From what I've tasted of desire*
>
> *I hold with those who favor fire.*
>
> *But if it had to perish twice,*
>
> *I think I know enough of hate*
>
> *To say that for destruction ice*
>
> *Is also great*
>
> *And would suffice.*

WHAT DO YOU THINK: FATE OF THE UNIVERSE

Some students are contemplating the end of the Universe.

- **Bo:** I think the Universe will keep expanding and cooling, galaxies will keep pulling farther apart from each other, and eventually everything will be dark.
- **Candice:** I don't think it makes sense to expand forever. I think the expansion will only go for so long, then it will begin to contract because of gravity . I think it's going to reach a maximum, then collapse back onto itself.
- **Dante:** It strikes me that the Universe ought to be timeless and that the Universe is just expanding and collapsing perpetually.
- **Emily:** No, I think Bo is right, except I think it's supposed to eventually expand so much that everything gets ripped apart.
- **Feng:** I don't think there's any way to know or predict the fate of the Universe.

Do you agree with any or all of these students and, if so, whom?

Bo

Candice

Dante

Emily

Feng

None

Explain.

17.3.1: THE FRIEDMANN EQUATION

The basis to understanding how the expansion of the Universe can be speeding up lies in the Einstein equations. General relativity can tell us physically what factors relate to the expansion. One version of the Einstein equations is the Friedmann equation, which describes how the expansion rate changes over time, and why, in a homogeneous and isotropic Universe.

$$H^2 - \frac{8\pi G\rho}{3} = \frac{kc^2}{S^2}$$

$$(expansion) - (density) = (curvature)$$

In this equation, H is the Hubble parameter (expansion rate), ρ is the average density of matter and energy in the Universe, S is the scale factor, and k is a number that describes the overall curvature of the Universe. That means the expansion is related to the contents of the Universe as well as the curvature of space.

The density term includes all of the types of matter and energy in the Universe.

$$\rho = \rho_{baryon} + \rho_{cdm} + \rho_{radiation} + \rho_{DE}$$

Here ρ_{baryon} is the density of regular matter (baryons), ρ_{cdm} is the density of cold dark matter, $\rho_{radiation}$ is the density of radiation, and ρ_{DE} is the dark energy density.

The Friedmann equation embodies the essence of the Einstein equation: matter and energy affect how spacetime bends. In general relativity, it is not only matter that produces curvature, but any sort of energy. This means that the gravitational attraction has a contribution not just from regular matter, but also from dark matter and radiation (massless particles like photons). The kinetic energy of particles is generally included in the Einstein equations as a pressure term, because pressure can be thought of as an energy density, or a kinetic energy per unit volume. There is also negative pressure from dark energy.

There is a special value of the density that causes the Universe to have zero curvature. It is called ρ_{crit}, the **critical density**:

$$\rho_{crit} = \frac{3H^2}{8\pi G}$$

If the current density is equal to the critical density, then the left side of the Friedmann equation equals 0, implying that k = 0. We say that such a Universe has no curvature, or that it is flat.

Math Exploration 17.1: Calculating the Critical Density

Worked Example:

1. What is the value of the critical density?

We can calculate ρ_c using a Hubble constant of H_0 = 70 km/s/Mpc. The value of the gravitational constant is G = 6.67 x 10^{-11} m^3/kg s^2.

- Given: H_0 = 70 km/s/Mpc, G = 6.67 x 10^{-11} m^3/kg s^2
- Find: ρ_c
- Concepts:
  - $\rho_c = 3H^2/(8\pi G)$
  - 1 Mpc = 3.086×10^{19} km
- Solve:

$$\rho_c = \frac{3(70\ \mathrm{km/s/Mpc})^2}{8\pi\left(6.67\times10^{-11}\ \mathrm{m^3/kg\ s^2}\right)}$$

Now we need to do some unit conversions:

$$\rho_c = \frac{3\left(70\ \frac{km}{s\cdot Mpc}\times\frac{1\ Mpc}{3.086\times10^{19}\ km}\right)^2}{8\pi\left(6.67\times10^{-11}\ \frac{m^3}{kg\cdot s^2}\right)}$$

Solving and simplifying units, we get:

$$\rho_c = 3.07\times10^{-27} kg/m^3$$

In the Friedmann equation, only the ratio of density to the critical density is important. This ratio is so important that it is given its own name: Ω (Greek letter “omega”):

$$\Omega \equiv \frac{\rho}{\rho_{crit}}$$

In general relativity, all of the sources of matter and energy are included and contribute to the total energy density. Thus we have:

$$\Omega = \Omega_{baryon} + \Omega_{cdm} + \Omega_{radiation} + \Omega_{DE}$$

Here Ω_{baryon} is the baryon content, Ω_{cdm} is the amount of cold dark matter, $\Omega_{radiation}$ is the radiation content, and Ω_{DE} is the contribution from dark energy.

If $\Omega = 1$, that means the density is equal to the critical density, so we have a flat Universe (k = 0).

We will generally discuss the density and geometry of the Universe in terms of Ω from now on rather than in terms of k or ρ_{crit}. One of the major efforts in cosmology over the past several decades has been to determine the value of Ω.

From the Friedmann equation, we can see that the interplay of the expansion of the Universe, density, and curvature lead to several possibilities for the fate of the Universe, depending on which term in the equation dominates:

1. **Critical Universe**: We have already discussed the case of a critical Universe, where the expansion and density terms are equal, $\Omega = 1$, and space overall is not curved. If there is no dark energy, the Universe will continue to expand, but increasingly slowly.
2. **Matter terms dominate**: If the Universe contains enough mass to counteract its expansion and there is no dark energy, it will eventually collapse. This is known as a closed Universe. In this case, $\Omega > 1$.
3. **Expansion term dominates**: If the Universe does not contain enough mass to counteract its expansion and there is no dark energy, it will expand forever. This is known as an open Universe. In this case, $\Omega < 1$.
4. **Dark energy dominates**: None of the three possibilities above lead to the accelerating expansion rate supported by supernova data. Dark energy can cause the expansion to speed up; therefore, we must examine possibilities that include dark energy.

We will consider each of these in detail next. We will explore how the expansion rate changes over time, what the matter-energy density (Ω) is like in each case, the resulting curvature of space, and what conditions in the Universe will be like in the future. First, we will consider the three scenarios with no dark energy, then we will explore scenarios that include dark energy.

17.3.2: THE UNIVERSE WITHOUT DARK ENERGY

The interplay between the terms of the Friedmann equation can be compared to the interplay between gravity and the energy of motion for a ball rising into the air using Newtonian physics.

Case 1: A Critical Universe

First, consider a scenario where a ball is launched at escape velocity from Earth's surface. The kinetic energy from the ball's motion exactly equals its gravitational potential energy. In other words, the sum of the terms for motion and gravity is zero. In this case, the ball will continue to fly away from Earth, but at an ever slower rate. This is analogous to a critical Universe. In a critical Universe the expansion (motion) and density (gravity) terms of the Friedmann equation sum to zero. The motion diagram for a ball at escape speed (position vs. time) and a diagram for the scale factor vs. time in a critical Universe are compared in Figure 17.4.

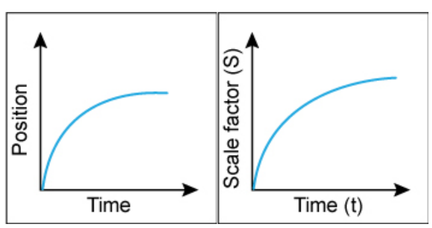


Figure 17.4: Left: a ball is launched from Earth at escape velocity (which is found by setting kinetic and potential energies equal to one another). The ball continues to move away from Earth, but at increasingly slower speeds, as shown by the flattening slope of the graph. Right: The scale factor of the Universe in the case of critical density. The expansion and gravity terms in the Friedmann equation are equal. The Universe continues to expand, but at an increasingly slower rate, as shown by the flattening slope. Credit: NASA/SSU/Aurore Simonnet.

In general relativity, the curvature of space as a whole is related to the expansion and gravity terms. In the case of a critical Universe, the curvature is zero. The expansion and geometry for a critical Universe are depicted in Figure 17.5.

https://phys.libretexts.org/@go/page/31454

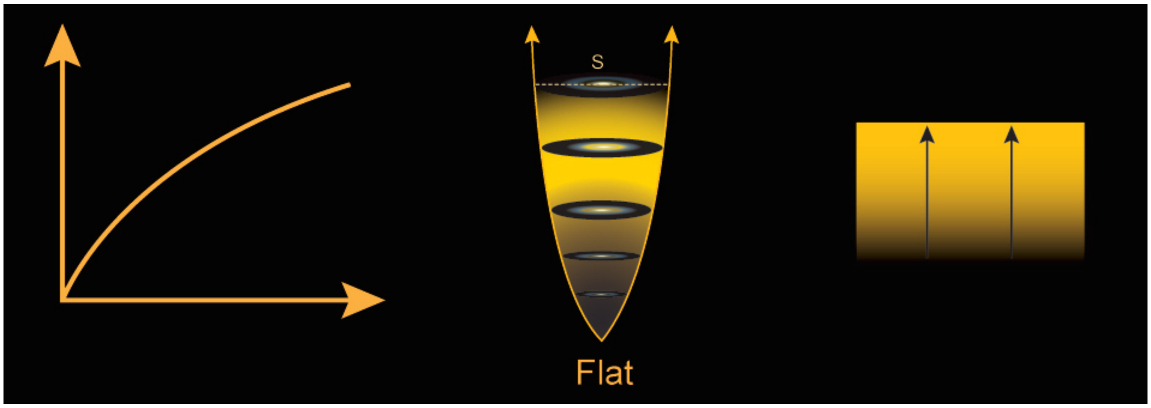
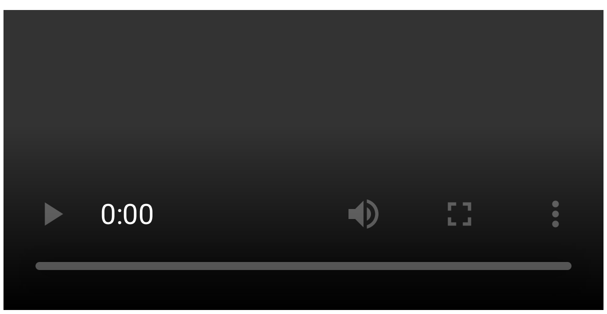


Figure 17.5: The expansion and geometry for a critical Universe. Left: A graph of the scale factor vs. time. Center: An illustration showing that the Universe continues to expand but ever more slowly. Right: The geometry of the Universe is flat, in other words, there is no curvature in any of the three spatial dimensions. Credit: NASA/SSU/Aurore Simonnet

In a critical Universe, space will continue to expand. As it expands, it will cool endlessly, approaching a temperature of 0 K (—273°C). It is at 2.7 K already, and as the temperature drops, the motions of particles and molecules will slow down. Stars will eventually burn out. Galaxies will run out of gas to make any new stars. No energetic photons will be produced to keep things warm. Collisions between particles will be too lethargic to excite electrons, and those collisions will become less and less frequent as the density drops. Black holes will eventually evaporate, and some particles thought currently to be stable might decay to simpler forms. Eventually, physical processes will simply cease, after unimaginably long timescales in excess of 10^{100} years. The fate of the Universe in this case is known as a big chill (Animated Figure 17.6).

@api/deki/files/54210/LinearExpansion.mp4?origin=mt-web < video src=”@api/deki/files/54210/LinearExpansion.mp4” type=”video/webm” controls > < /video >



Animated Figure 17.6: In a big chill scenario, galaxies continue to move away from each other forever, as the Universe expands and cools. Credit: NASA/SSU/Kevin John

Case 2: A Closed Universe

Now consider a case where a ball is launched from Earth’s surface, but with less kinetic energy. In that case, the ball will rise to a certain height and then fall back down to the ground because the gravitational potential energy overwhelms the energy of motion. This is analogous to what happens if the density of matter is greater than that required to produce a flat Universe. In that case, gravity will overwhelm the expansion of the Universe and it will eventually stop expanding and re-collapse. The density is greater than critical ($\Omega > 1$). The left side of the Friedmann equation is negative, implying that the curvature is positive. That is the case for a closed Universe. The motion diagram (position vs. time) for a ball launched at less than escape speed and a diagram for the scale factor vs. time in a closed Universe are compared in Figure 17.7. The expansion and geometry for a critical Universe are depicted in Figure 17.8.

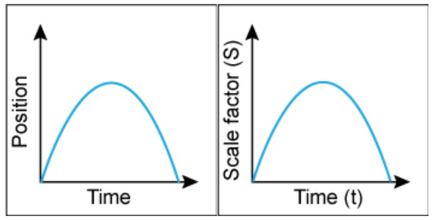


Figure 17.7 Left: A ball is launched from Earth at slower than escape velocity. At first it moves away from Earth, but then it reaches a maximum height and falls back down due to gravity. Right: The scale factor of the Universe in the case of a closed Universe. The gravity term in the Friedmann equation outweighs the expansion term. At first the Universe expands, but then it contracts due to gravity. Credit: NASA/SSU/Aurore Simonnet.

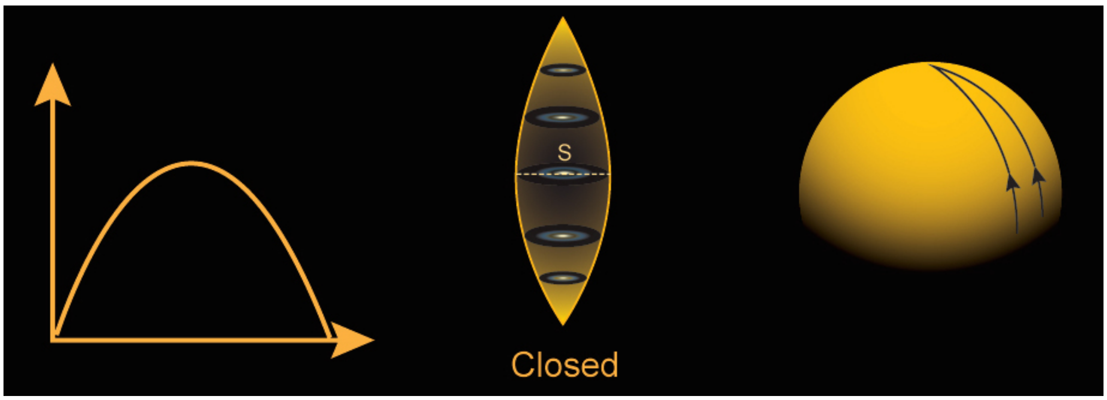
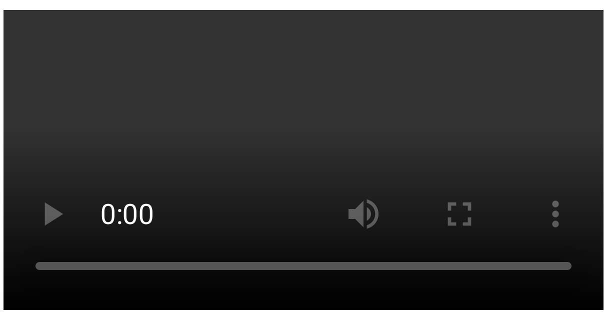


Figure 17.8 The expansion and geometry for a closed Universe. Left: A graph of the scale factor vs. time. Center: An illustration showing that the Universe expands then contracts. Right: The curvature of the Universe as a whole is positive, or in other words, the geometry of the Universe is spherical. Note that this means the global geometry is analogous to that of a sphere. It does not mean that the Universe is a sphere. Credit: NASA/SSU/Aurore Simonnet.

In a closed Universe, space will expand for a while but eventually stop and begin to collapse in upon itself, heating as it does so. Galaxies will move toward each other and the temperature of the Universe will increase. The particles in galaxies will eventually merge together into a state of high temperature and density similar to that found in the beginning of the Universe. The fate of the Universe in this case is known as a big crunch. It is depicted in Animated Figure 17.9.



Animated Figure 17.9 In a big crunch scenario, the gravitational attraction of matter causes galaxies to eventually start moving toward each other and the Universe becomes more dense. Credit: NASA/SSU/Kevin John.

Case 3: An Open Universe

At the opposite extreme, when the density of the Universe is less than critical ($\Omega < 1$), the gravity will never be able to halt the expansion. This is analogous to a ball launched from Earth with a speed greater than escape speed; the kinetic energy of the ball can overcome the gravitational pull from Earth. The motion diagram (position vs. time) for a ball launched with kinetic energy greater than its gravitational potential energy and a diagram for the scale factor vs. time in a closed Universe are compared in Figure 17.10.

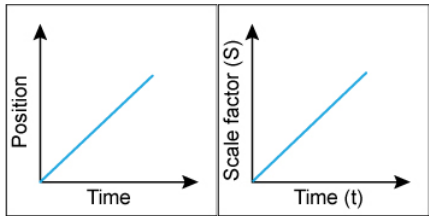


Figure 17.10 Left: A ball is launched from Earth at a speed greater than escape velocity. It flies away from Earth because its kinetic energy is greater than its potential energy. Right: The scale factor of the Universe in the case of an open Universe. The expansion term in the Friedmann equation outweighs the gravity term and the Universe expands forever at a constant rate. Credit: NASA/SSU/Aurore Simonnet.

In the case of an open Universe, the left side of the Friedmann equation is positive, so the curvature must be negative. That happens when the curvature of the Universe is saddle-shaped. The expansion and geometry for an open Universe are depicted in Figure 17.11. In an open Universe, space will continue to expand. In a fate similar to that of the critical case, galaxies will move farther apart from each other, the temperature of the Universe will cool, and stars in galaxies will eventually burn out. Again, we would have a big chill scenario.

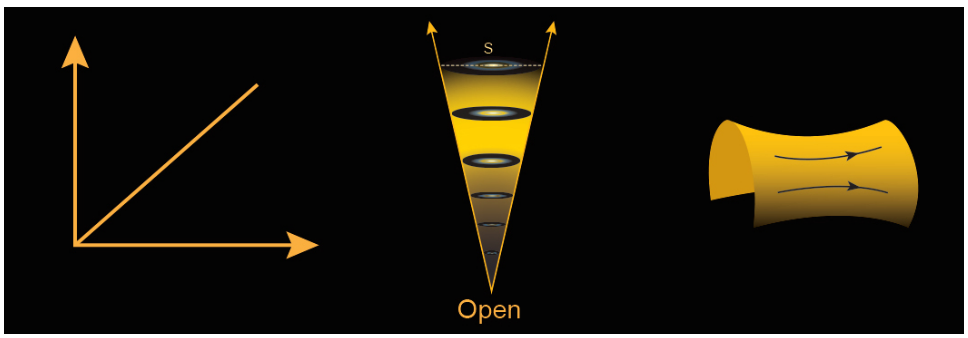


Figure 17.11 The expansion and geometry for an open Universe. Left: A graph of the scale factor vs. time. Center: An illustration showing that the Universe expands forever. Right: The curvature of the Universe as a whole is negative, or in other words, the geometry of the Universe is saddle-shaped (hyperbolic). Credit: NASA/SSU/Aurore Simonnet.

From these three cases—**open**, **closed**, and **critical**—we can see that changing the amount of matter in the Universe changes the way it expands. A lot of matter means that the cosmic expansion cannot overcome the gravitational attraction of the matter. The expansion could eventually slow until it stops, and then the Universe will re-collapse. On the other hand, if there is not enough matter in the Universe, then the expansion continues forever. These two cases are separated by a condition in which there is just enough matter for expansion to balance gravity—the critical case.

If no dark energy is present, once we determine the value of the curvature, we know it forever. It represents the geometry of the Universe as a whole. This is true whether the matter in the Universe is baryons or cold dark matter (or a combination of both). The Friedmann equation alone determines whether the Universe keeps expanding forever or eventually re-collapses; whether the scale factor S grows or shrinks is determined here only by the density of the Universe, which also determines its overall geometry.

EXPANSION OF THE UNIVERSE DEPENDS ON MATTER DENSITY

In the simplest cosmological models, the fate of the Universe depends on its rate of expansion, expressed by the Hubble parameter, and the amount of matter it contains. Qualitatively, the Hubble constant gives us an idea of the kinetic energy associated with the expansion of space. The matter content gives us an idea of the gravitational potential energy. In a Newtonian context, the two taken together give us the total energy. In general relativity, they produce the total curvature of the Universe.

In this activity, you will be allowed to adjust Ω and take note of how the expansion varies as this parameter changes.

Play Activity

1. In which scenario does the Universe eventually collapse?

$\Omega < 1$

$\Omega = 1$

1"> Ω > 1

2. In which scenario does the Universe expand at a constant rate forever?

Ω < 1

Ω = 1

1"> Ω > 1

3. In the Ω = 1 case, the graph of scale factor vs. time is a straight line.

True

False

EXPANSION AND GEOMETRY OF THE UNIVERSE

For the questions below, refer to Figure A.17.8, which shows three possible scenarios for the expansion of the Universe and the geometry that goes with them.

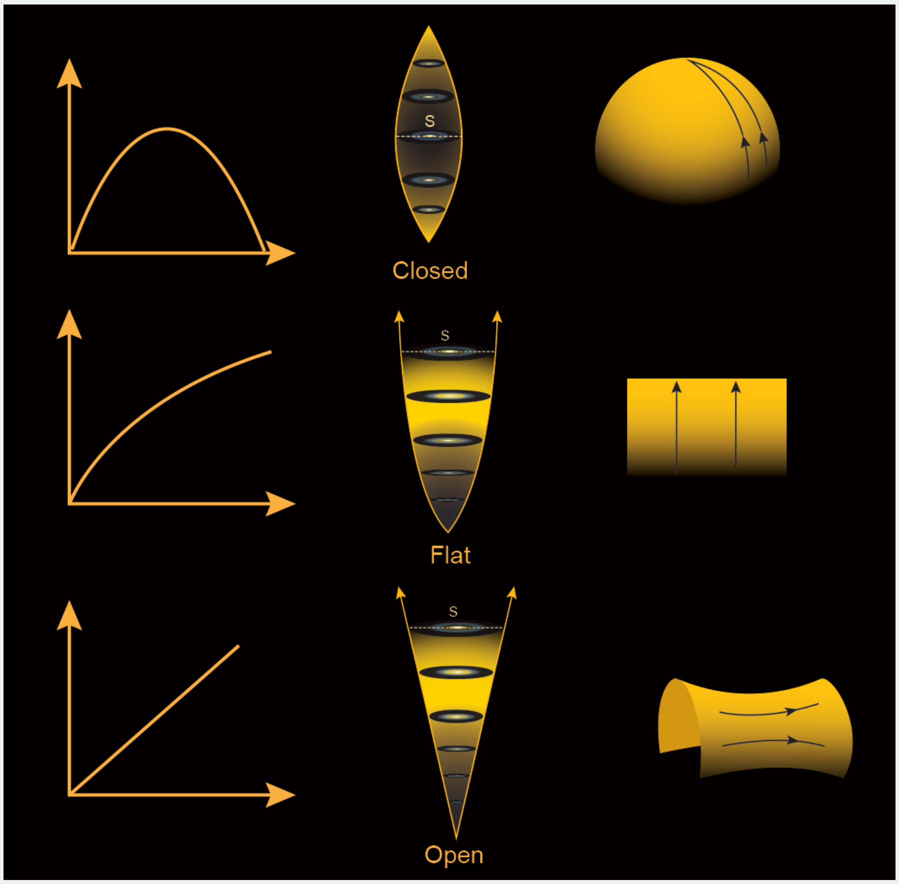


Figure A.17.8 Three different scenarios for the expansion of a matter-only Universe. On the left are graphs of the scale factor vs. time, in the center are illustrations depicting the scale factor of the Universe as time advances, and on the right are models for the curvature of space. Each scenario corresponds to a different curvature. Top row: positive curvature, the Universe expands for a while, then stops expanding and eventually collapses. Center row: zero curvature, the Universe expands forever but increasingly slowly. Bottom row: negative curvature, the Universe expands at a constant expansion rate forever. Credit: NASA/SSU/Aurore Simonnet.

1. Which scenario features negative curvature?

Top row

Center row

Bottom row

2. Which scenario shows a flat Universe?

Top row

Center row

Bottom row

3. Which scenario shows a Universe with positive curvature?

Top row

Center row

Bottom row

4. Which scenario features a Universe that expands forever, but increasingly slowly?

Positive curvature

Zero curvature

Negative curvature

5. In which scenario does the Universe expand at a constant rate forever?

Positive curvature

Zero curvature

Negative curvature

6. In which scenario does the Universe eventually collapse?

Positive curvature

Zero curvature

Negative curvature

In the previous activity, you saw that the Universe will either expand forever or eventually stop expanding, depending on how much mass it contains. In two of the cases, the expansion slows over time. The slowing is the result of the gravitational attraction that all the mass in the Universe has for all the other mass. The slowdown means that the age of the Universe is not simply the reciprocal of the Hubble constant, $1/H_{0,}$ as would be the case for a constant expansion rate. In the next activity, you will examine how the matter density and expansion rate affect the age of the Universe.

EXPANSION SCENARIO AND AGE OF THE UNIVERSE

In this activity, we will explore how changing the amount of matter and the expansion rate affects the age of the Universe for a closed Universe scenario.

The age of the Universe will be the amount of time that passes between the beginning of the Universe and the point at which the scale factor again becomes zero (blue line intersects the x-axis).

Play Activity

1. What happens to the age of the Universe if the Hubble constant (expansion rate) is greater?

The age is greater.

The age is less.

The age does not change.

2. What happens to the age of the Universe if the density (Ω) is greater?

The age is greater.

The age is less.

The age does not change

In the previous activity, you saw how changing the value of the expansion rate and density affected the age for a closed Universe. Figure 17.12 shows how the scale factor changes with time in the closed, open, and critical cases. It also shows what would happen in a case where the expansion is accelerating. We can see that the age of the Universe (time from the beginning until "now") is smallest for a closed Universe and increases as Ω gets smaller. Physically, this is because the age of the Universe is inversely proportional to the expansion rate; if the Universe expands more quickly, it takes less time to reach the state it is in today. If $\Omega > 1$, the expansion rate was greater in the past than it is today; thus the age will be younger compared to a Universe with constant expansion. The reverse is true if the expansion rate is speeding up.

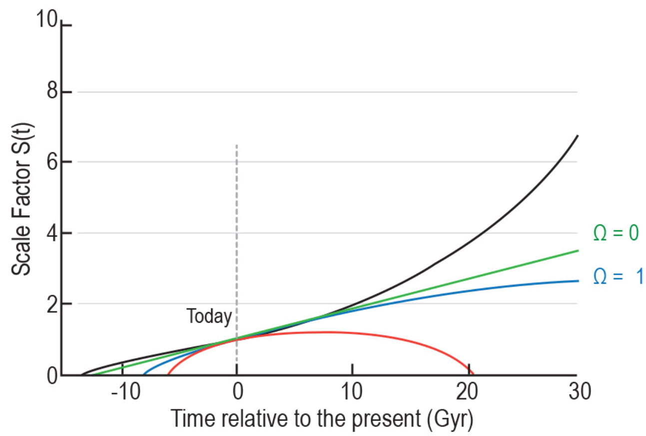


Figure 17.12 Several different possibilities for the age of the Universe. The red curve represents a closed Universe ($\Omega > 1$), which is the youngest possibility (about 7 billion years). As the density decreases, the age of the Universe today increases. For example, for a flat Universe with no dark energy as seen in the blue curve ($\Omega = 1$), the age of the Universe would be about 9 billion years; for a Universe devoid of matter (green curve) the age would be even greater. The greatest age for the Universe occurs in a Universe with an accelerating expansion due to dark energy (black curve). In this case, the age of the Universe matches observations (13.8 billion years old), further evidence that the Universe contains dark energy. Credit: NASA/SSU/Aurore Simonnet.

17.3.3: UNIVERSE WITH DARK ENERGY

In the previous section, we looked at how changing the amount of matter in the Universe relates to its expansion, age, and overall geometry. We saw a variety of outcomes, but none of them predicted an accelerating expansion. For that, we need dark energy. Including dark energy can cause dramatically different outcomes for the age, geometry, and eventual fate of the Universe, as compared to cases without dark energy. The exact outcome depends on the nature of the dark energy. There are three interesting dark energy scenarios: the dark energy could be a cosmological constant, it could grow stronger, or it could grow weaker.

In the case of dark energy being a cosmological constant (Λ), its density is constant—it is a property of space itself. Over time its contribution to the total energy density (Ω) increases because there is more and more space (containing a constant amount of Λ per unit volume). On the other hand, the total amount of matter and radiation is fixed. As the expansion progresses they become more and more dilute. In the next activity, you will see how in this scenario the total density to change over time, and how its constituent parts contribute different amounts at different eras in the evolution of the Universe.

BEHAVIOR OF DENSITY OVER TIME

In this activity you will explore how the proportions of radiation, matter, and dark energy (called Ω_R, Ω_M, and Ω_D, respectively) change over the history of the Universe. They are represented as fractions of the overall matter—energy budget in a cosmic pie chart. Here, we assume that the dark energy takes the form of a cosmological constant.

Use the slider bar to adjust the redshift (z) and answer the following questions.

Play Activity

1. When we look at higher redshift (z), we are (choose all that apply):

Looking farther back in time

Looking at the Universe earlier times in its history

Looking at later times in its history

Redshift is not related to time

2. What happens to the proportion of radiation as redshift gets bigger?

It is bigger

It is smaller

It remains the same

3. At what redshift were the proportions of radiation and matter approximately equal to each other?

z =

4. Was this before, after, or at the same time as the formation of the CMB (z ~ 1100)?

Before

After

Same time

5. Around what redshift do we start to see dark energy making up part of the matter-energy budget?

z =

6. Dark energy is what percentage of the matter-energy budget today?

%

As time advances from the past until now, the redshift gets smaller and the scale factor (*S*) gets larger. In the last activity, you should have seen that the energy contribution from matter and radiation as a percentage of the cosmic matter-energy budget both diminish, while the percentage of dark energy increases. The contribution from matter drops as $1/S^3$ because the particles are being spread over a greater volume as the Universe expands. The radiation contribution shrinks faster than matter because in addition to the number of photons per unit volume dropping (like the number of baryons and dark matter particles), the photon energy is also decreased due to the cosmological expansion; as the light shifts to longer wavelengths, the energy per photon drops. Thus, the total energy contribution from radiation decreases like $1/S^4$. Dark energy (if interpreted as a cosmological constant) remains a constant number as the scale factor increases and thus becomes a greater percentage of the matter-energy budget.

We can write this mathematically if we express the total density in terms of the scale factor.

$$\rho(S) = \rho_{crit}\left[\Omega_{DE} + \frac{\Omega_M}{S^3} + \frac{\Omega_R}{S^4}\right]$$

The scale factor here is S, ρ_{crit} is the current critical density, Ω_{DE} is the amount of dark energy, Ω_M is the amount of matter, and Ω_R is the amount of radiation. We can also demonstrate how this equation works in Figure 17.14, where we show how each of the densities (radiation, matter, and dark energy) drop as time advances.

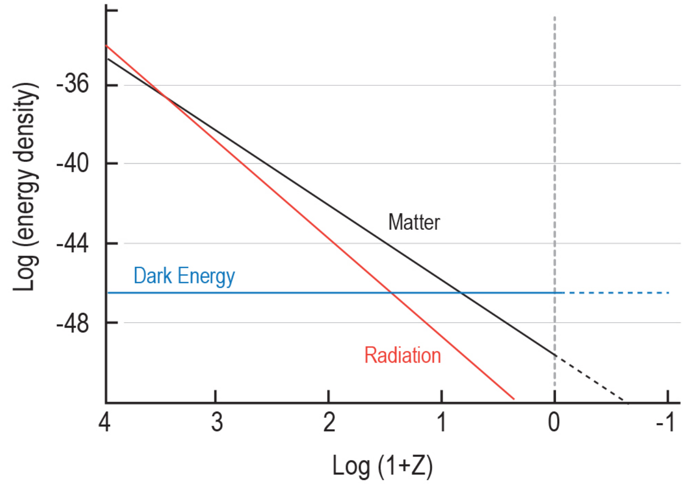


Figure 17.13 Energy density vs. time. The black line shows the decrease in the energy density of matter, and the red line shows the decrease in the energy density of radiation. The blue line shows the constant energy density of dark energy when it is assumed to be in the form of a cosmological constant (although other possibilities have not been completely ruled out). Earlier times in the history of the Universe are to the left and later times are to the right. Today is zero on the x-axis. Credit: NASA/SSU/Aurore Simonnet based on Frieman, Turner, and Huterer 2008. *Annual Reviews in Astronomy and Astrophysics*, 46, 385.

As we can see from Figure 17.13, dark energy (in the form of a cosmological constant) remains unaffected by the expansion. As dark energy becomes more dominant, it makes the expansion speed up, and objects in the Universe begin to accelerate away from each other. In a Universe with a cosmological constant, we will have a big chill scenario, plus an additional interesting effect. Galaxies will move away from each other increasingly quickly such that eventually an observer in any galaxy will see all galaxies outside her own small region disappear over her cosmic horizon. The only objects visible will be the ones in her own galaxy and possibly a few bound neighbors such as those in the Local Group of the Milky Way. This is a slow-motion recapitulation of what is thought to have occurred during the inflationary period in the early Universe.

The expansion of the Universe in a scenario with a cosmological constant is depicted schematically in Figure 17.14. Unlike a matter-only Universe, one with Λ will eventually speed up, no matter the initial curvature.

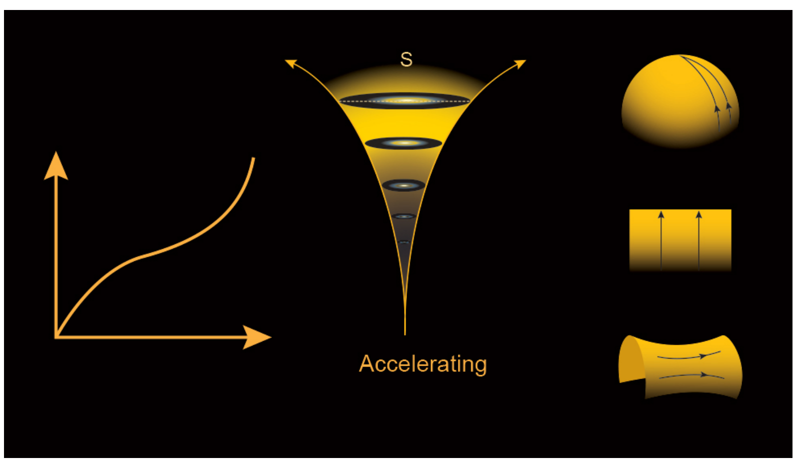
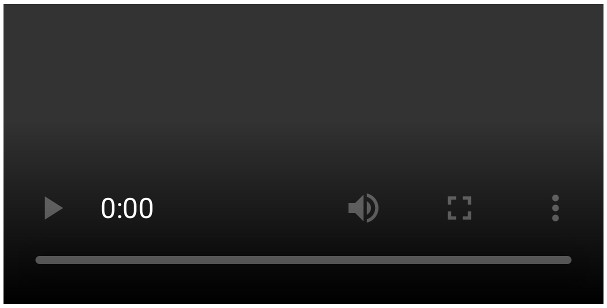


Figure 17.14 In a Universe that contains both matter and dark energy, it is possible for the expansion rate to accelerate with time. In some models, the dark energy can increase with time, in others it remains constant, while in others it may eventually decrease. In order to predict the eventual fate of the Universe, we need to understand more about the properties of dark energy. Credit: NASA/SSU/Aurore Simonnet.

The nature of the dark energy is still not settled. Increasing evidence indicates that it is indeed a cosmological constant, a constant property of space like the vacuum energy. However, other possibilities have not been completely ruled out; dark energy could change over time, becoming either stronger or weaker.

If the strength of the dark energy grows in time, and not only as a proportion of the total energy density like the cosmological constant, the dark energy concentration per unit volume increases. The increase in density will lead to an exponential expansion of space like the one that caused inflation early on. Eventually, the dark energy will become so strong that the expansion will tear galaxies apart. It will even rip molecules, atoms, and protons apart as its strength grows without bound. This scenario is referred to as the big rip (Animated Figure 17.15).



Animated Figure 17.15 In a big rip scenario, galaxies continue to move away from each other forever, as the Universe expands and cools. Dark energy that grows in strength causes the expansion to accelerate so strongly that eventually galaxies and even atoms to be ripped apart. Credit: NASA/SSU/Kevin John.

Finally, the dark matter might have a decreasing density and becomes less important with time. In this case, the expansion of the Universe would continue forever, but at a continually slowing rate, a scenario qualitatively similar to a Universe with no dark energy at all. In fact, that is what such a Universe eventually becomes as its dark energy density approaches zero.

Table 17.1 summarizes the possible scenarios for the fate of the Universe, both with and without dark energy.

Table 17.1 Possible scenarios for the fate of the Universe in the Big Bang model

SCENARIO	DOMINANT TERM	MATTER DENSITY	CURVATURE	OUTCOME
Critical (No Dark Energy)	None	$\Omega_m = 1$	Zero (flat)	Big Chill
Closed (No Dark Energy)	Matter	$\Omega_m > 1$	Positive (like a sphere)	Big Crunch
Open (No Dark Energy)	Expansion	$\Omega_m < 1$	Negative (like a saddle)	Big Chill
Dark Energy Constant	Dark Energy	could be any and still get acceleration at some point	could be any and still get acceleration	Big Chill
Dark Energy Increases	Dark Energy	could be any and still get acceleration at some point	could be any and still get acceleration	Big Rip
Dark Energy Decreases	Dark Energy and then Matter	depends on specific model	could be any and still get expansion	Big Chill

The Big Bang model not only tells us where we have come from, it also predicts where we are going — in the cosmic sense. The model explains observations of the state of the Universe in the past and the present, and it also predicts what will happen to the Universe in the future. The Big Bang model gives us insight into questions like: Will the Universe expand forever? Will it stop expanding and then collapse? The Big Bang model and general relativity tell us each of those scenarios is possible, and they predict specifically what the Universe will be like in each of them. They also tell us how to determine which scenario will occur based on quantities we can measure today.

So, what is the fate of our Universe? Will it expand forever, endlessly cooling to the point of a “big chill,” will it eventually reverse its direction and collapse into a “big crunch,” or will dark energy grow in strength enough to cause a “big rip?” In the next section, we will see how a combination of different observations has helped astronomers answer these questions.

This page titled 17.3: The Friedmann Equation and the Fate of the Universe is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.4: Cosmic Concordance and Cosmological Parameters

WHAT DO YOU THINK: WHAT IS THE MOST IMPORTANT EVIDENCE FOR THE BIG BANG?

Some students are studying for their cosmology class.

- **Gabe:** Here, in my notes, it says the main evidence for the Big Bang theory is Hubble's discovery of the expansion of the Universe.
- **Hans:** What about the cosmic microwave background? I think that's just as important. Maybe even more important.
- **Ilse:** And how much there is of each element. I think that's important for the Big Bang.
- **Jerome:** I think it's the large-scale structure, like how galaxies and the cosmic web formed.
- **Kitty:** I still think astronomers just make this stuff up.

Do you agree with any or all of these students and if so whom?

Gabe

Hans

Ilse

Jerome

Kitty

None

Explain.

17.4.1: THE CONCORDANCE MODEL: PUTTING THE MEASUREMENTS TOGETHER

Our current understanding of the formation and subsequent evolution of the Universe rests not on a single observation, but on a host of independent observations. It has extended beyond the original Big Bang model, which envisioned only that early on the Universe was in a hot and dense state. Our current theory of cosmology includes ideas like inflation, formation of large-scale structures, the cosmic microwave background (CMB) and its fluctuations, the creation of light elements, and other aspects of the Universe. The current overall theory of the formation and evolution of the cosmos is able to explain not only the early hot, dense state, but also how structures were able to form from that early state. Our best current model of the Universe is referred to as the concordance model. In this section we show how the various cosmological parameters in the theory fit together with multiple lines of observational evidence. We will also highlight areas where work still has to be done to complete our understanding.

In the early versions of the Big Bang theory, the quantities of concern were the expansion rate of the Universe (H_0), the existence of relic thermal radiation (the CMB), and the overall geometry of the Universe as described by the overall density parameter (Ω). As observations and our understanding have become more sophisticated, our theories have as well. Now our ideas encompass the baryon fraction (Ω_{baryon}), the baryon-to-photon ratio (η), the density of cold dark matter (Ω_M), the dark energy fraction (Ω_{DE}), and other quantities. Some model parameters are familiar and have already been discussed. Others are fairly technical for a non-expert; indeed, cosmologists add new parameters as the observations become more precise and are able to distinguish small differences between models.

What we would like to illustrate is how several independent observations can work together to place strong limits on the models, eliminating some possibilities and increasing our confidence that the concordance model is a good description of the Universe. We should repeat here that all of our conclusions have as their underpinnings the Friedmann—Robertson—Walker idea of a homogeneous and isotropic Universe described by general relativity. In this framework, small local deviations from the global homogeneity eventually lead to the structures we see in the present Universe.

Supernova data provided the first evidence that the Universe contained dark energy. Shortly afterward, the first precise measurements that determined the geometry of the Universe (and therefore Ω) were the measurements of the tiny temperature variations in the CMB. Repeated measurements by a variety of ground and space-based instruments have confirmed that the Universe is flat, that is, that the sum of all the matter-energy density in the Universe is the critical value, or in the language of cosmological researchers, $\Omega = 1$. (See **Going Further 17.2: The Flat Universe**).

GOING FURTHER 17.2: A FLAT UNIVERSE

Our best measurements of large-scale structure indicate that the total amount of matter - both baryonic and dark - accounts for about 30% of the critical amount. Therefore, the rest must be something else. Dynamically, we are missing about 70% of the energy content of the Universe, though these measurements do not give any hint about what that 70% might be. That is evidence for some energy density that we cannot see dynamically on the scales of galaxies and galaxy clusters. These observations point to, but do not necessarily demonstrate, the existence of dark energy.

The hot and cold spots we observe in the CMB, and the cosmic web of large-scale structures that we see throughout the Universe today, both formed from high- and low-density regions in the early Universe. The delicate looking cosmic web of galaxies, galaxy clusters, and filaments connecting them, form differently under the influence of dark energy than they do if no dark energy is present. Indeed, even different forms of dark energy can affect the eventual structures that exist later in the Universe.

While we have only one Universe to observe and study, we can model it using sophisticated computations. We can calculate many different instances of cosmic evolution to see how structure formation differs under different conditions. We then compare the output of the models to the structures that we observe. In these models, astronomers can vary, for example, the amounts of dark energy, dark matter, baryonic matter, radiation, the size of initial cosmic fluctuations and several other parameters.

One of the observable predictions of the models is the size and spacing of large-scale structures. Another is the spectrum of CMB variations. When cosmologists compare their models to the actual Universe, some of the most robust conclusions they find is that the structures we see are consistent with models that contain dark energy, and that they are inconsistent with models that do not contain dark energy. Furthermore, without dark matter there is not enough time for the structures we see to form. To learn more about why, see **Going Further 17.3: Why CMB and Large-Scale Structure Models Require Dark Matter and Dark Energy**.

Different measurements have different abilities to determine cosmological parameters. For instance, measurements of the CMB by itself are not able to determine the Hubble constant independently of the geometry (Ω) of the Universe. The two can be constrained together, but their individual uncertainties are correlated, as shown in Figure 17.16. As you can see, there are many combinations of values for the Hubble constant and Ω that are allowed by the data. This sort of behavior is called degeneracy.

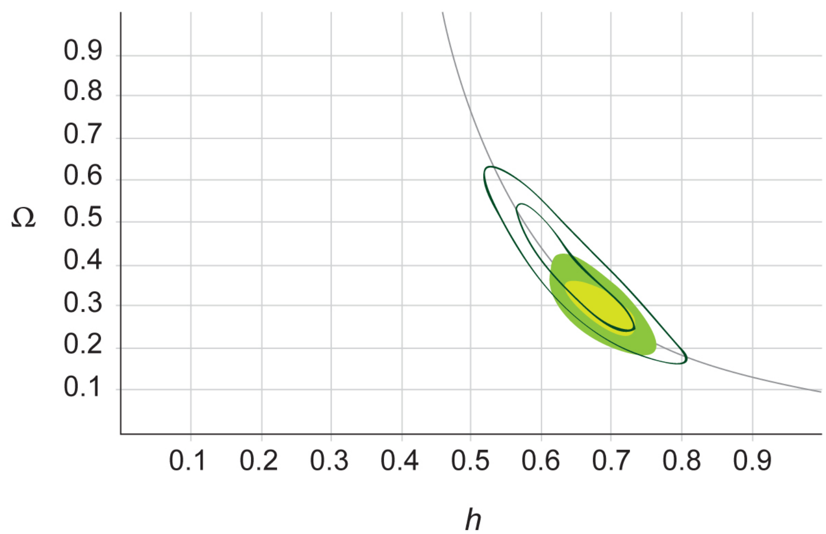


Figure 17.16 This figure shows the degeneracy in CMB data between measurements of the Hubble constant (in units of 100 km/s/Mpc: $h = H_0/100$) and the density of matter (Ω_m). The yellow ellipse denotes the region of 68% confidence and the green region is 95% confidence for WMAP data combined with all other CMB experiments; the inner solid contour is 68% confidence and the outer solid contour is 95% confidence for WMAP alone. (A 95% confidence contour is one in which the true value should lie outside the contour only 5% of the time. In other words, we are 95% confident that the true value lies inside the region bounded by the contour.) Credit: NASA/SSU/Aurore Simonnet based on Planck 2013 Results XVI: Cosmological Parameters, accepted to *Astronomy and Astrophysics.*

In order to break the degeneracy between two variables, we need additional information. Such information can be provided by a different technique that measures the same two parameters in a different way. Often this will break the degeneracy because the uncertainties for the other dataset are complementary to those of the first. Figure 17.17 illustrates the case where different measurements of the same parameters from different techniques can be combined to narrow the overall measurement uncertainties.

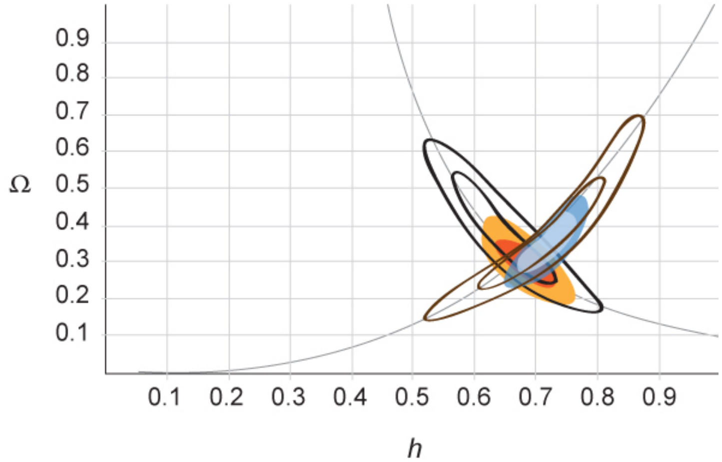


Figure 17.17 A single measurement of a cosmological parameter is often not able to constrain that parameter well. There may be dependencies on another parameter that create ambiguities between the two. By measuring the parameter via several methods these ambiguities, called degeneracies, can be removed in many cases. Credit: NASA/SSU/Aurore Simonnet based on Planck 2013 Results XVI: Cosmological Parameters, accepted to *Astronomy and Astrophysics*.

CONSTRAINING COSMOLOGICAL PARAMETERS

In this activity, you will practice simultaneously determining the fraction of matter in the Universe (Ω_m) and the Hubble constant from two different relations.

A. CMB temperature data

The pattern of temperature fluctuations in the CMB as measured by the Planck satellite can be used to measure the matter content and the Hubble constant. Unfortunately, there is a degeneracy in the two quantities such that:

$$\Omega_m h^3 = 0.0959 \pm 0.0006$$

Here, we are writing the Hubble constant such that H_0 = 100h km/s/Mpc (in other words, h = 0.7 for H_0 = 70 km/s/Mpc, h = 0.6 for H_0 = 60 km/s/Mpc, etc.). As Ω_m goes up, the value of h must go down to keep the product constant.

For now, we will ignore the uncertainty (±0.0006) and use this relation to plot Ω_m vs. h.

1. Fill in the table below using the relationship between Ω_m and h from CMB temperature fluctuations:

Worked example:

- Given: h = 0.5
- Find: Ω_m
- Concept: $\Omega_m h^3 = 0.0959$
- We can rewrite this as: $\Omega_m = 0.0959/h^3$
- Solve: $\Omega_m = 0.0959/(0.5)^3 = 0.767$

h	$\Omega_m = 0.0959h^{-3}$
0.5	0.7672
0.6	
0.7	
0.8	
0.9	
1.0	0.0959

2. Now enter the points in the table into the plotting tool.

- Under “Dataset Controls,” click the “+” to create a new dataset
- Set the labels for the x- and y-axes to h and omega_m, respectively
- Enter the points under “Points Control”

B. Complementary data

We can get additional information about the values of Ω_m and h by using other aspects of the data. This gives a different, and complementary, constraint:

$$\Omega_m h^{-3} = 1.03 \pm 0.13$$

To see what this means, you will plot these data on top of the data from the relationship in Part A. Again, we will ignore the uncertainty (±0.13).

1. Fill in the table below using the second relationship between Ω_m and h:

Worked example:

- Given: h = 0.5
- Find: Ω_m
- Concept: $\Omega_m/h^3 = 1.03$
- We can rewrite this as: $\Omega_m = 1.03\ h^3$
- Solve: $\Omega_m = 1.03(0.5)^3 = 0.129$

h	$\Omega_m = 1.03h^3$
0.5	.12875
0.6	
0.7	
0.8	
0.9	
1.0	1.03

3. Now enter the points in the table into the plotting tool.

- Under “Dataset Controls,” click the “+” to create a second dataset.
- Choose a new color for this dataset
- Set the labels for the x- and y-axes to h and omega_m, respectively
- Enter the points under “Points Control”

C. Combining the data

The two curves that you have plotted are somewhat perpendicular; they cross at a point with particular values of Ω_m and h. The point where the two curves intersect is the one consistent with both datasets.

1. From your graph, what is the value of the Hubble constant?

h =

2. From your graph, what is the value of the matter fraction?

Ω_m =

In this activity, the point where the two curves intersect is the only one consistent with both datasets. In a perfect world, this would be the only consistent point, and finding its positions in the $\Omega_m - h$ plane would determine the two parameters exactly. In the real world, we take the measurement uncertainties into account and get an ellipse that contains acceptable values, i.e., values that are consistent with both datasets.

All of our measurements contain uncertainties; it is the nature of measurement. We ignored them in the last activity, but it is possible to take them into account. To do this, we would plot a region in the $\Omega_m - h$ plane for each relation. Instead of a line for each, we would have a region, called an error ellipse (as seen in Figures 17.16 and 17.17). The overlap of the error ellipses for the two datasets would give us a region of consistency. In this case, it would not be a point, but would have a finite size. The uncertainty associated with the measurement of each parameter is reduced by using both datasets, even if it does not disappear completely.

Astronomers combine measurements from different datasets, such as the CMB, large-scale structure, supernova distances, ages of the oldest stars, the cosmic abundance of the lightest elements, and other measurements. Each of these independent measurements has its associated uncertainties, but in general they are independent of each other. They can be thought of as complementing one another. By combining many complementary measurements together, cosmologists can narrow the overall uncertainties in all of the cosmological parameters. Table 17.2 gives the current best measurements for some of these parameters.

Table 17.2 Cosmological parameters from a simultaneous analysis of CMB temperature data from the Planck satellite, CMB temperature and polarization data from other experiments, and large-scale structure data. (Adapted from Planck 2013 Results XVI: Cosmological Parameters, accepted to *Astronomy and Astrophysics*.)

PARAMETER	SYMBOL	VALUE
Hubble Constant	H_0	67.80 +/- 0.77 km/s/Mpc
Cold Dark Matter density	Ω_{cdm}	0.1187 +/- 0.0017/h^2
Baryon density	Ω_{baryon}	0.02214 +/-0.00024/h^2
Cosmological constant density	Ω_{DE}	0.692+/-0.010

GOING FURTHER 17.3: WHY CMB AND LARGE-SCALE STRUCTURE MODELS REQUIRE DARK MATTER

17.4.2: FURTHER EVIDENCE FOR DARK ENERGY

The concordance model is often called ΛCDM, because multiple measurements point to a Universe that contains not just regular matter, but also cold dark matter and dark energy; the dark energy is represented as the cosmological constant Λ, though dark energy might not actually be constant.

As we have discussed earlier, the first evidence for the presence of dark energy came from observations of Type I supernovae. But because dark energy creates an effect opposite to that of gravity, its presence should also be detectable in the way that structures form in the Universe. As a result, there should be small, but discernable, differences between a Universe composed of only matter, which condenses via gravity to form clumps, and a Universe with matter and dark energy. This is because dark energy opposes gravitational attraction and impedes the formation of clumps. If the dark energy is extremely strong, then no clumps will form at all. If it is quite weak, we will not be able to observe its effects. Clearly, in our Universe, the dark energy has not prevented the formation of structure, and astronomers have been able to search for its imprint on the structures they observe.

In Figure 17.18, we show the error ellipses from three independent data sets. The diagonal orange region is derived from CMB measurements. The blue region comes from supernova measurements. The nearly vertical green region shows results from large-scale structure measurements, specifically baryon acoustic oscillations (see Going Further 17.4: Baryon Acoustic Oscillations). Together, the three data sets place strong limits on the cosmological model. In this case the values of Ω_m (total matter content) and Ω_{DE} (dark energy content) are strongly constrained. The diagonal line running from upper left to lower right indicates a flat Universe, in which $\Omega = \Omega_m + \Omega_{DE} = 1$.

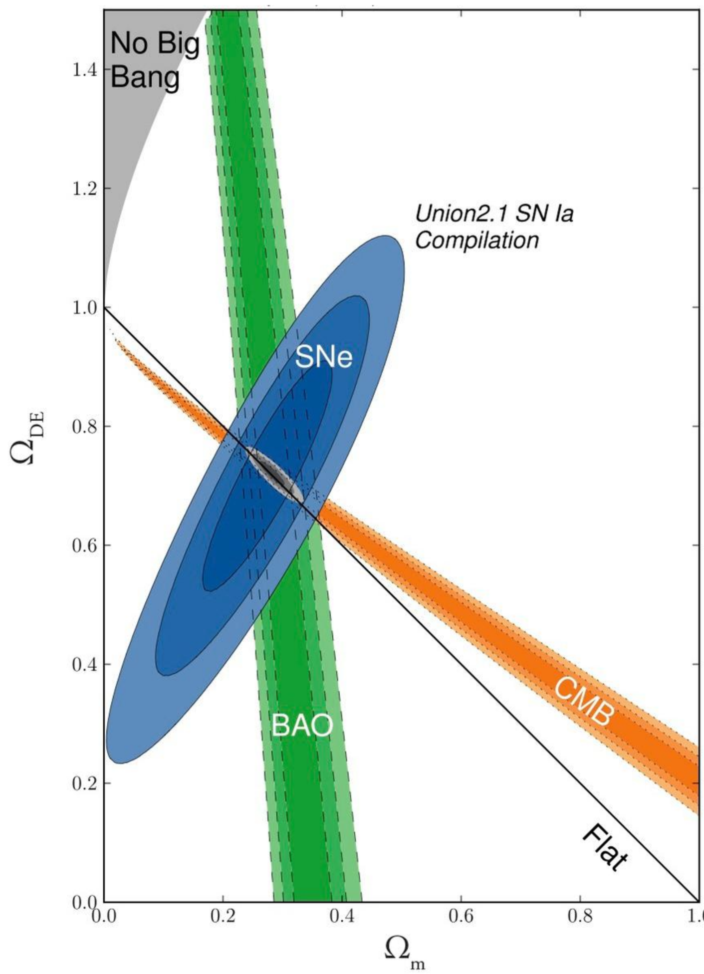


Figure 17.18 This plot shows constraints on the matter fraction (Ω_m) and dark energy fraction (Ω_{DE}) as derived from measurements of large-scale-structure baryon acoustic oscillations (BAO, green), the cosmic microwave background (CMB, orange), and supernovae (SNe, blue). None of the measurements alone does a very good job of constraining either parameter, but together they place very strong limits on both, as indicated by the gray region at their intersection. Credit: Supernova Cosmology Project, Suzuki et al. 2012, *Astrophysical Journal.* 746, 85.

There is something truly remarkable about this process. Different sorts of measurements — each using different kinds of instruments to look at completely different kinds of objects, all involving different kinds of physical processes — give completely consistent results. Their error ellipses converge to a small region in the parameter space. One could imagine them giving radically different answers, totally inconsistent with one another. In that case we would know that our model cosmology—the concordance model—was completely wrong. The fact that all the disparate measurements are instead convergent in a small region of parameter space lends confidence that the model is actually a fair description of the Universe and its evolution.

There are many complementary sets of parameters of this sort. Figure 17.19 shows an example of constraints on the dark energy from measurements of the CMB. CMB measurements are sensitive to not only the dark energy, but also the amount of normal matter (baryons, Ω_{baryon}), the amount of cold dark matter (Ω_{cdm}), and other parameters. These variables are shown along with the dark energy fraction, here called Ω_Λ. Error ellipses are determined from the Planck and WMAP missions. Notice how the baryon graph complements the cold dark matter graph, as their uncertainty ellipses are nearly perpendicular. Together they provide much tighter constraints on the cosmological model. Plots with other variables can also be constructed, as can plots using datasets other than those from the CMB.

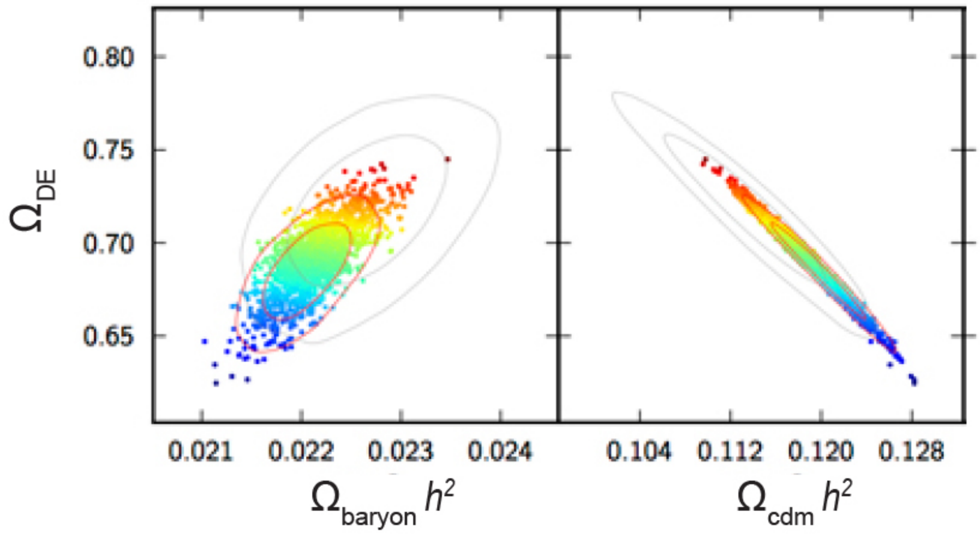


Figure 17.19 The baryon fraction ($\Omega_b h^2$) and cold dark matter fraction ($\Omega_c h^2$) are plotted vs the dark energy fraction (here called Ω_Λ) from CMB data. The smaller red contours are from Planck and WMAP polarization data. The larger gray contours are from WMAP alone. The colored dots depict different values of the Hubble constant, running from 64 km/s/Mpc (blue) to 72 km/s/Mpc (red). The horizontal band running through the most likely values for the various parameters shows that they are all consistent with dark energy making up about 70% of the total energy density. The two plots complement one another because the uncertainties in one are perpendicular to the uncertainties in the other. Credit: NASA/SSU/Aurore Simonnet. —Adapted from Planck 2013 Results XVI.

GOING FURTHER 17.4: BARYON ACOUSTIC OSCILLATIONS

17.4.3: OBSERVATIONAL TESTS: WHAT IS THE MOST LIKELY FATE OF THE UNIVERSE?

Current observational tests of cosmology indicate that the matter in the Universe accounts for approximately 30% of the critical density. Only one-sixth of that amount is baryons. The rest is dark matter, which so far is completely undetectable aside from its gravitational influence on the baryonic material we are able to see. In addition, 70% of the matter—energy density of the Universe is dark energy. The balance of observational evidence suggests that the dark energy is a cosmological constant, though other models have not been completely ruled out. The overall curvature of the Universe is consistent with zero to high precision. Thus, the flat Universe predicted by inflation is borne out by observations. In this section, we will recap these observations with an emphasis on their cosmological ramifications.

Astronomers have long known that there is not enough matter in the Universe to create a flat geometry. They know this from their census of galaxies and galaxy clusters. From these counts they can compute a baryon density, since galaxies glow from the light of their stars. Clusters of galaxies glow from stars too, as well as from x-rays emitted by the hot intracluster gas they contain. The mass contained in this gas actually exceeds that in stars as a proportion of the total cluster mass. Of course, the bulk of the total mass in galaxies and clusters is not baryons at all, it is cold dark matter.

Outside of galaxies and clusters, a few baryons are found in the intergalactic medium. This material can be seen because it absorbs some of the light from background quasars.

When you add all of the mass - baryonic and dark matter - you find that you have $\Omega_m = 0.3$, just under a third of what is needed for a flat geometry. This was the best that could be done until the late 1990s.

While many cosmologists were busy searching for matter in the Universe, a small group was engaged in trying to piece together the matter content by determining its expansion history. As we have already discussed, they used white dwarf supernovae to create Hubble diagrams. By using extremely bright standard candles, like Type I supernovae, these diagrams could stretch back billions of years.

To their astonishment, and almost everyone else's, what they found was that the Universe is expanding more rapidly now than it did in the past. It therefore contains some sort of dark energy, a substance that has negative pressure. In the context of general relativity, such a substance creates a repulsive gravitational effect. The early supernova studies were not able to determine the amount of the dark energy, nor could they discern its type. But they did show that it was present in amounts significant enough to have an effect on the expansion.

About a year after the discovery of dark energy, a balloon-borne CMB experiment in Antarctica (BOOMERANG) was able to measure the first peak of the power spectrum of temperature fluctuations in the cosmic microwave background. From these data it was clear that we live in an $\Omega = 1$ Universe. This result did not come as a surprise; inflation had made a strong prediction that the Universe must be flat, and the combination of previous CMB measurements pointed in that direction. However, directly measuring a flat geometry was an important observational test.

The $\Omega = 1$ result, coupled with $\Omega_m = 0.3$ from earlier studies, was an independent, though indirect, detection of the dark energy. It showed that the dark energy comprised around 70% of the critical density. Since that time, CMB measurements have also constrained the fraction that baryons (4.5%) and cold dark matter (25%) can contribute to the overall density. Subsequent CMB experiments of much higher sensitivity have also confirmed and strengthened the $\Omega = 1$, $\Omega_{DE} = 0.7$ finding.

In addition to these measurements, the first decade or so of this century has seen cosmologists increase their efforts to understand the details of the Big Bang paradigm and the concordance model. One of the most powerful methods they have employed has been to use galaxy redshift surveys to extract the three-dimensional distribution of matter in the Universe. Just like the CMB, the pattern of fluctuations in the data is intimately related to the constituents of the Universe: it is the same density fluctuations in the early Universe that formed the hot and cool spots in the CMB and the high- and low-density regions that are seen as galaxy clusters, filaments, and voids today.

Unlike the CMB, the galaxy distribution can be studied at different times by observing galaxies at different redshifts. When this is done, it allows cosmologists to observe effects like the dark energy as it begins to dominate the expansion. These sorts of studies have not only confirmed that the dark energy exists, but they strongly imply that it is a constant in time, a cosmological constant rather than other models in which dark energy varies in time.

Thus, our current understanding suggests that the Universe will eventually cool in a big chill, with an expansion rate that is ever increasing. Dark energy began to dominate the energy content of the Universe around six to eight billion years after the Universe began, and the overall composition is increasingly dominated by dark energy. This is depicted in Figure 17.20.

The conclusion of a dark energy-dominated cosmology is still not completely certain, but most data suggest that is where the Universe is headed. The expansion and geometry of the Universe according to our current best measurements is depicted schematically in Figure 17.21 and Animated Figure 17.22. Time will tell if this is the correct picture, or if important modifications will be necessary as new data are collected and analysed.

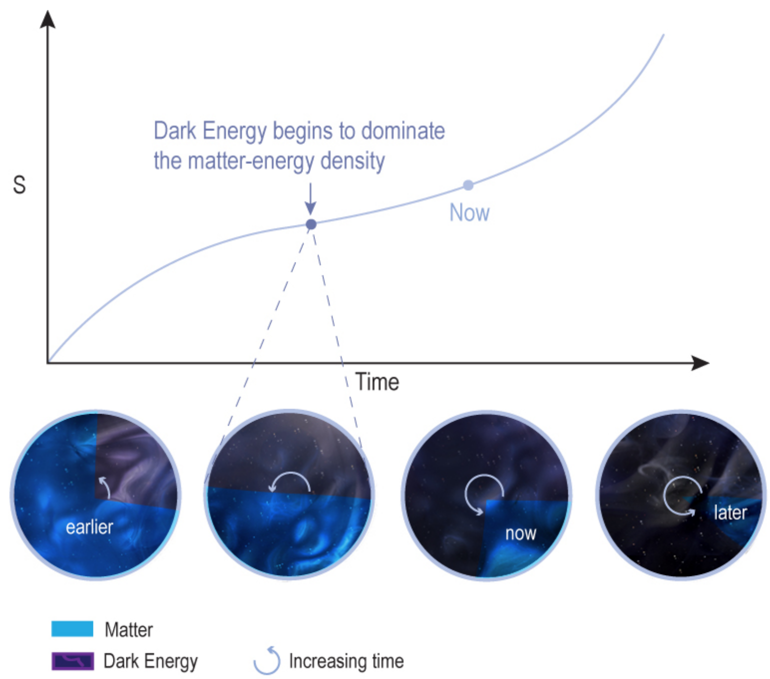


Figure 17.20 Current data suggests that the dark energy is a cosmological constant. Therefore, as the Universe continues to expand, the dark energy will come to dominate more and more, driving the expansion to ever-faster rates. Under these conditions, the Universe eventually reaches a big chill scenario. Credit: NASA/SSU/Aurore Simonnet.

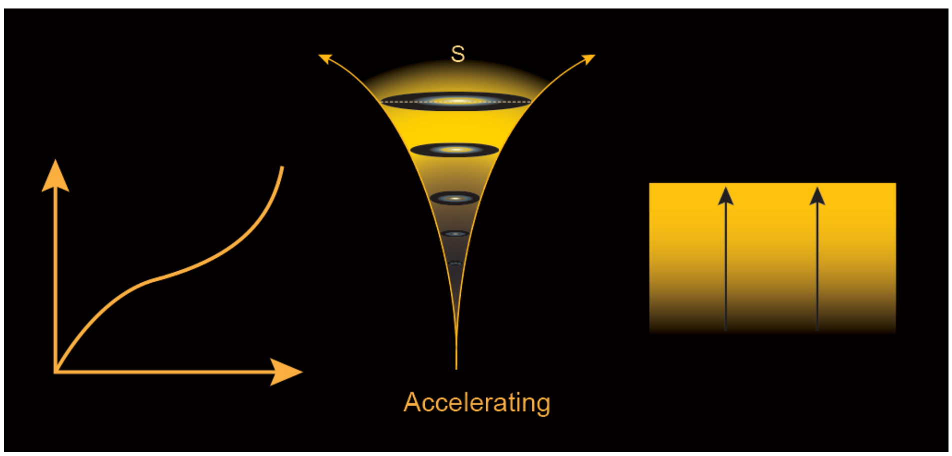
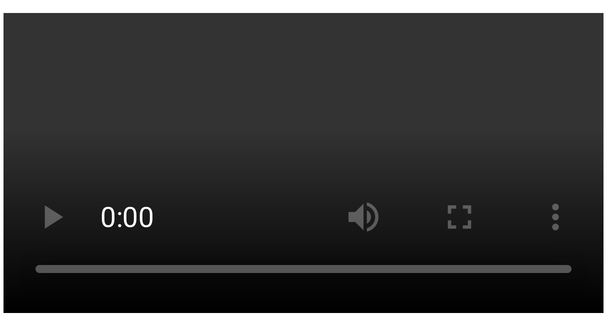


Figure 17.21 The concordance model of the Universe. With the inclusion of dark energy, a flat Universe will undergo accelerating expansion. The graph on the left is the scale factor vs. time, which is illustrated in the center. The illustration on the right indicates that the geometry of the Universe is flat, in other words, that the curvature of space overall is zero. Credit: NASA/SSU/Aurore Simonnet.



Animated Figure 17.22 Current data indicates that expansion of the Universe will speed up due to dark energy, and that the fate will be a big chill scenario. Credit: NASA/SSU/Kevin John.

This page titled 17.4: Cosmic Concordance and Cosmological Parameters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.5: Summary

Learning Objectives

- You will reflect on what you have learned about cosmology

Helped are those who love the entire cosmos... for to them will be shown the unbroken web of life and the meaning of infinity...

- Alice Walker

WHAT DO YOU THINK: THE FATE OF COSMOLOGY STUDENTS

Some students are relaxing after their final exam in cosmology, reflecting on what they learned.

- **Julianna:** I never knew the Universe was so complicated. I'm glad I don't have to keep all that stuff straight anymore.
- **Teresa:** I don't know. I think it all fit together pretty well. I wish I could keep learning about it.
- **Luca:** I think it's cool that I can understand some of the news reports about space now. I never could before. Maybe I'll take another astronomy class sometime.

Do you agree with any or all of these students and if so whom?

Teresa

Julianna

Luca

None

Explain.

With this chapter we end our exploration of the modern ideas in cosmology. Our species' ideas concerning the cosmos began with attempts to explain the nature and motions of the stars and planets. Eventually those early notions led to mathematical laws describing the motion of falling bodies, the nature of electromagnetic waves and the structure of atoms. Over the past century we have seen our grasp of Nature extend to theories of the ultimate nature of matter, in which all of its interactions and constituent parts are explained by a single standard model of particle physics. Our understanding of the three of the four forces of nature fit into an orderly, if as yet incomplete, structure based on quantum theory. Yet the most familiar force, gravity, is understood in terms of the geometry of a four-dimensional spacetime within the framework of general relativity, a decidedly non-quantum theory. Unifying gravity with the other three forces in a single mathematical framework, a "Theory of Everything," is still an unrealized ambition of many physicists.

In these modules we have presented the story of cosmology, how our understanding of the cosmos developed, and how early developments led to later ones. But more than that we have attempted when possible to let you discover the story for yourself, feeling that it is best if you understand what the data have to say about the nature of our Universe. It is the data that are the ultimate authority in science, and it is an understanding of these data, both their strengths and their limitations, that is required by anyone who wants to know how scientists go about exploring reality.

We hope you have come to appreciate both the simplicity and complexity of the underpinnings of the Big Bang theory. At its most fundamental level, it is composed of some very simple ideas: expanding gases cool, so the Universe was hotter (and denser) in the past and will be cooler (and less dense) in the future; the Universe was simpler in the past, and has evolved complex structures over time as the result of small deviations from homogeneity early in its history; basic gravitational, thermal, atomic and nuclear physics acted on these early seeds to form the structures we see now; this process of evolution continues, and will continue, into the future.

These ideas have led to some truly startling conclusions. The first, that the material Universe that we see is only a small fraction of the matter that exists, has its origins in the study of the dynamics of galaxies and galaxy clusters, and it is bolstered by Big Bang explanations of structure formation. Furthermore, the matter in the Universe makes up less than a third of its total energy content, and that fraction is shrinking as the expansion continues; the Universe is now dominated by a form of energy, called dark energy for lack of a better name, that is completely mysterious at the moment, and it is also almost completely unexpected.

We do not yet know how the story will end. Current evidence suggests that the dark energy will become more dominant over time. In any event, its mere presence demonstrates that the Universe is likely destined to expand forever, cooling and becoming ever less dense. The stars will go out and darkness will reign. Eventually black holes will evaporate, and perhaps even some of the "fundamental" particles might decay into simpler forms. On the longest timescales it seems that the Universe ends not with a bang, but a whimper. In the meantime, there is a vast amount to explore, learn about and come to understand. Our explorations have barely started.

This page titled 17.5: Summary is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.6: Wrapping It Up 17 - Cosmological Parameter Estimation

Scientists working with supernova, CMB, and large-scale structure data have determined to high precision the ingredients making up the Universe. From these data, scientists have constructed a pie chart depicting the matter and energy budget of the Universe (Figure A.17.9).

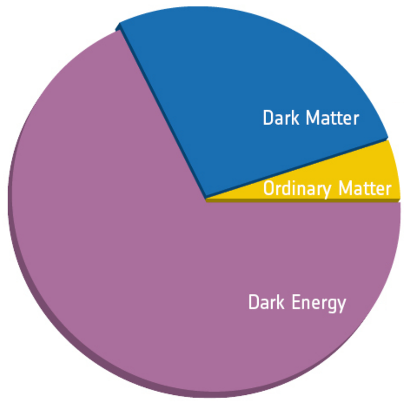


Figure A.17.9: Ordinary matter represents only a small portion of the total mass-energy inventory of the Universe. Dark energy, which is responsible for accelerating the expansion of space, accounts for the majority of the budget. Credit: ESA/Planck

In this activity, you will use data from various cosmological measurements in order to determine the values of several cosmological values: the amount of matter (Ω_m), which is comprised of regular baryonic matter and cold dark matter, and the amount of dark energy (Ω_{DE}).

You should notice that there are constraints on Ω_m and Ω_{DE} from three different measurements:

- Supernovae (SNe, blue region)
- The cosmic microwave background (CMB, orange region)
- Large-scale structure baryon acoustic oscillations (BAO, green region)

Click on each cosmological source to pull up confidence regions for the measurement.

For a given source, the combination of Ω_m and Ω_{DE} has a 68% chance of falling within the smallest region, 95% chance of falling within the next larger region, and a 99% chance of falling within the entire shaded region. A 99% confidence region means that the true value will lie outside the region only 1% of the time. In other words, we are 99% confident that the true value lies inside the region bounded by the contour.

Play Activity

Part I: Individual Measurements

1. Plot the BAO measurement. With 99% confidence, what is the range of values for Ω_m?

Cannot determine from this data

0.15 – 0.30

0.15 – 0.45

0.20 – 0.40

0.25 – 0.45

2. With 68% confidence, what is the range of values for Ω_m from BAO alone?

Cannot determine from this data

0.15 – 0.30

0.15 – 0.45

0.20 – 0.40

0.25 – 0.45

3. Within 68% confidence, what is the range of values for Ω_{DE} from BAO alone?

Cannot determine from this data

0.15 – 0.45

0.20 – 0.40

0.60 – 0.80

4. Overall, with 99% accuracy, do measurements from BAO alone constrain Ω_m, Ω_{DE}, or both?

m"> constrains only Ω_m

DE"> constrains only Ω_{DE}

constrains both

5. Now click the measurements from the CMB only to constrain the region of possible values for Ω_m and Ω_{DE}. Overall, with 99% accuracy, does this measurement alone constrain Ω_m, Ω_{DE}, or both?

m"> constrains only Ω_m

DE"> constrains only Ω_{DE}

constrains both

6. Now click the measurements from the supernovae (SNe) only to constrain the region of possible values for Ω_m and Ω_{DE}. Overall, with 99% accuracy, does this measurement alone constrain Ω_m, Ω_{DE}, or both?

m"> constrains only Ω_m

DE"> constrains only Ω_{DE}

constrains both

Part II: Combining measurements

1. As you click to add additional cosmological measurement sources, the size of the overlapping region

increases

decreases

stays the same

2. Combing the confidence regions of both the BAO and CMB, with 68% confidence, what is the range of values for Ω_m?

cannot determine from this data

0.22 – 0.39

0.29 – 0.32

3. Combing the confidence regions of both the BAO and CMB measurement, are Ω_m and Ω_{DE} constrained to within 99% confidence?

m"> constrains only Ω_m

DE"> constrains only Ω_{DE}

constrains both

4. Click on various combinations of BAO, CMB, and SNe. Which pair of measurements gives the smallest range of values for Ω_m?

BAO+CMB

BAO+SNe

CMB+SNe

5. Which pair of measurements gives the smallest range of values for Ω_{DE}?

BAO+CMB

BAO+SNe

CMB+SNe

6. With all three measurements checked, you should see a set of gray/black confidence regions. What is your best estimate of the matter density fraction, Ω_m from all measurements combined?

7. What is the best estimate of the dark energy density fraction, Ω_{DE} from all measurements combined?

8. You should also see a line labeled “flat,” which denotes the region of this plot where the total density, Ω, is equal to 1. In other words, $\Omega_m + \Omega_{DE} = 1$. If the combined values of Ω_m and Ω_{DE} fall below and to the left of this line, what type of Universe would that indicate?

a flat universe

one that will expand forever

one that will collapse eventually

9. The gray region indicates that the best fit for the geometry of the Universe is:

flat ($\Omega = 1$)

1)"> spherical ($\Omega > 1$)

saddle-shaped ($\Omega < 1$)

This page titled 17.6: Wrapping It Up 17 - Cosmological Parameter Estimation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

17.7: Mission Report 17 - Cosmological Parameter Estimation

A. Please rate on a scale of 1 to 5 how effective you think the Wrapping It Up activity was in helping you understand the material. (1: not effective at all → 5: very effective) * 1 2 3 4 5

B. What were the main ideas that you learned in conducting the Wrapping It Up activity? Be specific and detailed in your response. Please address the following questions: What did you learn? How did you learn it? What is still unclear? (At least 150–200 words.)

*

C. If the Wrapping It Up activity included measurements or data, please describe what factors influenced the accuracy of your results. (Do *not* include mistakes, only unavoidable measurement imprecision.) If you obtained any numerical values for the accuracy of your measurements during the activity, note those here. If there were no measurements or data, say so explicitly.

*

D. Questions to be graded for accuracy. Show your work!

1. Describe two types of observational evidence that support the Big Bang theory and the concordance model.

*

2. Give the approximate percentages of baryonic matter, dark matter, and dark energy as currently measured.

*

3. Describe at least four possible fates of the Universe, and what the contents, geometry, and value of Ω will be in each case.

*

4. Summarize the current understanding of what the fate of the Universe will be based on observational data.

*

This page titled 17.7: Mission Report 17 - Cosmological Parameter Estimation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

SECTION OVERVIEW

Knowls

18: Big Ideas JS code

Pop ups

Knowls is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

18: Big Ideas JS code

test

This page titled 18: Big Ideas JS code is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

0.5: Meters To Kilometres

1.4 km
×
1000 m
1km
=
1400 m

Play

Reset

0.5: Meters To Kilometres is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

0.5: Sun Moon Distance

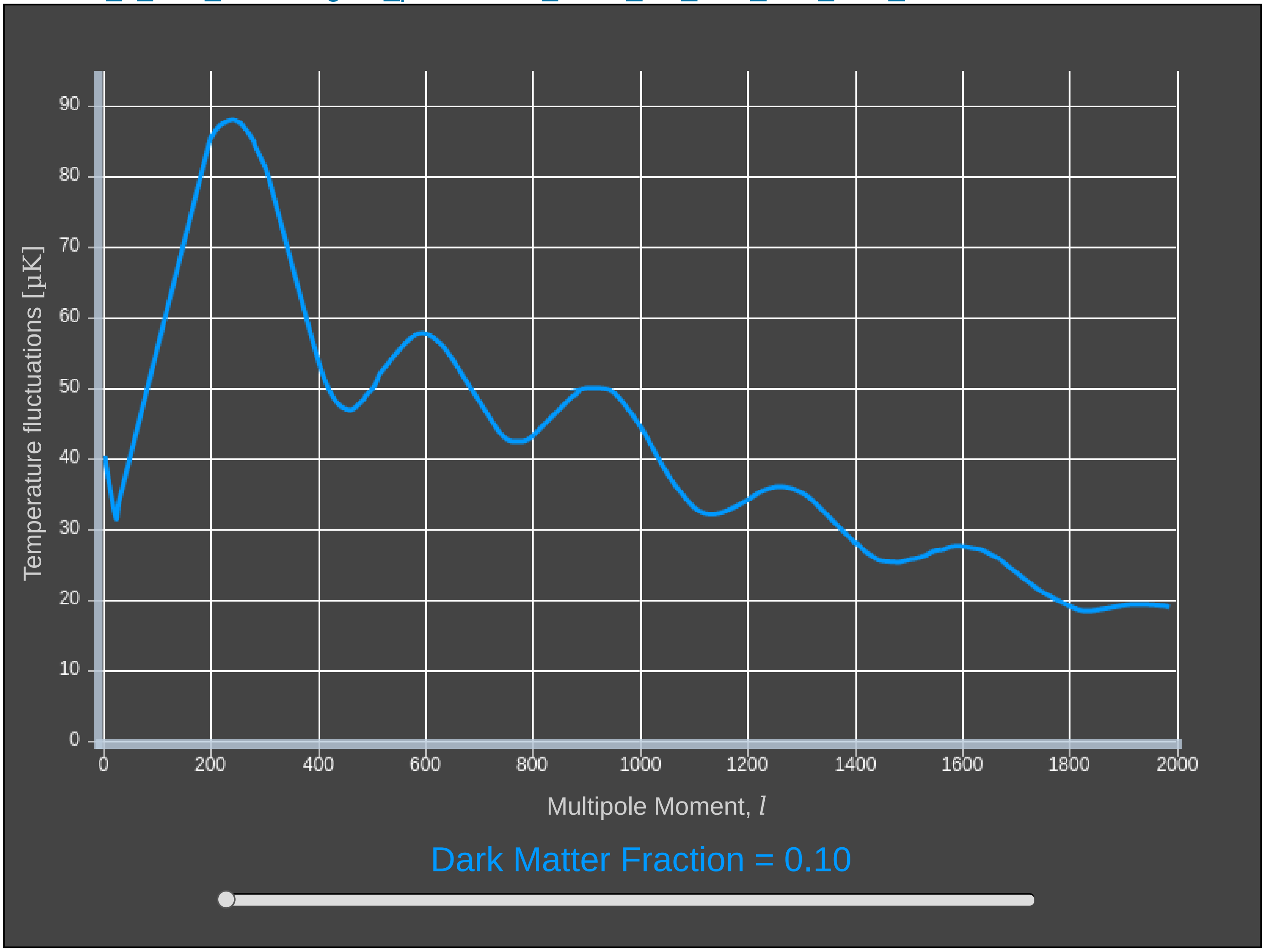
Distance to Sun
Distance to Moon
=
1.5×10^8 km
3.8×10^5 km
=
395 times further

Play

Reset

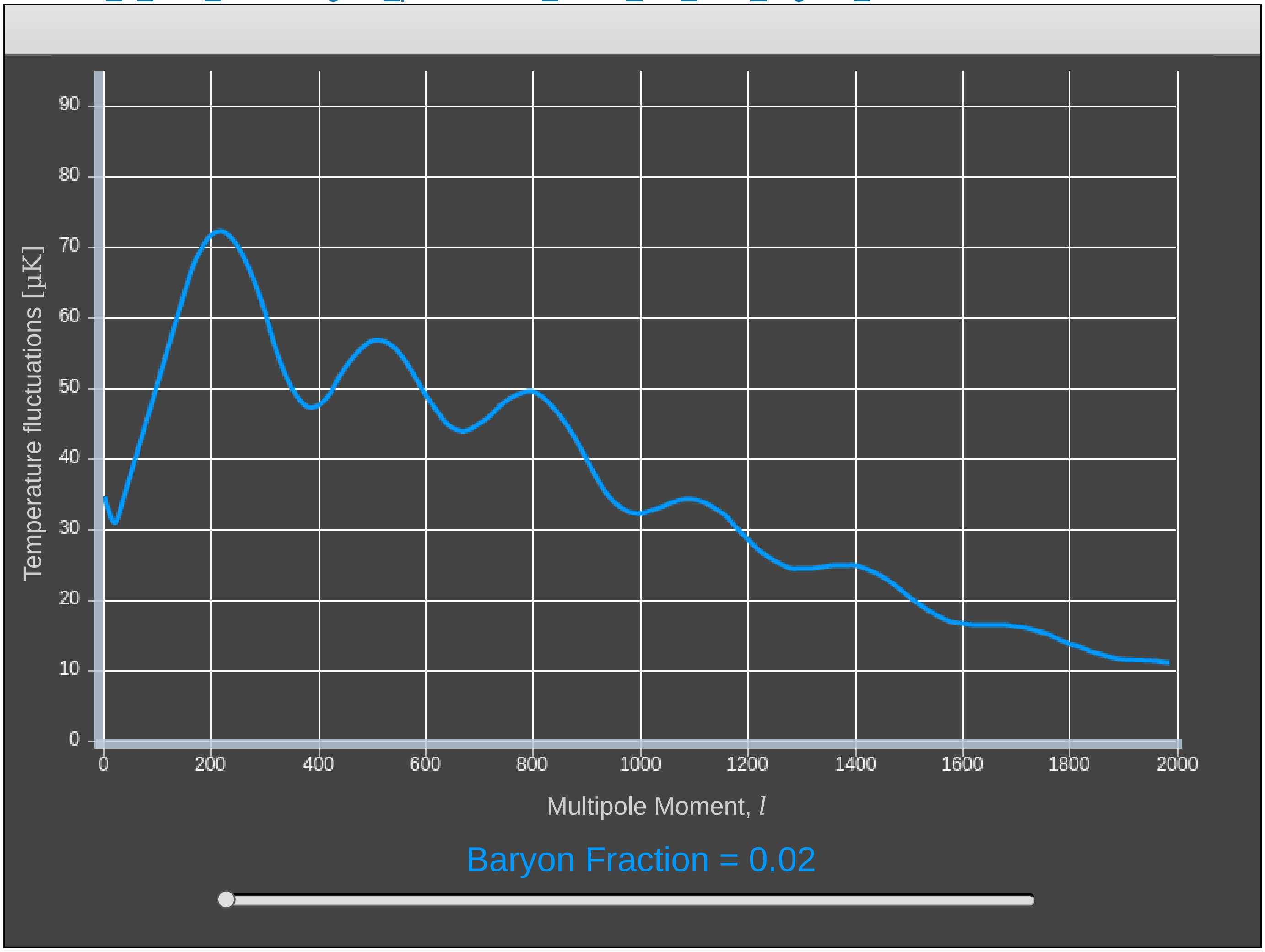
0.5: Sun Moon Distance is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

18.0: 15_4_how_cosmological_parameters_affect_the_cmb_cold_dark_matter



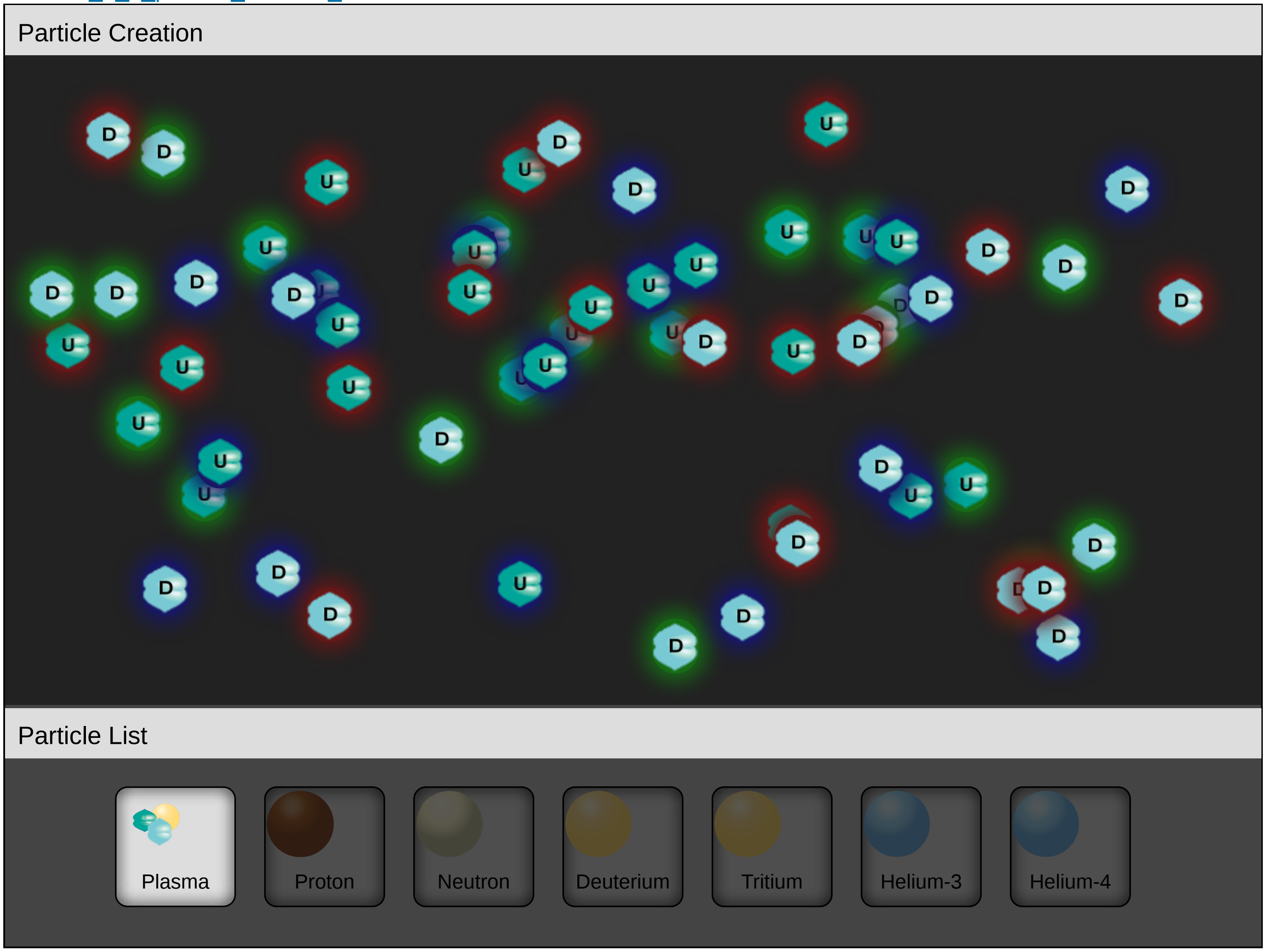
This page titled 18.0: 15_4_how_cosmological_parameters_affect_the_cmb_cold_dark_matter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.1: 15_4_how_cosmological_parameters_affect_the_cmb_regular_matter



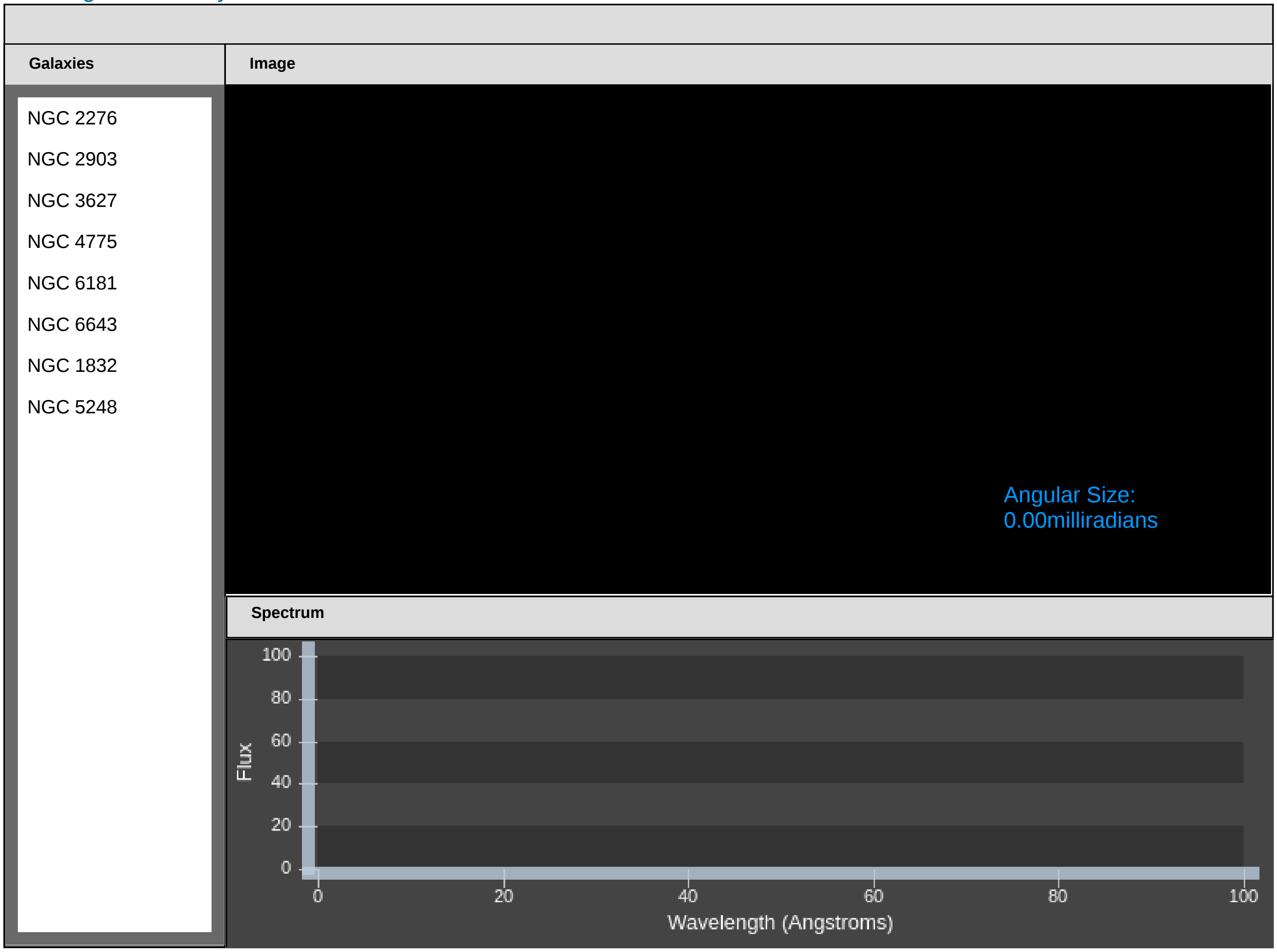
This page titled 18.1: 15_4_how_cosmological_parameters_affect_the_cmb_regular_matter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.2: 16_2_4_particle_creation_timeline



This page titled 18.2: 16_2_4_particle_creation_timeline is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.3: bigIdeasActivityOne



18.4: ch16_particle_classification

Particles

P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13

Categories

+ -

Export

Data

Label	Mass	Electric Charge	Spin	Color Change
P1	Yes	2/3	1/2	Yes
P2	Yes	-1	1	No
P3	Yes	0	1	No
P4	Yes	-1/3	1/2	Yes
P5	Yes	1	1/2	No
P6	Yes	2/3	1/2	Yes
P7	Yes	-1	1/2	No
P8	Yes	-1/3	1/2	Yes
P9	Yes	-1/3	1/2	Yes
P10	Yes	2/3	1/2	Yes
P11	Yes	-1	1/2	No
P12	No	0	1	Yes
P13	No	0	1	No

18.5: ch4pg1lightminute

This page titled 18.5: ch4pg1lightminute is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.6: ch4pg1lightyear

This page titled 18.6: ch4pg1lightyear is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.7: ch4pg1powersoftentimescales

Time for half of Uranium 238 to decay		1.4×10^{17} seconds
Average time to hard boil an egg		4×10^{2} seconds
Time for the Sun to orbit the center of the Galaxy		7.1×10^{15} seconds
Time it takes to blink your eye		3.5×10^{-1} seconds
Time for Earth to rotate on its axis	24 hours	
Lifetime of the Sun		3.15×10^{17} seconds
Time for lightning flash to reach you from 5 km away	0.0167 seconds	
Time for half of Iodine 131 to decay	8 days	
Time for Earth to orbit the Sun		3.15×10^{7} seconds
Average daily sleep time		3×10^{4} seconds
Time for Neptune to orbit the Sun	165 years	5.2×10^{9} seconds

6.91×10^{5} seconds

230 million years

1.67×10^{-2} seconds

4.47 billion years

0.35 seconds

8.64×10^{4} seconds

8.3 hours

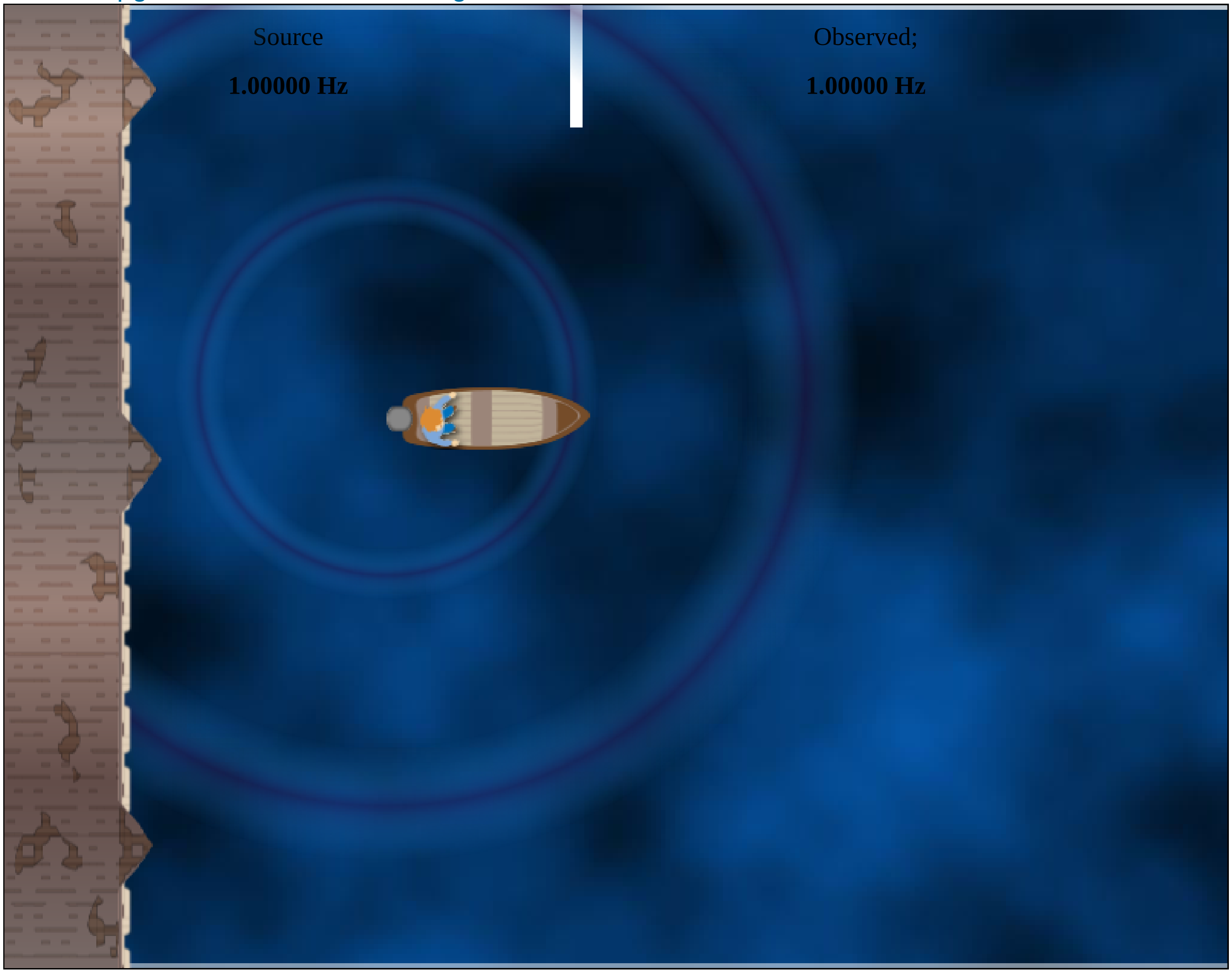
10 billion years

400 seconds

365 days

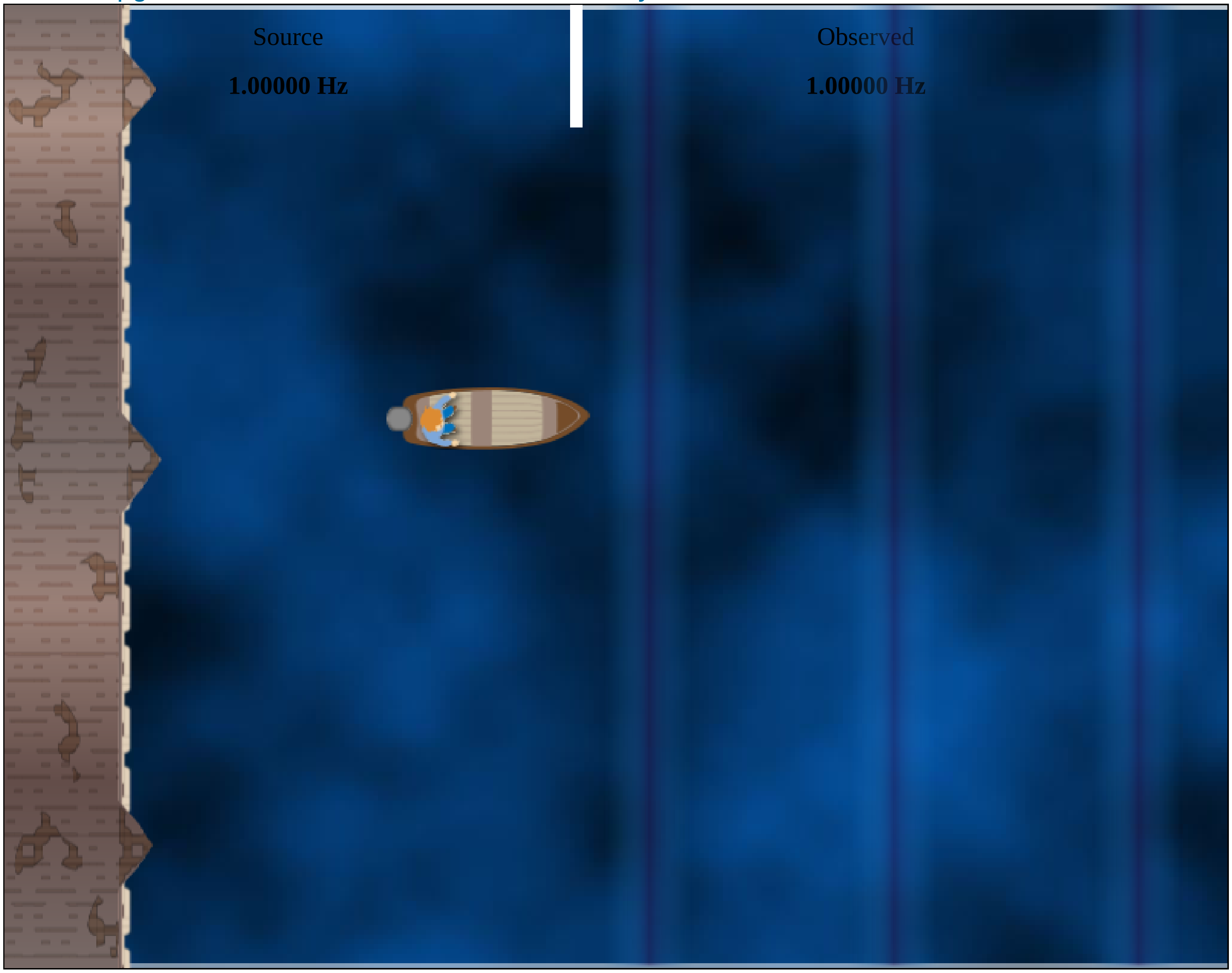
This page titled 18.7: ch4pg1powersoftentimescales is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.8: ch4pg2boatwavedemobcmovingsource



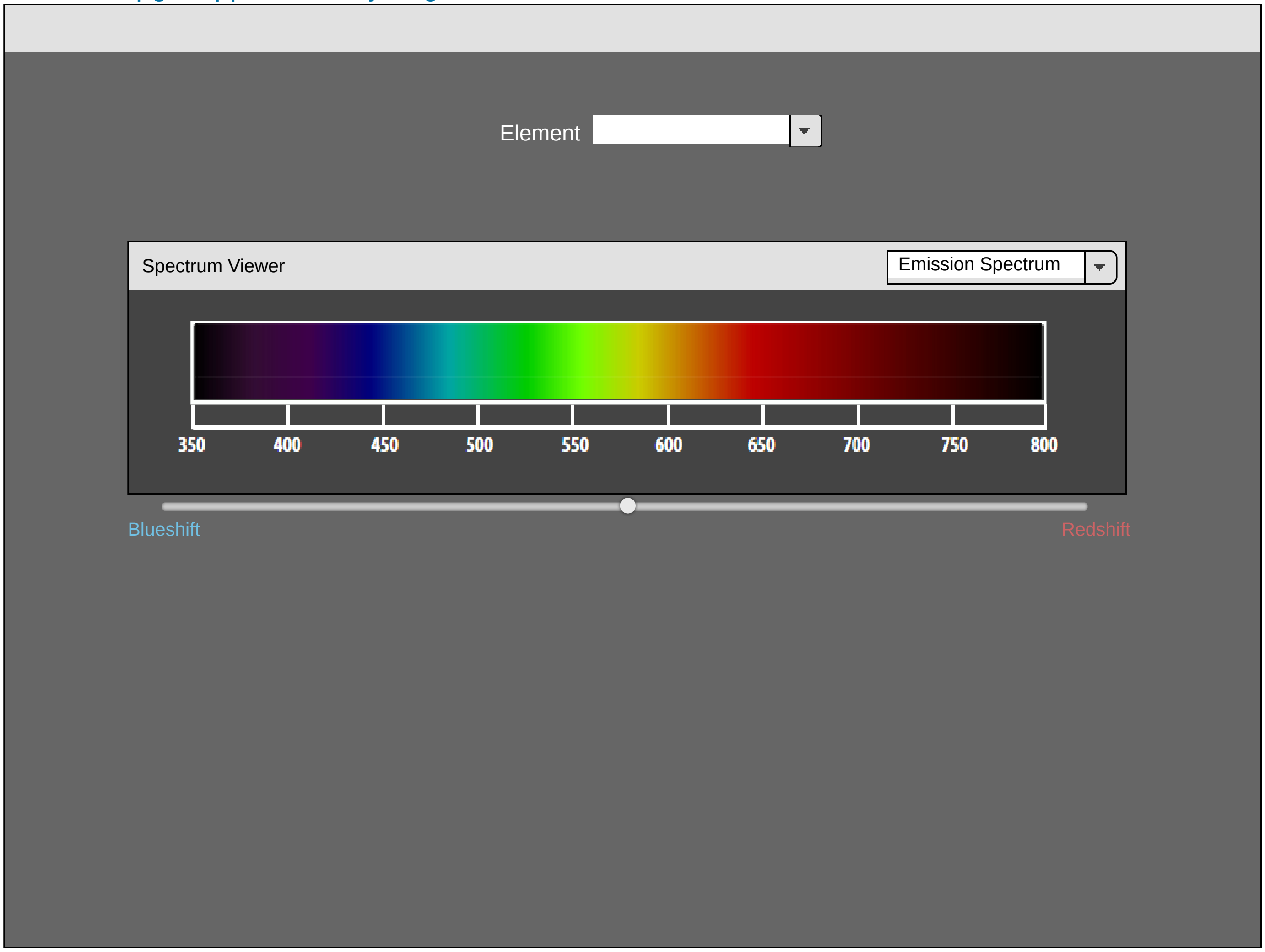
This page titled 18.8: ch4pg2boatwavedemobcmovingsource is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.9: ch4pg2boatwavedemosourceofwavesstationary



This page titled 18.9: ch4pg2boatwavedemosourceofwavesstationary is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.10: ch4pg2dopplershiftofhydrogen

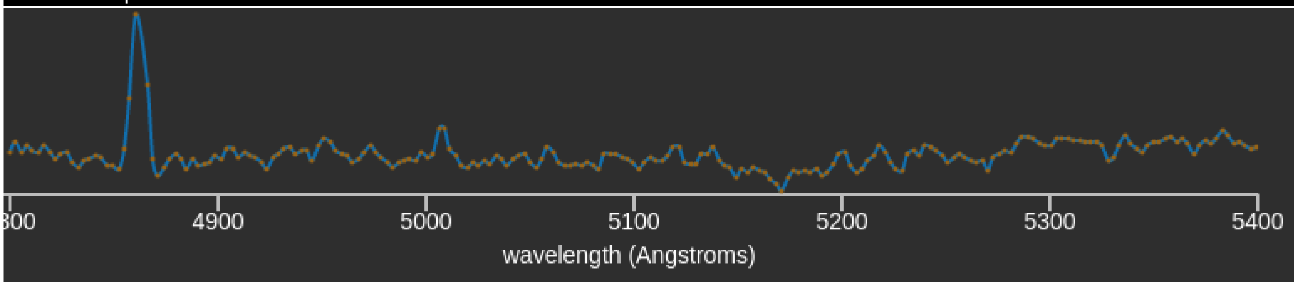


This page titled 18.10: ch4pg2dopplershiftofhydrogen is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.11: ch4pg2quantitativedopplershift

Galaxy:

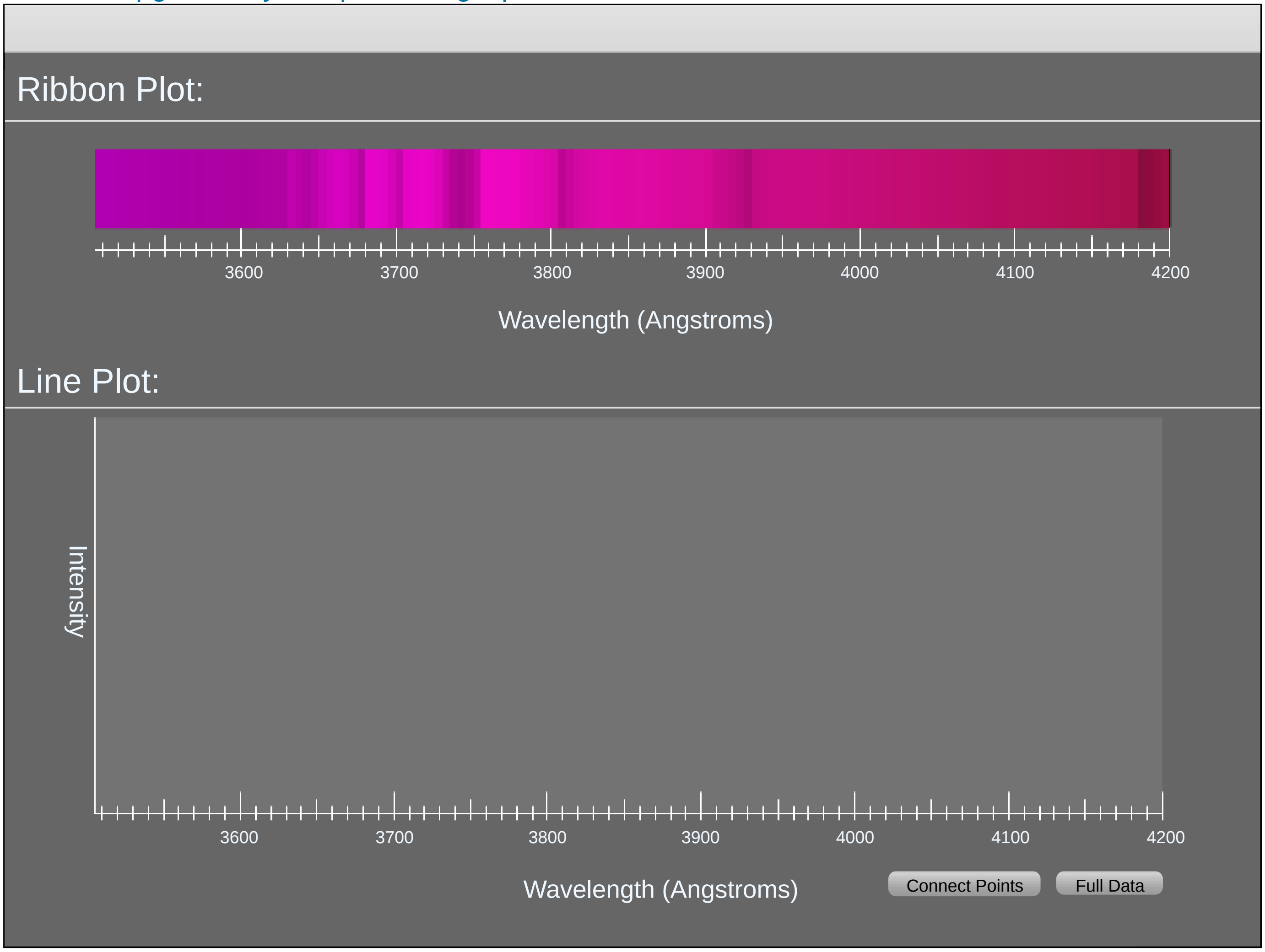
Reference Spectrum:



Observed Spectrum:

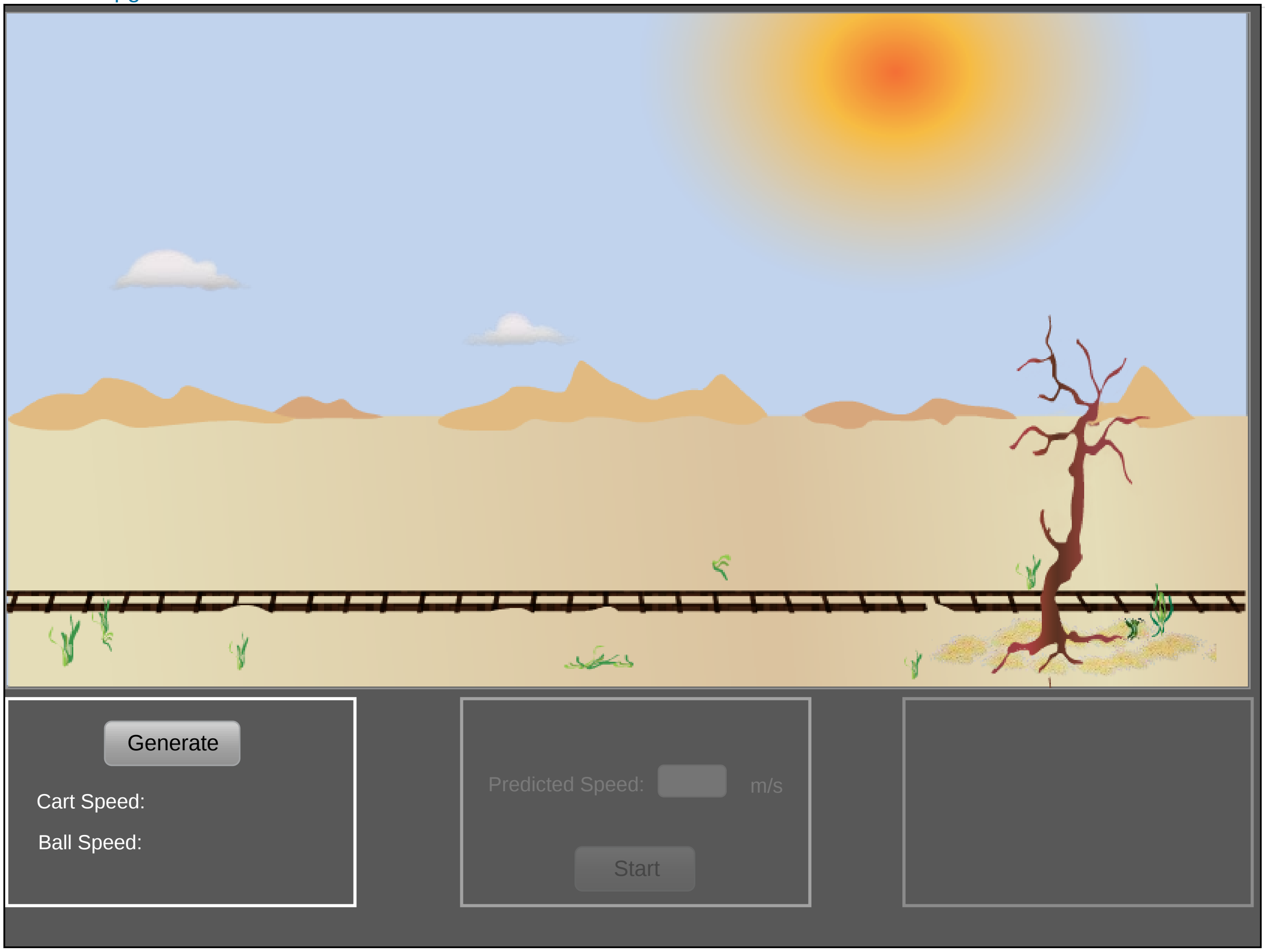
This page titled 18.11: ch4pg2quantitativedopplershift is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.12: ch4pg2twowaysofrepresentingaspectrum



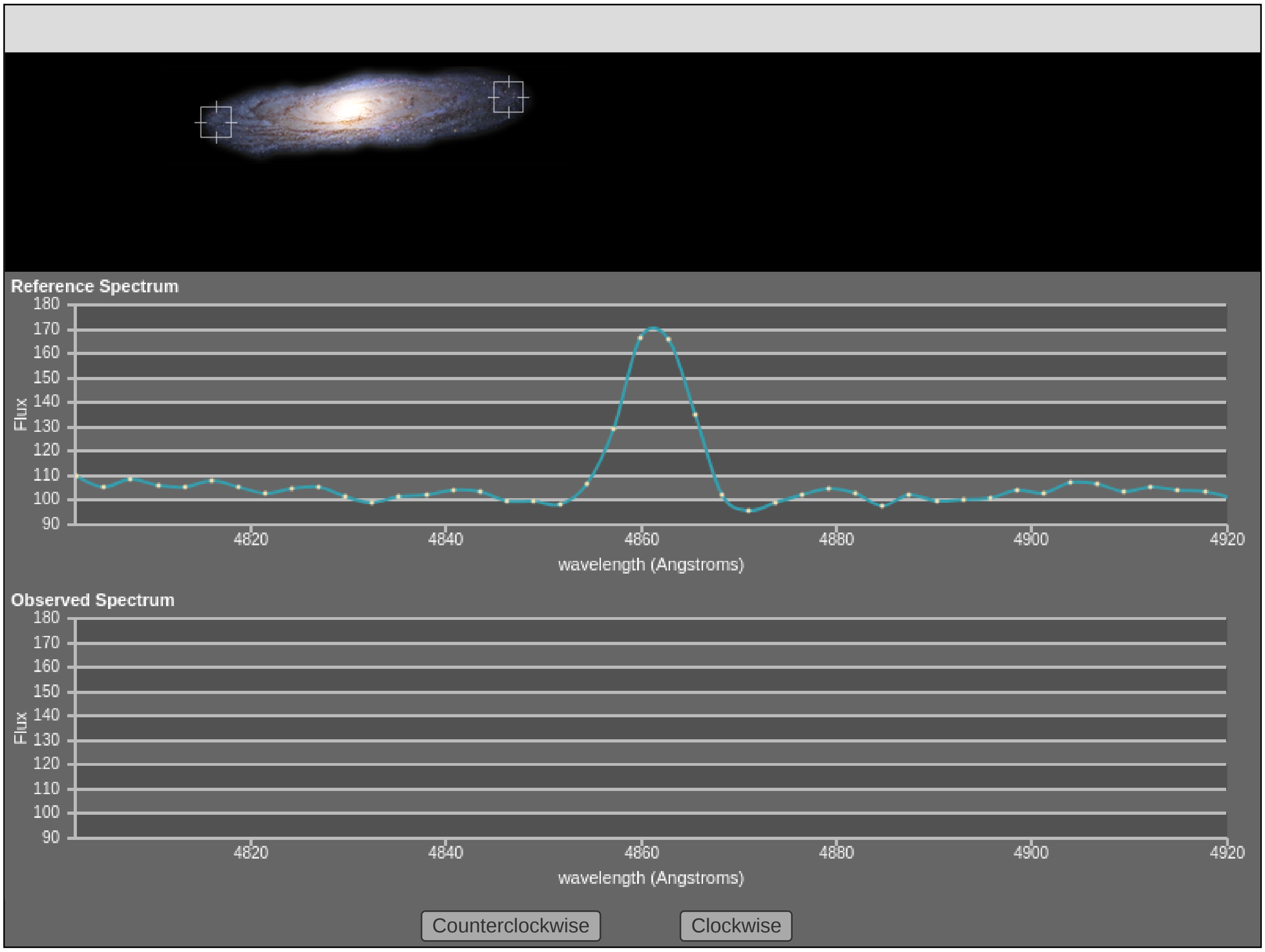
This page titled 18.12: ch4pg2twowaysofrepresentingaspectrum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.13: ch4pg3relativemotion



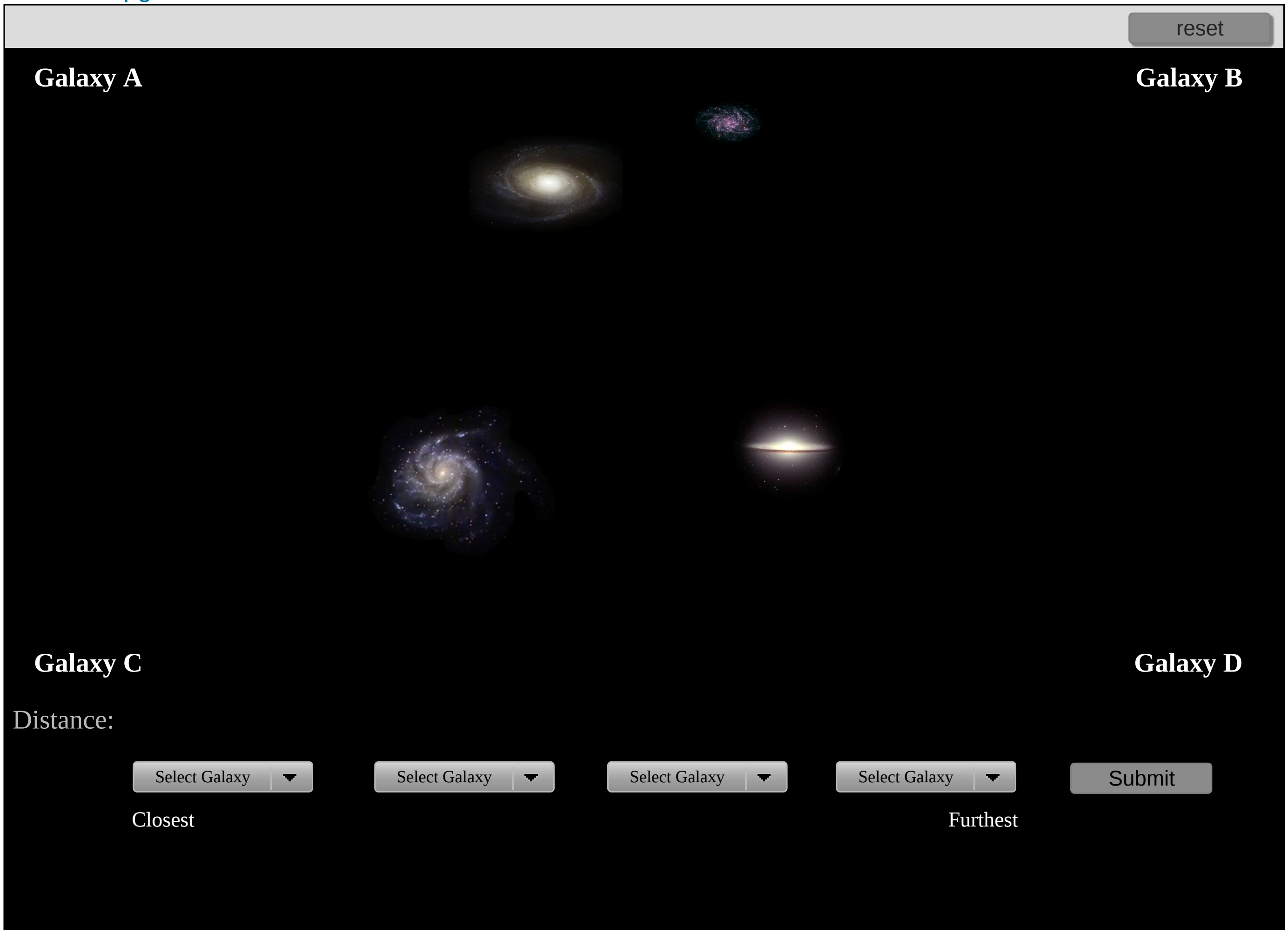
This page titled 18.13: ch4pg3relativemotion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.14: ch4wiu4theandromedashift



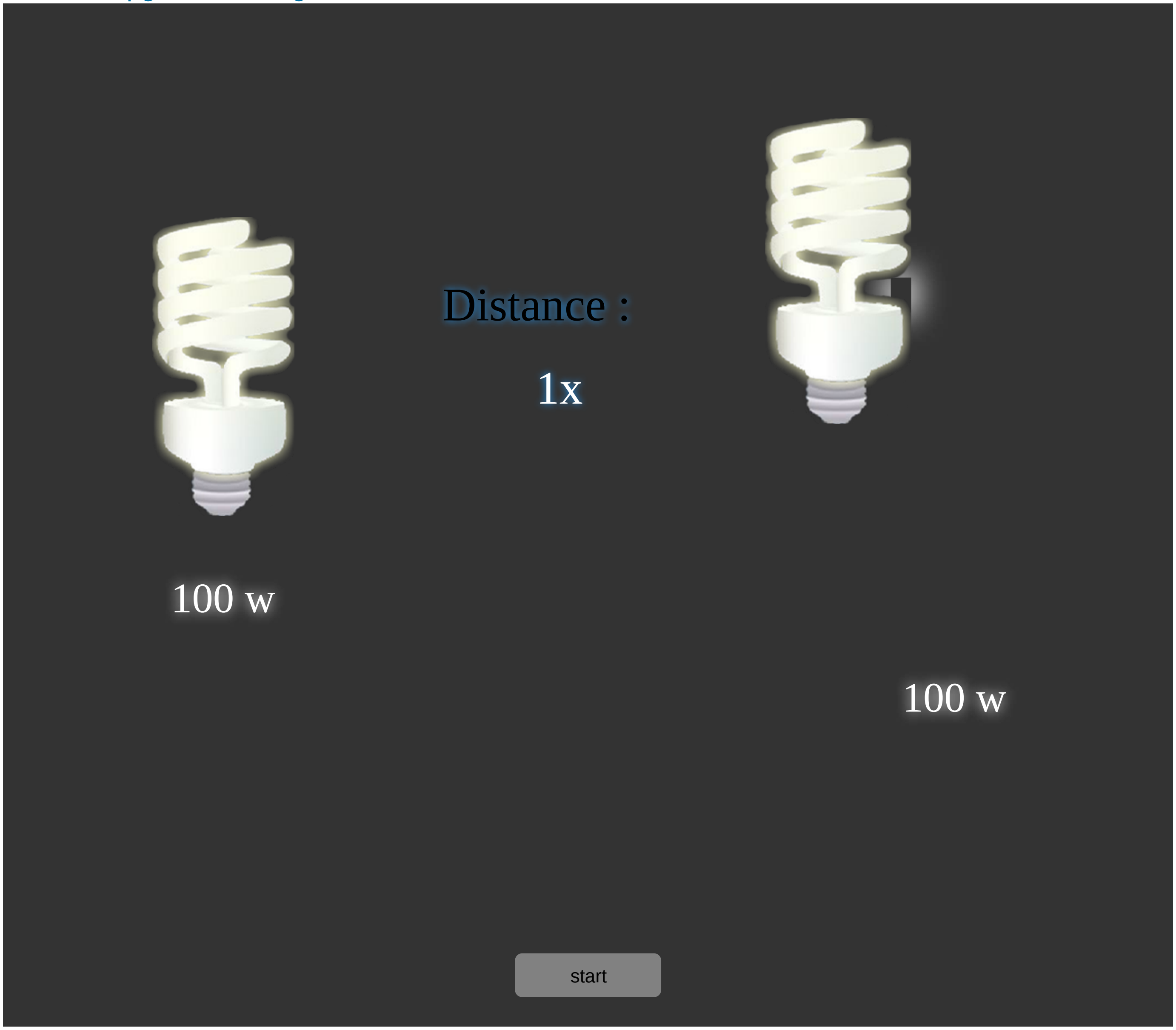
This page titled 18.14: ch4wiu4theandromedashift is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.15: ch6pg2standardruler



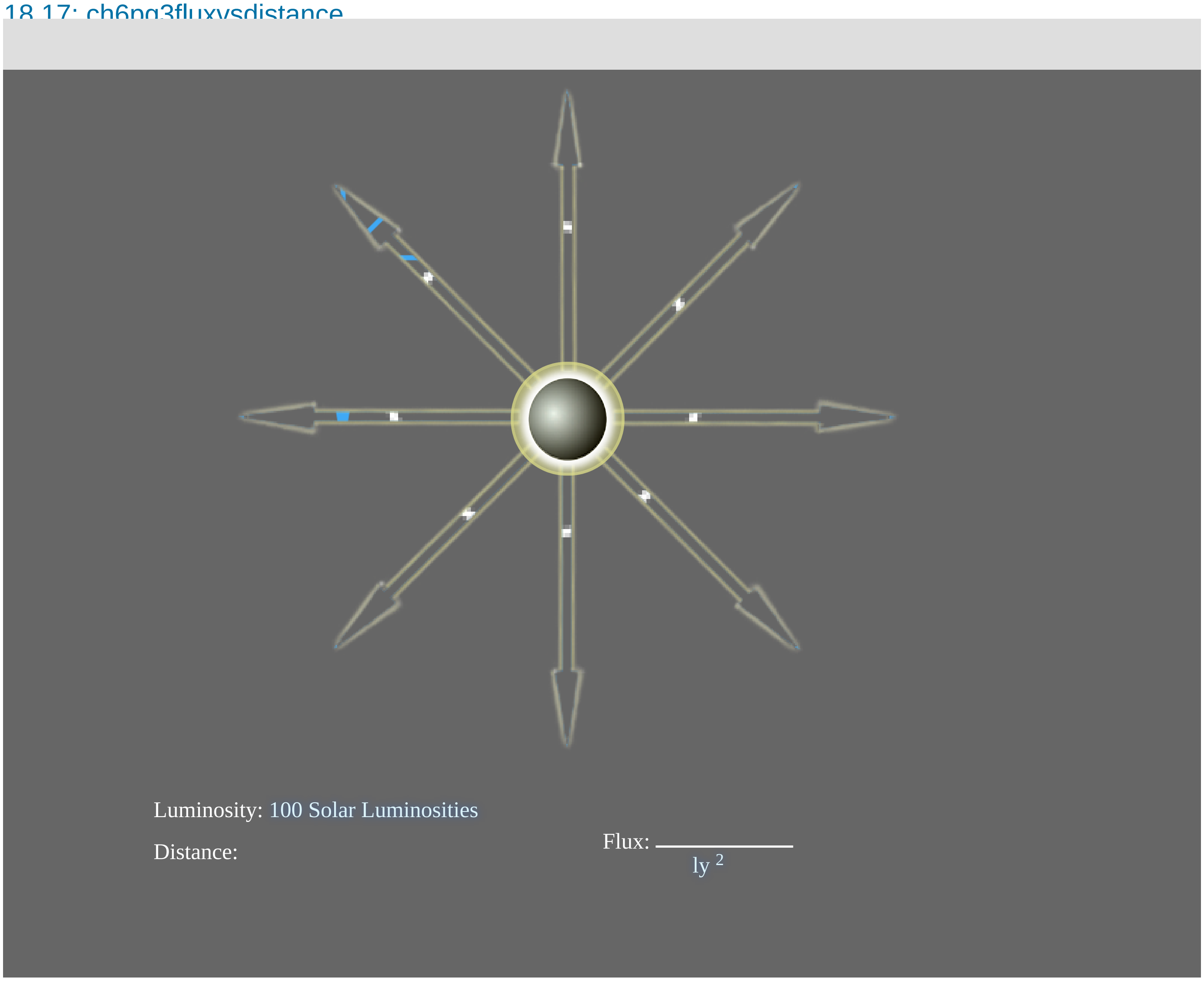
This page titled 18.15: ch6pg2standardruler is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.16: ch6pg3animatedfigure612



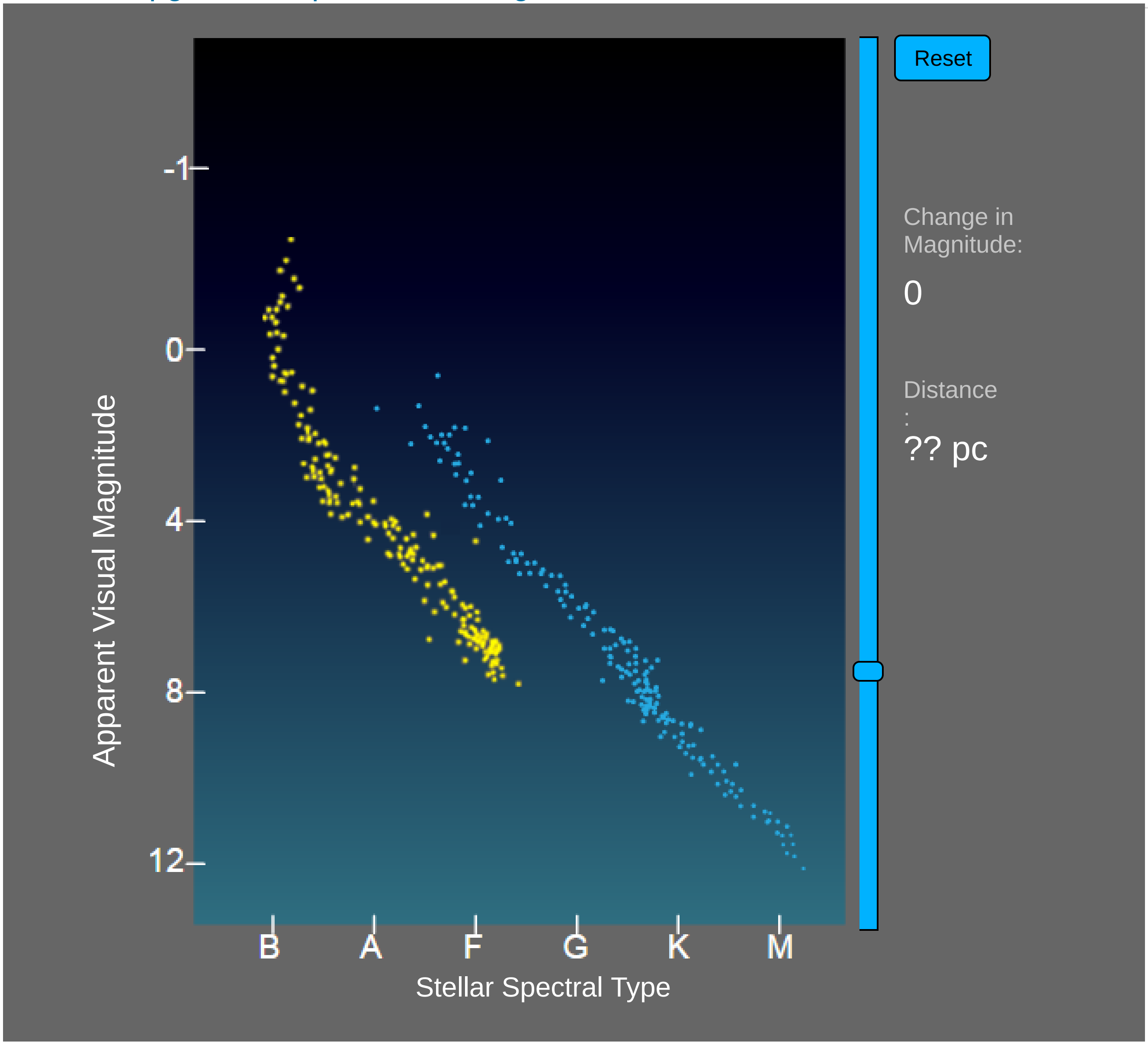
This page titled 18.16: ch6pg3animatedfigure612 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.17: ch6pg3fluxvsdistance



This page titled 18.17: ch6pg3fluxvsdistance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

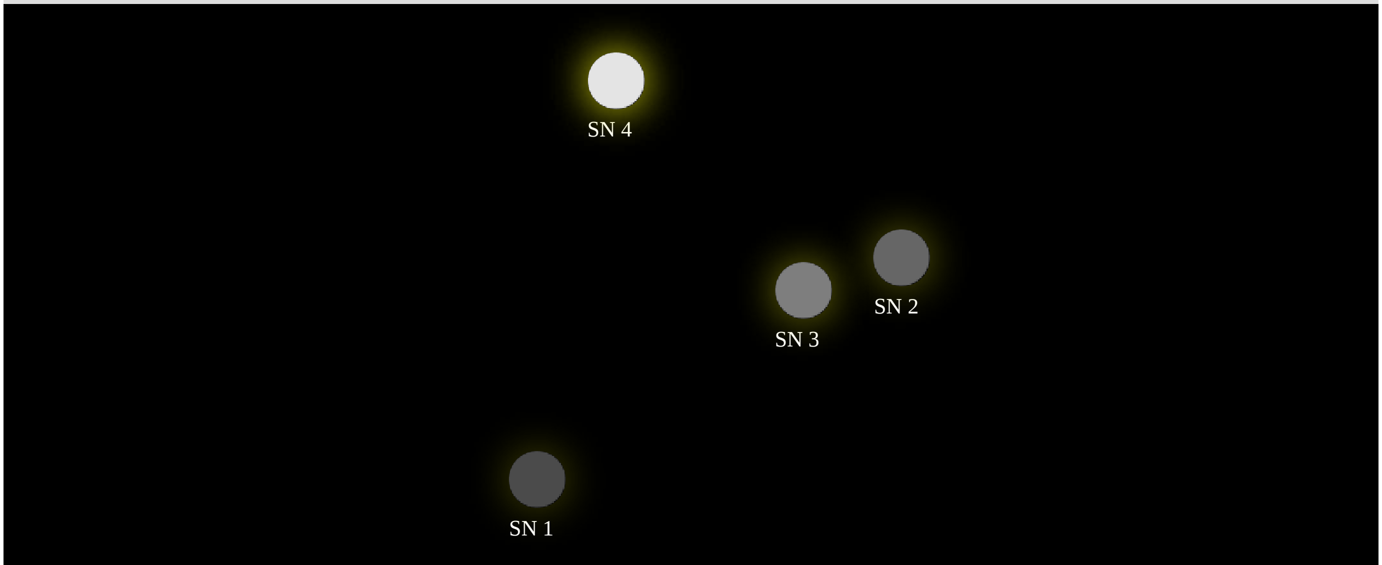
18.18: ch6pg3mainsequencematchingforclustersofdifferentdistances



This page titled 18.18: ch6pg3mainsequencematchingforclustersofdifferentdistances is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.19: ch6pg3standardcandle

reset



Reference Supernova:

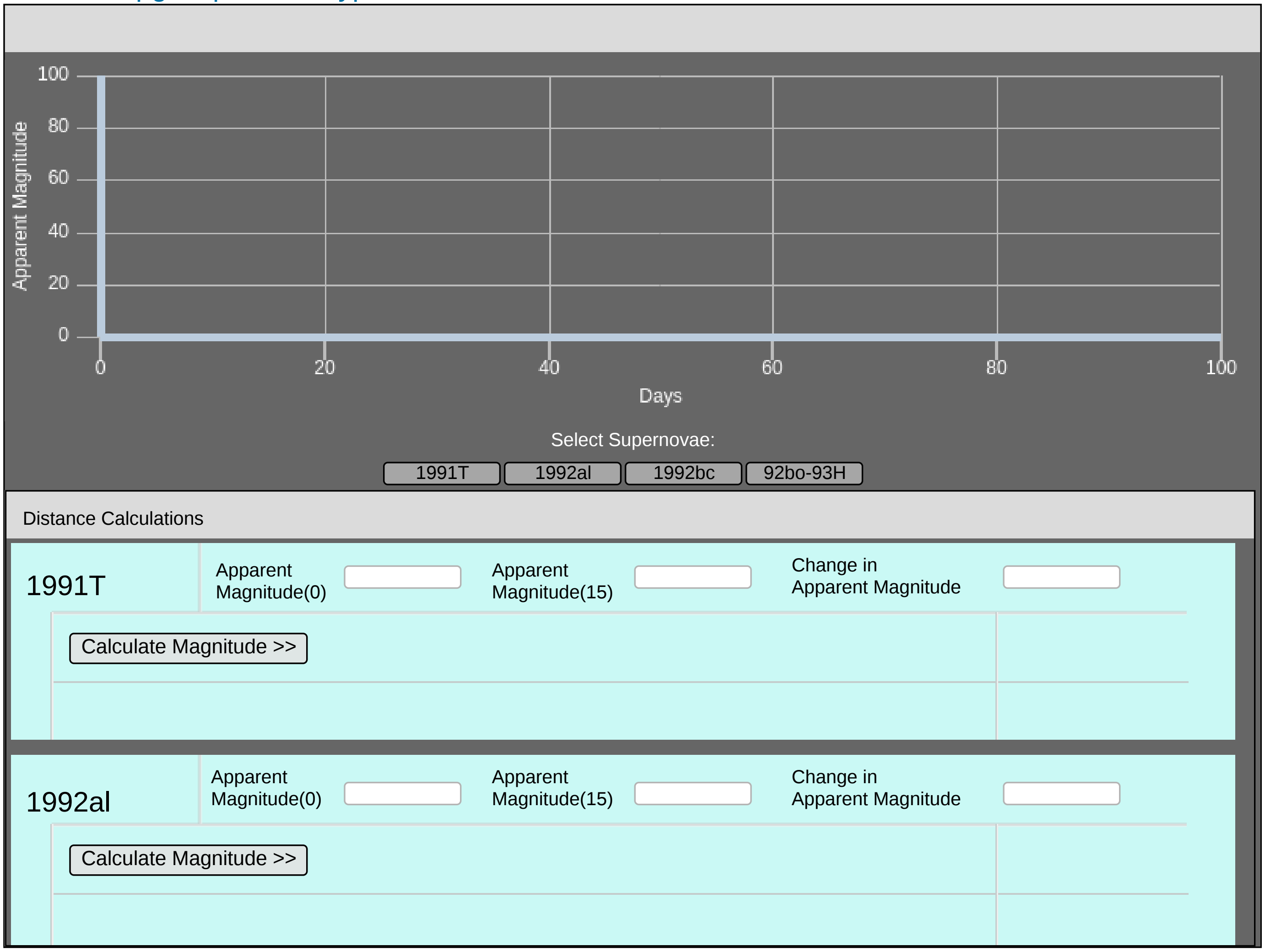
Type 1A Supernova
Flux: 100 w/m^2

SN1 is select times as far as the baseline	SN2 is select times as far as the baseline
SN3 is select times as far as the baseline	SN4 is select times as far as the baseline

Submit

This page titled 18.19: ch6pg3standardcandle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.20: ch6pg3supernovaetypelaanddistances

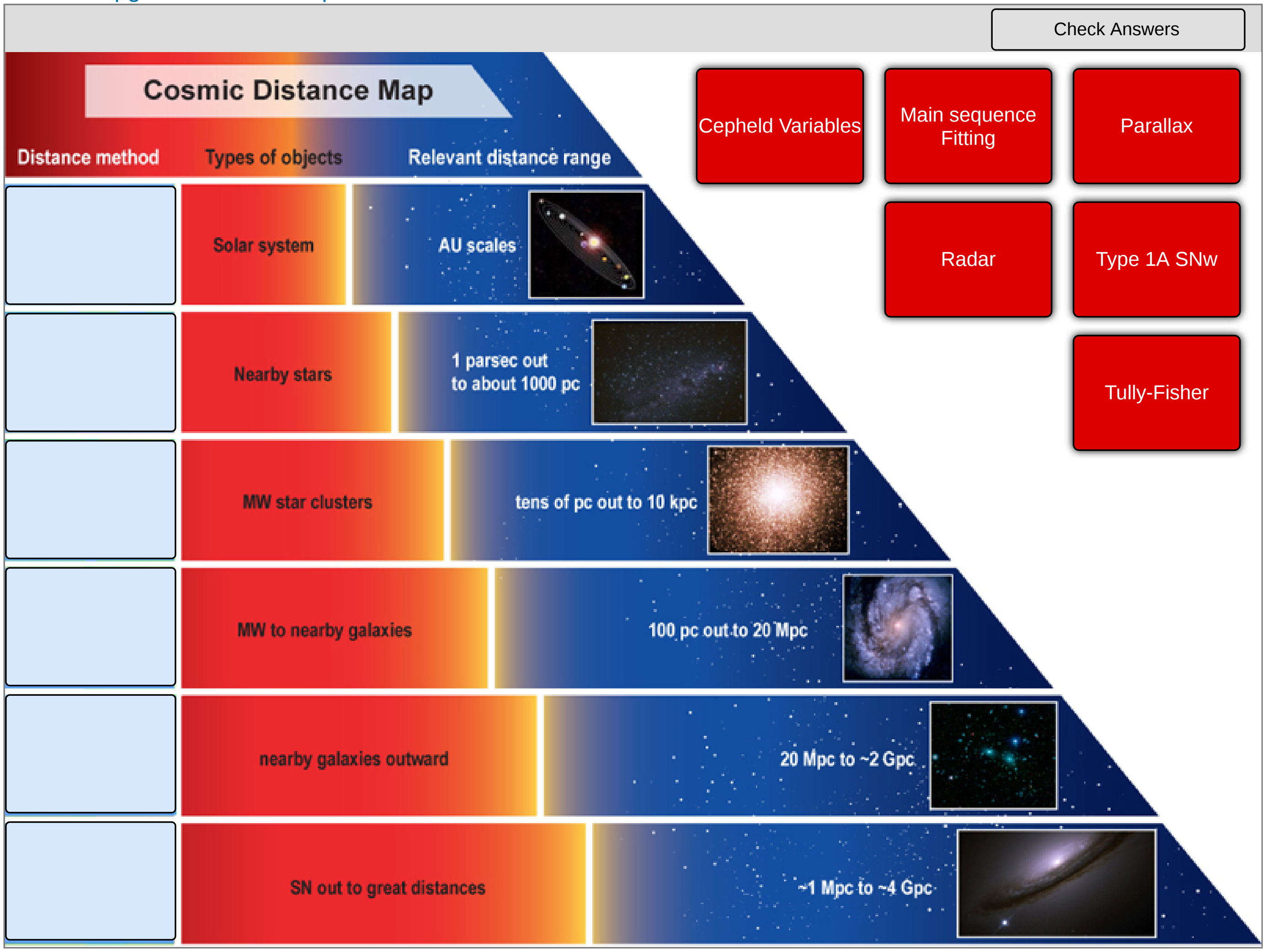


This page titled 18.20: ch6pg3supernovaetypelaanddistances is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

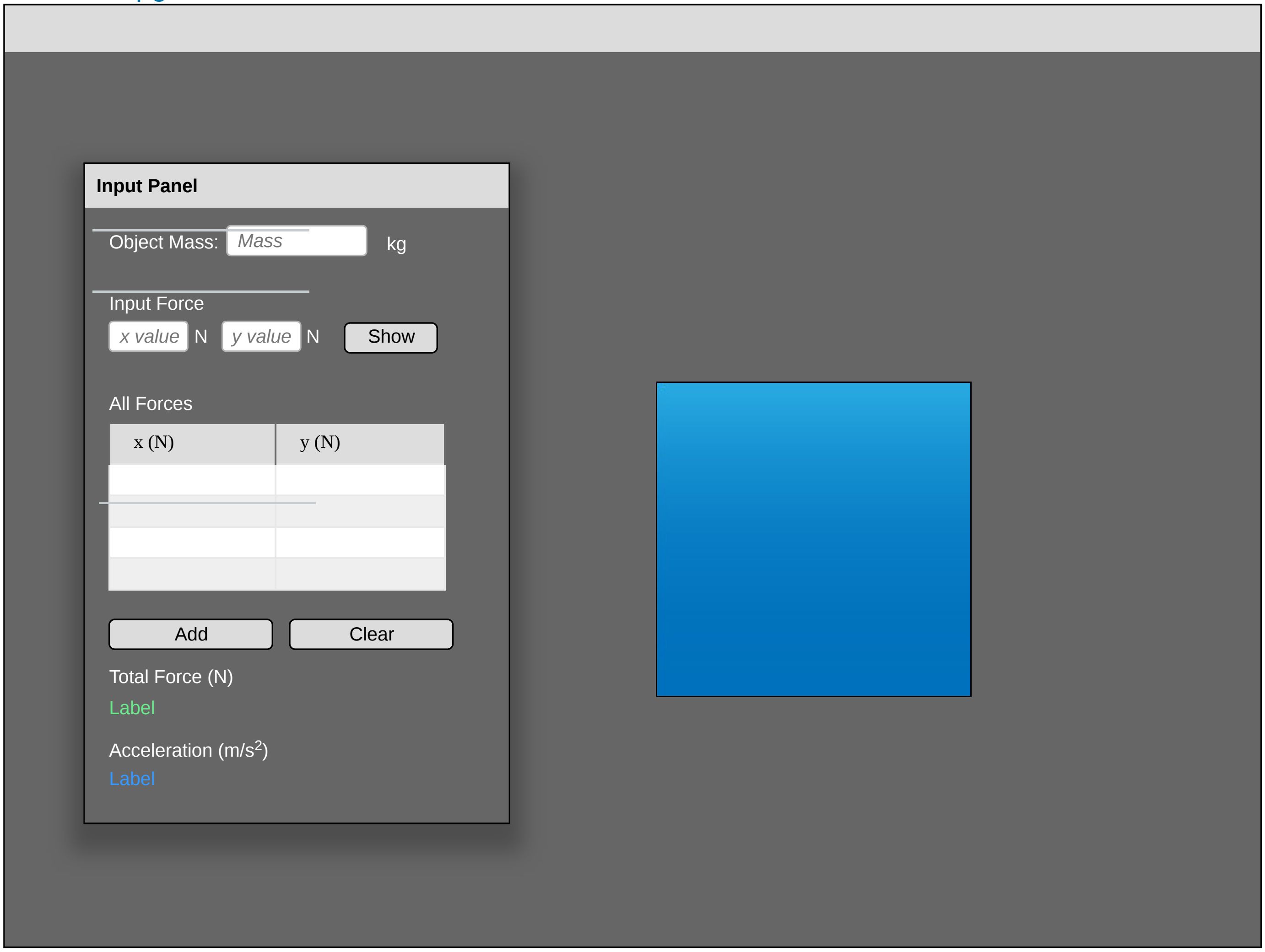
18.21: ch6pg3tullyfisher

This page titled 18.21: ch6pg3tullyfisher is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.22: ch6pg4matchthetechniquetothedistance

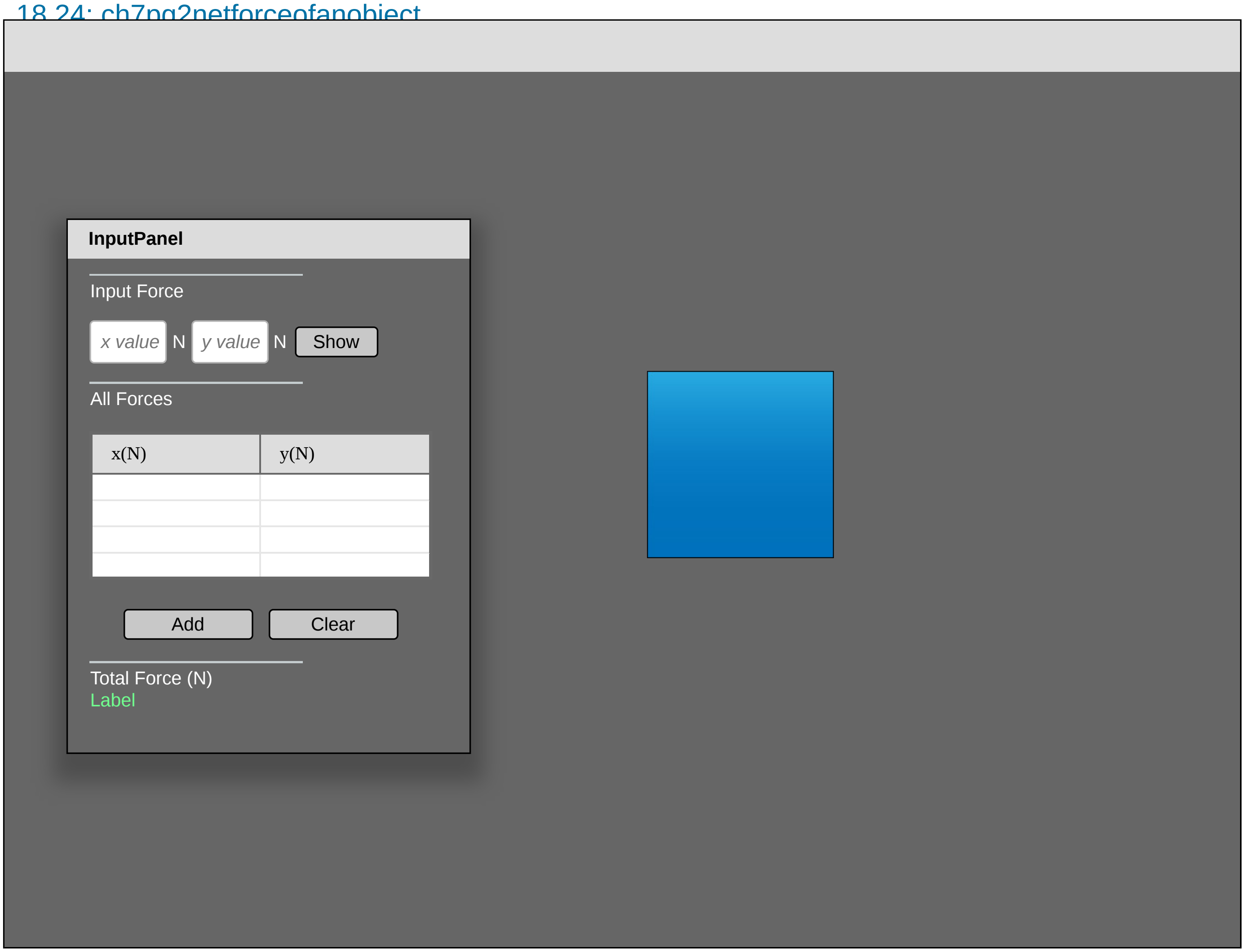


18.23: ch7pg2accelerationandmass



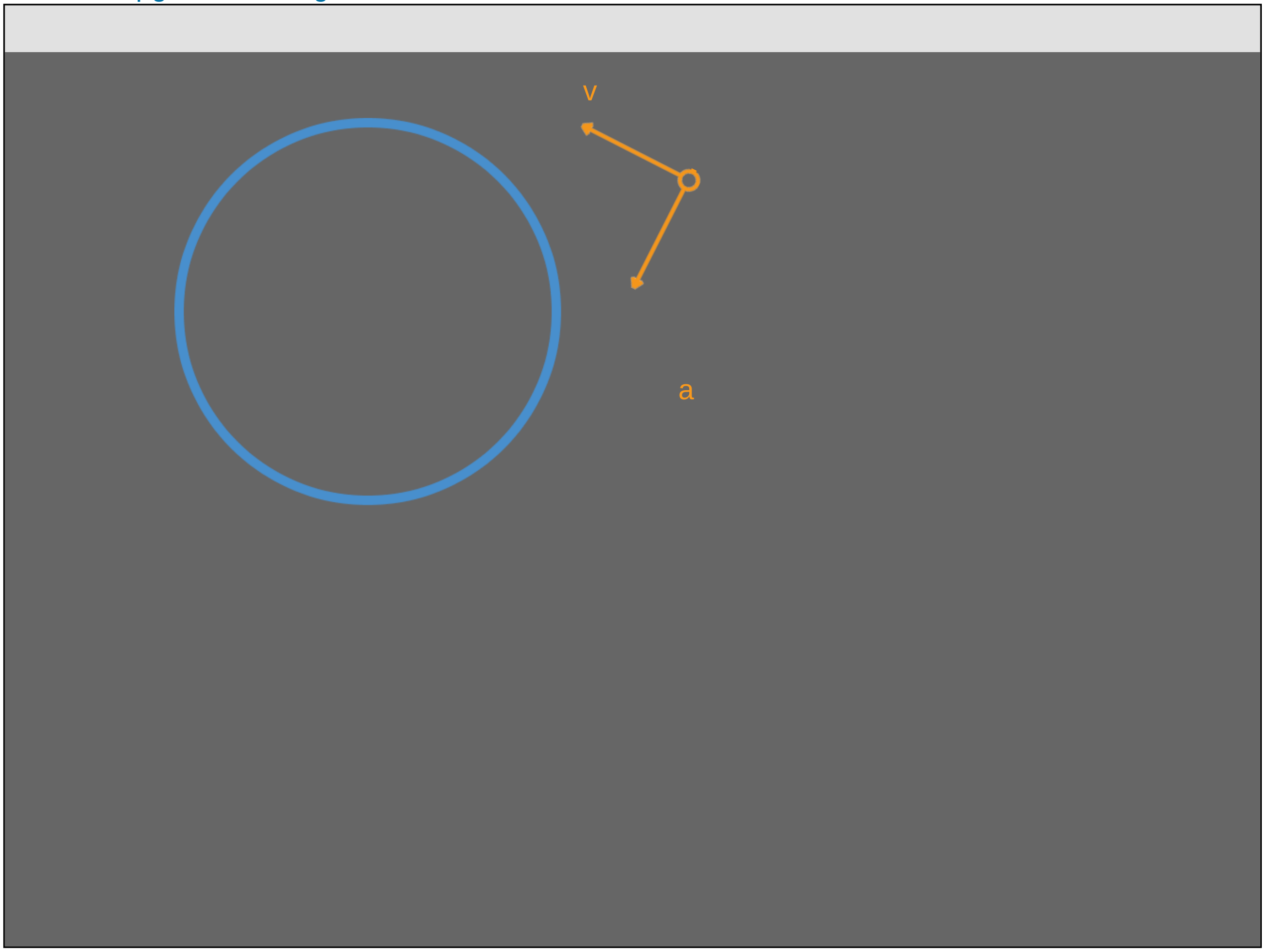
This page titled 18.23: ch7pg2accelerationandmass is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.24: ch7pg2netforceofanobject



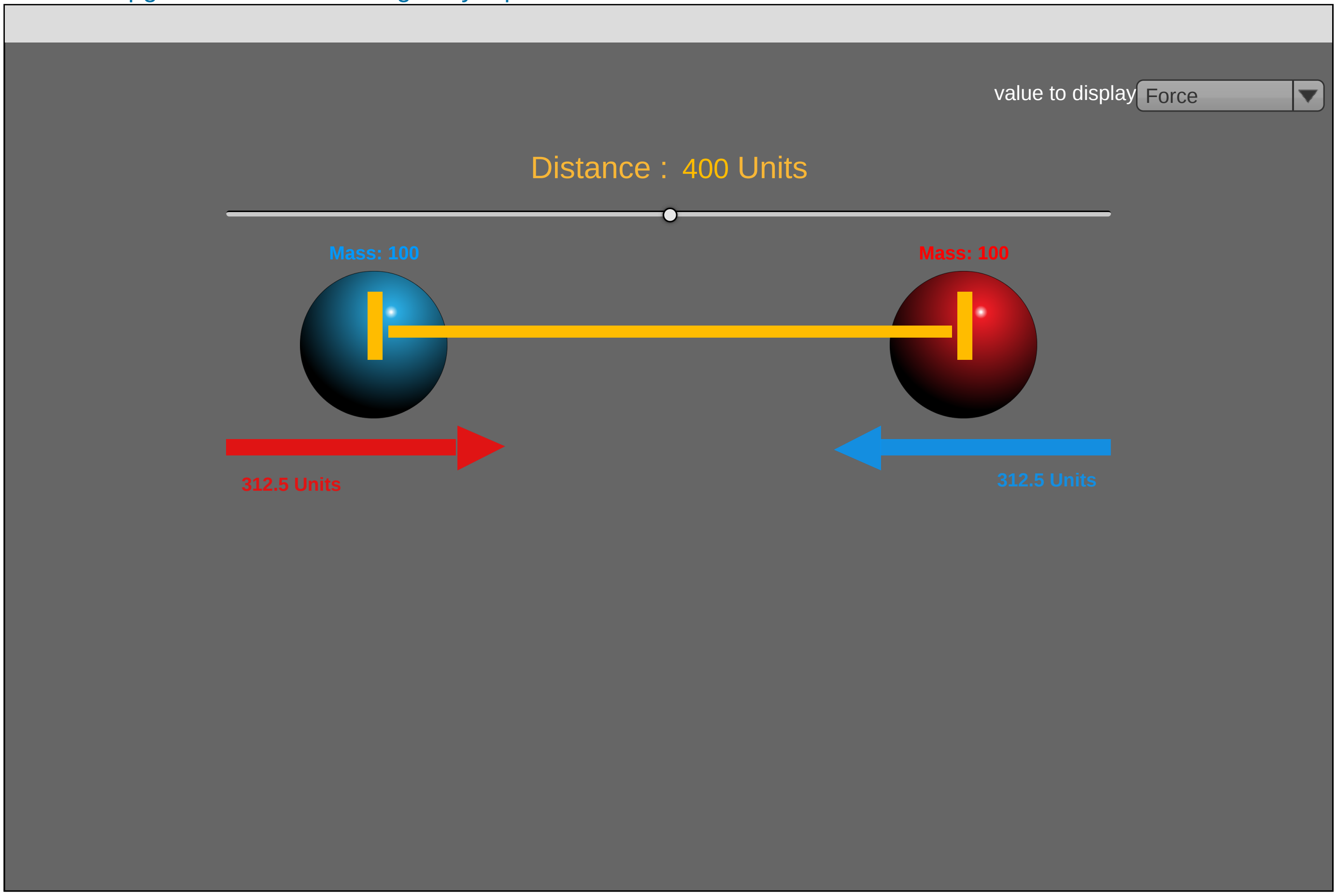
This page titled 18.24: ch7pg2netforceofanobject is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.25: ch7pg3animatedfigure75



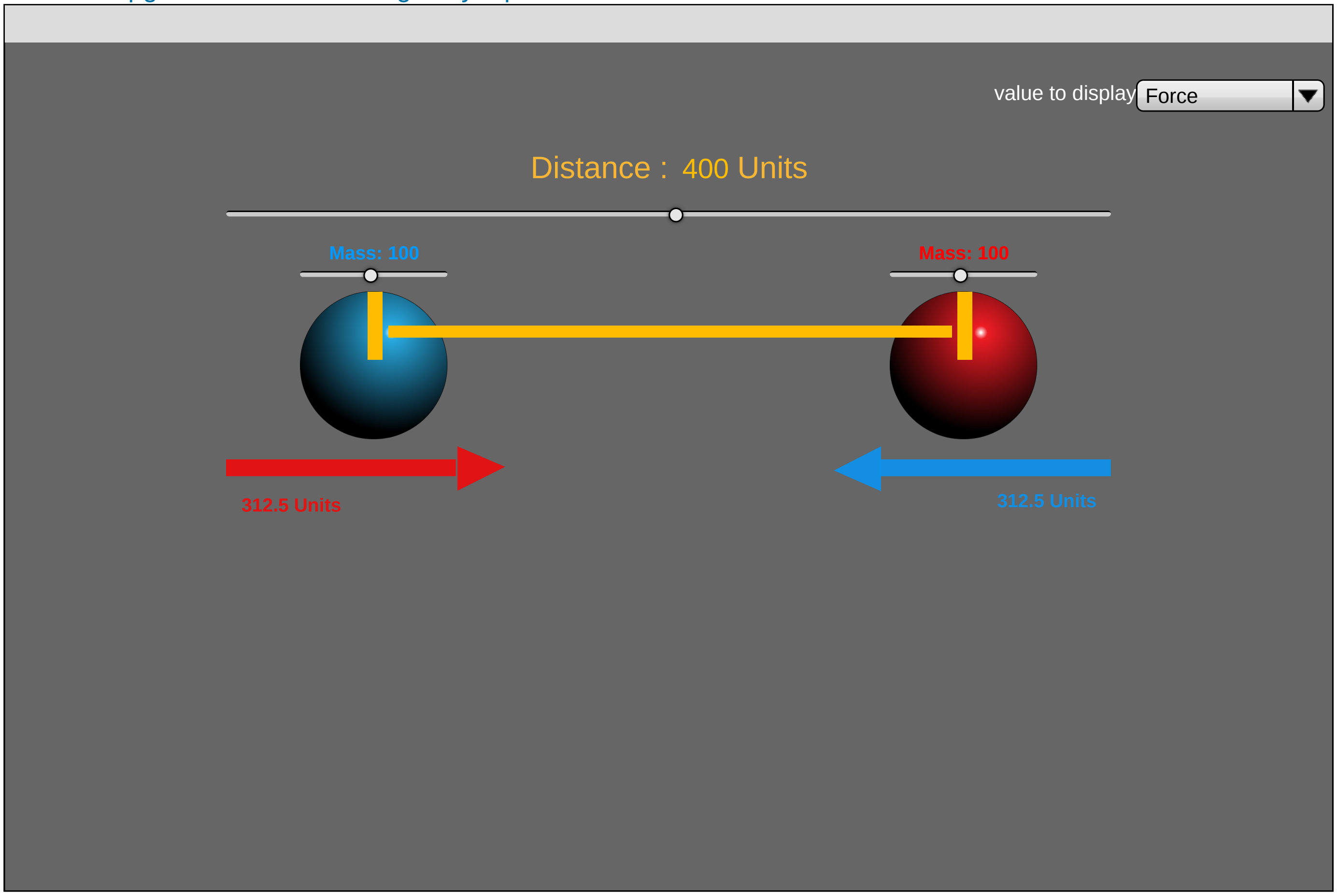
This page titled 18.25: ch7pg3animatedfigure75 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.26: ch7pg3howdoestheforceofgravitydependondistance



This page titled 18.26: ch7pg3howdoestheforceofgravitydependondistance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.27: ch7pg3howdoestheforceofgravitydependonmass



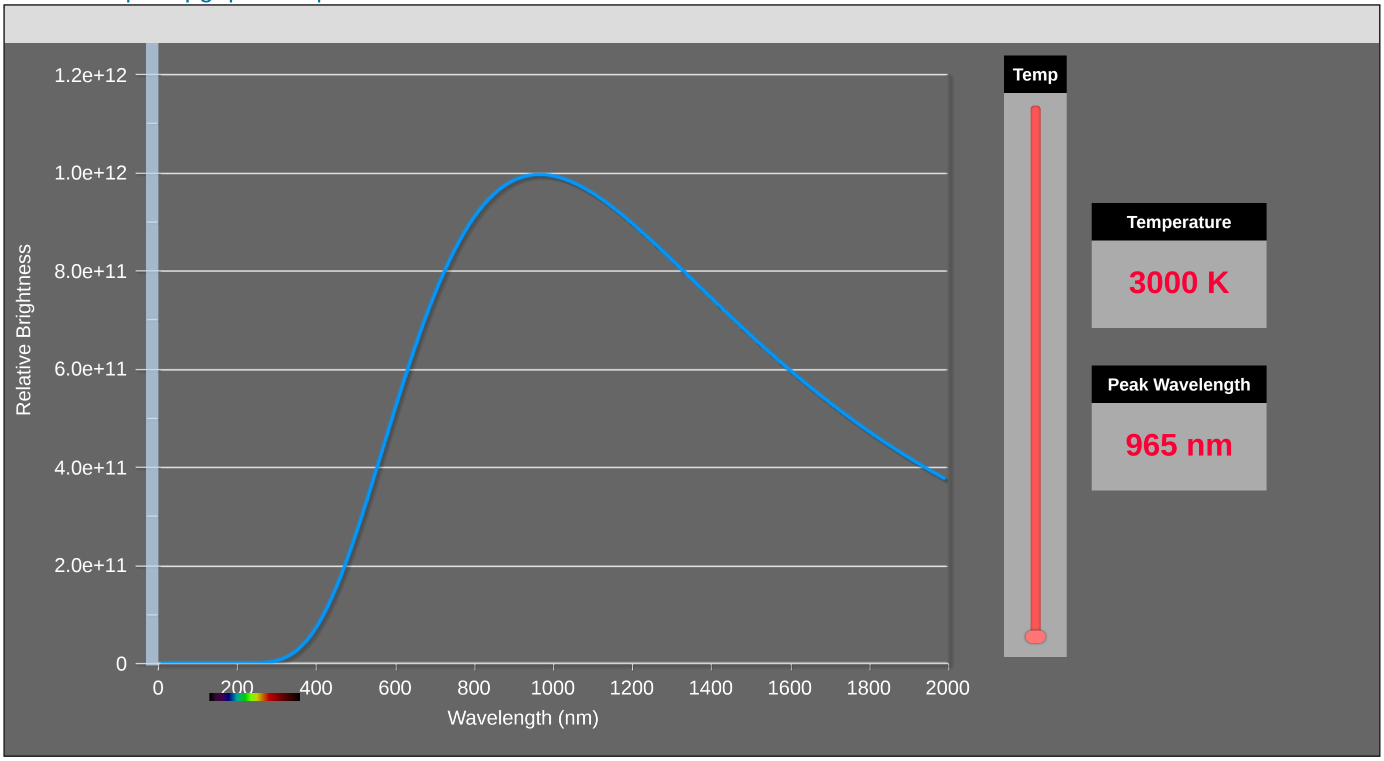
This page titled 18.27: ch7pg3howdoestheforceofgravitydependonmass is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.28: ch7wiu7thegalileanmoonsofjupiter

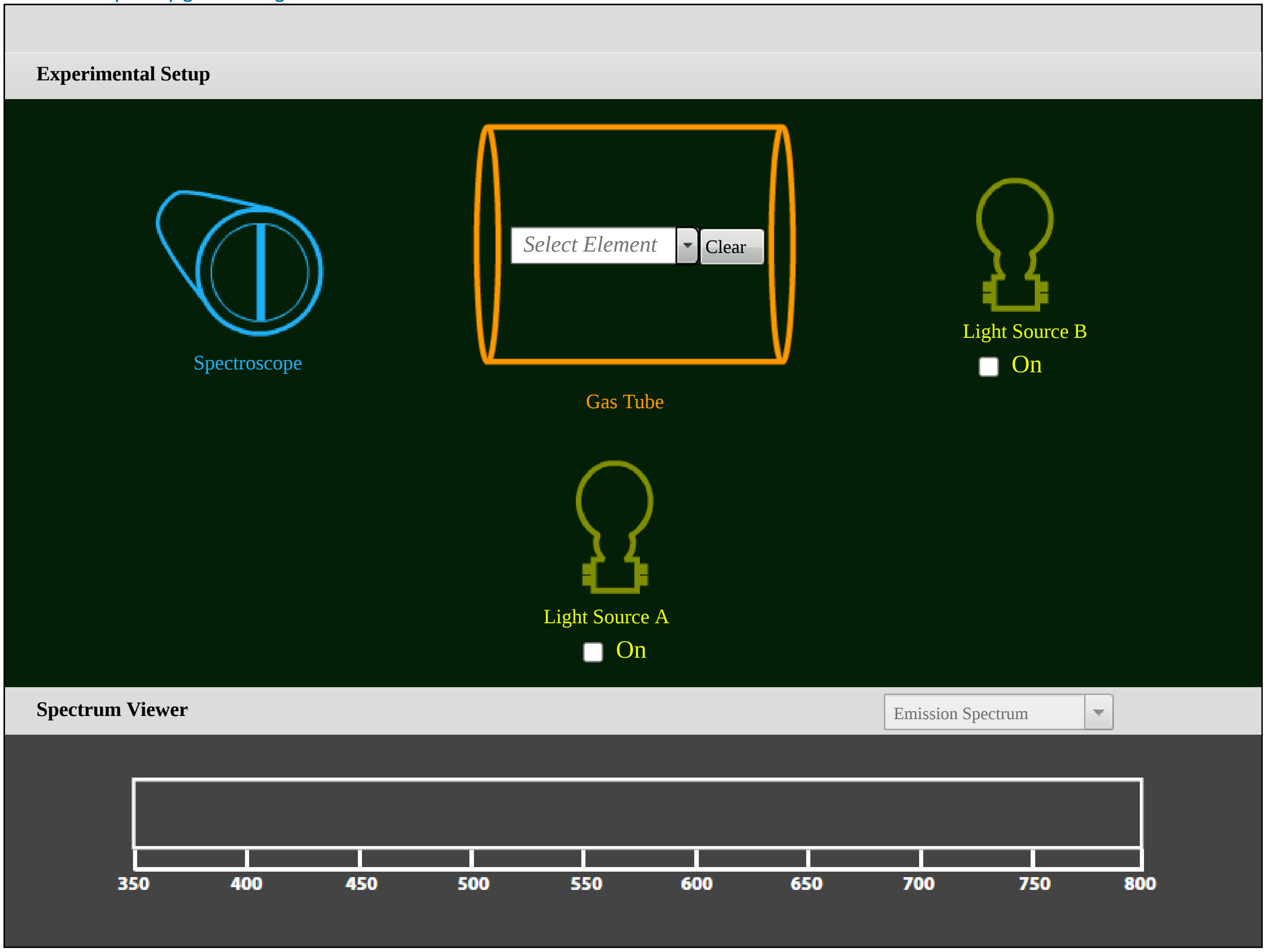
This page titled 18.28: ch7wiu7thegalileanmoonsofjupiter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.29: chapter2pg1photonraceroundone

18.30: chapter2pg5planckspectrum

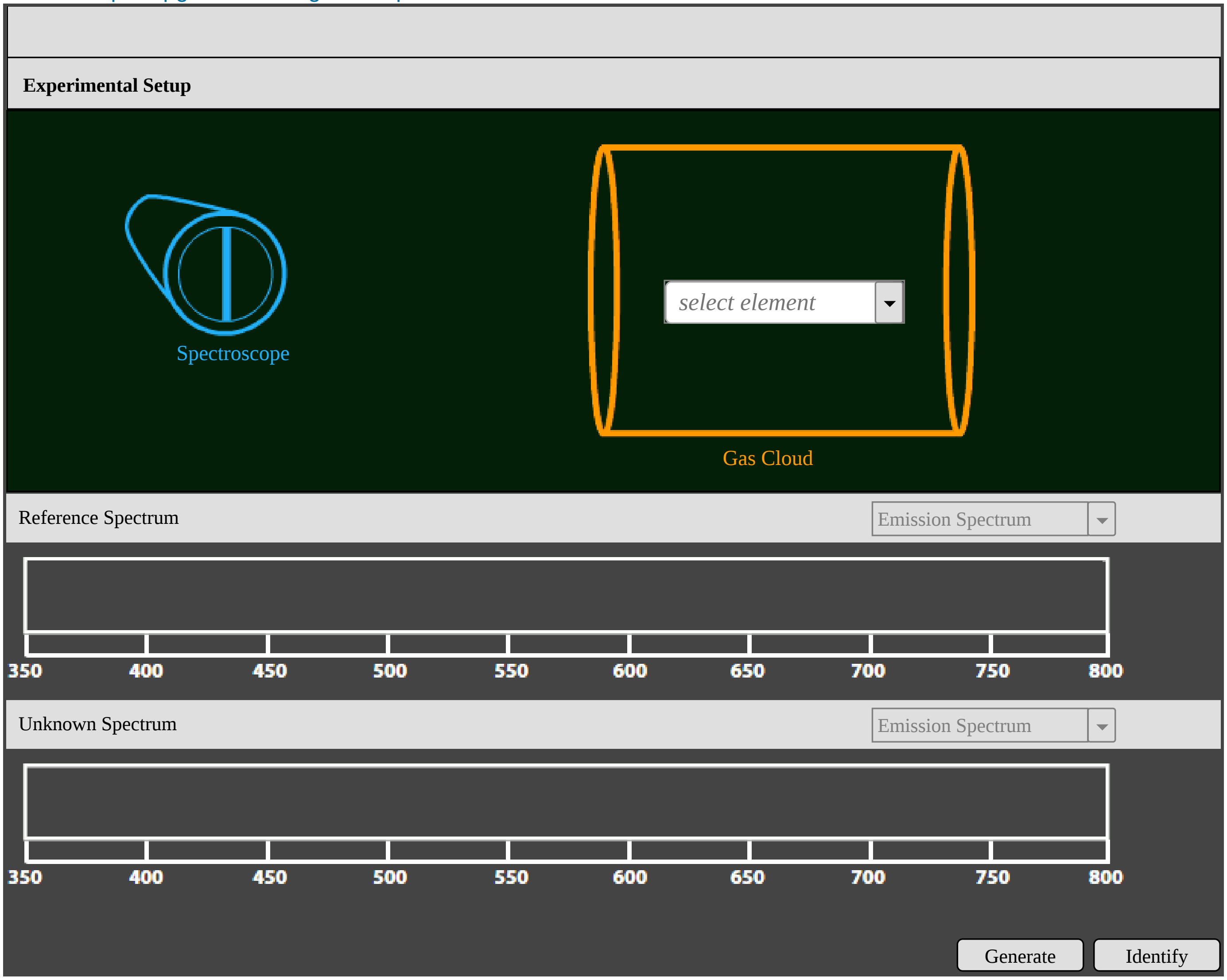


This page titled 18.30: chapter2pg5planckspectrum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.



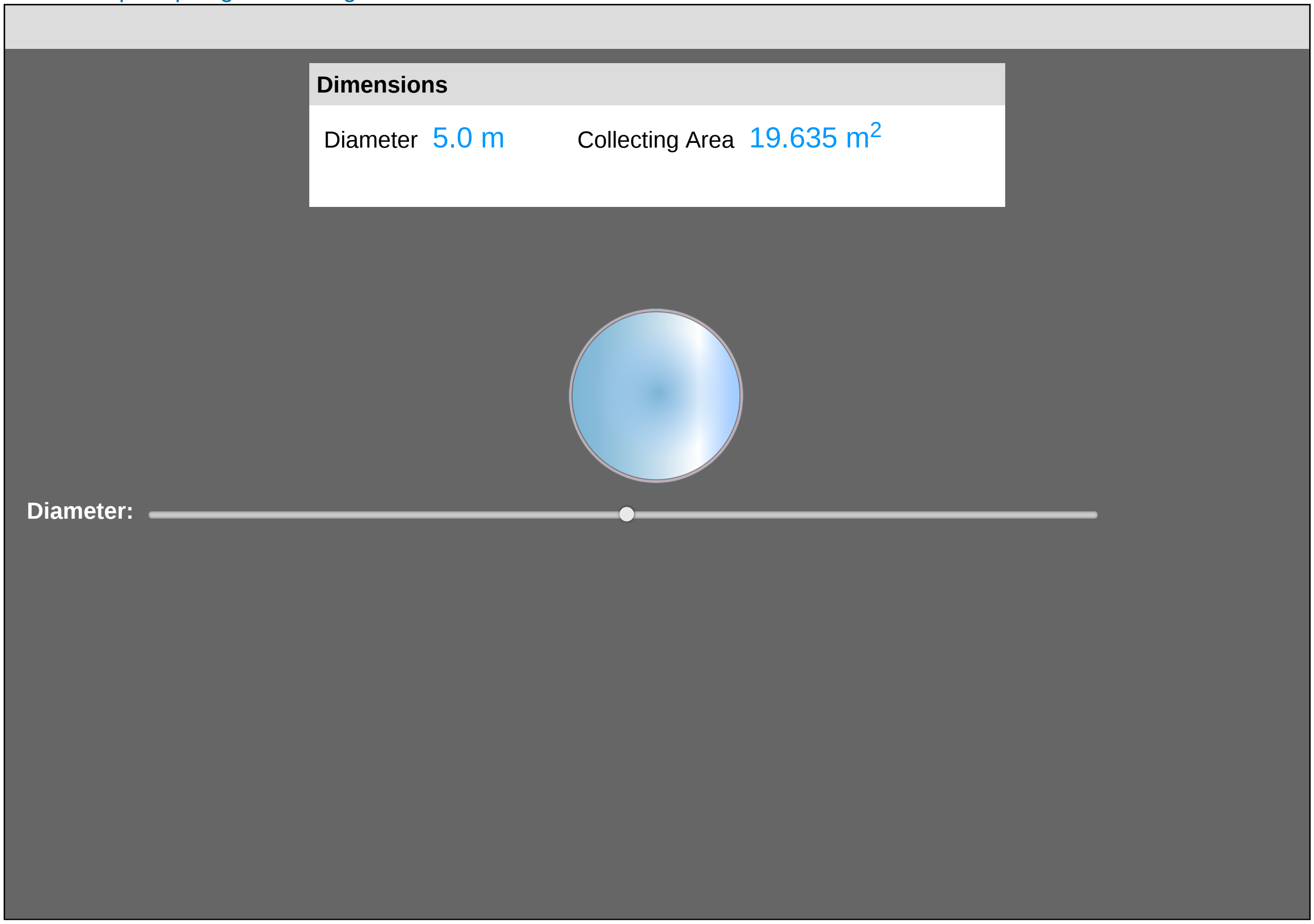
https://phys.libretexts.org/@go/page/32325

18.32: chapter2pg7determiningthecomposition



This page titled 18.32: chapter2pg7determiningthecomposition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.33: chapter3pt1lightcollectingarea



This page titled 18.33: chapter3pt1lightcollectingarea is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.34: chapter3pt4afindingthesourceandsubtractingthebackground

6	1	4	2	4	2	2	7	13	13	3	9	5	4	2
7	9	0	11	8	8	6	7	14	15	12	6	3	8	2
7	9	0	11	8	8	6	7	14	15	12	6	3	8	2
10	7	16	0	7	13	9	18	8	14	11	13	8	11	9
1	10	16	0	7	13	9	18	8	14	11	13	8	11	9
7	5	90	19	3	14	11	18	10	11	13	8	4	7	15
5	0	13	12	9	24	32	37	28	21	3	8	7	4	11
6	4	2	10	22	39	73	105	84	20	9	7	8	15	10
1	6	13	16	13	35	83	140	105	42	9	11	5	9	0
11	12	1	13	23	35	75	115	91	30	12	9	9	9	4
8	10	12	8	18	21	41	57	46	21	8	7	0	0	10
7	11	12	12	4	11	21	42	20	12	5	7	4	5	7
10	2	2	9	2	9	15	16	16	6	4	0	7	1	10
7	9	7	4	11	3	9	10	8	6	7	11	3	11	8
1	4	12	8	8	13	11	9	8	4	10	8	16	19	8
0	10	30	4	14	5	0	3	20	17	16	9	14	11	0

Select Pixels

Source Pixels:

Number of pixels selected:

0

Sum of all selected Photons:

0

Background Pixels:

Number of pixels selected:

0

Sum of all selected Photons:

0

Sum of Source Pixels: 0 - **Average Background Count** 0 X **Number of Source Pixels** 0 = **Star Intensity** 0

This page titled 18.34: chapter3pt4afindingthesourceandsubtractingthebackground is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

6	1	4	2	4	2	2	7	13	13	3	9	5	4	2
7	9	0	11	8	8	6	7	14	15	12	6	3	8	2
7	9	0	11	8	8	6	7	14	15	12	6	3	8	2
10	7	16	0	7	13	9	18	8	14	11	13	8	11	9
1	10	16	0	7	13	9	18	8	14	11	13	8	11	9
7	5	90	19	3	14	11	18	10	11	13	8	4	7	15
5	0	13	12	9	24	32	37	28	21	3	8	7	4	11
6	4	2	10	22	39	73	105	84	20	9	7	8	15	10
1	6	13	16	13	35	83	140	105	42	9	11	5	9	0
11	12	1	13	23	35	75	115	91	30	12	9	9	9	4
8	10	12	8	18	21	41	57	46	21	8	7	0	0	10
7	11	12	12	4	11	21	42	20	12	5	7	4	5	7
10	2	2	9	2	9	15	16	16	6	4	0	7	1	10
7	9	7	4	11	3	9	10	8	6	7	11	3	11	8
1	4	12	8	8	13	11	9	8	4	10	8	16	19	8
0	10	30	4	14	5	0	3	20	17	16	9	14	11	0

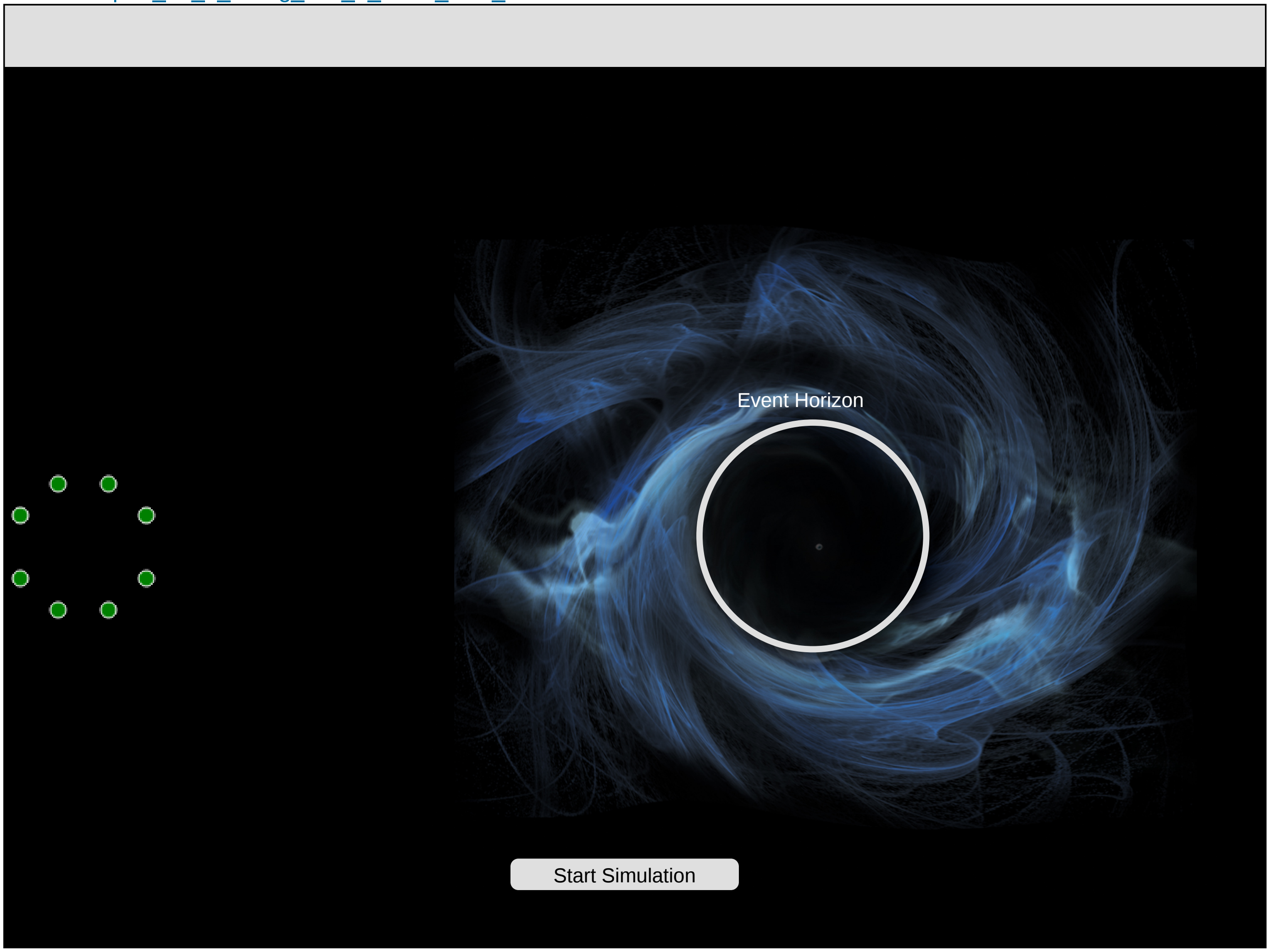
Remove Color

Min Value	Max Value	Color

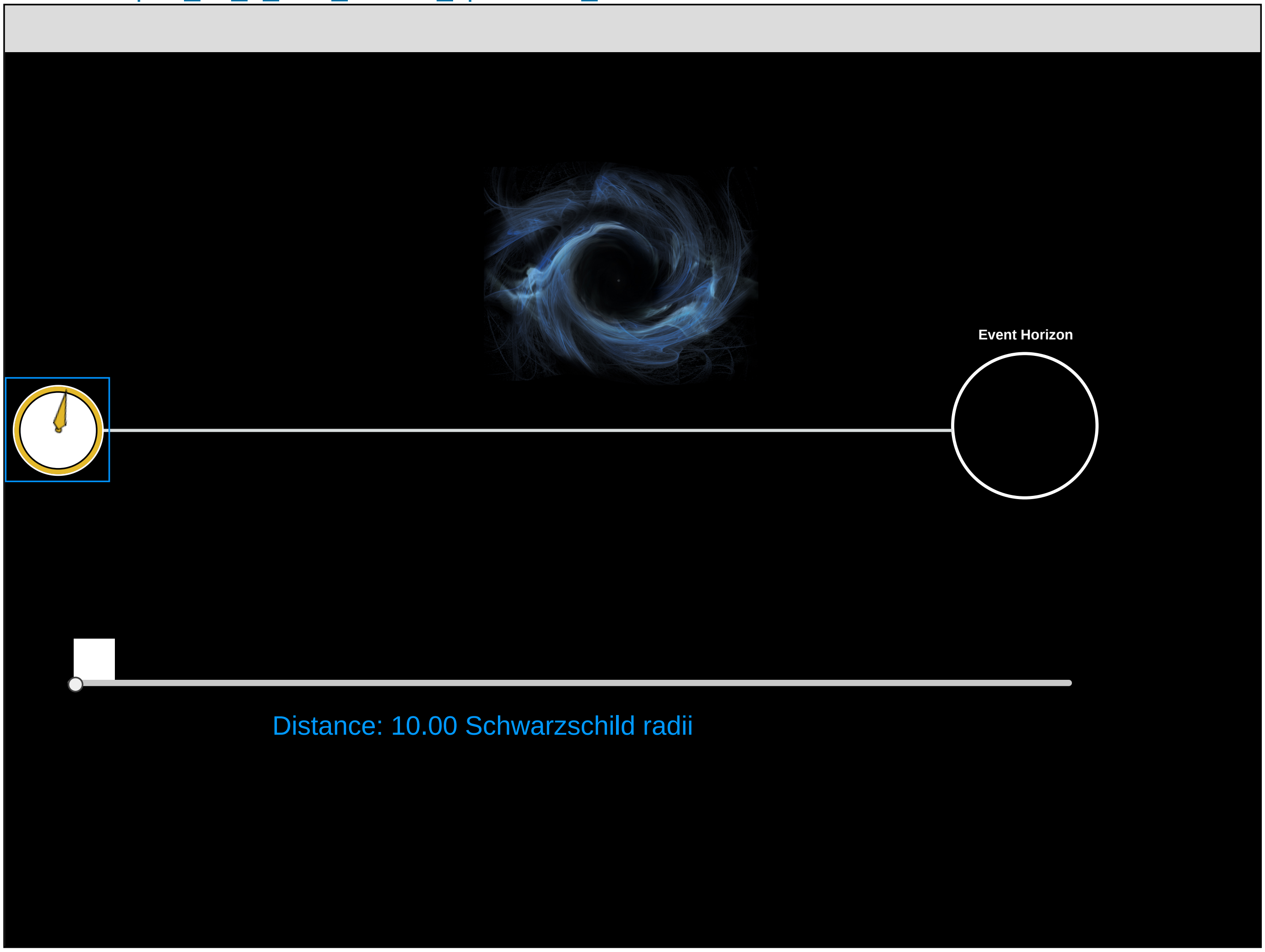
Add Color

T lorizingccdimages is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn C

18.36: chapter_11_2_falling_into_a_black_hole_tidalforces

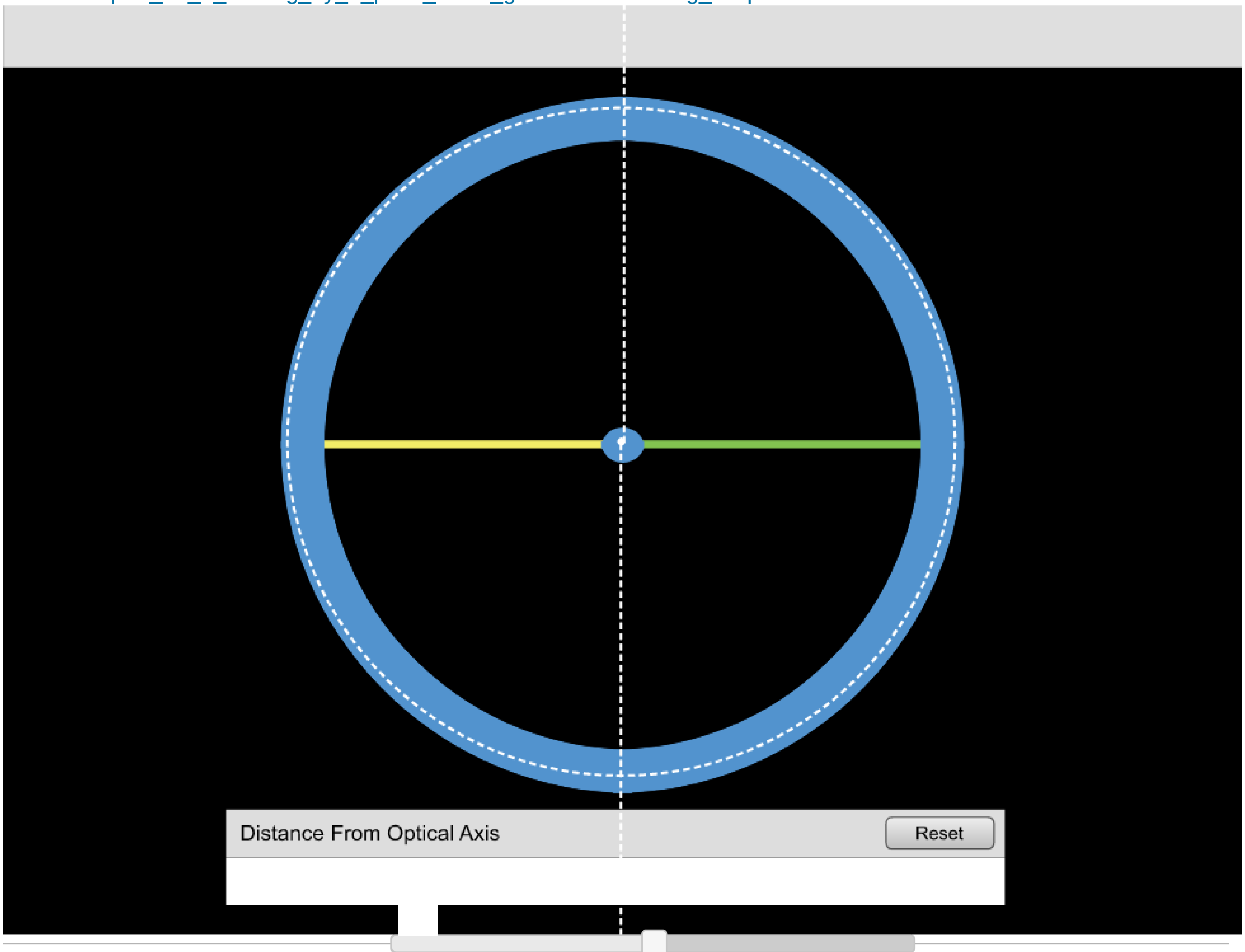


18.37: chapter_11_2_time_dilation_spacetime_dialation



This page titled 18.37: chapter_11_2_time_dilation_spacetime_dialation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.38: chapter_12_2_lensing_by_a_point_mass_gravitationallensing_simple

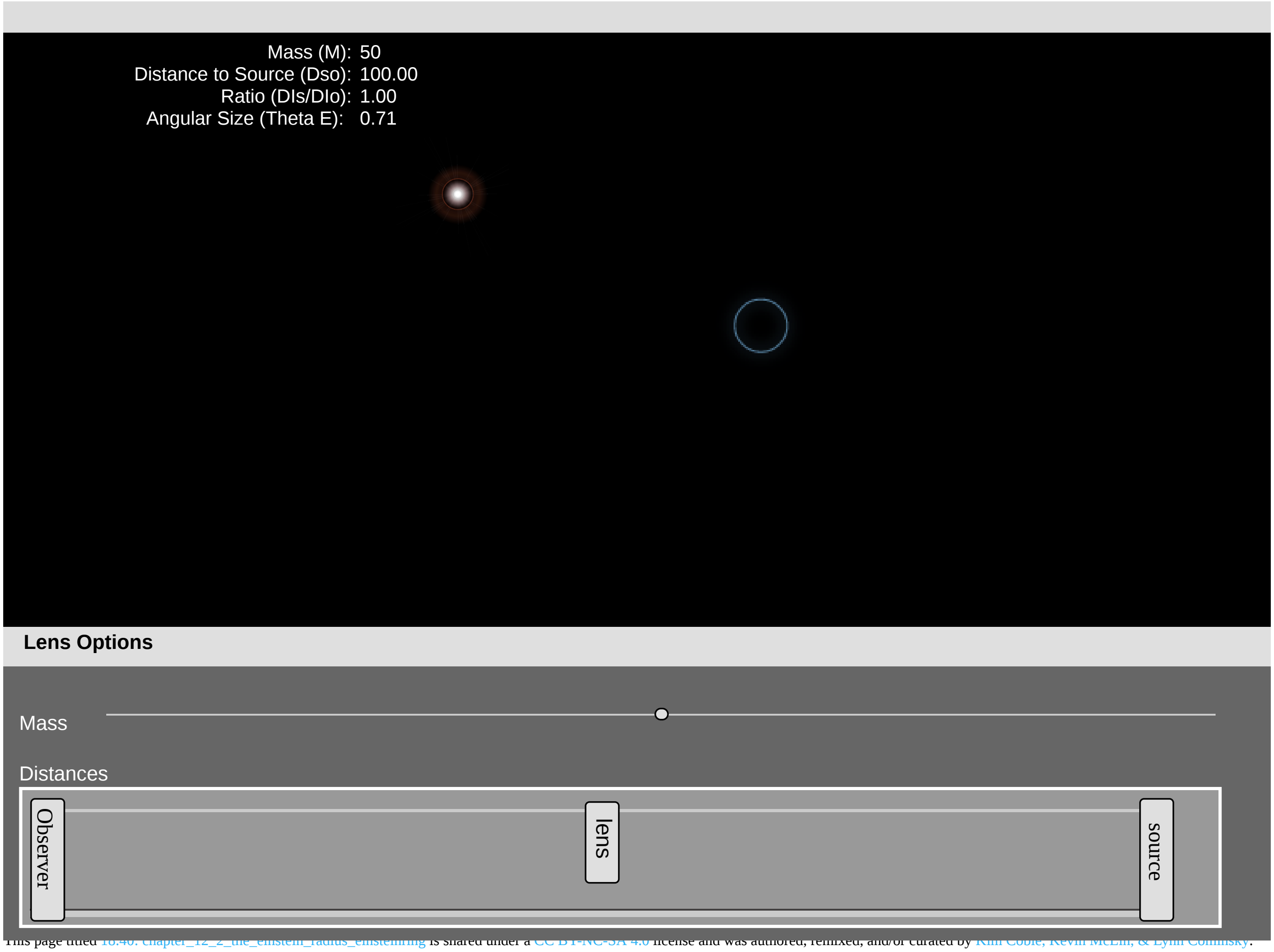


This page titled 18.38: chapter_12_2_lensing_by_a_point_mass_gravitationallensing_simple is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.39: chapter_12_2_macho_light_curve

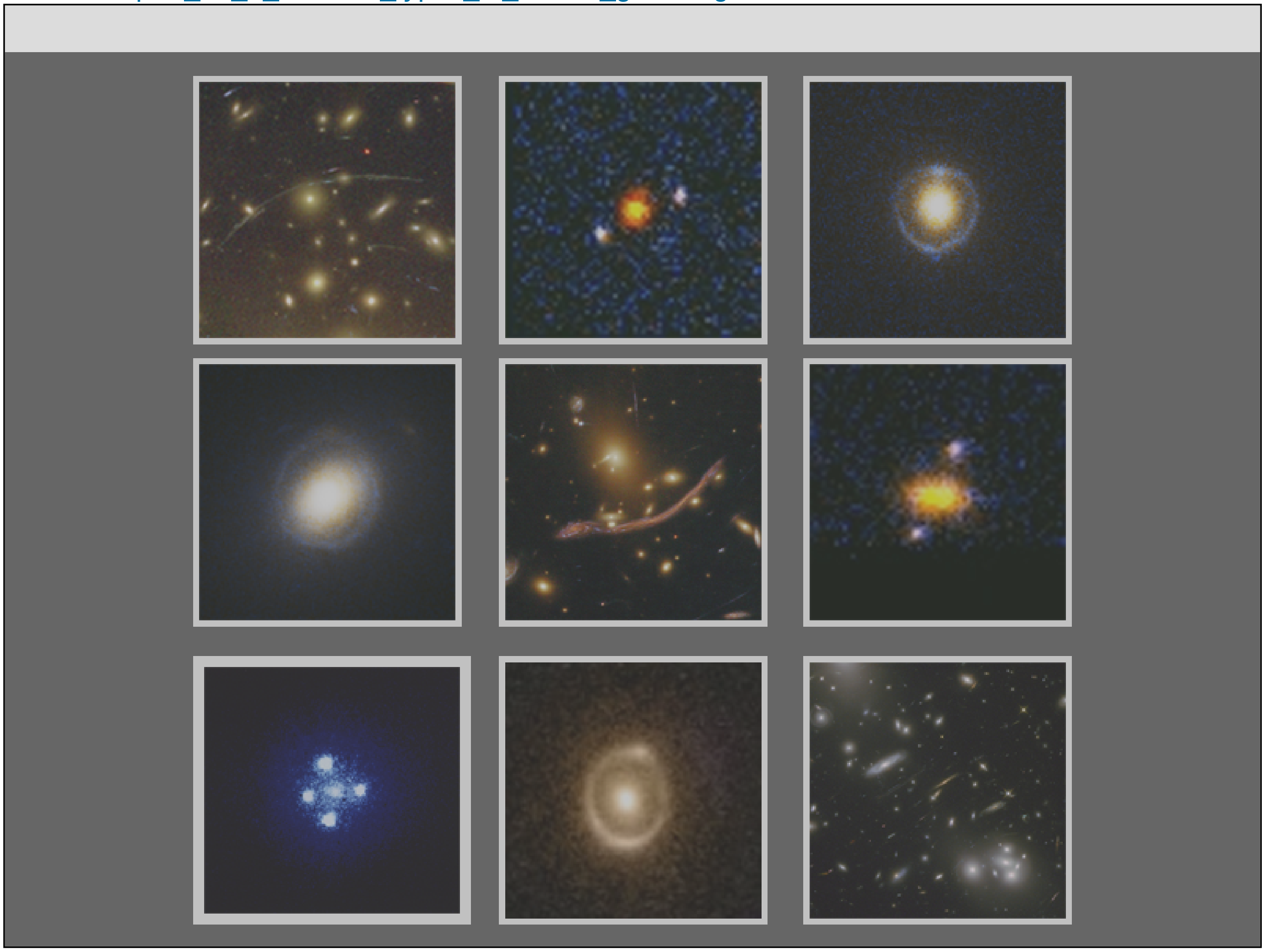
This page titled 18.39: chapter_12_2_macho_light_curve is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.40: chapter_12_2_the_einstein_radius_einsteinring



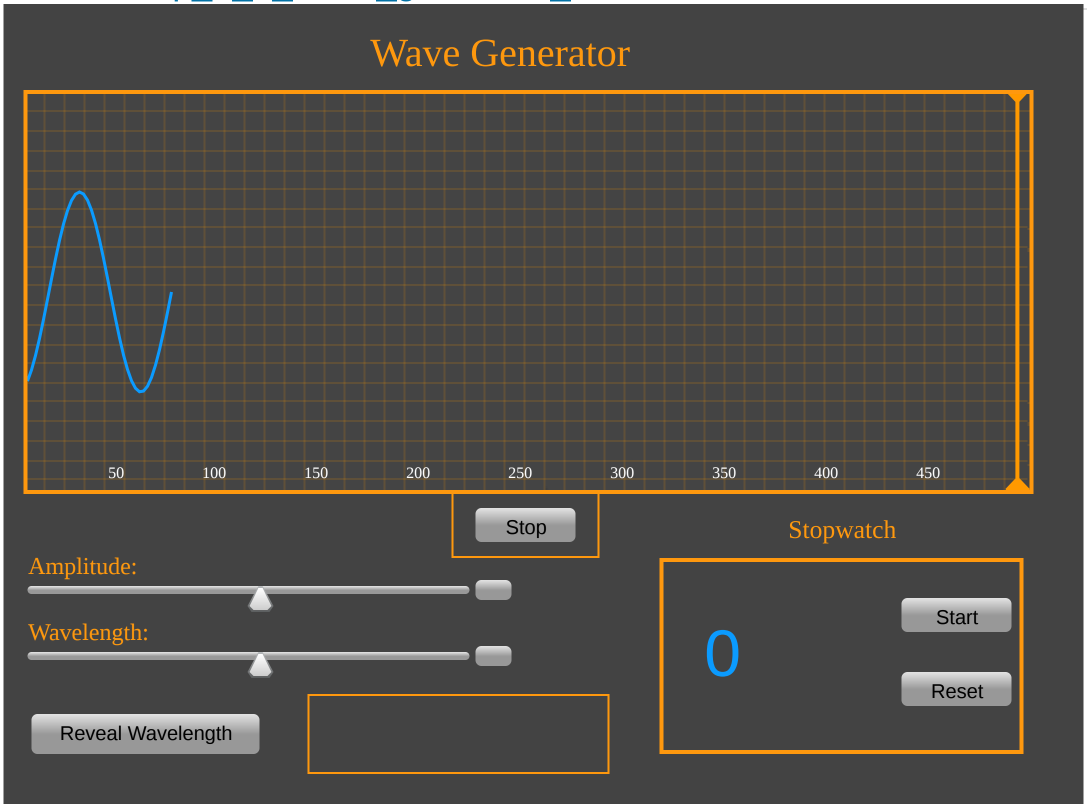
This page titled 18.40: chapter_12_2_the_einstein_radius_einsteinring is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.41: chapter_12_3_different_types_of_lenses_glbinning



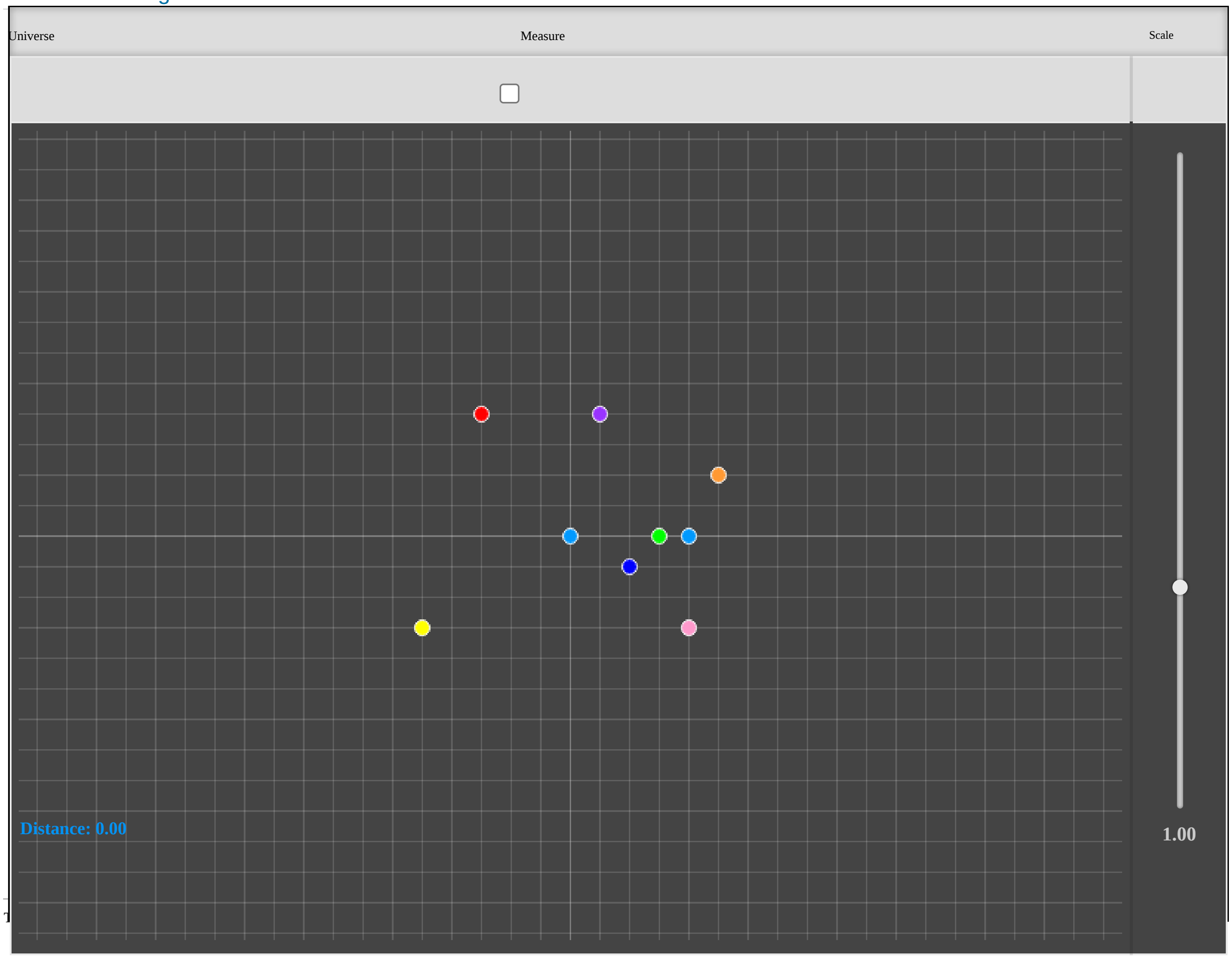
This page titled 18.41: chapter_12_3_different_types_of_lenses_glbinning is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.42: chap_2_1_wave_generator_NEW



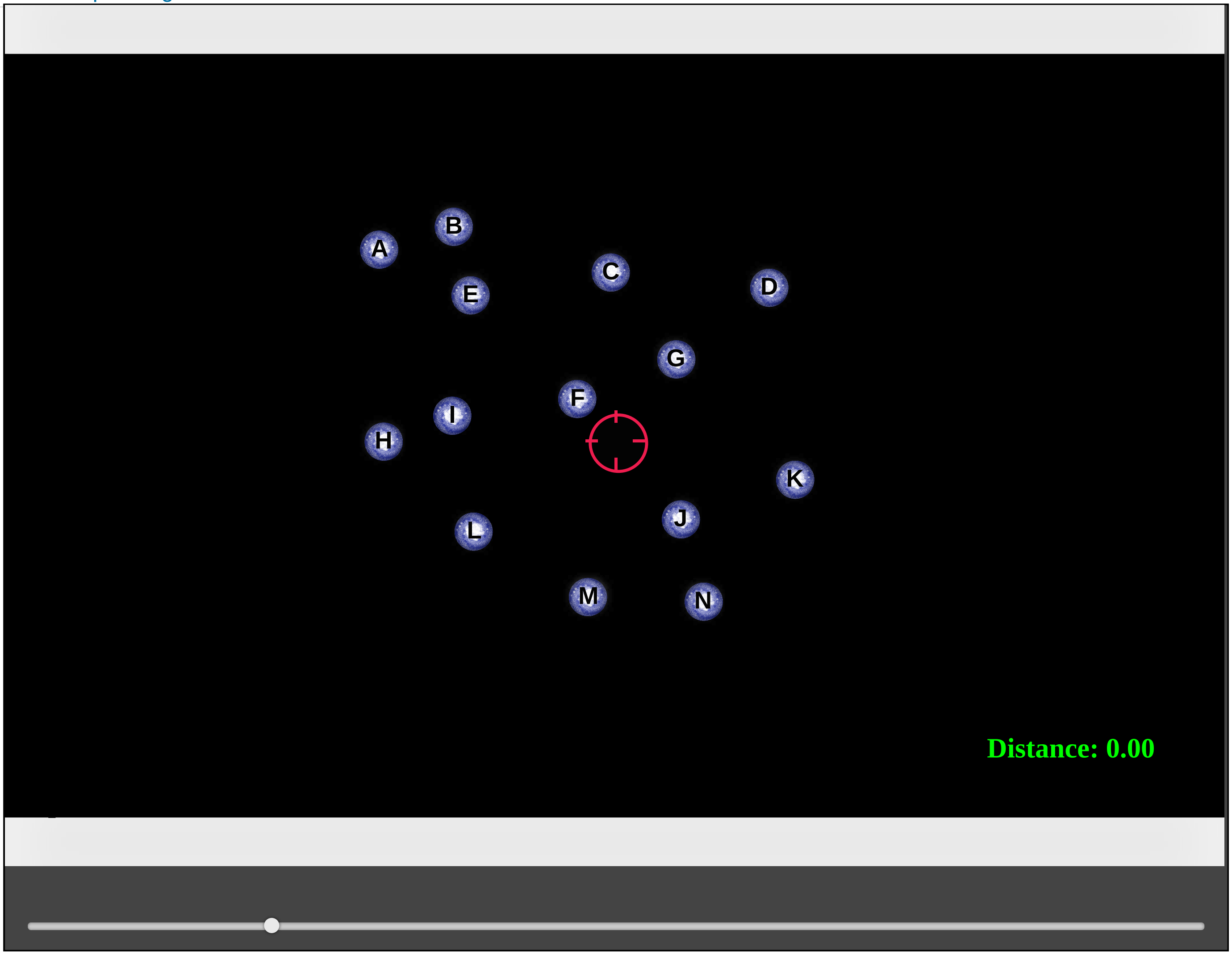
This page titled 18.42: chap_2_1_wave_generator_NEW is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.43: comovingcoords



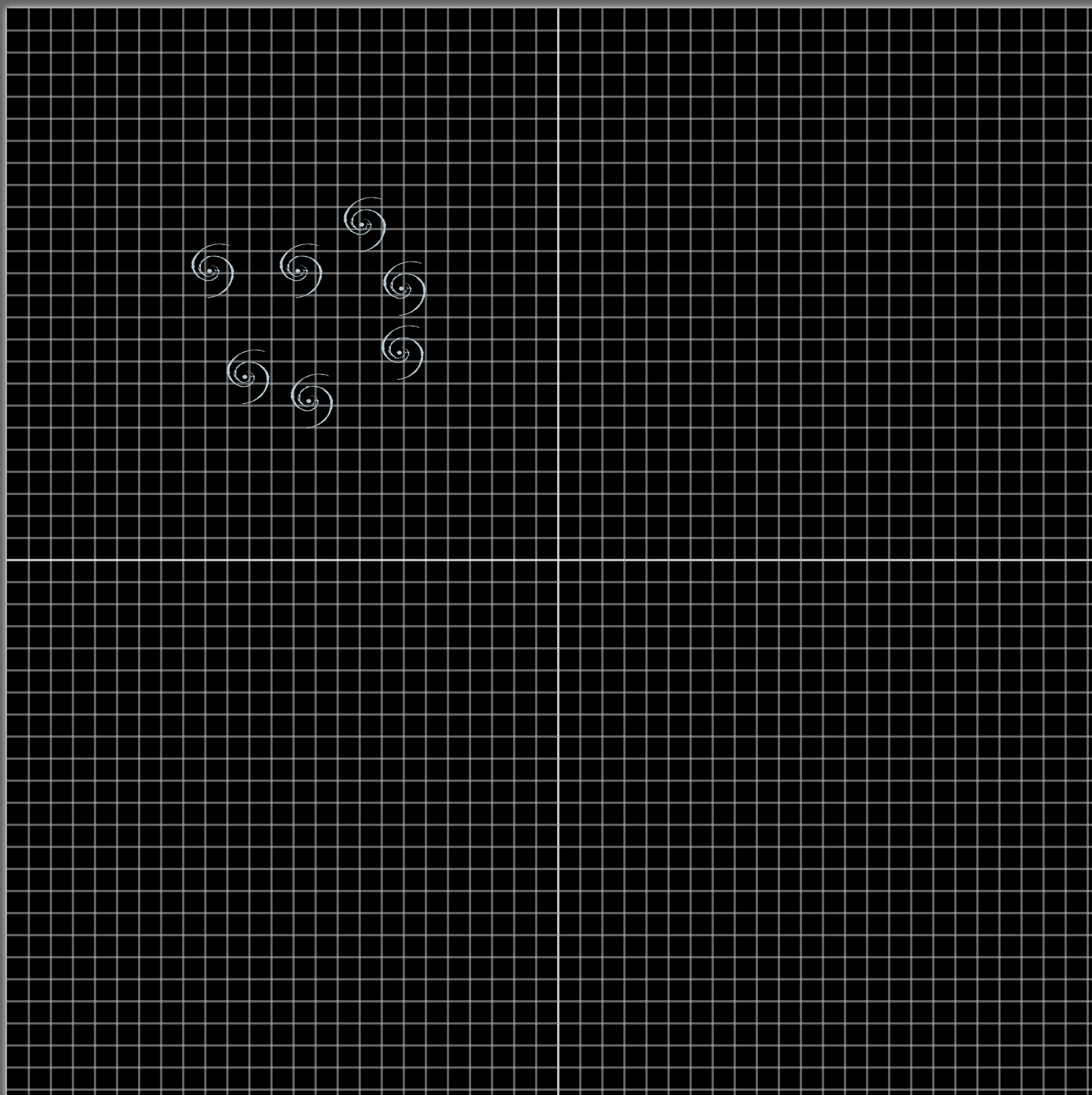
18.44: expandinguniverse1

Galaxies



This page titled 18.44: expandinguniverse1 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

Universe



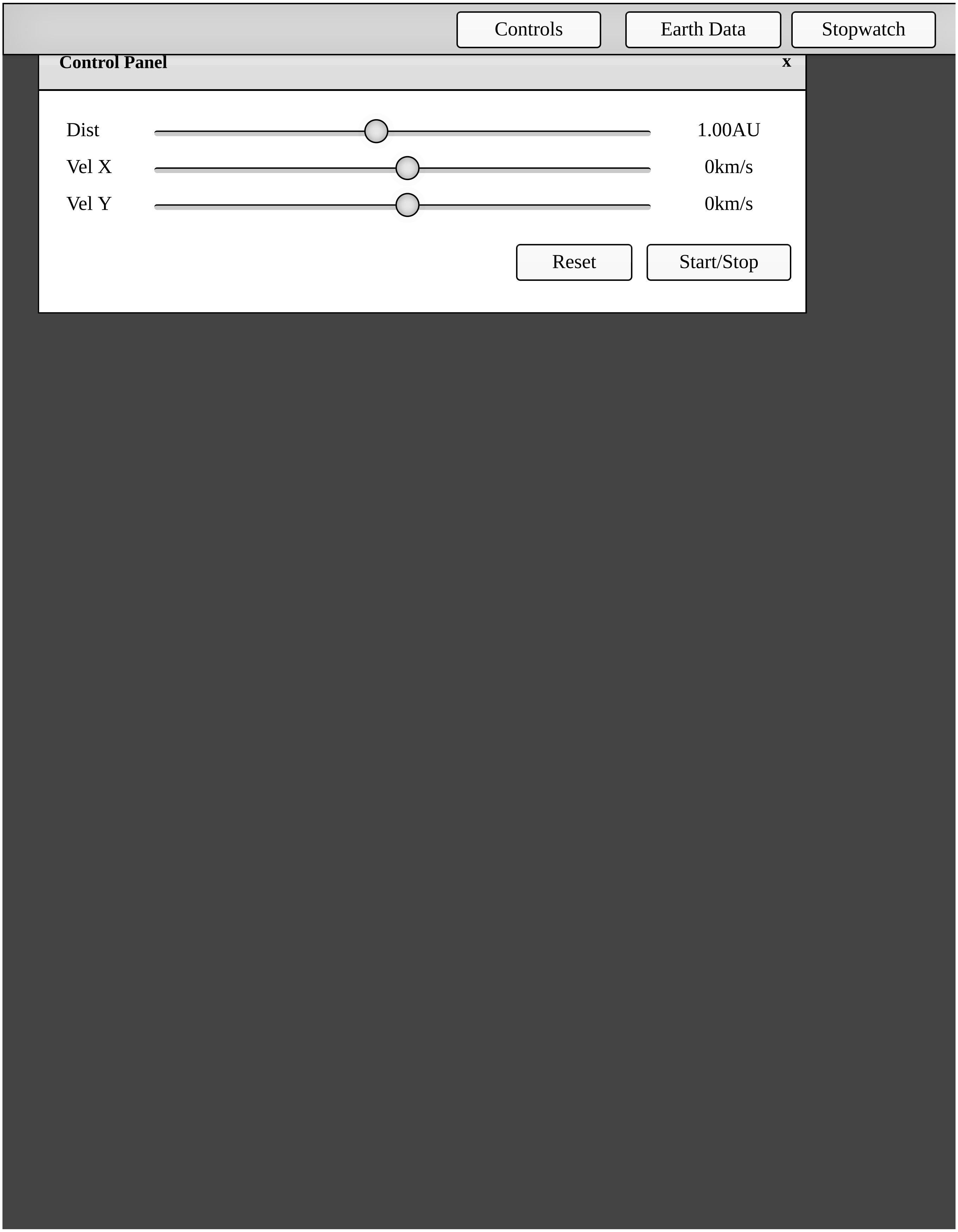
Explosion

This page

18.46: explosionvsexpansiondensity

This page titled 18.46: explosionvsexpansiondensity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.47: orbitsimulator


Controls
Earth Data
Stopwatch
Control Panel
x
Dist
1.00AU
Vel X
0km/s
Vel Y
0km/s
Reset
Start/Stop

This page titled 18.47: orbitsimulator is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.48: testjQueryUI

Loading ...

Element

Hydrogen (1)
Helium (2)
Carbon (6)
Nitrogen (7)
Oxygen (8)
Neon (10)
Iron (26)

Spectrum Viewer

Emission Spectrum

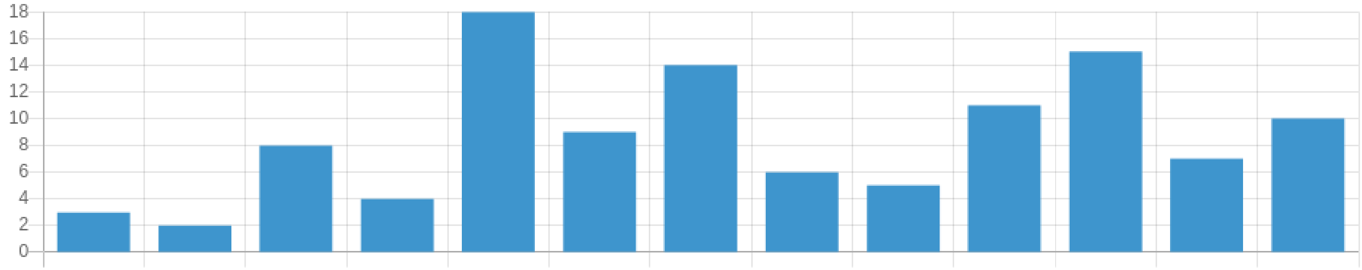
Emission Spectrum
Absorption Spectrum
Blueshift
Redshift

5.89

$^3 \times 4\pi^2$

This page titled 18.48: testjQueryUI is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.49: var1

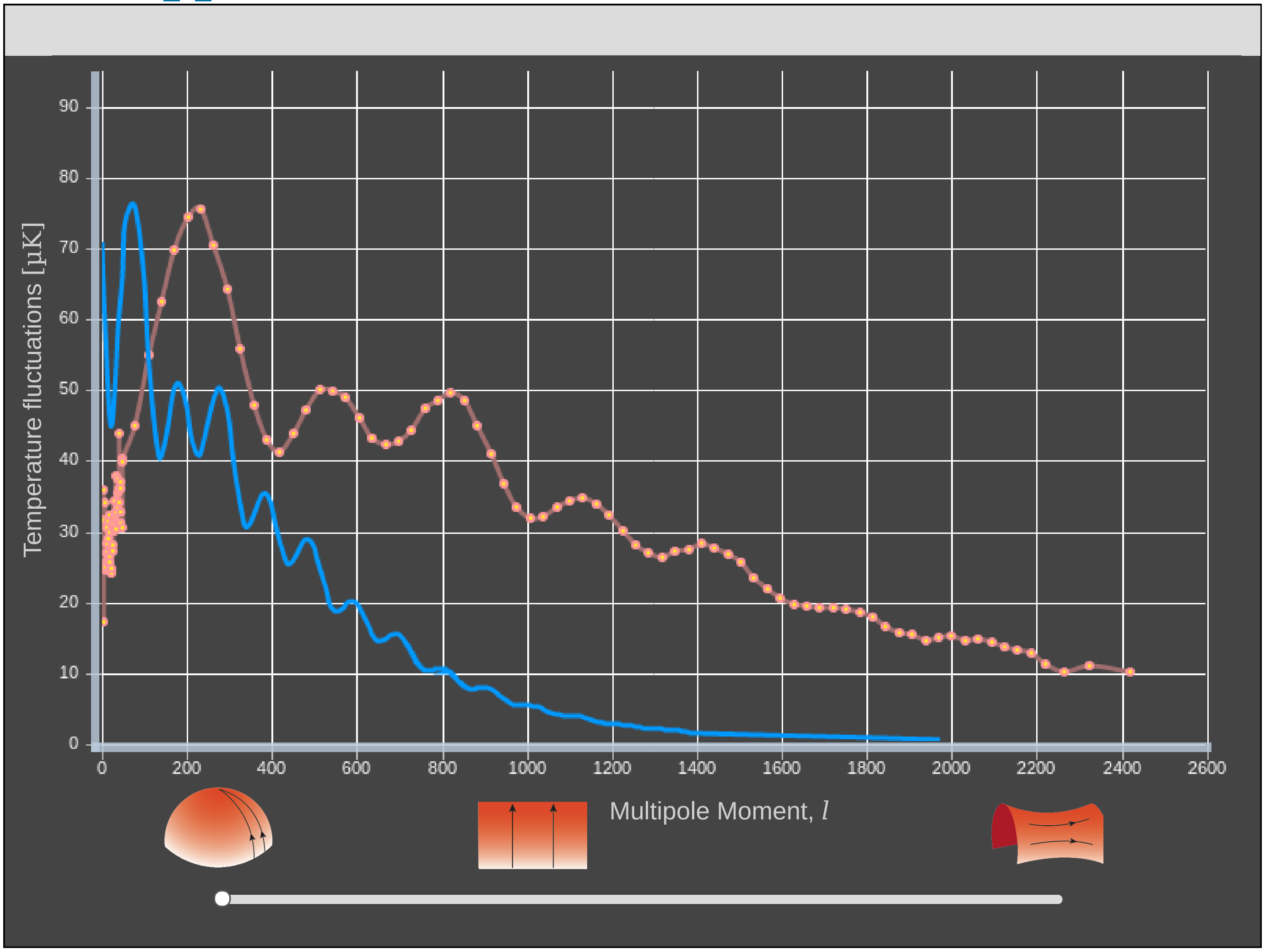


	A	B	C	D
1	3			
2	2		Minimum:	2
3	8		Mean:	8.714285714285714
4	4		Samp. Variance:	22.065934065934073
5	18		Samp.Stand. Dev:	4.697439096564645
6	9		Median	8.5
7	14		Maximum:	18
8	6		Q3:	10.75
9	5		n:	14
10	11		S&L Left: 10^?:	1
11	15			
12	7		Q1	5.25
13	10		Box: No Outliers?:	No
14	10			
15			Confidence Level	95
16			Conf. Int. Lower	6.002065133610323
17			Conf. Int. Upper	11.426506294961104
18			H1: mu ? mu0	<

19			mu0	45
20			t	-28.902708071468695
21			p-value	7.7639940494178e-14
22				
23			Other Statistics	
24			Trimmed %	5
25			Trimmed Mean	8.5
26				
27			x Value	0
28			Z-Score (x)	-1.855114145207053
29				
30			Coeff of Var	[illegible]

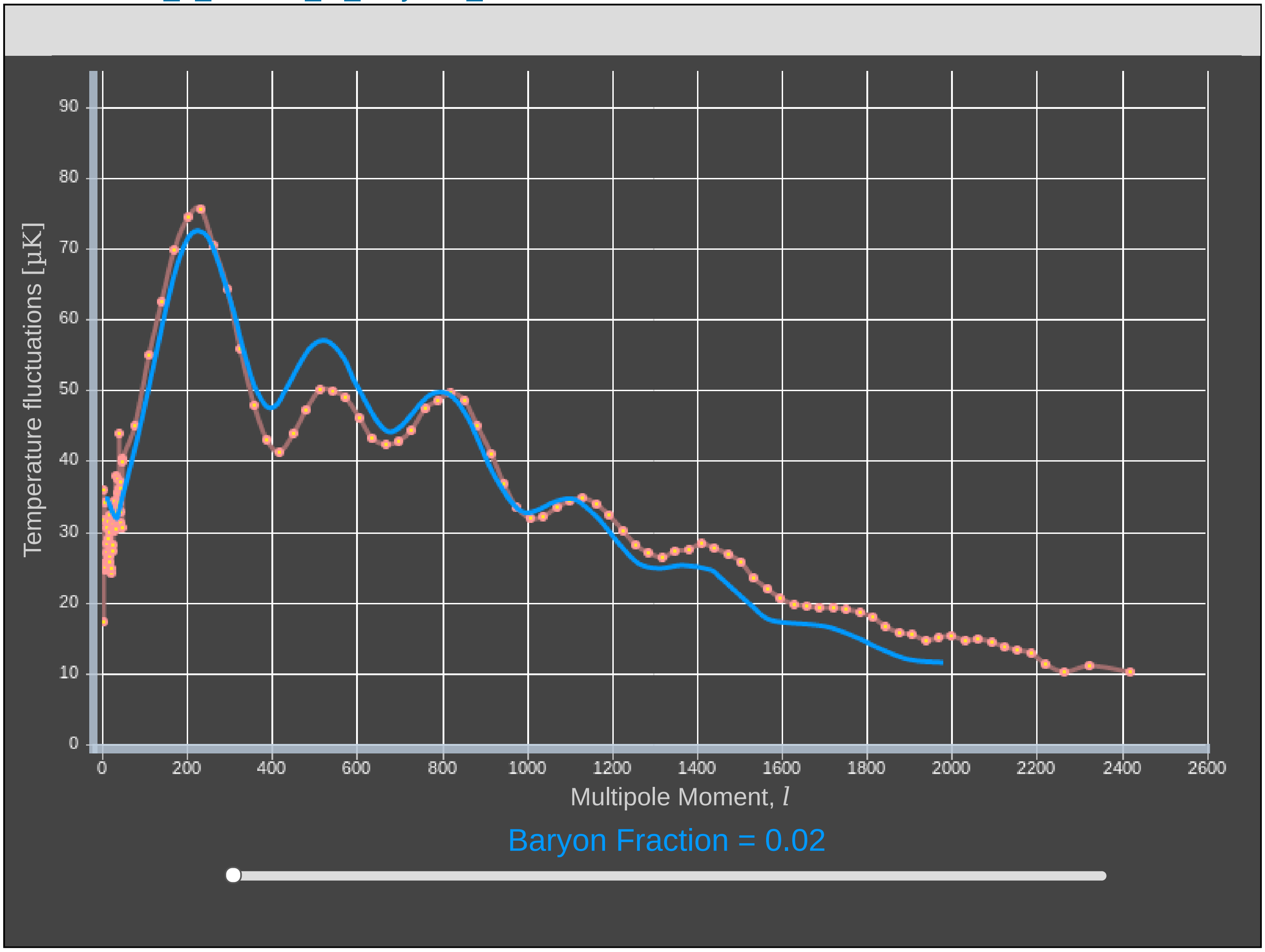
This page titled 18.49: var1 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.50: wiu15_a_curvature



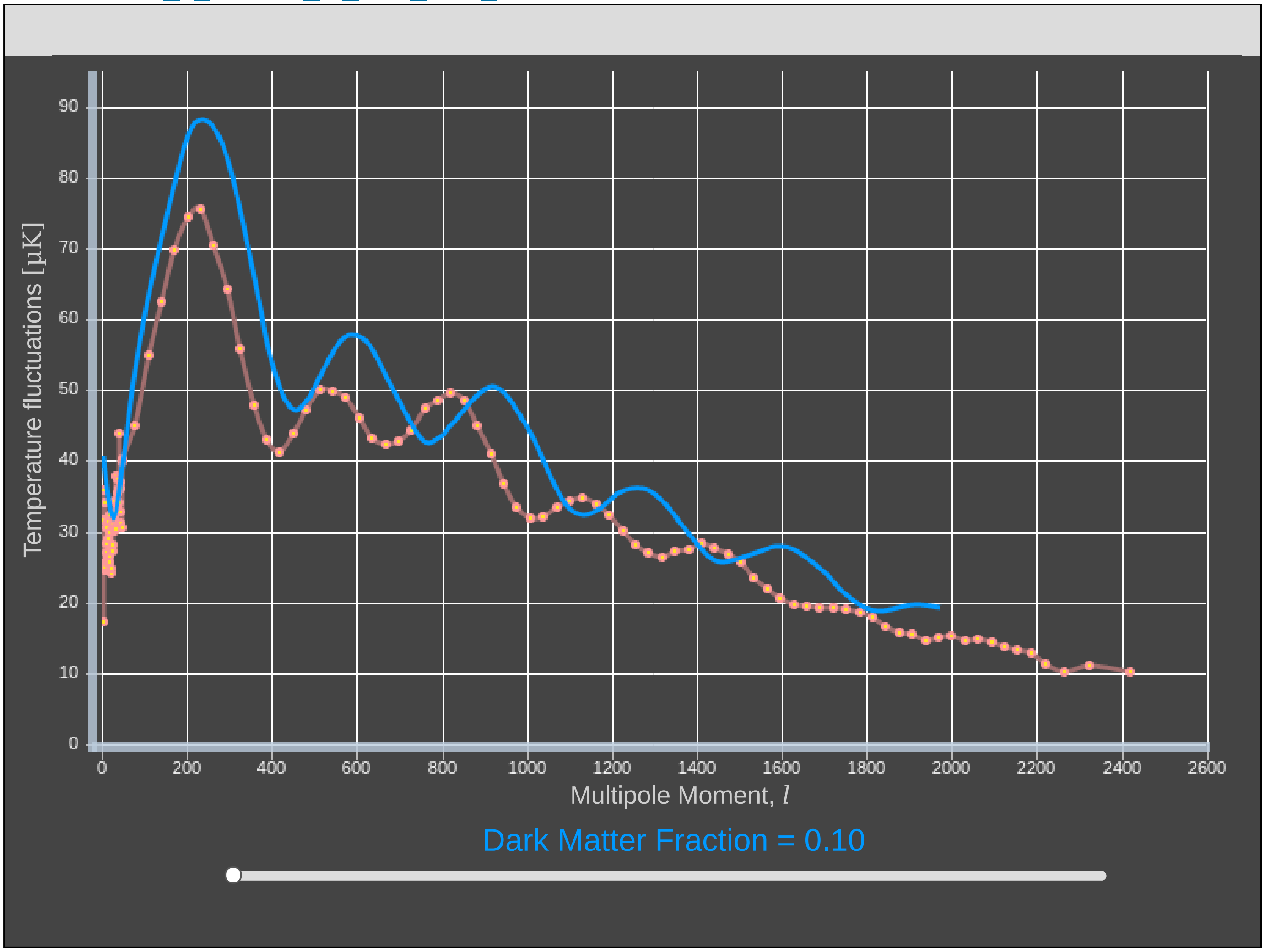
This page titled 18.50: wiu15_a_curvature is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.51: wiu15_b_amount_of_baryonic_matter



This page titled 18.51: wiu15_b_amount_of_baryonic_matter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.52: wiu15_c_amount_of_cold_dark_matter

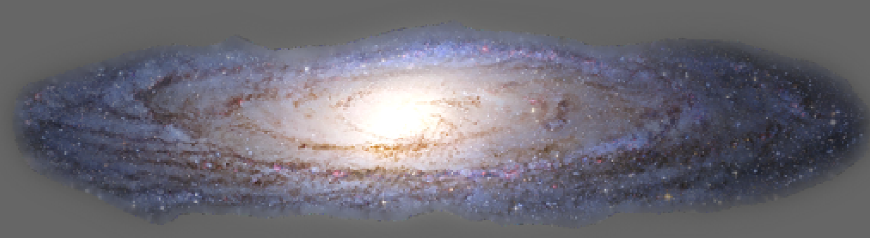


This page titled 18.52: wiu15_c_amount_of_cold_dark_matter is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.53: wiu6thesupernovaof1885

Show Data

Introduction



Steps:

1. Use the image to measure the angular size of the remnant.
2. Use the spectrum to estimate the expansion velocity of the cloud.
3. Use the age of the remnant and its expansion velocity to determine the physical size of the cloud.
4. Use the small angle formula to determine a distancr from the angulat size and physical size.

Next

This page titled 18.53: wiu6thesupernovaof1885 is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.54: wiu_12_1_determining_masses_of_lenses_grav_len_mass_measurement

Image | Select Galaxy Image

Measurements

Lens Distance : 0 Mpc

Source Distance : 0 Mpc

Line Length :

0.00 "

18.55: wiu_12_1_determining_masses_of_lenses_mass_calculator

Einstein Ring radius: x 10 radians

Distance Source to Observer (Dso): x 10 meters

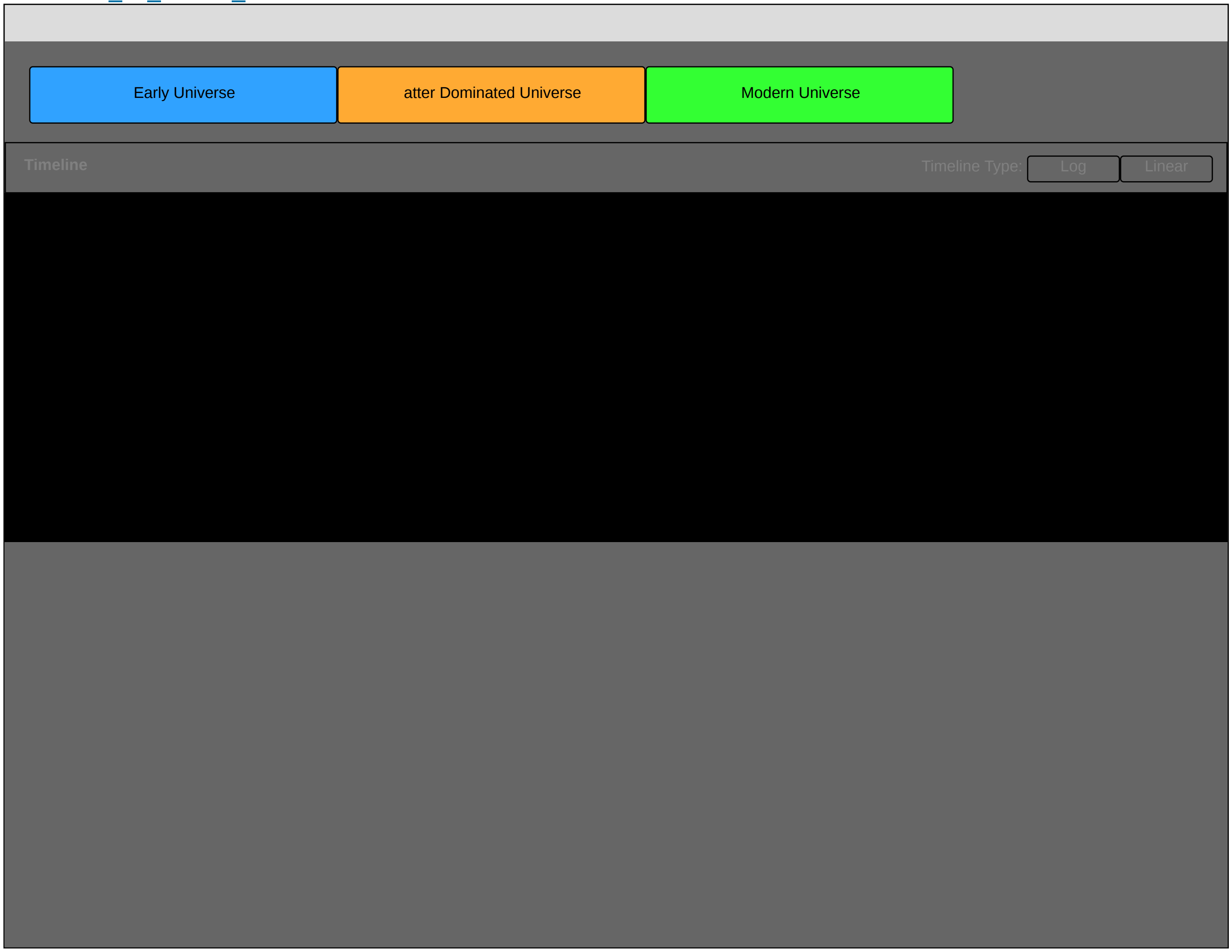
Ratio (Dlo/Dso) x 10

$$\text{Mass} = \frac{0.00\text{e-}16 \text{ radians}^2\ C^2}{4G}\ 0.00\text{e-}16\ 0.00\text{e-}16 \text{ m}$$

Mass = 0.000e-16 Kg

This page titled 18.55: wiu_12_1_determining_masses_of_lenses_mass_calculator is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

18.56: wiu_16_cosmic_timeline



This page titled 18.56: wiu_16_cosmic_timeline is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kim Coble, Kevin McLin, & Lynn Cominsky.

Kilometres To Miles

1.4 km
×
0.62 mi
1km
=
0.87 mi
Play
Reset

Kilometres To Miles is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

Pop ups

Topic hierarchy

Pop ups is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Math Exploration 13.1

The merging of two neutron stars produces a spectacular explosion called a **gamma-ray burst**. The annihilation of electrons and positrons causes the emission of an enormous amount of high energy radiation at 511 keV (a wavelength of 0.00242 nm). One of the most distant gamma-ray bursts has occurred at a redshift of $z = 4.5$.

a. At what wavelength would we observe the emission line now?

- **Given**: $z = 4.5$ and $\lambda_{emitted} = 0.00242\,\mathrm{nm}$
- **Find**: $\lambda_{observed}$
- **Concept**: The observed and emitted wavelengths are related by:

$$z = \frac{\lambda_{observed} - \lambda_{emitted}}{\lambda_{emitted}}$$

or equivalently

$$\frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = 1 + z$$

- **Solution**:

$$\begin{aligned}\lambda_{\text{observed}} &= 1 + z(\lambda_{\text{emitted}}) \\ &= (1 + 4.5)(0.00242\,nm) \\ &= (5.5)(0.00242\,nm) \\ &= 0.0133\,nm\end{aligned}$$

b. By what factor has the Universe stretched since the neutron stars merged and emitted the light that we are just seeing now?

- **Given**: $z = 4.5$
- **Find**: The ratio $S(t_{\text{observed}})/S(t_{\text{emitted}})$
- **Concept**:

$$1 + z = \frac{S(t_{\text{observed})}}{S(t_{\text{emitted})}}$$

- **Solution**:

$$\frac{S(t_{\text{observed}})}{S(t_{\text{emitted}})} = 1 + 4.5 = 5.5$$

- The Universe has expanded by a factor of 5.5 since the light we are seeing now was emitted from Galaxy A.

Math Exploration 13.1 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Math Exploration 13.2

When we calculate **escape velocity**, we set the total energy equal to zero. That is equivalent to setting the curvature term in the Friedmann equation to zero:

$$\frac{kc^2}{S^2} = 0$$

The Friedmann equation then becomes:

$$H^2 - \frac{8\pi G\rho}{3} = 0$$

The only two adjustable quantities in the equation now are ρ, the average density of the Universe, and the expansion rate, H. Solving for ρ in terms of H we get:

$$\rho_{crit} = \frac{3H}{8\pi G}$$

Math Exploration 13.2 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Index

T

V

W

LibreTexts

Glossary

absolute magnitude | The apparent magnitude an object would have if it were at a distance of 10 parsecs. cf magnitude.

absolute zero | An idealized temperature at which there is no energy left in a given system. 0 Kelvin is absolute zero, which is -273.15 ^oC, or -459.67 ^oF. Modern quantum physics precludes real systems from reaching absolute zero.

absorption | The process by which light or other electromagnetic radiation is absorbed by an atom, giving its energy to the atom in the process.

absorption line spectrum | A spectrum showing dark lines in some narrow color regions (wavelengths). The lines are formed when atoms absorb the light at specific wavelengths.

accelerate | A change in uniform motion, either from slowing down or speeding up, or changing direction.

accretion disk | A disk of matter that forms when material is transferred to a small gravitating body, such as a black hole or protostar. For black holes, the disks form outside the event horizon. For other objects, such as neutron stars or protostars, the disks can extend down to the stellar surfaces. Friction (collisions) within the disks heat them and allow material to flow inward while angular momentum flows outward. Accretion disks often emit a wide range of different types of electromagnetic radiation including infrared, UV, and x-rays.

accuracy | In science, the closeness of a measurement of some quantity to its true value. This differs from precision.

active galaxy | A galaxy with a very bright, energetic nucleus. The evidence suggests that they are powered by the release of gravitational energy as material falls onto a central black hole. The range in mass from several million to several billion times the mass of the Sun. Sometimes active galaxies are called AGN, for active galactic nuclei.

alpha radiation | a type of radiation emitted during radioactive decay, an alpha particle is a ^4He nucleus

anti-particle | The antimatter complement to a particle, mostly having identical properties to the particle, but with opposite electric charge.

apparent magnitude | How bright an object appears as seen by an observer. cf magnitude.

arcminute | A measure of angular size based on a circle. A full circle has 360^o, which can be divided into 60 equal parts, each part being 1 degree. An arcminute, in turn, is one sixtieth (1/60) of a degree.

arcsecond | One sixtieth (1/60) of an arcminute.

asteroid | A rocky object in space that can be a few meters to hundreds of kilometers wide.

astronomical unit | The average distance between Earth and the Sun, equal to 149,597,870.7 kilometers. Abbreviated AU.

atom | A basic physical building block of matter in the Universe, which is composed of electrons, protons, and neutrons.

atomic number | Indicates the number of protons in the nucleus of an atom. The atomic number defines a chemical element.

AU | See Astronomical Unit.

background counts | Sources of light detected by a CCD that are extraneous to the celestial object being studied.

band-pass | A specific range of electromagnetic frequencies, often used to describe a satellite's viewing capabilities.

BBN | See Big Bang Nucleosynthesis.

beta radiation | a type of radiation emitted during radioactive decay, an electron or a positron is emitted

Big Bang Nucleosynthesis (BBN) | The processes in the early Universe that created the lightest chemical elements. Also, the period of time over which these processes were active.

Big Bang Theory | The theory that suggests that the Universe at early times was compressed into an extremely hot, dense state. Then, for reasons currently unknown, the universe expanded and cooled to its present condition, and it continues to expand and cool at the present time.

big chill | An end-of-Universe scenario where galaxies continue to move away from each other and the temperature of the universe continues to cool.

big crunch | An end-of-Universe scenario where gravitational attraction causes galaxies to eventually start moving toward each other.

big rip | An end-of-Universe scenario where dark energy becomes so strong that the expansion rips apart galaxies, even at the atomic level.

binary | The word binary simply means there are two of something. When applied to a star system, it means that two stars orbit their common center of mass.

binary star system | A system of two stars, very close to each other, that orbit around their common center of mass. Such systems are quite common.

binding energy | The energy required to break up an atomic nucleus into its constituent protons and neutrons.

black hole | A region of space within which the force of gravity (space-time curvature) is so strong that nothing, not even light, can escape from it.

blackbody | A theoretical object that is a perfect absorber and emitter of electromagnetic radiation. Such an object would emit so-called blackbody radiation.

blackbody radiation | Radiation produced by a blackbody. The intensity at each wavelength follows a distribution that depends on the temperature of the object.

blackbody spectrum | The spectrum of the radiation emitted by a blackbody depends upon the temperature of the black body. cf. blackbody radiation.

blueshift | An apparent shift of spectral lines in the radiation emitted by an object toward shorter wavelengths. Often caused by motion of the object toward the observer, or vice versa. See also Doppler effect.

brightness | The amount of light an observer sees emitted from an object like a star. Brightness is measured in watts per square meter (W/m^2).

brown dwarf | A cosmic object that is too small to be a star and too large to be a planet. Brown dwarfs have the same composition as stars, but because of their low mass are unable to sustain nuclear fusion at their cores (assuming that they ever manage to get fusion started).

bulge | The central region of a spiral galaxy.

causality | the idea that some events are the direct cause of other events, and that the cause must precede its effect.

CCD | See charge-coupled device.

chameleon particles | A hypothetical particle which is postulated as a dark energy candidate. The particle is chameleon-like in that its mass can change.

charge-coupled device (CCD) | An instrument that can act as a camera sensor. It works by converting the light that falls onto it into electrical signals in individual picture elements (pixels), which are generally arranged in a square array.

chronometer | A watch that has been specifically designed to keep very accurate time.

coefficient | A number that multiplies another number or expression.

coma | The cloud-like ball of gas and dust surrounding a comet's nucleus.

comet | Dusty bodies of ice that orbit a star. We typically imagine comets with their characteristic tails, but the tails only form when comets approach close to a star. Comets have three distinct tails: the dust tail is composed of dust pushed out by radiation pressure from the star, the ion tail is composed of particles evaporated by solar winds and pushed back, and a tail of sodium escapes from the dust. The sodium tail is not visible to the naked eye. These tails point in slightly different directions but always away from the star.

concordance model | Our currently accepted and most commonly used cosmological model.

conservation of energy | a physical law that says that energy cannot be created or destroyed. Because of this law, any naturally occurring physical process can only transfer energy from one part of a system to another part; the total energy must remain the same.

constellation | A region in the sky that has been officially defined by the International Astronomical Union. Constellations are used by astronomers to designate a region in the sky, similar to the way countries are used to designate regions on Earth. Most constellations, especially in the Northern Hemisphere, have historical origins related to the myths of Ancient Greece.

continuous spectrum | A spectrum that is an unbroken range of wavelengths. An object emitting some light in every wavelength, such as a blackbody, produces a continuous spectrum.

continuum | See continuous spectrum.

convergence point | The point on the horizon where parallel lines seem to converge. In astronomy, it is the point to which an extended object such as a star cluster seems to converge as it moves away from the observer. If the motion is toward the observer, then the object seems to expand away from the convergence point.

cosmic distance ladder | A hierarchal process used by astronomers to determine the distances to very distant astronomical objects based on known distances of similar objects that are closer.

Cosmic Microwave Background (CMB) | This is the radiation left over from the big bang. It was produced early in the age of the universe, when the average density and temperature were much higher than today. The expansion of the universe has cooled the radiation to its current temperature of about 2.7 Kelvin.

cosmic rays | high-energy charged particles from outer space. These include protons, neutrons, atomic nuclei, and subatomic particles.

cosmology | The astrophysical study of the history, structure, and evolution of the universe.

critical density | A special value of density that causes the Universe to have zero curvature.

dark current | The small but detectable charge that accumulates in a CCD, even when it is not exposed to light. This is a source of noise in astronomical instruments that must be removed.

dark energy | a hypothetical form of energy or property of space that causes the expansion rate of the universe to accelerate.

dark matter | Making up approximately 80% of the matter in the Universe, dark matter does not appear to radiate or absorb any light and is only detectable indirectly by the effect its gravity has on visible objects.

dark nebula | cold, relatively dense clouds of interstellar gas and dust

data | The factual information a scientist collects that is related to the hypothesis he or she is testing. Data can be direct measurements of properties, for example, the height of plants one month after seeds have been planted. Data can also be observations of patterns, for example, the behavior of animals when they encounter a predator. To determine whether a particular measurement or observation is a rule rather than an exception, a scientist will often repeat measurements or observations to gather additional data.

degeneracy | A combination of values having the same properties.

density | A measure of how much matter is packed into a given volume. High-density objects have more material packed into a given volume than low-density objects.

detector | A device, or devices, used to capture photons, or in some cases, other particles like protons, electrons, etc.

diffraction grating | Parallel slits or grooves etched on an optical surface that cause light to bend and create an interference pattern after passing through the grating. This spreads the light out into a rainbow of colors.

disk | A thin, roughly circular plane in which the majority of stars, gas, and dust of a spiral galaxy are contained.

distance | The length of space between two objects.

Doppler effect | The Doppler effect is the apparent change in the wavelength and frequency of sound or light depending on whether the source is moving toward or away from you. The faster you or the source is moving, the more profound the impact. You may have experienced an example of this with sound waves if you have ever heard a higher pitch sound from a car moving toward you, and a lower pitch noise when it is moving away.

Doppler shift | The shift in the frequency or wavelength of a wave due to the Doppler effect.

dust | mixture of molecules, such as silicates, graphite, iron, and other compounds

dwarf galaxy | A galaxy that contains somewhere around several billion stars, as opposed to the hundreds of billions of stars found in a large galaxy like the Milky Way.

dwarf planet | An astronomical object that orbits its parent star and that has enough mass for gravity to make it spherical in shape, but has not cleared its orbital path of debris.

ecliptic | The name given to the plane in which the Earth travels as it orbits the Sun.

electromagnetic (EM) radiation | Another term for light, including visible light and invisible forms from radio up through gamma-rays. See electromagnetic spectrum.

electromagnetic (EM) spectrum | This is the continuum of waves of light, which range from very low-frequency and low-energy radio waves to very high-frequency and high-energy gamma rays. The kind of light we are familiar with is visible light, which is a tiny sliver in the middle of the EM spectrum.

electromagnetic force | The electromagnetic (or EM) force is one of the four known universal forces, along with gravity and the strong and weak nuclear forces. The EM force holds all the molecules and cells in your body together and is the result of interactions between charged particles (protons and electrons) within the atoms and molecules.

electromagnetic waves | Another term for light. Light waves are fluctuations of electric and magnetic fields in space.

electron | A fundamental subatomic particle that is commonly found in the outer regions of an atom and is negatively charged. Electrons are a type of lepton.

electron cloud | The region around the nucleus of an atom that the electrons occupy.

electron volt | A unit of energy. 1 eV = 1.602×10^{-19} joules. It is the energy gained by an electron (or a proton) falling through an electrical potential difference of one volt. Visible light photons have energies of about one electron volt.

element | A substance that is composed of a single type of atom.

elementary particle | A particle which is not made up of any other particles or substructure, such as quarks and leptons.

elliptical galaxy | These galaxies range in shape from nearly spherical to flattened disks. They are characterized by an old population of stars and have very low rates of star formation, meaning very few stars are being born in them. Ellipticals contain little or no cool gas or dust.

emission | The production of light, or more generally, electromagnetic radiation, by an atom or other object.

emission line spectrum | A spectrum consisting of bright lines at certain wavelengths separated by dark regions in which there is no light.

emission nebula | clouds of hot interstellar gas that emit light

energy | Energy is the ability to do work. The SI unit for energy is the joule (J), where 1 J = 1 kg $(m/s)^2$ = 1 N m. The eV (see electron volt) is another common unit of energy: 6.242×10^{18} eV = 1 J. Joules are also related to watts (W), the SI unit for power, via: 1 W = 1 J/s. One exploded ton of TNT is equivalent to 4.2×10^9J, or 2.6×10^{28} eV.

entropy | a physical measure of the disorder of a system, often computed by measuring the number of possible configurations of the system.

error ellipse | A region on a plot that covers all allowable values for the specified cosmological parameters given the measurement uncertainties.

escape speed | The minimum speed required to escape the gravitational pull from a massive object.

eV | See electron volt.

event | in special relativity, an event is a location (t, x, y, z), in spacetime, that describes the position and time of an occurrence.

event horizon | The region near a black hole where the escape speed becomes the speed of light. Anything that crosses the event horizon cannot escape the gravitational pull of the black hole.

evolve | To change over time.

excited state | An energy state in a quantum system that lies above the lowest available state, or ground state. In an atom, an electron that has gained energy and moved from its lowest state into a higher energy state is said to be in an excited state. The atom must release energy, often by emitting a photon, for the electron to return to the ground state.

exponent | The value to which a base number is raised. For example, in the number 10^4, the 4 is the exponent, 10 is the base.

extra-solar planet | A planet that orbits a star other than the Sun and is therefore not within our Solar System.

field bosons | Fundamental particles of integral spin (0, +/-1, +/-2, etc) that carry force between other particles. Sometimes called *field boson*s. One example is the photon, which carries the electromagnetic force. Another is the gluon, which carries the strong nuclear force.

filter | A device that can be placed in front of a camera that lets certain wavelengths of light pass through and blocks other wavelengths.

flux | A measure of the amount of energy given off by an astronomical object per unit time per unit area. Because the energy is measured per time and area, flux measurements make it easy for astronomers to compare the relative energy output of objects with very different sizes or ages.

frame (of reference) | in physics, a frame of reference refers to the coordinate system used to conduct measurements, actual or imagined. These coordinates could be fixed to a laboratory that is stationary on Earth, or they could be on a moving object like a plane flying through the air. They could even be on some other planet, or even on an imagined spaceship that travels between the planets or stars. All measurements are only meaningful if referred to some standard of measure in some frame of reference.

frequency | A property of a wave that describes how many wave patterns, or cycles, pass by a point in a given time. Frequency is often measured in hertz (Hz), where one hertz is one cycle per second.

fusion | The process of merging multiple objects into one. In nuclear fusion, light atoms are forced together to form heavier atoms. For instance, hydrogen can be fused to form helium. Helium in turn can be fused into the heavier atom carbon, and so on. This process releases large amounts of energy and is what powers stars. However, fusion requires very high temperatures. That is why a tube of hydrogen or helium gas in a laboratory will not spontaneously undergo nuclear fusion.

galaxy | A gravitationally bound system of stars, their satellites, dust, gas, and dark matter that contains a supermassive black hole at its center. There are three general types of galaxies; spiral, elliptical, and irregular.

galaxy cluster | A group of two or more galaxies that are bound by gravity.

galaxy halo | A spherical distribution of stars, with very low density, in which galaxies are embedded. Different than dark matter halo.

gamma | The factor in special relativity by which time and space are stretched or compressed by relative motion. Also, the ratio of the energy of a particle to its rest energy. The gamma factor depends on the relative velocity between reference frames approximately equal to x for small angles x (when x is measured in radians).

gamma rays | The very highest energy end of the electromagnetic spectrum, with the shortest wavelengths. Gamma rays typically have energies a few hundred times larger than low-energy x-rays and wavelengths shorter than a few hundred picometers (pm, 10^{-12} m).

gas | atoms and small molecules, primarily hydrogen

gas giant | A planet that is at least five times as massive as Earth and comprised mostly of gases, such as hydrogen and helium.

Gedankenexperiment | German, "thought-experiment." These are illustrative scenarios used to guide one's thinking when considering physical situations, especially when thinking about relativistic systems. These were often employed by Albert Einstein when developing his physical intuition regarding problems related to relativity.

GeV | Gigaelectron volt, or 1 billion electron volts.

Giga (prefix) | A billion, 10^9. Denoted as G, e.g., GeV.

globular clusters | Exceptionally dense, spherical clusters of old stars that are gravitationally bound, located in galaxy halos.

gluon | A subatomic particle that carries the strong nuclear force between quarks.

gravitational constant | See Universal Constant of Gravitation.

gravity | The universal force of attraction between all matter.

Great Red Spot | A giant cyclone that has existed on Jupiter for at least 300 years. It is not known whether it will ever disappear.

greatest elongation | When the Sun, Earth and a planet interior to Earth's orbit are positioned such that the planet appears as far as possible from the Sun in the sky as seen from Earth. Only Venus and Mercury can attain such a configuration.

ground state | The ground state is the lowest energy state available in an atom or other quantum system.

H-R diagram | See Hertzsprung-Russell diagram.

half-life | The average time required for half of a sample of radioactive atoms to decay

halo | See dark matter halo or galaxy halo.

helium-3 | A form of helium that has 2 protons and 1 neutron. He-3, which is used in nuclear fusion research, is rare on Earth.

hertz | Abbreviated "Hz". The derived SI unit of frequency, defined as cycles per second.

Hertzsprung-Russell diagram | A plot of the brightness of stars versus their surface temperature or spectral class. Abbreviated to H-R diagram.

histogram | A graph that displays the density of data, generally the frequency of occurrence of a range of events, such as the frequency of photons of specific wavelengths striking a detector per second (example histogram below). A familiar use of a histogram is the "bell curve" used to show the distribution of grades on a test, for example.

HST Key Project | Hubble Space Telescope (HST) had several chief missions, called Key Projects. One such mission was to accurately determine the extragalactic distance scale, beginning with galaxies in the Virgo Cluster.

Hubble's Law | The relationship between the recession velocity of a galaxy and its distance from the observer. The farther away a galaxy is from an observer, the faster it is moving away. The law is named for Edwin Hubble, who first published it in 1929 (Proc. Nat. Acad. of Sci, Vol 15, March 15, 1929).

hydrostatic equilibrium | The state in a fluid, such as the gas in a star, or the atmosphere or ocean of a planet, in which the compressional force of gravity is everywhere offset by the outward force of pressure in the fluid.

Hz | See hertz.

image processing | The process by which data gathered from a CCD or other electronic imaging detector are converted into images that can be interpreted by an astronomer.

imaging | Scientific technique that results in photographic or computerized representations of data.

imprecision | Lack of precision or repeatability. cf. precision.

inertial frame | a frame of reference that moves at constant velocity. An inertial frame does not change its speed or direction.

infrared light | Abbreviated "IR". The band of the electromagnetic spectrum intermediate between optical and microwaves, with wavelengths in the micron (μm, 10^{-6}m) range. Infrared is correspondingly more energetic than microwaves, but less energetic than optical.

intensity | See flux.

International System of Units | Abbreviated SI Units, also called the metric system, it is the scientific standard of carefully defined units of measurement. The SI unit of length is the meter, of time is the second, and of mass is the kilogram. Many other units are used to measure other quantities.

invariance (invariant) | see Principle of Invariance.

Inverse Square Law | A relationship that states that the flux, or apparent brightness, of an object decreases as the inverse of the square of its distance to the observer.

ionization | The process of stripping electrons from an atom.

ionized gas | see Plasma.

irregular | A galaxy that does not fit into any of the other categories (spiral or elliptical).

isotopes | Atoms with the same number of protons but different numbers of neutrons.

jovian planets | Jupiter-like planets that do not have solid surfaces. They are composed primarily of helium and hydrogen. They typically have radii larger than 10,000 km (6,213.7 miles) with a mass over 1×10^{25} kg. These gaseous planets also have rings and many moons.

k | See kilo.

K | See Kelvin.

Kelvin | The SI unit for temperature. The temperature at which water freezes on the surface of Earth is 273.15 kelvin, and the temperature that water boils is 373.15 kelvin. Zero on the Kelvin scale is the theoretical point where all motion ceases in classical thermodynamics (absolute zero). The name honors the 19th century Scottish physicist William Thomson, who is more commonly known as Lord Kelvin.

keV | 1 kilo electron volt is equal to 1,000 electron volts, and is a unit of energy convenient for describing x-ray energies.

kg | See kilogram.

kilo (prefix) | A thousand; 10^3. Denoted as k, e.g., keV.

kilogram | Abbreviated "kg". The SI unit of mass. The kilogram is the only SI unit still maintained by a physical artifact (a platinum-iridium bar) kept in the International Bureau of Weights and Measures at Sevres, France. One kilogram is equivalent to 1,000 grams or about 2.2 pounds; the mass of a liter of water.

kilometer | A unit for measuring length, one thousand meters. Abbreviated km.

Kuiper belt | The region of the solar system past Neptune, from approximately 30 AU to 100 AU. Objects in this region are called Trans-Neptunian Objects (TNO).

length contraction | the distortion of measured lengths between inertial frames. Observers in different inertial frames will measure distances along the direction of relative motion between their frames to be longer in their own frame than in any frame moving with respect to their own.

lepton | A fundamental particle with little mass, or possibly no mass in some cases. Electrons, muons and tauons are all leptons, as are their associated neutrinos.

light speed | See Speed of Light.

light-hour | The distance that light travels in a vacuum in one hour, approximately equal to one billion kilometers (1×10^9 kilometers).

light-minute | The distance that light travels in one minute, which is approximately 18 million kilometers (1.8×10^7 km).

light-year | The distance that light travels in one year, which is about 10 trillion kilometers (9.45×10^{12} km). Light-years are a convenient unit of measure for most astronomical distances.

lookback time | The delay of time, due to the finite speed of light, required for light to travel from its source to its observer. The lookback time for terrestrial objects is negligible, but for astronomical objects, it grows with their distance, from about 8 minutes for light from the Sun to billions of years for distant galaxies.

luminosity | The amount of energy an object, like a star, radiates per unit time. This is usually measured in watts, just like a light bulb.

magnification | The process of enlarging the appearance of an object through various optics and lenses.

magnitude (astronomy) | A measure of brightness. Counterintuitively, the brighter an object is, the lower its magnitude. A first magnitude star is about 2.5 times as bright as a second magnitude star, and so on. The brightest object in the sky is, of course, the Sun, with a magnitude of-26.73. The full moon is -12.6, and with the naked eye, we can see all the way down to about a magnitude of 6. The brighter stars in the sky are around magnitude zero, with the brightest, Sirius, having a magnitude of about -1.5.

magnitude system | See magnitude.

main sequence | Abbreviated "MS". The region of a Hertzsprung-Russell diagram, running diagonally from hot and bright to cool and dim, stars that appear in this region derive their energy solely from hydrogen fusion. The MS contains roughly 90% of all stars.

main sequence fitting | Determining cosmic distances by comparing the main sequence regions in H-R diagrams of different star clusters.

main sequence star | A star that is actively fusing hydrogen into helium in its core. The inward gravitational force due to the mass of the star is balanced by the outward thermal pressure generated by nuclear fusion.

main sequence turn-off point | Refers to the point at which stars leave the main sequence in the H-R diagram as they exhaust the hydrogen in their core.

mass number | The total number of protons and neutrons in an atom.

masses | A measure of the inertia of an object and also of the strength of its gravitational interactions, with larger masses having greater inertia and stronger gravitational interactions. Mass is related to how much "stuff"—in the form of protons and neutrons—an object is made of, and is only changed by changing the amount of this stuff.

meteor | The flash of light we see when a solid object falls into our atmosphere and disintegrates. These objects vary in size from as small as sand to many meters in diameter. The largest reach the ground before burning up and create impact craters when they land. The pieces of rock or metal that remain are called meteorites.

meteor shower | When we see a large number of meteors in a relatively short time, created when Earth passes through a cloud of dust left over from the pass of a comet. For example, during its peak, observers might notice one or two meteors a minute from the Perseid meteor shower that occurs each year around August 12.

meter | Abbreviated "m". The fundamental SI unit of length, defined as the length of the path traveled by light in vacuum during a period of 1/299,792,458 s. A unit of length equal to about 39 inches. A kilometer is equal to 1000 meters.

metric system | See International System of Units. The metric system uses Celsius rather than Kelvin for temperature, but is otherwise the same as the International System of Units.

MeV | Megaelectron volt, or 1 million electron volts.

micro (prefix) | One-millionth; 10^{-6}. Denoted as μ (Greek lowercase mu), e.g., μm.

microwave | A region of the electromagnetic spectrum between infrared and radio. The energy of microwaves is a bit higher than radio waves. Their wavelengths are therefore shorter and are typically measured in centimeters (cm or 10^{-2} m).

Milky Way | Common name for the galaxy in which our Solar System is located.

milli (prefix) | One-thousandth; 10^{-3}. Denoted as m, e.g., mm.

model | A simplified explanation of how a natural system works that is based on empirical evidence and logic. To be useful, a model should make testable predictions. See also theory.

molecular cloud | A giant region of diffuse gases that can be several hundred light years across. They are composed mostly of molecular hydrogen, with helium and a few other elements dispersed throughout. Internal gravitation in colder denser regions of the cloud can trigger collapse and star formation. In addition to molecular hydrogen, molecular clouds contain molecules like CO, CH_4, NH_3, HCN, CH_2O and others.

molecule | Two or more atoms held together by chemical bonding.

moon | A celestial body that orbits a planet or smaller body. See also satellite.

moving cluster method | A method for determining the distances to clusters of stars that employs geometry and trigonometry to determine the cluster distances. This method would be useful just outside the boundary of using the parallax method.

multi-wavelength astronomy | The study of the Universe in all ranges of the electromagnetic spectrum, from radio waves to gammarays.

muon | a subatomic particle similar to, but more massive than, an electron.

natural satellite | See moon.

nebula | Plural, nebulae. Interstellar clouds of dust and gas, from the Greek, for *cloud*.

nebulae | Singular, nebula. Interstellar clouds of dust and gas, from the Greek, for *cloud*.

neutrino | An elementary particle that has an extremely small mass and only very weakly interacts with matter. The neutrino is part of the lepton family of particles. The majority of neutrinos detected on Earth come from the Sun.

neutron | One of the particles that makes up the nucleus (center) of atoms and has no charge. Neutrons are composed of three quarks.

neutron star | The collapsed core of a massive star, composed mostly of neutrons. Neutron stars are very small, with a diameter of about 10 kilometers. They have an enormous mass for their size, ranging from 1.4 solar masses to a bit more than twice that.

Newton's constant | See Universal Constant of Gravitation.

nuclear fission | The process by which heavy elements split apart into lighter ones. For instance, uranium nuclei can be split into two nuclei, each of which is roughly half the mass of the original uranium nucleus.

nuclear fusion | The process by which lighter elements like hydrogen and helium fuse together to make heavier elements like lithium, carbon, oxygen, etc.

nuclear reaction | The process by which the nucleus of an atom gains or loses neutrons and protons.

nuclei | Singular, *nucleus*. The central core of an atom, composed of neutrons and protons.

nucleus | Plural, nuclei. The central core of an atom, composed of neutrons and protons.

observatory | A facility that includes a telescope, either on the ground or in space.

Oort cloud | A region of space where long period comets originate, approximately 50,000 AU from the Sun.

open cluster | Loosely associated stars, numbering in the hundreds, that have formed together in the same cloud, but have not yet had time to drift apart.

open universe | A universe with no dark energy that does not contain enough mass to counteract its expansion; Omega is less than 1.

optical | The band of electromagnetic radiation that we can see with our eyes. It is intermediate in terms of energy and wavelength, between ultraviolet and infrared. Wavelengths range from approximately 400 to 750 nm, and energies are about one eV.

orbit | The path followed by a moon, planet, artificial satellite or other body, as dictated by gravity.

osmological constant | A constant term that can be added to Einstein's equations; works in the opposite direction to the gravity due to mass-energy; causes space to expand rather than contract.

oxidation | The combination of a chemical element with oxygen.

parallax | The apparent shift in position of a relatively nearby object compared to a more distant background as the location of the observer changes. Astronomically, it is half the angle that a star appears to move as Earth orbits from one side of the Sun to the other.

parsec | A unit of distance used by astronomers. An object one parsec away will exhibit a parallax of one arcsecond. One parsec equals about 3.3 light-years.

particle | See subatomic particle.

period | Time required for cyclic motion to repeat. For instance, the period of Earth to turn once around its axis is 24 hours, while the period for Earth to travel once around the Sun is 365 days. We would say that Earth has a *rotation period* of 24 hours and an *orbital period* of 365 days.

photo-excitation | The process by which an electron in an atom absorbs the energy from a photon and is excited to a higher energy state.

photoelectric effect | An effect whereby materials are induced to emit electrons when light shines onto them. The effect was explained in 1905 by Einstein by employing a particle theory of light.

photometry | The measurement of the brightness of astronomical objects. A standard result of photometry is the light curve (a plot of brightness versus time).

photon | A quantum (particle) of light or electromagnetic energy. Photons are have zero rest-mass and no electric charge.

pico (prefix) | One-trillionth; 10^{-12}. Denoted as p, e.g., pm.

Planck spectrum | See blackbody spectrum.

Planck's constant | A fundamental physical constant denoted by h. It has the value 6.626196×10^{-34} J s.

planet | Meaning "wanderer" in Greek, a celestial body that is massive enough for its own gravity to form itself as a spheroid but is not massive enough to begin thermonuclear fusion. Planets orbit a star or stellar remnant and have cleared their orbital paths of debris.

planetary nebula | Plural, nebulae. The expelled outer layers of low-mass stars, ionized by the ultraviolet radiation of a central white dwarf.

planetary nebulae | Singular, nebula. The expelled outer layers of low-mass stars, ionized by the ultraviolet radiation of a central white dwarf.

plasma | A gas that contains charged particles. It is composed of electrons that have been stripped from atoms, and the resulting positively charged particles called ions.

precision | The expected range of uncertainty of a physical measurement. Repeatability of that measurement. Precision differs from accuracy.

Principle of Invariance | this principle states that the spacetime separation of two events is a constant in Special Relativity, or in other words, that it is the same for all inertial frames of reference.

proper motion | The angular change in position of an astronomical object over time as seen from Earth. Measured in arcseconds per year.

proton | One of three subatomic particles that make up an atom. Protons are positively charged and located in the nucleus of an atom. Protons are composed of three quarks.

protostar | A young star that is still accreting matter from an accretion disk and that is enshrouded in a cloud of gas. A protostar is the earliest stage of a star's life, before it has even grown large enough to start nuclear fusion in its core.

pulsar | A type of magnetized spinning neutron star that emits a flash of light from a bright spot, like a lighthouse. Since the bright spot on the star's surface only points at us some of the time, it looks to us like it is pulsing on and off, hence the name pulsar.

Pythagorean Theorem | a theorem that relates the sides of a right triangle, stating that the square of the hypotenuse of the triangle (the side opposite the right angle) is equal to the sum of the squares of the other two sides. The theorem takes its name from the Ancient Greek philosopher and mathematician Pythagoras.

quantized | Discrete. For example, electrons in atoms, rather than having continuous energies, can only have a set of discrete or quantized energies, but not others.

quantum | Plural, quanta. A discrete minimal unit that is valid for physical systems. Photons, for example, are quanta of light.

quantum mechanics | The branch of physics that deals with the properties and behaviors of atoms and subatomic particles.

quantum system | A system that must be analyzed using quantum mechanics.

quark | A type of elementary particle that combines to make neutrons, protonsand other types of particles. There are six types of quarks, along with their anti-particle pairs. Three quarks combine to make neutrons and protons.

quintessence | A hypothetical form of non-constant dark energy postulated as an explanation of the observation of an accelerating rate of expansion of the Universe; involves a decaying energy field.

radial velocity | The velocity of an object along the observer's line of sight.

radian | A unit of angular measure, 1 radian = 57.3 degrees.

radiation | Energy emitted in the form of waves (example: light) or particles (example: electrons).

radio waves | The name given to the lowest energy region of the electromagnetic spectrum. Radio waves have wavelengths of meters (m), or even kilometers (km).

radioactive | An atom that is unstable and will break apart to become a new element, releasing energetic particles.

radioactivity | The natural or artificial process by which the nucleus of an atom is unstable and thereby breaks apart (decays) to become a new element. The decay process is accompanied by emission of energetic particles.

redshift | This is the name given to the apparent change in the wavelength of light due to the Doppler effect. Scientists know what the regular spectrum of a galaxy should look like (based on the spectrum of light emitted from known elements). If the light waves from a galaxy appear to have shifted towards higher frequency (blue), it is moving towards us, and if they have shifted toward a lower frequency (red), it means the object is moving away.

redshift (cosmological) | This is the name given to the apparent change in the wavelength of light due to the expansion of the Universe. The cosmological redshift is denoted by the letter z, and it is defined such that the Universe has expanded by an amount 1+z over the time the light has traveled to us. So an object with redshift z=1 is seen when the Universe was half its present size (it is twice as big as when the light was emitted), if z=2 the Universe is three times bigger than when the light was emitted, if z=3 the Universe is four times bigger, and so on.

reference frame | see frame.

relativistic | systems in which relativity is important, generally because the velocities are an appreciable fraction of the speed of light.

relativistic gamma | The factor in special relativity by which time and space are stretched or compressed by relative motion. Also, the ratio of the energy of a particle to its rest energy. The gamma factor depends on the relative velocity between reference frames approximately equal to x for small angles x (when x is measured in radians).

resolution | The fine-ness of a measurement. For example, a camera with a high resolution has the capability of capturing a more detailed image than a camera with a lower resolution.

rest energy | See rest mass.

rest energy | the energy a particle has in its rest frame, which depends on the particle's mass and the speed of light.

rest frame | a frame that is not moving with respect to an observer making measurements in it. We say that the observer is at rest in such a frame.

rest mass | The mass of an object measured when it is at rest relative to the observer measuring it.

rotate | turns on its own axis

satellite | A natural or man-made object that orbits a planet or other object.

scale | A ratio between the measurement of an object or event and its representation within a model. The scale can be represented as 1:X, where one is "one unit on the model" and is equal to X units in the actual system. On some maps, there is a scale that reads "one inch equals 10 miles" or something similar.

scientific notation | Using base-ten exponential form, e.g. 2.04×10^4 kg for 20,400 kg, to write numbers, especially very large or very small numbers.

second | Abbreviated "s". The fundamental SI unit of time, defined as the period of time equal to the duration of 9,192,631,770 periods of the radiation corresponding to the transition between two hyperfine levels of the ground state of the cesium-133 atom. There are 60 seconds in each minute and 3600 seconds in each hour of time.

Second Law of Thermodynamics | a physical law that says that the total entropy of a system must always increase.

SI Units | Abbreviated SI Units, also called the metric system, it is the scientific standard of carefully defined units of measurement. The SI unit of length is the meter, of time is the second, and of mass is the kilogram. Many other units are used to measure other quantities.

singularity | A place in the center of a black hole where the equations describing the mass density and gravitational force become infinite.

sky brightness | Extraneous light in the sky that is a result of scattered light from ground sources, light emitted from the atmosphere itself (perhaps due to its reaction with cosmic rays), and light from unresolved background sources.

small angle approximation | A mathematical approximation which amounts to tan(x) and sin(x) being approximately equal to x for small angles x (when x is measured in radians).

small angle formula | See small angle approximation.

SNR | See supernova remnant.

solar wind | The term given to the stream of charged particles that are ejected from the Sun's atmosphere. One of the effects of the interaction of solar wind and Earth's atmosphere is a creation of beautiful light patterns in sky known as auroras.

solar/star system | Defined as a system of celestial objects such as planets and asteroids orbiting one or more stars. When capitalized (Solar System), indicates the system associated with Earth and the Sun.

sources of error | Factors that can introduce errors into measurements and experiments. One of the most common sources of error originates in the instruments themselves.

spacetime | the four-dimensional system employed in special relativity that merges three spatial dimensions with one dimension of time and describes all events as points in that system: *(t, x, y, z)*. Three coordinates, *x*, *y*, and zdescribe the position of an event in space, and one, *t*, describes its position in time.

spacetime diagram | a simplified schematic representation of spacetime that generally shows the time axis as vertical with one axis of space being horizontal. The other two spatial dimensions are usually suppressed for simplicity. The diagrams are useful for understanding the relationship between events as seen by different observers in special relativity.

Special Theory of Relativity | Commonly called special relativity, the theory that predicts the behavior of objects moving at speeds close to the speed of light. One assumption of Special Relativity is that the speed of light in a vacuum is always constant.

spectra | Singular, *spectrum*. The distribution of intensity (i.e., number) of photons as a function of energy. Equivalently, it could be photon intensity distribution versus either wavelength or frequency.

Spectral Class | See stellar spectral classification.

spectrograph | Scientific instrument used to measure a spectrum.

spectroscopy | The scientific technique in which the intensity of light of different colors or wavelengths is measured. Comparing the measurements at different wavelengths can help to determine, for example, which elements are present in the light source.

spectrum | Plural, spectra. The distribution of intensity (i.e., number) of photons as a function of energy. Equivalently, it could be photon intensity distribution versus either wavelength or frequency.

speed | The distance (d) covered by a moving object in a given time (t). Mathematically, speed, s, is given by s = d/t. Differs from velocity in that velocity takes account of direction, not just how fast something moves.

speed of light | In a vacuum, denoted as "c", the speed of light is 299,792,458 m/s in all frames of reference, regardless of their relative states of motion.

spin | The quantum mechanical property of a particle that is analogous to the classical angular momentum (like a spinning top). For particles, the spin is always quantized and is either integral (for bosons) or half-integral (for fermions).

spiral arms | See spiral galaxy.

spiral galaxy | A galaxy whose primary feature is a flattened disk with long spiraling arms that extend out from a central core or bulge. The arms contain many massive young stars, a sign of high rates of star formation there.

standard candle | A celestial object that has a known inherent luminosity. Because these objects have a known luminosity output, their distances can be measured using their apparent brightness and the fact that brightness of an object falls as the inverse-square of distance from the object.

standard ruler | An astronomical object whose physical size is known. Using the known size and the small angle approximation, it is possible to determine the distance to the object.

stellar nurseries | Giant molecular clouds in galaxy, in which the density and temperature of the gas are such that parts of the clouds collapse under their own gravity to form new stars.

stellar spectral classification | A system in which stars are given a classification of O, B, A, F, G, A, K, or M based on their pattern of spectral absorption lines, which is related to their surface temperatures. O stars are the hottest and M stars are the coolest.

strong nuclear force | The force between quarks that keeps protons and neutrons bound within the nucleus. The force is also responsible for binding the nuclei themselves.

subatomic particles | Particles smaller than an atom, such as neutrons, protons and electrons, as well as other smaller particles like quarks, neutrinos, etc.

supermassive black hole | A black hole with mass on the order of millions or billions of solar masses. There is strong evidence that all large galaxies contain such black holes in their cores, including our own Milky Way, which contains a 4-million solar mass black hole at its center.

supernova | Plural, supernovae. The explosive collapse of the core of an evolved star. In massive stars, this collapse forms a neutron star or black hole. Under some circumstances low mass stars can also undergo supernovae as the result of the explosion of a particular type of white dwarf.

supernova remnant | The gas expelled during a supernova explosion, as well as the material swept up by that gas.

supernovae | Singular, *supernova*. The explosive collapse of the core of an evolved star. In massive stars, this collapse forms a neutron star or black hole. Under some circumstances low mass stars can also undergo supernovae as the result of the explosion of a particular type of white dwarf.

tangential velocity | In astronomy, the velocity of a star perpendicular to our line of sight, i.e., its velocity in the plane of the sky.

telescope | Optical instruments used to see great distances. Although originally telescopes were handheld, today they include both ground- and space-based varieties. Some examples are 10-meter motor-driven optical instruments, such as the Keck telescopes in Hawaii; the 27 antenna Very Large Array (VLA) radio observatory in New Mexico; and the orbiting Fermi Gamma ray Space Telescope some 550 km above Earth.

temperature | Most commonly a measure of the average energy of a particle in a system. Measured in kelvin (K) in the SI system.

Tera (prefix) | A trillion; 10^{12}. Denoted as T, e.g., TeV.

terrestrial | "Earth-like" planets. They are made mostly of rock, have solid surfaces, and typically have a mass comparable to Earth's. They generally have only one or two moons, if any. From the Latin, *terra*, meaning Earth.

theory | A conceptual framework that has explanatory and predictive power related to some aspect of the world. Theories generally encapsulate many experimental results and observations of the world into a coherent logical structure that makes testable predictions about related phenomena. For instance, Newton's theory of gravity explains the motions of the moon and falling objects on Earth, and makes predictions about the motions of other planets in the solar system as well as stars and galaxies.

time | A measure of how long it takes something to happen (an event's duration).

time dilation | The slowing of clocks that are in motion relative to an observer when compared to clocks at rest with respect to the observer. Predicted by Einstein's Special Theory of Relativity.

time dilation | the distortion of time between inertial frames. Observers in different inertial frames will measure time to pass more quickly in their own frame than in any inertial frame moving relative to their own.

Trans-Neptunian Objects | Trans-Neptunian Objects—Minor planets that orbit the Sun interior to the Kuiper Belt.

transit | When a planet passes in between Earth and the Sun such that the planet is seen to cross the face of the Sun. Only Mercury and Venus can undergo transits. Also, the passage of an extrasolar planet across the face of its star.

turbulent flow | Chaotic or disorganized motions within a fluid.

Type 1A supernova | The explosion of a carbon-oxygen (C-O) white dwarf that has accumulated enough material from a companion star to exceed the Chandrasekhar mass limit.

ultraviolet (UV) light | Electromagnetic radiation intermediate between the blue/violet end of visible light and x-rays. Ultraviolet radiation is more energetic than visible light but less than x-rays. Ultraviolet wavelengths are typically about 100 to 3800 nanometers (nm, 10^{-9} m).

uncertainty | The range of likely errors based on measurements of a given quantity, generally denoted as plus/minus (±) error-range and depending upon the experiment and measurement techniques. For example, a measurement listed as 6 ± 1 mm might be expected to have a true value of anywhere from 5 to 7 mm.

uncertainty principle | The position and the velocity of an object cannot both be known to perfect precision simultaneously; similarly, the energy and lifetime of virtual particles cannot both be known to perfect precision simultaneously.

uniform motion | Moving with a constant speed and direction.

Universal Constant of Gravitation | Denoted as capital G. The constant of proportionality in Newton's law of universal gravitation. It plays an analogous role in Einstein's general relativity. It is equal to $6.67428 \times 10^{-11}\ m^3 / kg\text{-}sec^2$.

Universe | Everything that exists, including Earth, planets, stars, galaxies, and all that they contain; the entire cosmos.

vacuum energy | The energy content of "empty" space; one possible explanation for the cosmological constant.

variable | A value that can change. For instance, the brightness of a pulsar changes depending on whether or not its beam of light is pointing toward us when we are looking at it, so its brightness is variable. Often in mathematical expressions, some parameters are allowed to change, and are thus variable.

velocity | How fast an object moves in a given direction, i.e., the speed of an object in a given direction. Velocity differs from speed because speed is how fast something moves without regard to direction.

visible light | Electromagnetic radiation at wavelengths which the human eye can see. We perceive this radiation as colors ranging from red (longer wavelengths; ~ 700 nanometers) to violet (shorter wavelengths; ~400 nanometers.) Also called optical light.

wavelength | The distance between adjacent peaks in a series of periodic waves. Also see electromagnetic spectrum.

worldline | The path taken by a particle through spacetime. This line connects all the events in the particle's history and future.

x-ray | High-energy electromagnetic radiation. X-rays are more energetic than ultraviolet light but less energetic than gamma rays. The energy of x-rays ranges roughly from 1 keV up to a few hundred keV. Their wavelengths are from about 10 nm down to about 10 pm.

± | "Plus/Minus," indicates the range of uncertainty of a value, e.g. 10.2 ± 0.4 kg indicates the mean value of the experiment was 10.2 kg, but there is a possibility that the true mass lies somewhere between 9.8—10.6 kg.

Detailed Licensing

Overview

Title: Big Ideas in Cosmology (Coble et al.)

Webpages: 240

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 95% (228 pages)
- Undeclared: 5% (12 pages)

By Page

- Big Ideas in Cosmology (Coble et al.) - *CC BY-NC-SA 4.0*
  - Front Matter - *CC BY-NC-SA 4.0*
    - TitlePage - *CC BY-NC-SA 4.0*
    - InfoPage - *CC BY-NC-SA 4.0*
    - Table of Contents - *Undeclared*
    - About the Book - *CC BY-NC-SA 4.0*
    - Licensing - *Undeclared*
    - About the Authors - *CC BY-NC-SA 4.0*
  - 0: Review of Mathematics - *CC BY-NC-SA 4.0*
    - 0.0: Scientific Notation - *CC BY-NC-SA 4.0*
    - 0.1: Manipulating Numbers in Scientific Notation - *CC BY-NC-SA 4.0*
    - 0.2: The Metric System - *CC BY-NC-SA 4.0*
    - 0.3: Special Units Used in Astronomy - *CC BY-NC-SA 4.0*
    - 0.4: Comparing Sizes and Converting Units - *CC BY-NC-SA 4.0*
    - 0.5: Problem Solving and Thinking About Your Answers - *CC BY-NC-SA 4.0*
    - 0.6: Measurement Accuracy and Uncertainty - *CC BY-NC-SA 4.0*
    - 0.7: Wrapping It Up - *CC BY-NC-SA 4.0*
  - 1: Size and Scope - *CC BY-NC-SA 4.0*
    - 1.0: Size and Scope Introduction - *CC BY-NC-SA 4.0*
    - 1.1: Your Model of the Universe - *CC BY-NC-SA 4.0*
    - 1.2: Our Solar System - *CC BY-NC-SA 4.0*
    - 1.3: Our Galaxy - The Milky Way - *CC BY-NC-SA 4.0*
    - 1.4: Other Galaxies and Large-scale Structures in the Universe - *CC BY-NC-SA 4.0*
    - 1.5: The Smallest Stuff- Particles, Atoms, and Molecules - *CC BY-NC-SA 4.0*
    - 1.6: Hierarchy- How the Objects Are Arranged - *CC BY-NC-SA 4.0*
    - 1.7: Scale the Universe - *CC BY-NC-SA 4.0*
    - 1.8: Wrapping It Up 1 - Size, Scope, and Units - *CC BY-NC-SA 4.0*
    - 1.9: Mission Report 1 - Size, Scope, and Units - *CC BY-NC-SA 4.0*
  - 2: Light - *CC BY-NC-SA 4.0*
    - 2.0: Light Introduction - *CC BY-NC-SA 4.0*
    - 2.1: The Wave Nature of Light - *CC BY-NC-SA 4.0*
    - 2.2: The Particle Nature of Light - *CC BY-NC-SA 4.0*
    - 2.3: The Electromagnetic Spectrum - *CC BY-NC-SA 4.0*
    - 2.4: What a Spectrum of Light Can Tell Us About Matter - *CC BY-NC-SA 4.0*
    - 2.5: Continuous Spectra - a Planck Spectrum Tells us the Temperature of Objects - *CC BY-NC-SA 4.0*
    - 2.6: Lines Spectra- Emission and Absorption Lines - *CC BY-NC-SA 4.0*
    - 2.7: Determining the Composition of an Unknown Gas - *CC BY-NC-SA 4.0*
    - 2.8: Wrapping It Up 2 - The Properties of Light - *CC BY-NC-SA 4.0*
    - 2.9: Mission Report 2 - The Properties of Light - *CC BY-NC-SA 4.0*
  - 3: Telescopes - *CC BY-NC-SA 4.0*
    - 3.0: Telescope Introduction - *CC BY-NC-SA 4.0*
    - 3.1: Designing Telescopes Across the Spectrum - *CC BY-NC-SA 4.0*
    - 3.2: Location, Location, Location - *CC BY-NC-SA 4.0*
    - 3.3: Imaging Astronomical Objects - *CC BY-NC-SA 4.0*
    - 3.4: Determining the Brightness of Astronomical Objects - *CC BY-NC-SA 4.0*
    - 3.5: Wrapping It Up 3 - Light, Telescopes, and Astronomical Images - *CC BY-NC-SA 4.0*
    - 3.6: Mission Report 3 - Light, Telescopes, and Astronomical Images - *CC BY-NC-SA 4.0*
  - 4: Moving Through Space - *CC BY-NC-SA 4.0*
    - 4.0: Moving Through Space Introduction - *CC BY-NC-SA 4.0*

- 18.1: 15_4_how_cosmological_parameters_affect_the_cmb_regular_matter - *CC BY-NC-SA 4.0*
- 18.2: 16_2_4_particle_creation_timeline - *CC BY-NC-SA 4.0*
- 18.3: bigIdeasActivityOne - *CC BY-NC-SA 4.0*
- 18.4: ch16_particle_classification - *CC BY-NC-SA 4.0*
- 18.5: ch4pg1lightminute - *CC BY-NC-SA 4.0*
- 18.6: ch4pg1lightyear - *CC BY-NC-SA 4.0*
- 18.7: ch4pg1powersoftentimescales - *CC BY-NC-SA 4.0*
- 18.8: ch4pg2boatwavedemobcmovingsource - *CC BY-NC-SA 4.0*
- 18.9: ch4pg2boatwavedemosourceofwavesstationary - *CC BY-NC-SA 4.0*
- 18.10: ch4pg2dopplershiftofhydrogen - *CC BY-NC-SA 4.0*
- 18.11: ch4pg2quantitativedopplershift - *CC BY-NC-SA 4.0*
- 18.12: ch4pg2twowaysofrepresentingaspectrum - *CC BY-NC-SA 4.0*
- 18.13: ch4pg3relativemotion - *CC BY-NC-SA 4.0*
- 18.14: ch4wiu4theandromedashift - *CC BY-NC-SA 4.0*
- 18.15: ch6pg2standardruler - *CC BY-NC-SA 4.0*
- 18.16: ch6pg3animatedfigure612 - *CC BY-NC-SA 4.0*
- 18.17: ch6pg3fluxvsdistance - *CC BY-NC-SA 4.0*
- 18.18: ch6pg3mainsequencematchingforclustersofdifferentdistances - *CC BY-NC-SA 4.0*
- 18.19: ch6pg3standardcandle - *CC BY-NC-SA 4.0*
- 18.20: ch6pg3supernovaetypelaanddistances - *CC BY-NC-SA 4.0*
- 18.21: ch6pg3tullyfisher - *CC BY-NC-SA 4.0*
- 18.22: ch6pg4matchthetechniquetothedistance - *CC BY-NC-SA 4.0*
- 18.23: ch7pg2accelerationandmass - *CC BY-NC-SA 4.0*
- 18.24: ch7pg2netforceofanobject - *CC BY-NC-SA 4.0*
- 18.25: ch7pg3animatedfigure75 - *CC BY-NC-SA 4.0*
- 18.26: ch7pg3howdoestheforceofgravitydependondistance - *CC BY-NC-SA 4.0*
- 18.27: ch7pg3howdoestheforceofgravitydependonmass - *CC BY-NC-SA 4.0*
- 18.28: ch7wiu7thegalileanmoonsofjupiter - *CC BY-NC-SA 4.0*
- 18.29: chapter2pg1photonraceroundone - *CC BY-NC-SA 4.0*
- 18.30: chapter2pg5planckspectrum - *CC BY-NC-SA 4.0*
- 18.31: chapter2pg6viewingorientation - *CC BY-NC-SA 4.0*
- 18.32: chapter2pg7determiningthecomposition - *CC BY-NC-SA 4.0*
- 18.33: chapter3pt1lightcollectingarea - *CC BY-NC-SA 4.0*
- 18.34: chapter3pt4afindingthesourceandsubtractingthebackground - *CC BY-NC-SA 4.0*
- 18.35: chapter3pt4bcolorizingccdimages - *CC BY-NC-SA 4.0*
- 18.36: chapter_11_2_falling_into_a_black_hole_tidalforces - *CC BY-NC-SA 4.0*
- 18.37: chapter_11_2_time_dilation_spacetime_dialation - *CC BY-NC-SA 4.0*
- 18.38: chapter_12_2_lensing_by_a_point_mass_gravitationallensing_simple - *CC BY-NC-SA 4.0*
- 18.39: chapter_12_2_macho_light_curve - *CC BY-NC-SA 4.0*
- 18.40: chapter_12_2_the_einstein_radius_einsteinring - *CC BY-NC-SA 4.0*
- 18.41: chapter_12_3_different_types_of_lenses_glbinning - *CC BY-NC-SA 4.0*
- 18.42: chap_2_1_wave_generator_NEW - *CC BY-NC-SA 4.0*
- 18.43: comovingcoords - *CC BY-NC-SA 4.0*
- 18.44: expandinguniverse1 - *CC BY-NC-SA 4.0*
- 18.45: explodinguniverse - *CC BY-NC-SA 4.0*
- 18.46: explosionvsexpansiondensity - *CC BY-NC-SA 4.0*
- 18.47: orbitsimulator - *CC BY-NC-SA 4.0*
- 18.48: testjQueryUI - *CC BY-NC-SA 4.0*
- 18.49: var1 - *CC BY-NC-SA 4.0*
- 18.50: wiu15_a_curvature - *CC BY-NC-SA 4.0*
- 18.51: wiu15_b_amount_of_baryonic_matter - *CC BY-NC-SA 4.0*
- 18.52: wiu15_c_amount_of_cold_dark_matter - *CC BY-NC-SA 4.0*
- 18.53: wiu6thesupernovaof1885 - *CC BY-NC-SA 4.0*

Resources: Formulae, Constants, and Conversions

Formulae	Relationship	Initial Chapter
$\lambda = c/f$	wavelength, frequency, and speed	Chapter 2
$E = hf$	energy and frequency	
$E_{photon} = \Delta E_{electron}$	energy of photon emitted/absorbed by atom	
$E_n = -13.6/n^2$	energy levels of atom	
$T = 2.9e\text{-}3/\lambda$	temperature and peak wavelength	
$F = \sigma T^4$	brightness and temperature	
$A = \pi r^2$	area and radius	Chapter 3
$\theta = 1.22\, \lambda / D$	resolution	
$v = d/t$	speed, distance, and time	Chapter 4
$z = \Delta\lambda/\lambda$	redshift	
$v = cz$	velocity and redshift	
$t = E/L$	stellar lifetime	Chapter 5
$E = mc^2$	energy and mass	
$t \sim m^{-2}$	stellar lifetime and mass	
$L \sim m^3$	stellar mass and luminosity	
$d = 1/a$	parallax angle and distance	Chapter 6
$d = S/\theta$	small angle formula	
$F= L/4\pi d^2$	inverse square law	
$F = ma$	Newton's second law	Chapter 7
$F_g = mg$	Weight and mass	
$a_c = v^2/r$	Centripetal acceleration	
$F_g = Gm_1m_2/r^2$	Newton's law of gravity	
$PE = mgh$	Potential energy (non-relativistic)	
$KE = ½\, mv^2$	Kinetic energy (non-relativistic)	
$E_{final} = E_{initial}$	Conservation of energy	

Formulae	Relationship	Initial Chapter
$v_{escape} = (2GM/R)^{½}$	Escape velocity	
v ∝ r	Rotation of rigid disk	Chapter 8
v ∝ 1/r	Rotation of water around drain	
v ∝ constant	Rotation of cars on roundabout	
$v \propto 1/r^{½}$	Keplerian rotation (planets)	
$v^2 = GM/r$	Relationship of enclosed mass to velocity and distance	
M = ρV	Mass, density, and volume	
$\gamma = 1/\sqrt{(1-v^2/c^2)}$	gamma factor	Chapter 9
$\Delta t' = \gamma \Delta t$	time dilation	
$L' = L/\gamma$	length contraction	
$d^2 = \Delta x^2 + \Delta y^2$	Pythagorean Theorem	
$s^2 = \Delta x^2 – c(\Delta t)^2$	spacetime interval	
$E = mc^2$	mass and rest energy	
$E = \gamma E_0$	total energy and rest energy	
$g=\frac{GM}{R^2}$	Surface gravity	Chapter 10
$d=v_0t+\frac{1}{2}at^2$	Distance and acceleration	
$v=at$	Velocity and acceleration	
$t=\frac{t_0}{\left(1-\frac{gH}{c^2}\right)}$	Time dilation (weak field approximation)	
$f=f_0\left(1-\frac{gH}{c^2}\right)$	Gravitational redshift (weak field approximation, frequency, photon traveling upward)	

Formulae	Relationship	Initial Chapter
$$\lambda=\frac{\lambda_0}{\left(1-\frac{gH}{c^2}\right)}$$	Gravitational redshift (weak field approximation, wavelength)	
$$f=f_0\sqrt{1-\frac{2GM}{Rc^2}}$$	Gravitational redshift (full expression, frequency)	
$$\lambda=\frac{\lambda_0}{\sqrt{1-\frac{2GM}{Rc^2}}}$$	Gravitational redshift (full expression, wavelength)	
$$d^2=(\Delta x)^2+(\Delta y)^2$$	Pythagorean theorem	
$$d^2=(R\Delta\theta)^2+\cos^2\theta(R\Delta\alpha)^2$$	Distance on a sphere	
$$d^2=(\Delta x)^2+(\Delta y)^2+(\Delta z)^2$$	Pythagorean Theorem in 3-D	
$$s^2=(\Delta x)^2+(\Delta y)^2+(\Delta z)^2-(c\Delta t)^2$$	Spacetime interval in flat space	
$$s^2=\left(1-\frac{2GM}{rc^2}\right)^{-1}\left[(\Delta x)^2+(\Delta y)^2+(\Delta z)^2\right]-\left(1-\frac{2GM}{rc^2}\right)(c\Delta t)^2$$	Spacetime interval in spherically curved space	
$$\theta=\frac{2GM}{bc^2}$$	Angle of deflection of light	
$$P_{gw}=\frac{2}{5}\left(\frac{GM}{Rc^2}\right)^5\left(\frac{m}{M}\right)^2\left(\frac{c^5}{G}\right)$$	Power emitted by gravitational waves	
$$\mathbf{G}=8\pi G\,\mathbf{T}/c^4$$	Einstein equation	

Formulae	Relationship	Initial Chapter
$$R_S = \frac{2GM}{c^2}$$	Schwarzschild radius	Chapter 11
$$s^2 = \left(1 - \frac{R_S}{d}\right)^{-1} (\Delta d)^2 - \left(1 - \frac{R_S}{d}\right) (c\Delta t)^2$$	Spacetime interval in a spherically symmetric space (Schwarzschild interval)	
$$T_{bh} = \frac{1.23 \times 10^{23}}{M}$$	Temperature of a black hole	
$$\Delta E \Delta t \geq \frac{h}{4\pi}$$	Uncertainty principle	
$$L = \frac{3.56 \times 10^{32}}{M^2}$$	Luminosity of a blackhole	
$$t \approx 2.5 \times 10^{-16} M^3$$	Evaporation time	
$$\frac{\Delta m}{\Delta t} = -\frac{2L}{c^2}$$	Accretion rate	
$$L_{edd} = 6.3 M_{BH}$$	Eddington luminosity	
$$\alpha = \frac{4GM}{bc^2}$$	Deflection angle (full)	Chapter 12
$$\theta_E = \sqrt{\left(\frac{4GM(b)}{c^2}\right)\left(\frac{D_{LS}}{D_{LO} D_{SO}}\right)}$$	Einstein radius	
$$\theta^2 - x\theta - \theta_E^2 = 0$$	Lens equation	

Formulae	Relationship	Initial Chapter
$$m=\frac{1}{\left[1-\left(\frac{\theta_E}{\theta}\right)^4\right]}$$	Magnification for a point-mass lens	
$$v=H_0 d$$	Hubble law	Chapter 13
$$v=cz$$	Cosmological redshift	
$$d_{\text{physical}}(t)=d_{\text{comoving}}(t)S(t)$$	Comoving coordinates	
$$1+z=\frac{S(t_{\text{observed}})}{S(t_{\text{emitted}})}$$	Ratio of scale factors	
$$t=\frac{1}{H_0}$$	Hubble time (age)	
$$H^2-\frac{8\pi G\rho}{3}=-\frac{kc^2}{S^2}$$	Friedman equation	
$d = cz/H_0$	distance and redshift	Chapter 14
$T_e / T_o = 1 + z = S_o / S_e$	Temperature, redshift, and scale factor	Chapter 15
E ~ kT	energy and temperature	Chapter 16
$T \sim mc^2/k$	temperature of Universe and mass of particle in reaction	

Constants	Name
$c = 3 \text{ x } 10^8$ m/s = $3 \text{ x } 10^5$ km/s	speed of light
$h = 6.63 \text{ x } 10^{-34}$ J s = 4.136e-15 eV s	Planck’s constant
$G = 6.67 \text{ x } 10^{-11}$ N m^2/kg^2	Universal gravitational constant
$k_B = 1.38 \times 10^{-23}$ J/K	Boltzmann's constant
$t_{planck} \sim 10^{-43}$ s	Planck time
$t_{planck} \sim 4 \times 10^{-35}$ m	Planck length

Formulae	Relationship	Initial Chapter
$m_e = 9.1 \times 10^{-31}$ kg	mass of an electron	
Conversions	Units	
1 km = 1000 m	km and meters	
1 km = 0.6 mi	km and miles	
1 AU = 1.5 x 10^{11} m = 1.5 x 10^{8} km	AU and meters and km	
1 ly = 9.5 x 10^{15} m = 9.5 x 10^{12} km	light-years and meters and km	
1 ly = 6.3 x 10^{4} AU	light-years and AU	
1 eV = 1.6 x 10^{-19} J	eV and joules	
1 degree = 60 arcmin	degrees and arcminutes	
1 arcmin = 60 arcsec	arcminutes and arcseconds	
1 angstrom (Å) = 1 x 10^{-10} meters	angstroms and meters	
1 pc = 3.26 ly	parsecs and light-years	
1 N = 0.2248 pounds	newtons and pounds	
1 kpc = 3.086 x 10^{19} m	kiloparsecs and meters	
1 solar mass = 2 × 10^{30} kg	solar masses and kg	
1 Mpc = 3.09 x 10^{22} m	megaparsecs and meters	
1 radian = 2.06 x 10^{5} arcsecond	radians and arcsec	
1 Mpc = 3.09 × 10^{19} km	megaparsecs and km	
Units (abbreviation)	Type of quantity	
meters (m)	length (SI)	
kilograms (kg)	mass (SI)	
second (s)	time (SI)	
meters per second (m/s)	speed (SI)	
kelvin (K)	temperature (SI)	
miles (mi)	length	
astronomical unit (AU)	length	
year (yr)	time	
light-year (ly)	length	
light-minutes	length	
light-seconds	length	
g/cm^3	density	

Formulae	Relationship	Initial Chapter
solar masses	mass	
hertz (Hz) = cycles/s = 1/s = s^{-1}	frequency (SI)	
joules (J)	energy (SI)	
electron volts (eV)	energy	
watts (W) = J/s	power	
radians	angle	
degrees	angle	
arcmin	angle	
arcsec	angle	
angstrom (Å)	length	
parsec (pc)	length	
m/s^2	acceleration (SI)	
newton (N) = kg m/s^2	force (SI)	
joules (J) = N m	energy (SI)	
μK micro Kelvin = 10^{-6} K	temperature	

Prefix	Meaning (in USA)	Exponent	Symbol
Tera	trillion	10^{12}	T
Giga	billion	10^{9}	G
Mega	million	10^{6}	M
kilo	thousand	10^{3}	k
centi	one-hundredth	10^{-2}	c
milli	one-thousandth	10^{-3}	m
micro	one-millionth	10^{-6}	μ
nano	one-billionth	10^{-9}	n
pico	one-trillionth	10^{-12}	p